Situation Semantics and Computational Linguistics:
towards Informational Ecology.*

A semiotic perspective for cognitive information processing systems.

Burghard B. Rieger**
Dept. of Computational Linguistics — University of Trier
D-54286 TRIER, Germany
rieger@ldv01.uni-trier.de

Abstract

Other than the clear-cut realistic division between information processing systems and their surrounding environments employed sofar in models of natural language understanding by machine, it is argued here that a semiotic approach based on an ecological understanding of informational systems is feasible and more adequate. Characterizing such systems’ performance in general and the pragmatics of communicative interaction by real language users in particular, a critical evaluation of cognitive approaches in knowledge-based computational linguistics together with the seminal notions of situation and language game are combined to allow for a procedural modelling and numerical reconstruction of processes that simulate the constitution of meanings and the interpretation of signs prior to any predicative and propositional representations which dominate traditional formats in syntax and semantics. This is achieved by analysing the linear or syntagmatic and selective or paradigmatic constraints which natural language structure imposes on the formation of (strings of) linguistic entities. A formalism with related algorithms and test results of their implementation are produced in order to substantiate the claim for a model of a semiotic cognitive information processing system (SCIPS) that operates in a language environment as some meaning acquisition and understanding device.

1 Framing the background

A man, viewed as a behaving system, is quite simple. The apparent complexity of his behavior over time is largely a reflection of the complexity of the environment in which he finds himself ... provided that we include in what we call man’s environment the cocoon of information, stored in books and long-term memory, that man spins about himself.1

As long as the concept of meaning was conceived as some independent and pre-existing entity, very much like that of physical objects in the real world, meaning appeared to

---


**The author is indebted to discussion of central ideas of this paper with Petra Badry, Kathrin Gieseking, 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.

1Simon (1982), p. 127
be analysable and representable accordingly, i.e. as sets of features and as entries in a knowledge base whose elements and their mutual relations stand for certain other entities which are taken to be represented by these data base entries due to the labels or signs attached whose functions or meanings are known and understood. Thus, semantic net models depict meanings as a system of relations between signs and designata whose representations as labeled nodes and arcs make up structured sets of elements whose meanings are signalled by language symbols attached to them. Within this frame of semantics as a theory about the nature of reality (Lakoff 1988) the problem of semantic grounding cannot be addressed, the related questions of how a symbolic expression may serve as a label and on what grounds it gets associated with a node in order to let this node be understood to stand for the entity (meaning or object) it is meant to represent, cannot even be posed. However, these questions have to be realized, explored, and eventually answered:

- it has to be realized that there are certain entities which—beyond their physical existence in the world—are (or become) signs and have (or acquire) interpretable meanings that can be understood in the sense of knowing what the signs signify and stand for (whereas other entities in the world do not).
- it has to be explored how such (semiotic) entities may be constituted and how the meaning relation be established on the basis of which observable regularities (uniformities), controlled by what constraints (syntagmatic and paradigmatic), and under which boundary conditions of situational configurations for communicative interactions (pragmatics).
- it has to be answered why some entities may signify others by serving as labels or representations for them (or rather by the labeling and representational functions these entities serve), instead of being just named according to their positions, load values and/or patterns exhibited in a representational system of semiotic/non-semiotic entities.

In doing so, a semiotic paradigm based upon some ecological concepts of the theory of information processing systems will be followed which hopefully may allow to avoid (if not to solve) a number of spin-off problems, which originate in the traditional distinction and/or the methodological separation of the meaning of a language’s term (or rather, its format of representation) from the way it is employed in discourse. It appears that failing to mediate between these two sides of natural language semantics, phenomena like creativity, dynamism, efficiency, vagueness, and variability of meaning—to name only the most salient—have fallen in between, have stayed (or be kept) out of the foci of interest, or have been overlooked altogether (Rieger 1991a, 1991b). Moreover, classical formal theoretic approaches to natural language semantics which are based upon constraints of propositional constructions and confined to the sentence boundary are badly in want of operational tools that may allow to bridge the gap between formal theory of language description (competence) and empirical analysis of language usage (performance) felt to be responsible for the unwarranted abstractions from fundamental properties of natural languages.
1.1 An ecological approach to the semiotics of representation

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 experiencial results of environmental adaptation only biogenetically 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 (intraneous) 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. 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—for short: space-time-dispensation—is not only conditional for the possible distinction of (mutually and relatively independent) systems from their environments, but establishments 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

---

2. According to standard theory there is no direct genetic coding of experiencial 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.

