

# Meaning Acquisition by Semiotic Agents

## Semiotic Cognitive Information Processing in a language environment\*

Burghard B. Rieger\*\*

Dept. of Computational Linguistics — University of Trier

D-54286 TRIER, Germany

rieger@ldv01.uni-trier.de

### Abstract

Anything we know or believe about the world can (more or less precisely) be communicated verbally. We do so by using words, forming sentences and producing texts whose meanings are understood to stand for, represent, or deal with the topics and subjects, the domains and structures in the real world they are meant to refer to. Natural language texts (still) are the most flexible and as that a highly efficient form to represent knowledge for and convey learning to others. Traditional approaches to the study of language understanding in CL and AI employ rule based formats of *linguistic knowledge* and symbol representations of *world knowledge* structures to model language processing by machine. Providing these initial knowledge bases and allowing them to be modified by system designers (external change), or dynamically as a function of processing (internal learning) proved to be everything from enormously laborious to error prone, from extremely difficult to virtually impossible. Computational Semiotics (CS) neither depends on rule-based or symbolic formats for (linguistic) knowledge representations, nor does it subscribe to the notion of (world) knowledge as some static structures that may be abstracted from and represented independently of the way they are processed. Consequently, knowledge structures and the processes operating on them are to be modelled procedurally and have to be implemented as algorithms which determine *SCIP* systems. As a collection of cognitive information processing devices these systems' *semiotic* character consists in their multi-level representational performance of (working) structures emerging from and being modified by such processing. The emergence of semantic structure as a self-organizing process is studied on the basis of word usage regularities in natural language discourse, whose linearly agglomerative (or *syntagmatic*) and whose selectively interchangeable (or *paradigmatic*) constraints are exploited by text analysing algorithms. They accept natural language discourse as input and end up to produce a vector space structure as output. This may be interpreted as an (internal) representation of the semiotic system's states of adaptation to the (external) structures of its environment as mediated by the natural language discourse processed. In order to evaluate the internal picture which the system computes from the natural language texts according to its processing capabilities against the external reality whose structure and properties are described by natural language discourse only, 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, on the right — in front, behind*) and *hedges* (like: *extremely, very, rather — nearby, faraway*). Processed during the system's training phase, the corpus reveals structural constraints which the system's hidden structures or internal meaning representations apparently reflect. Compared with a two-dimensional representation of the external reality – as described by the texts and specified by the underlying syntax and semantics – a two-dimensional transform of the system's internal view of its environment proves to be surprisingly adequate.

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) referential predicates. Test results from an experimental setting with varying fuzzy interpretations of hedges will be produced to illustrate the *SCIP* system's miniature (cognitive) language understanding and meaning acquisition capacity without any initial explicit syntactic and semantic knowledge.

---

\*Paper presented at the ICAS/ German Federal Forces University Workshop on *Agents, Cooperation, and Communication (AC&C)*, San Marco di Castellabate, SA, Italy, June 18–24, 1995. To appear in: Becker, J.D. (Ed.): *Agents, Communication, and Cooperation*, [Lecture Notes in Artificial Intelligence], Berlin/Heidelberg/New York (Springer).

\*\*The author is indebted to discussion of central ideas of this paper with *Petra Badry, Kathrin Gieseke, Beate Oerder, Maria Reichert* and *Ralph Wagner* whose substantial contributions (of varying intensity and uneven distribution during different phases of the project) in converting procedural models to operational programs are highly appreciated. The errors are his own as always.

# 1 An ecological approach to semiotics

Life may be understood as the ability to survive by adapting to changing requirements in the real world. Living systems do so by way of processing information they receive or derive from relevant portions of their surrounding environments, of learning from their experience, and of changing their behaviour accordingly. In contrast to other living systems which transmit experiential results of environmental adaptation only biogenetically<sup>1</sup> to their descendants, human information processing systems have additional means to convey their knowledge to others. In addition to the *vertical* transmission of system specific (*intraeous*) experience through (biogenetically successive) generations, mankind has complementally developed *horizontal* means of mediating specific and foreign (*extraneous*) experience and knowledge to (biogenetically unrelated) fellow systems within their own or any later generation. This is made possible by a *semiotic* move that allows not only to distinguish *processes* from *results* of experience but also to convert the latter to *knowledge* facilitating it to be re-used, modified and improved in *learning*. Vehicle and medium of this move are *representations*, i.e. complex sign systems which constitute *languages* and form structures, called *texts* which may be realized in communicative processes, called *actualisation*.

In terms of the theory of information systems, *texts*—whether internal or external to the systems—function like virtual environments<sup>2</sup>. Considering the system-environment relation, *virtuality* may be characterized by the fact that it dispenses with the identity of space-time coordinates for system-environment pairs which normally prevails for this relation when qualified to be indexed *real*.

It appears, that this dispensation of identity (space-time-dispensation, for short) is not only conditional for the possible distinction of (mutually and relatively independent) *systems* from their *environments*, but establishes also the notion of *representation*.