3. Simon's (1982) 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. Primarily concerned with where to place the boundary, he does not seem to see its placing in need to be justified or derived as a consequence of some possibly representational processes we call semiotic. As will become clear in what follows, Simon's distinction of inner (memory structure) and outer (world structure) environments misses the special quality of language signs whose twofold environmental embedding (textual structure) cuts across that distinction, resolving both in becoming representational for each other.
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 es-

tablish according to its own evolved adaptedness or dispositions (i.e. innate and acquired
structuredness, processing capabilities, represented knowledge).

The actualisation of environments, however, does not merely add to the amount of experi-

ciential 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 re-
alized 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 experien-
cial 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 experi-
mental 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 actual-
ized 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 adap-
tive information processing immensely, will have to be constrained simultaneously by dy-
namic structures, corresponding to knowledge. The built-up, employment, and modifica-
tion of these structural constraints 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 learning) 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 called semiosis. 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).

---

4What Simon (1982) calls Memory in accordance with his questioning of the inner-outer-distinction of
cognitive systems and their environments.

5By semiosis I mean [...] an action, or influence, which is, or involves, a cooperation 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. (Peirce 1906, p. 282)
1.2 A semiotic approach to cognitive information processing

For the majority of researchers in knowledge representation and natural language semantics the common ground and widely accepted frame for their modelling may be found in the dualism of the rationalistic tradition of thought as exemplified in its matter-mind notion of an independent (objective) reality and some (subjective) conception of it. According to the realistic view, the meaning of any portion of language material (like e.g. discourse, utterance, word(token), morph, phone, etc.) is interpreted as being an instantiation of (or as partly derivable from) certain other entities, called linguistic categories (like e.g. text, sentence, word(type), morpheme, phoneme, etc.), with the understanding that these categories structure natural language material according to their compositional functions. It is by these functions that language material (strings of terms) appear to be composed of linguistic entities (aggregates of categories) to form structures and it is also by these functions that the quality of language structures (having meaning) is conceived as being part of both, the physical reality of language material and the semantic significance of linguistic signs. Illustrating this twofold membership are the graph-theoretical formats which have become standard representations for natural language meanings, both as relational structure and as referential denotation. Thus, relating arc-and-node configurations with sign-and-term labels in graphs like trees and nets appears to be but another aspect of the traditional mind-matter-duality according to which a realm of meanings is presupposed very much like the assumption of the pre-given structures of the real world related by signs. Accepting this duality has neither allowed to explain where the structures or where the labels come from nor how their mutual relatedness as meanings of signs can be derived. The emergence of the meaning relation, therefore, never occurred to be in need of some explanatory modelling because the existence of signs, objects and meanings were taken for granted and hence seemed to be out of all scrutiny. Under this presupposition, fundamental semiotic questions of semantics simply did not come up, they have hardly been asked yet, and are still far from being solved.

Following an attempt to classify approaches in cognitive science, we may discern four categories of approaches in modelling cognition:

- the cognitive approach presupposes the existence of the external world, structured by given objects and properties and the existence of internal representations of (fragments of) this world, so that cognitive systems’ (observable) behaviour of action and reaction may be modelled by processes operating on these structures;

- the associative approach is described as a dynamic structuring based on the model concept of self-organization with the cognitive system constantly adapting to changing environmental conditions by modifying its internal representation of them.

Whereas both these approaches apparently draw on the traditional rationalistic paradigm of mind-matter-duality—static the former, dynamic the latter—in presupposing the external world structure and an internal representation of it, the third and fourth category do not:

- the enactive approach may be characterized as being based upon the notion of structural coupling. It disposes of the assumption—whatever else the semantics of coupling might suggest—of an external world and a system’s internal representation of it, but

---

6 There were only the first three of these four approaches distinguished by Varela/Thompson/Rosch (1991).
considers instead some unity of structural relatedness to be fundamental of—and the (only) condition for—any abstracted or acquired duality that philosophical realism has in the past (or might in future) come up with;

▷ the semiotic approaches focus on the notion of semiosis and may be characterized by the process of enactment too, supplemented, however, by the representational impact.

It is considered fundamental to the distinction of e.g. cognitive processes from their structural results which—due to the traces these processes leave behind—will emerge in some form of knowledge whose different representational modi comply with the distinction of internal or tacid knowledge (i.e. memory) on the one hand and of external or declarative knowledge (i.e. texts) on the other.

According to these categories of cognitive modelling, computational semiotics can be characterized to be aiming at the dynamics of modelling natural language meaning within the frame of ecological information processing.

Taking human beings as the most efficient semiotic cognitive information processing systems (SCIPS) whose outstanding sign and symbol manipulation and understanding capabilities are but a consequence of the very efficient knowledge organisation, representation, and modification processes they apply, then the observable structures resulting from these processes, namely natural language discourse provides a cognitively interesting meaning representation system whose outstanding structuredness in the aggregated form of text corpora from communicative situations may serve as a guideline (Rieger 1977) rarely followed in research yet. In doing so, however, it will be necessary to pass on from traditional approaches in linguistics proper analysing introspectively the propositional contents of singular sentences as conceived by ideal speakers/writers to semiological approaches based upon the empirically well founded observation and rigorous mathematical description of global regularities in masses of texts produced by real speakers/writers in actual situations of either performed or intended communication. Only such new approaches will give a chance to reconstruct procedurally both, the significance of entities and the meanings of signs as a function of a first and second order semiotic embedding relation of situations (or contexts) and of language games (or cotexts) which corresponds to the two-level actualisation of SCIPSs.

2 Opening the topic

It has been outlined above that any 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

---

\[7\] Whereas tacid knowledge will not be represented other than by the immediate system-environment’s corresponding status, explicit knowledge is bound to acquire some formal properties in order to be representable and become part of mediate system-environments. Natural languages obviously provide these formal properties—as identified by research in linguistic competence (knowledge and acquisition of language)—whose enactment—as investigated in studies on natural language performance (production and understanding of texts)—draws cognitively on both bases of (explicit and tacid) knowledge.
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\(^8\) as soon as a semiotic line of approaches to cognition will be followed.

It is the semiotic foundation (Rieger 1985e, 1989) that adds an ecological dimension to cognitive models (Rickheit/Strohner 1993) of natural language processing extending them explicitly to comply with those conditions which sign quality object domains produce, whether in processes of natural language performance (intention, production, reception, and understanding of discourse) or in the (explanatory, derivational, or procedural) modelling of these processes. Thus, for semiotic models the ecological paradigm (Bateson 1979, Maturana 1978, Maturana/Varela 1980) essentially translates to an obligation according to which the entities any system model starts with are to be kept (conceptually) as simple and (in number) as small as possible in order to avoid presupposing complex entities (structures or processes of some kind) where these could also be derived or their emergence be simulated by the model. This obligation applies very much to natural language meaning and in the case of SCIPSs boils down to the necessity to include the situational context of language performance from the very start of the modelling process instead of extending or modifying some model result by adding a set of contextual parameters to it. Therefore, the present approach is based upon a phenomenological (re-)interpretation of the formal concept of situation 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 semiosis.

2.1 Semantics: the notion of situation

According to Situation Semantics (Barwise/Perry 1983), 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

---

\(^8\)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 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—Zadeh (1965), (1975), (1981)—and their logical derivate were embraced 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.
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.

In semiotic sign systems like natural languages, such uniformities also appear to be signalled more basically 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 across 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.