Accordingly, *immediate* or space-time-identical system-environments existing in their space-time-identity may well be distinguished from *mediate* or space-time-dispensed system-environments whose particular representational form (*texts*) corresponds to their particular status both, as language material (being *signs*), and as language structure (having *meaning*). This double identity calls for a particular modus of actualisation (*understanding*) that may be characterized as follows:

For systems appropriately adapted and tuned to such environments *actualisation* consists essentially in a twofold embedding to realize

- ▷ the space-time-identity of pairs of *immediate* system-environment coordinates which will let the system experience the material properties of texts as *signs* (i.e. by functions of *physical access* and *mutually homomorphic* appearance). These properties apply to the percepts of language structures presented to a system in a particular *discourse situation*, and
- ▷ the representational identity of pairs of *mediate* system-environment parameters which will let the system experience the semantic properties of texts as *meanings* (i.e. by functions of *emergence*, *identification*, *organisation*, *representation* of structures). These apply to the comprehension of language structures recognized by a system to form the *described situation*.

Hence, according to the theory of information systems, functions like *interpreting* signs and *understanding* meanings translate to processes which extend the fragments of reality accessible to a living (natural and possibly artificial) information processing system. This extension applies to both, the *immediate* and *mediate* relations a system may establish according to its own evolved adapt- edness or dispositions (i.e. innate and acquired *structuredness*, processing *capabilities*, represented

---

<sup>1</sup>According to standard theory there is no direct genetic coding of experiential results but rather indirect transmission of them by selectional advantages which organisms with certain genetic mutations gain over others without them to survive under changing environmental conditions.

<sup>2</sup>Simon's [15] remark "There is a certain arbitrariness in drawing the boundary between inner and outer environments of artificial systems. ... Long-term memory operates like a second environment, parallel to the environment sensed through eyes and ears" (pp. 104) is not a case in point here. As will become clear in what follows, his distinction of *inner* (memory structure) and *outer* (world structure) environments of a system misses the special *semiotic* quality of natural language signs whose twofold environmental embedding (textual structure) cuts across the *inner/outer* distinction, resolving both, memory and world structures in becoming representational for each other.

knowledge).

The *actualisation* of environments, however, does not merely add to the amount of experiential results, but constitutes instead a significant change in experiential modus. This change is characterized by the fact that only now the *processes* of experience may be realized as being different and hence be separated from the *results* of experience which may thus even be represented, other than in *immediate* system-environments where *result* and *process* of experience appear to be indistinguishable. Splitting up experience in experiential processes and experiential results—the latter being representational and in need for actualisation by the former—is tantamount to the emergence of *virtual* experiences which have not to be *made* but can instead just be *tried*, very much like hypotheses in an experimental setting of a testbed. These *results*—like in *immediate* system-environments—may become part of a system’s adaptive knowledge but may also—different from *immediate* system-environments—be neglected or tested, accepted or dismissed, repeatedly actualized and re-used without any risk for the system’s own survival, stability or adaptedness.

The experimental quality of textual representations which increases the potentials of adaptive information processing immensely, will have to be constrained simultaneously by *dynamic* structures, corresponding to *knowledge*. The built-up, employment, and modification of these structural constraints<sup>3</sup> is controlled by procedures whose processes determine *cognition* and whose results constitute *adaptation*. Systems properly adapted to textual system-environments have acquired these structural constraints (language knowledge) and can perform certain operations efficiently on them (language understanding). These are prerequisites to recognizing *mediate* (textual) environments and to identify their need for and the systems’ own ability to *actualize* the mutual (and trifold) relatedness constituting what Peirce [3] called *semiosis*<sup>4</sup>.

Systems capable of and tuned to such knowledge-based processes of actualisation will in the sequel be referred to as *semiotic cognitive information processing systems (SCIPS)*.

## 2 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<sup>5</sup> as

---

<sup>3</sup>What Simon [15] calls *memory* in his questioning the inner-outer-distinction of cognitive systems and their environments.

<sup>4</sup>”By *semiosis* I mean [...] an action, or influence, which is, or involves, a coöperation of *three* subjects, such as sign, its object, and its interpretant, this tri-relative influence not being in any way resolvable into actions between pairs.” (p.282)

<sup>5</sup>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 [18] [19] and their logical derivatives 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.

soon as a *semiotic* line of approaches to cognition will be followed [4].

## 2.1 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 expression 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 *referential*) 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.

## 2.2 Communicating: language games

The notion of *language games* [17] "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 [4] 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 cognitive 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 [5] [7] [9] [14] 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.

### 3 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 [16] 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 [6] 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 [10] [11] 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.