Owing to BARWISE/PERRY’s new approach—and notwithstanding its traditional (mis)-conception as duality (i.e. the independent sign-meaning-view) of an information processing system on the one hand which is confronted on the other hand with a prefixed external reality whose accessible fragments are to be recognized as its environment—this notion of situation proves to be pivotal for an empirical extension to their theory of semantics. Not only can it be employed to devise a procedural model for the situational embeddedness of cognitive systems as their primary means of mutual accessability (RIEGER/THIOPOULOS 1989, RIEGER 1991b), but also does it allow to capture the semiotic unity as specified by the idea of language games.

2.2 Communication: the notion of language game

The WITTGENSTEINian notion of language game9 is the core of a semantic theory which identifies the meanings of words with the regularities these words are used by interlocuters in situations of communicative interaction (Wittgenstein 1969). Operationalizing this notion according to which a great number of texts analysed for the terms’ usage regularities will reveal essential parts of the concepts and hence the meanings conveyed has produced some evidence that an analytical procedure appropriately chosen could well be identified

---

9"There are ways of using signs simpler than those in which we use the signs of our highly complicated everyday language. Language games are the forms of language with which a child begins to make use of words. […] We are not, however, regarding the language games which we describe as incomplete parts of a language, but as languages complete in themselves, as complete systems of human communication.” (Wittgenstein 1958, pp.17 and 81; [my italics])
also with solving the representational task if based upon the universal constraints known to be valid for all natural languages.

Trying to model language game performance along traditional lines of cybernetics by way of, say, an information processing subject, a set of objects surrounding it to provide the informatory environment’s input, and some positive and/or negative feedback relations between them, would hardly be sufficient to capture the cognitive dynamism that enactive systems of knowledge acquisition and meaning understanding are capable of due to their elastic constraints, i.e. the restrictions which hold for the agglomerative or syntagmatic and the selective or paradigmatic structuring of language units in discourse.

The philosophical concept of language games as specified by the formal notion of situations, not only allows for the formal identification of both, the (internal) structure of the cognitive subject with the (external) structure of its environment. Being tied to the observables of actual language performance, communicative language usage 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 communicative interaction by use of language items.

As has been outlined in some detail elsewhere (Rieger 1989, 1990; Rieger/Thiopoulos 1993), 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.

Implemented, such a system eventually may well lead to something like machine-simulated cognition, letting information be processed as a means of constituting a (system-dependent) view of reality from the system’s (linguistically structured) environment. It is argued here that its faculty of representing reality may be reconstructed as a complex function of those regularities which elements of communicative sign usage exhibit on different levels of linguistic description from morphs to words and discourse.

\[\text{Feedback is a method of controlling a system by reinserting into it the results of its past performance. If these results are merely used as numerical data for the criticism of the system and its regulations, we have the simple feedback of control engineers. If, however, the information which proceeds backward from the performance is able to change the general method and pattern of performance, we have a process which may well be called learning.”} \text{(Wiener 1956, p.60)}\]
3 Setting the stage

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 (Winograd 1983) 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 (Rieger 1981) 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.

Under the notion of lexical relevance and semantic disposition (Rieger 1985b, 1985d) a corresponding meaning representation 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 Language material: 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 (called $\alpha$- and $\delta$-abstraction) on the set of fuzzy subsets of the vocabulary—providing the word-types’ usage regularities or corpus points—and on the set of fuzzy subsets of these—providing the corresponding meaning points as a function of those 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. 5) to specify intensities of co-occurring lexical items in texts, and a measure of similarity (or rather, dissimilarity) (Eqn. 8) to specify differing distributions of correlation values. Simultaneously, these measures may be interpreted semiotically as providing for set theoretical constraints or formal mappings (Eqns. 6 and 9) which model the meanings of words as a function of differences of usage regularities.

As a first mapping function $\alpha$ allows to compute the relational interdependence of any two lexical items from their textual frequencies. For any text corpus

$$K = \{k_t\} ; \; t = 1, \ldots, T$$  (1)
of pragmatically homogeneous discourse, having an overall length

$$L = \sum_{t=1}^{T} l_t ; \quad 1 \leq l_t \leq L$$

measured by the number of word-tokens per text, and a vocabulary

$$V = \{x_n\} ; \quad n = 1, \ldots, i, j, \ldots, N$$

of \(n\) word-types of different identity \(i, j\) whose frequencies are denoted by

$$H_i = \sum_{t=1}^{T} h_{it} ; \quad 0 \leq h_{it} \leq H_i$$

the modified correlation-coefficient \(\alpha_{i,j}\) 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

$$\alpha(x_i, x_j) = \frac{\sum_{t=1}^{T} (h_{it} - h^*_it)(h_{jt} - h^*_jt)}{\left(\sum_{t=1}^{T} (h_{it} - h^*_it)^2 \sum_{t=1}^{T} (h_{jt} - h^*_jt)^2\right)^{1/2}} ; \quad -1 \leq \alpha(x_i, x_j) \leq +1$$

where \(h^*_it = \frac{H_i}{L} l_t\) and \(h^*_jt = \frac{H_j}{L} l_t\).

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

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

where the tupels \(((x_{n,1}, \tilde{\alpha}(n, 1)), \ldots, (x_{n,N}, \tilde{\alpha}(n, N)))\) represent the numerically specified, syntagmatic usage regularities that have been observed for each word-type \(x_i\) against all other \(x_n \in V\) and can therefore be abstracted over one of the components in each ordered pair, thus, by \(\alpha\)-abstraction defining an element

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

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.

### 3.2 Language processing: distributed meaning representation

Considering \(C\) as representational structure of abstract entities constituted by syntagmatic regularities of word-token occurrences in pragmatically homogeneous discourse,

11
then the similarities and/or dissimilarities between these abstract entities will capture the corresponding word-types’ paradigmatic regularities. These can be modelled by the $\delta$-abstraction which is based on the numerically specified evaluation of differences between any two of such points $y_i, y_j \in C$. They will be the more adjacent to each other, the less the usages (tokens) of their corresponding lexical items $x_i, x_j \in V$ (types) differ. These differences may be calculated by a distance measure $\delta$ of, say, Euclidean metric.

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

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

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

$$\tilde{\delta} \mid y_n : C \rightarrow S; \{z_n\} =: S$$

where the tuples $\langle (y_{n,1}, \tilde{\delta}(n, 1)), \ldots, (y_{n,N}, \tilde{\delta}(n, N)) \rangle$ 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 again analogous to Eqn. 7, 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

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

By 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

$$\zeta : S \times S \rightarrow I$$

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

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 $z_n \in S$.

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 understood 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.

4 Sketching the plot

So far, the system of word meanings (lexical knowledge) has been modelled as a relational data structure whose linguistically labelled elements (meaning points) and their mutual distances (meaning differences) form a system of potential stereotypes. Although on first sight these points appear to be symbolic meaning representations it is worth mentioning that in fact each such point is determined by a fuzzy distribution of wordtype-value
pairs. Meaning representation via points in SHS is a matter of the position a point takes among others, and it is this position which interprets the lexical label attached to it, not vice versa. Therefore, based upon SHS-structure, the meaning of a lexical item may be described either as a fuzzy subset of the vocabulary, or as a meaning point vector, or as a meaning point’s topological environment delimiting the central point’s position indirectly as its stereotype (Rieger 1985a, 1985c).

This variability of representational formats complies with the semiotic notion of understanding and meaning constitution, according to which the SHS may be considered the core or base of a multi-level conceptual knowledge representation system (Rieger 1989). Essentially, it separates the format of a basic (stereotype) word meaning or concept representation from its latent relational forms of organization for particular cognitive purposes. Whereas the former may be thought of as a rather stable (or long-term), topologically structured (associative) memory, the latter can be characterized as a collection of structuring procedures which re-organize the memory data according to cognitive tasks to be solved under situational (or short-term) conditions.11

As we have separated cognitive processes from their resultant structures above12, so we may distinguish here between the long-term structure as an addressable representation of knowledge (stereotype or concept) and its short-term process in a situational embedding (employment or activation) with the semiotic implication that the structures depend on the processes and vice versa to let addressable representations emerge and cognitive processes be enacted. Thus, the duality of the inner-outer distinction or the system-environment opposition above13 may be mediated by processes operating on some common, basal representational structures14 whose efficient reorganisation can be modelled procedurally to result in a—more or less subjective—internal (or endo-)view the system develops, a n d in a—more or less objective—external (or exo-)view of the surrounding environment that constitutes reality.