#### 3.1 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  $\alpha$ -abstraction) on the set of *fuzzy* subsets of the vocabulary provides the word-types' usage regularities or *corpus points*, the second (called  $\delta$ -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  $K = \{k_t\}; t = 1, \dots, T$  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  $V = \{x_n\}; n = 1, \dots, i, j, \dots, N$  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 co-occur 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,  $\tilde{\alpha} : V \times V \rightarrow I$  can be conditioned on  $x_n \in V$  which yields a crisp mapping

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

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

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

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 co”-or”-di”-na”-tes— 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.

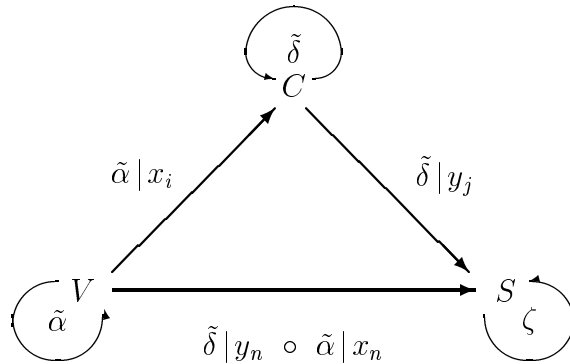


Figure 1: Fuzzy mapping relations  $\tilde{\alpha}$  and  $\tilde{\delta}$  between the structured sets of vocabulary items  $x_n \in V$ , of corpus points  $y_n \in C$ , and of meaning points  $z_n \in S$ .

### 3.2 Distributed meaning representation

Considering  $C$  as representational structure of abstract entities [12] 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  $\delta$  of, say, EUCLIDIAN metric

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

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

Thus,  $\delta$  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,  $\tilde{\delta} : C \times C \rightarrow I$  can be conditioned on  $y_n \in C$  which again yields a crisp mapping

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

where the tuples  $\langle (y_{n,1}, \tilde{\delta}(n,1)), \dots, (y_{n,N}, \tilde{\delta}(n,N)) \rangle$  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

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

$V \times V$	$\alpha$ -abstraction	$C \times C$	$\delta$ -abstraction	$S \times S$
$\tilde{\alpha} \mid \begin{array}{c ccc} x_1 & \dots & x_N \\ \hline x_1 & \alpha_{11} & \dots & \alpha_{1N} \\ \vdots & \vdots & \ddots & \vdots \\ x_N & \alpha_{N1} & \dots & \alpha_{NN} \end{array}$	$\Downarrow$	$\tilde{\delta} \mid \begin{array}{c ccc} y_1 & \dots & y_N \\ \hline y_1 & \delta_{11} & \dots & \delta_{1N} \\ \vdots & \vdots & \ddots & \vdots \\ y_N & \delta_{N1} & \dots & \delta_{NN} \end{array}$	$\Downarrow$	$\zeta \mid \begin{array}{c ccc} z_1 & \dots & z_N \\ \hline z_1 & \zeta_{11} & \dots & \zeta_{1N} \\ \vdots & \vdots & \ddots & \vdots \\ z_N & \zeta_{N1} & \dots & \zeta_{NN} \end{array}$
	$\tilde{\alpha} \mid x_i$		$\tilde{\delta} \mid y_j$	
	$\Uparrow$		$\Uparrow$	
	<i>Syntagmatic</i>	<i>C o n s t r a i n t s</i>	<i>Paradigmatic</i>	

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

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

$$\zeta : S \times S \rightarrow I \quad (7)$$

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 ( $\delta$ - or distance-values) of all regularities of usage ( $\alpha$ - 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 (Fig. 2) 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.

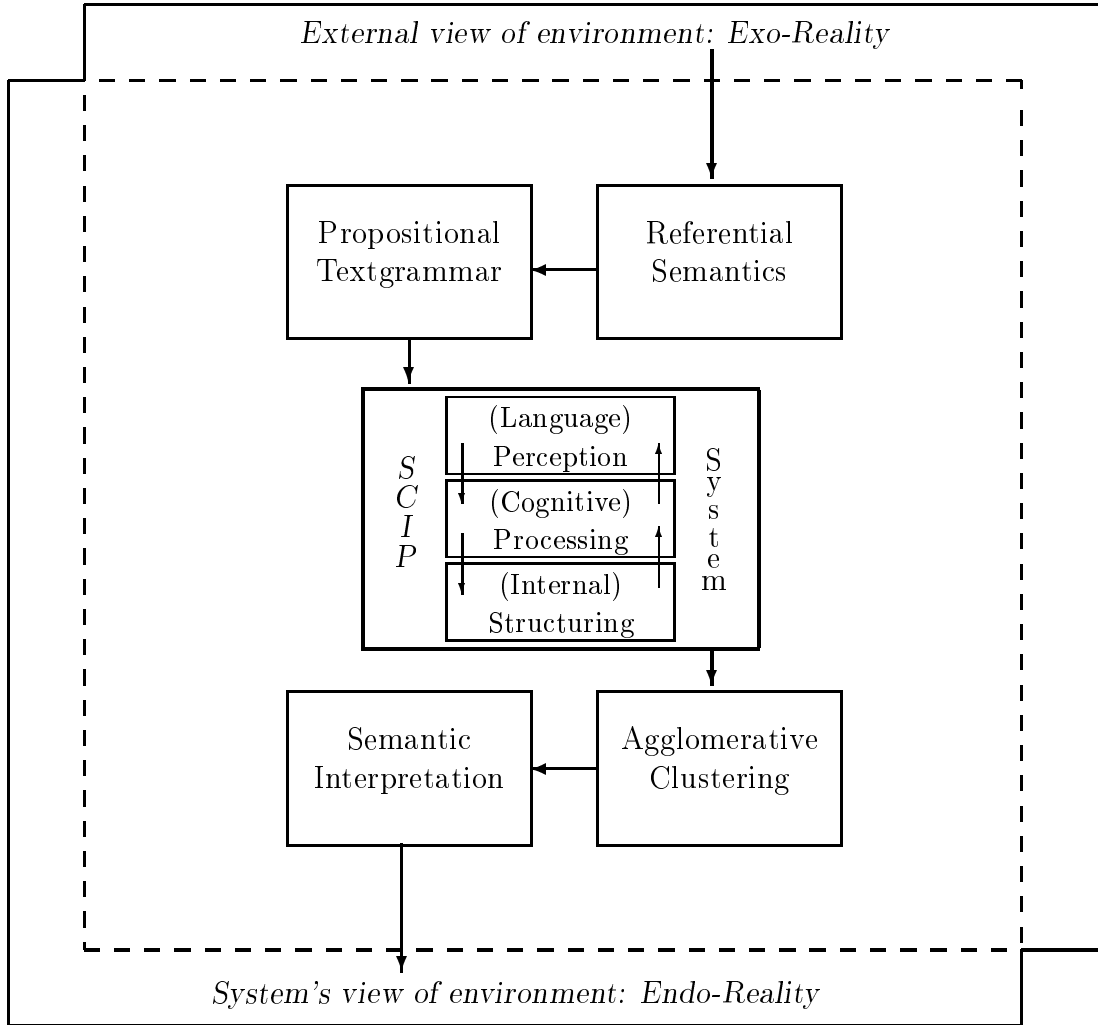


Figure 2: Situational setting of *SCIP* system within its environment which is defined to allow for the system's view (*Endo-Reality*) to differ from the external observer's view (*Exo-Reality*) by keeping the system's (non-propositional) faculties of language processing strictly apart from the (propositional) way of generating the environmental language data as textual descriptions. Note, that *grammar* (lexicon, syntax) and *semantics* are not part of the system's knowledge base but are introduced to specify and formally control the language environment the system is exposed to as "true" descriptions of the *external* reality. Thus, the the system's processing of these language data and its independently built-up *internal* representations allow for a semantic interpretation and visible imaging of the structures the system might have acquired.

## 4 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<sup>6</sup>. To achieve this, it had to be guaranteed



<i>SCIP</i> – System – properties	
<i>SCIPS</i> :	$\{\mathcal{O}, \mathcal{B}, \mathcal{W}, \mathcal{F}, \mathcal{K}\}$
<i>Orientation</i> :	$\mathcal{O} := \{\vec{N} = (0, 1), \vec{O} = (1, 0), \vec{S} = (0, -1), \vec{W} = (-1, 0)\}$
<i>Mobility</i> :	$\mathcal{B} := \{k(0, 1), k(1, 1), k(1, 0), k(1, -1),$ $k(0, -1), k(-1, -1), k(-1, 0), k(-1, 1) : k = 1\}$
<i>Perception</i> :	$\mathcal{W} := \{K := \{k_t\}, L := \sum_{t=1}^T l_t, V := \{x_i\},$ $H_i := \sum_{t=1}^T h_{it} : i = 1, \dots, j, \dots, N\}$
<i>Processing</i> :	$\mathcal{F} := \{\alpha, \delta, \zeta, \dots\}; \quad \mathcal{K} := \{\tilde{\alpha}   x, \tilde{\delta}   y, \dots\}$
<i>Semantics</i> :	<b>none</b>
<i>Syntax</i> :	<b>none</b>

Table 2: Collection of *SCIP*-systemic properties and capabilities. Note, that there is no knowledge of *Syntax* or *Semantics*.

<i>SCIP</i> – Environment – properties	
<i>SCIPE</i> :	$\{\mathcal{R}_E, \mathcal{R}_O, \mathcal{R}_R, \mathcal{D}, \ell_{\mathcal{R}}\}$
<i>Reference plane</i> :	$\mathcal{R}_E := \{P_{n,m} : \exists R_{n,m} \in \mathcal{R}_R(n_0, m_0, g), P_{n,m} \in R_{n,m}\}$
<i>Reference objects</i> :	$\mathcal{R}_O := \{\square, \triangle, \circ, \dots\}$
<i>Referential grid</i> :	$\mathcal{R}_R(n_0, m_0, g) := \{R_{n,m} = [(n-1)g, ng] \times [(m-1)g, mg] :$ $1 \leq n \leq n_0, 1 \leq m \leq m_0, g > 0\}$
<i>Direction</i> :	$\mathcal{D} := \{\vec{N} := (0, 1), \vec{O} := (1, 0), \vec{S} := (0, -1), \vec{W} := (-1, 0)\}$
<i>Object – location</i> :	$\ell_{\mathcal{R}} : \mathcal{R}_O \rightarrow \mathcal{R}_E$

Table 3: Collection of *SCIP*-environmental properties and definitions with environmental *directions* corresponding to systemic *orientations*.

<i>SCIP</i> – Language material – entities
<b>Word:</b> the sign-object identified as vocabulary element (type) whose occurrences in (linear) sets of sign-objects (tokens) are countable
<b>Sentence:</b> 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
<b>Text:</b> 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
<b>Corpus:</b> 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: Restrictions on concepts of language material employed to define the language environment entities.

- ▷ 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 surroundings (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.

## 4.1 Positions and locations

The experimental setting consists of a two dimensional environment with some objects at certain places (Fig. 4) that a *SCIP*-system will have to identify on 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 developed so far 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*.

Illustrating an example situation, the reference plane (Fig. 4) shows two object-locations. These have (automatically) been described in a corpus of language 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.

## 4.2 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.

Syntax of Textgrammar	
$T(\text{ext})$	$:= \{S_i \mid S_i \rightarrow S_{i+1} : \mathcal{B} \wedge \{KP_1, KP_2\} \in S_i \wedge \{KP_1, KP_2\} \in S_{i+1} \wedge \forall KP_j \in S_i \cup S_{i+1} ; j = 1, 2 \ ; \ i = 1, \dots, I\}$
$\mathcal{B}$	$:= \{k(0, 1), k(1, 1), k(1, 0), k(1, -1), k(0, -1), k(-1, -1), k(-1, 0), k(-1, 1) : k = 1\}$
$S_i$	$\rightarrow$ NP VP
NP	$\rightarrow$ N
VP	$\rightarrow$ V PP
PP	$\rightarrow$ HP KP
N	$\rightarrow$ A $\langle$ <i>triangle</i>   <i>square</i>   <i>circle</i> $\rangle$
V	$\rightarrow$ <i>lies</i>
HP	$\rightarrow$ $\langle$ <i>extremely</i>   <i>very</i>   <i>rather</i> $\rangle$ $\langle$ <i>near by</i>   <i>far away</i> $\rangle$
KP	$\rightarrow$ $\langle$ <i>on the left</i>   <i>on the right</i> $\rangle$   $\langle$ <i>in front</i>   <i>behind</i> $\rangle$

Table 5: Syntax of textgrammar for the generation of strings of correct sentences, constituting (situational) descriptions of possible system-*position* and object-*location* relations (for moving systems only).

Semantics of Textgrammar					
<b>Core-predicates (KP)</b>					
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 $x, y$	<i>in front</i>	<i>behind</i>	EAST $x, y$	<i>in front</i>	<i>behind</i>
<i>on the left</i>	>m, <n	>m, >n	<i>on the left</i>	<m, <n	>m, <n
<i>on the right</i>	<m, <n	>m, <n	<i>on the right</i>	<m, >n	>m, >n
SOUTH $x, y$	<i>in front</i>	<i>behind</i>	WEST $x, y$	<i>in front</i>	<i>behind</i>
<i>on the left</i>	<m, >n	<m, <n	<i>on the left</i>	>m, >n	<m, >n
<i>on the right</i>	>m, >n	<m, >n	<i>on the right</i>	>m, <n	<m, <n
<b>Hedge-predicates (HP)</b>					
as distances between system-position and object-location (under <i>crisp</i> - and <i>fuzzy</i> -interpretation): for numbers of grid-points $ x - n $ and $ y - m $					
1 2 3 4 5 6 7 8 9 10	Crisp <sub>1.0</sub>	Fuzzy <sub>1.1</sub>	1 2 3 4 5 6 7 8 9 10		
1 1 0 0 0 0 0 0 0 0	extremely	nearby	1 1 .7 .2 0 0 0 0 0 0		
0 0 1 1 0 0 0 0 0 0	very	nearby	.2 .7 1 1 .7 .2 0 0 0 0		
0 0 0 0 1 0 0 0 0 0	rather	nearby	0 0 .2 .7 1 .7 .2 0 0 0		
0 0 0 0 0 1 0 0 0 0	rather	faraway	0 0 0 .2 .7 1 .7 .2 0 0		
0 0 0 0 0 0 1 1 0 0	very	faraway	0 0 0 0 .2 .7 1 1 .7 .2		
0 0 0 0 0 0 0 0 1 1	extremely	faraway	0 0 0 0 0 0 .2 .7 1 1		

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

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	2	2	2	2	4	4	4	4	1	1	1	1	1	1	0	0	0	
	0	0	2	2	2	2	4	4	4	4	1	1	1	1	1	1	0	0	0	
	0	0	2	2	2	2	4	4	4	4	1	1	1	1	1	1	0	0	0	
	0	0	5	5	5	5	2	2	2	2	1	1	1	1	1	1	0	0	0	
	0	0	5	5	5	5	2	2	2	2	1	1	1	1	1	1	0	0	0	
	0	0	5	5	5	5	2	2	2	2	1	1	1	1	1	1	0	0	0	
R	0	0	4	4	4	4	2	2	2	2	1	1	1	1	1	1	0	0	0	
I	0	0	4	4	4	4	2	2	2	2	1	1	1	1	1	1	0	0	0	
G											▽									
H	0	0	1	1	1	1	1	1	1	1	6	6	3	3	3	3	0	0	0	
T	0	0	1	1	1	1	1	1	1	1	6	6	3	3	3	3	0	0	0	
	0	0	1	1	1	1	1	1	1	1	3	3	7	7	5	5	0	0	0	
	0	0	1	1	1	1	1	1	1	1	3	3	7	7	5	5	0	0	0	
	0	0	1	1	1	1	1	1	1	1	3	3	4	4	4	4	0	0	0	
	0	0	1	1	1	1	1	1	1	1	3	3	4	4	4	4	0	0	0	
	0	0	1	1	1	1	1	1	1	1	3	3	4	4	4	4	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	
I N F R O N T																				

Table 7: Intermediate representation matrix  $\mathbf{Endo} - \mathbf{1}_{i,j}$ ;  $1 = 20, j = 20$  which determines regions of likely object locations around the centrally positioned system ( $i, j = 10$ ) by number of marks that grid squares received according to pairs of (*crisply* 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	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: Representation matrix  $\mathbf{Endo} - \mathbf{2}_{n,m}$ ;  $n=10, m=10$  showing regions of object location likelihood as determined by superimposing (line- and column-wise) locality patterns assembled from value distributions in the intermediate representation matrix  $\mathbf{Endo} - \mathbf{1}_{i,j}$ ;  $i=20, j=20$ .

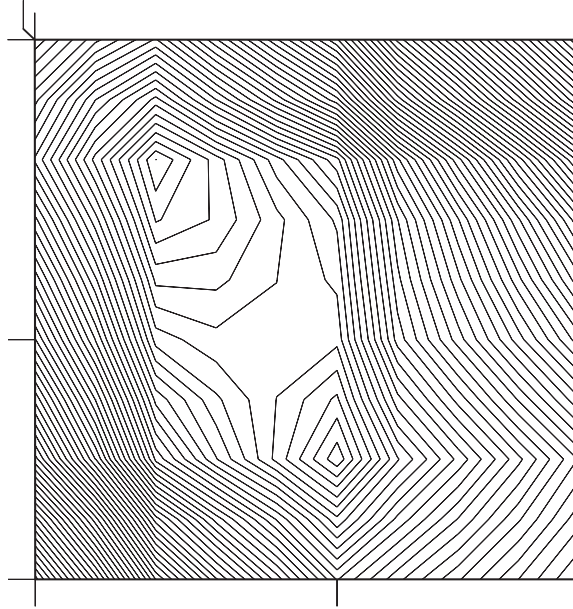


Figure 3: External 2-dim-image of the *SCIP* system's *endo*-view showing regions of potential object locations under *crisp*<sub>1.0</sub> hedge interpretation as computed from the training corpus of texts describing these locations relative to system positions in the reference plane.

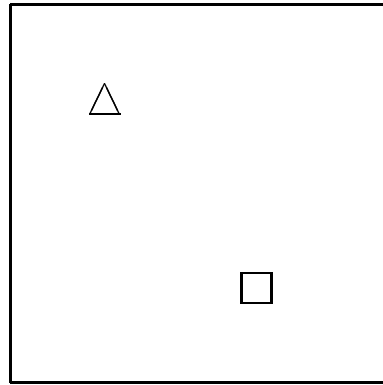


Figure 4: Reference plane with location of objects (  $\Delta$  and  $\square$  ) propositionally described by texts in the training corpus.

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 (Kohonen 1989) [2] or—with comparable results—average linkage cluster analysis [8] allows to identify structurally adjacent word-types (like object label and predicate label candidates) [13],
- ▷ 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 independent 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 referential 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

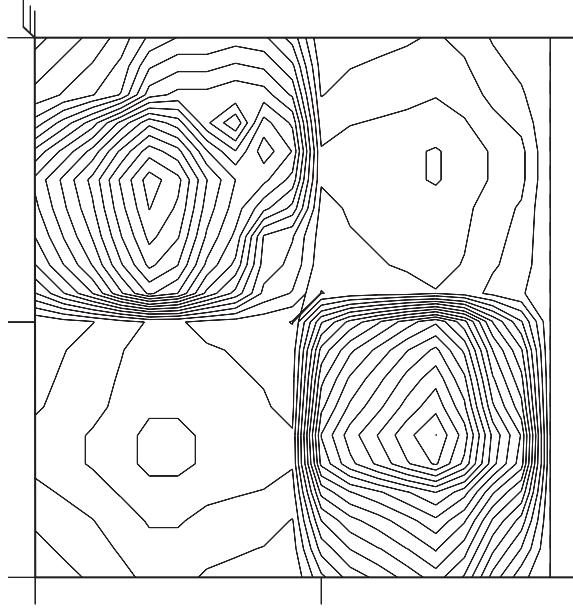


Figure 5: External 2-*dim*-image of the *SCIP* system's *endo*-view showing regions of potential object locations under *fuzzy*<sub>1.1</sub> hedge interpretation.

transform which results in a new mapping  $Endo2_{m,n}$  (Tab. 8) according to the summation equation

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

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) scattergram of  $Endo2$  gives an overall picture of even referential likelihood by *isoreferentials* denoting potential object locations quite clearly, under *crisp*<sub>1.0</sub> (Fig. 3) and under *fuzzy*<sub>1.1</sub> interpretation (Fig. 5). The corresponding 3-dimensional profile representations of the same patterns (sl Fig. 6 and sl Fig. 7) show in an even more detailed illustration the higher referential resolution which *fuzzy* interpretations of descriptive hedged core predications gain over *crisp* ones.

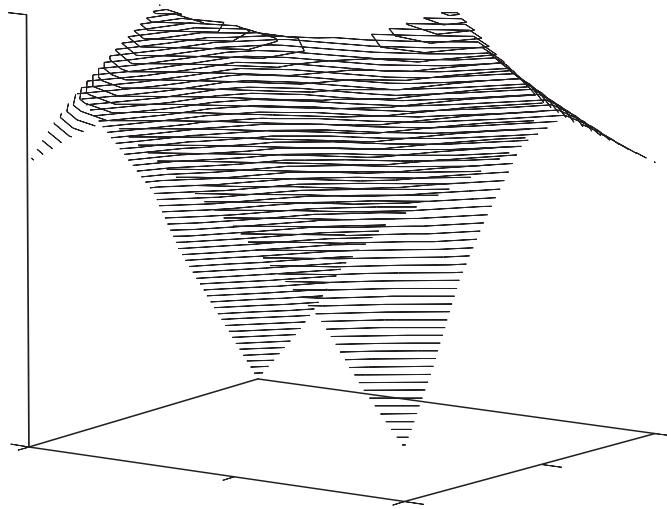


Figure 6: 3-*dim*-image of the *SCIP* system's *endo*-view of its *reality* showing highest potentials for object locations under *crisp* hedge interpretation, as computed from the training corpus of texts describing these locations relative to system positions in the reference plane.

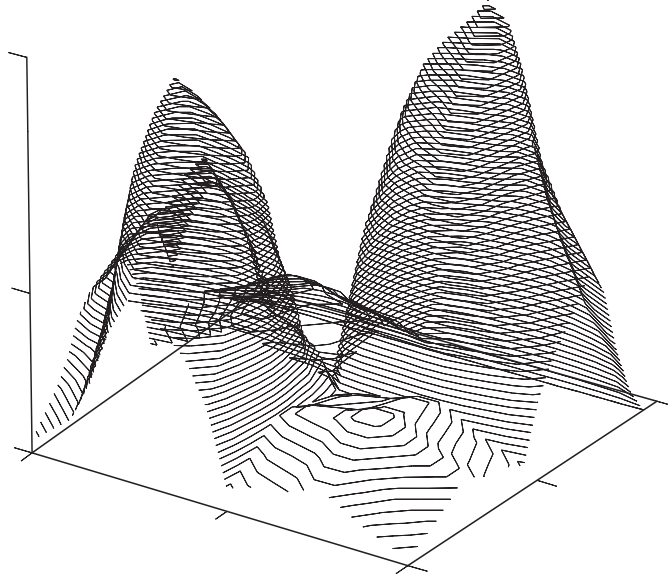


Figure 7: 3-dim-image of the *SCIP* system's *endo*-view of regions showing highest potentials for object locations under *fuzzy*<sub>1.1</sub> hedge interpretation, as computed from the training corpus of texts describing these locations relative to system positions in the reference plane.

## 5 Conclusion

The paradigm of agentic systems seems to be particularly suited for any multivariate form of dynamic interaction that leaves traces of and is dependent on the results of such mutually triggered activity. Natural language communication certainly is an example for such a phenomenon whose traces in the form of texts have to be actualized in order to let the processes believed to be responsible for their production be inverted and experienced as understanding. The concepts of *situation* and *language game* have proved to be seminal in elucidating the structure and compounds involved in the constitution of meaning.

The development of the above model of *semiotic cognitive information processing*, however, will have to be elaborated in at least three directions before the present *SCIP*-systems may justifiably be named *semiotic agents*:

- ▷ As the language environments the systems are exposed to sofar have been generated on a stratum which is external to modeling the semiotic process of meaning constitution (namely to allow for the testing of semiotically derived structures against the linguistically determined syntactic and semantic structures), it is highly desirable to have more than one system be modelled. Having two would provide for the possibility to let one part of the language environment be tied to one system, the other part to the other with the understanding that each system is producing language representations of their mutual situational embedding forming the respective other system's language environment.
- ▷ As the processing of language material by the system sofar is not timed other than by the implemented model structures' internals, the exchange of language material produced by the one system and received by the other calls for the introduction of time cycles corresponding to or even emulating the frequency of interchange in processing mutual environmental (language) data. This is a prerequisite for the re-construction of movement and the differentiation of subject- or object-movement.
- ▷ Allowing for more than one *SCIP*-system in a single reference plane, each producing his own and processing the others' language material generated, it will prove to be necessary to introduce a non-language data channel in order to derive situational as well as physical correlates for language representations. Additional receptive channels of the systems may open up new ways of iterating language representations towards mutual reliability (correctness), towards differencing

between self and other (subject-object boundary), and even towards the more efficient structuring of environmental and systemic data (space-time relation) in accordance with changing states of the manifold, constituted by the systems' adaptation to language representations (learning), to the resolutive power of the language material employed (understanding), and the interplay of cognitive and representational activity with the dynamics of what is cognized and represented (knowledge).

Some of these ideas are being followed and discussed just now in an very early state of development however which still is characteristic of all *computational semiotics* sofar.

## References

- [1] J. Barwise / 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] C. Peirce. Pragmatism in Retrospect: A Last Formulation. In J. Buchler, editor, *The Philosophical Writings of Peirce*, pages 269–289. Dover, New York, 1906.
- [4] 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.
- [5] 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 / K. Fuchi, editors, *COLING-80 8th International Conference on Computational Linguistics*, pages 76–84, Tokyo, 1980. International Committee on Computational Linguistics, ICCL.
- [6] B. B. Rieger. Feasible Fuzzy Semantics. On some problems of how to handle word meaning empirically. In H. Eikmeyer / 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.
- [7] B. B. Rieger. Fuzzy Representation Systems in Linguistic Semantics. In R. Trappl / N. Findler / W. Horn, editors, *Progress in Cybernetics and Systems Research*, Vol. XI, pages 249–256. McGraw-Hill Intern., Washington/ NewYork/ London, 1982.
- [8] 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.
- [9] B. B. Rieger. Semantic Relevance and Aspect Dependency 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.
- [10] B. B. Rieger. Lexical Relevance and Semantic Disposition. On stereotype word meaning representation in procedural semantics. In G. Hoppenbrouwes / P. Seuren / T. Weijters, editors, *Meaning and the Lexicon*, pages 387–400. Foris Publications, Dordrecht, 1985.
- [11] B. B. Rieger. Stereotype representation and dynamic structuring of fuzzy word meanings for contents-driven semantic processing. In J. Agrawal / P. Zunde, editors, *Empirical Foundations of Information and Software Science*, pages 273–291. Plenum Press, NewYork/ London, 1985.



- [12] B. B. Rieger. Distributed Semantic Representation of Word Meanings. In J. D. Becker / I. Eisele / F. W. Mündemann, editors, *Parallelism, Learning, Evolution. Evolutionary Models and Strategies, WOPLOT-89*, Lecture Notes in Artificial Intelligence 565, pages 243–273. Springer, Berlin/ Heidelberg/ NewYork, 1991.
- [13] B. B. Rieger. Situation semantics and computational linguistics: towards informational ecology. a semiotic perspective for cognitive information processing systems. In K. Kornwachs / K. Jacoby, editors, *Information. New Questions to a Multidisciplinary Concept*, page [im Druck], Berlin, 1995. E.W.-Heraeus-Stiftung, Deutsche Physikalische Gesellschaft, Deutsche Gesellschaft für Systemforschung, Akademie.
- [14] B. B. Rieger / C. Thiopoulos. Semiotic Dynamics: a self-organizing lexical system in hypertext. In R. Köhler / B. Rieger, editors, *Contributions to Quantitative Linguistics. Proceedings of the 1st Quantitative Linguistics Conference – QUALICO-91*, pages 67–78. Kluwer Academic Publishers, Dordrecht, 1993.
- [15] H. A. Simon. *The Sciences of the Artificial*. MIT Press, Cambridge, MA, 2nd2nd edition, 1982.
- [16] T. Winograd. *Language as a Cognitive Process. Vol. I: Syntax*. Addison-Wesley, Reading, MA, 1983.
- [17] L. Wittgenstein. *The Blue and Brown Books. (Ed. by R.Rhees)*. Blackwell, Oxford, 1958.
- [18] 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.
- [19] 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.