To find out (and preferably be able to test) what of the structural information inherent in natural language discourse—defined a n d structured by the text analytical processes described above15—might be involved in mediating or constituting that duality, an experimental setting has been designed. It is based on the assumption that some deeper representational level or core structure—similar to the one modelled by SHS—might be identified which could be considered a common base for different notions of meaning developed by theories of referential and situational semantics as well as some structural or stereotype semantics.

---

11This notion corroborates and extends ideas also expressed within the theories of priming and spreading activation (Lorch 1982) allowing e.g. for the dynamic and task-oriented generation of paths (along which activation might spread) allowing priming to be modelled as cognitive function rather than one of its presupposed conditions.

12cf. p. 3

13cf. p. 3, footnote3

14Representational formats will be called basal if they can provide a frame for the formal unification of categorial-type, concept-hierarchical, truth-functional, propositional, phrasal, or whatever other representations.

15v. pp. 13
4.1 Experimental descriptions: positions and locations

For the purpose of testing semiotic processes of meaning constitution, the 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, undoubtedly the common basis of traditional meaning theories, will be used here only to control the format of the natural language training material, not, however, to determine the way it is to be processed.

Figure 2: Reference plane with location of objects (△ and □) propositionally described by texts in the training corpus which serves as language environment to the SCIP-System.

To give a general view of the approach first, the experimental setting is pictured by a mobile system in a two dimensional environment with some objects at certain places to identify. The system’s channels of perception to form its own or endo-view of its surroundings are extremely limited, and its ability to act (and react) is heavily restricted compared to natural or living information processing systems. What makes such an artificially abstracted system a semiotic one is that—whatever it might gather from its environment—it will not be the result of some decoding processes which would necessarily call for that code to be known by the system, but will instead be constituted according to the system’s (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).

This postulate, obviously, rules out immediately any traditional form of cognitive modellings which do not see the need of such a differentiation. They are satisfied instead with well established formalisms (in syntax, semantics, predicate logics) not only to distinguish, describe, and represent different levels of natural language structures but consider them also appropriate to determine and control the processing of what these levels are meant to model. Consequently, corresponding knowledge bases representations are conceived which cognitive processes are believed to operate on when employed to analyse and interpret...
**SCIP – System – properties**

<table>
<thead>
<tr>
<th><strong>SCIPS</strong></th>
<th>{O, B, W, F, K}</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Orientation</strong></td>
<td>( O := { \vec{N} = (0,1), \vec{O} = (1,0), \vec{S} = (0,-1), \vec{W} = (-1,0) } )</td>
</tr>
<tr>
<td><strong>Mobility</strong></td>
<td>( 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 } )</td>
</tr>
<tr>
<td><strong>Perception</strong></td>
<td>( W := { K := { k_t }, L := \sum_{t=1}^T l_t, V := { x_i }, ) ( H_i := \sum_{t=1}^T h_{it} : i = 1, \ldots, j, \ldots, N } )</td>
</tr>
<tr>
<td><strong>Processing</strong></td>
<td>( F := { \alpha, \delta, \zeta, \ldots }; \ K := { \tilde{\alpha}</td>
</tr>
<tr>
<td><strong>Semantics</strong></td>
<td>none</td>
</tr>
<tr>
<td><strong>Syntax</strong></td>
<td>none</td>
</tr>
</tbody>
</table>

Table 2: Collection of systemic properties.

**SCIP – Environment – properties**

<table>
<thead>
<tr>
<th><strong>SCIP E</strong></th>
<th>{RE, RO, RR, D, \ell_R}</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Referential plane</strong></td>
<td>( RE := { P_{n,m} : \exists R_{n,m} \in RR(n_0, m_0, g), P_{n,m} \in R_{n,m} } )</td>
</tr>
<tr>
<td><strong>Referential objects</strong></td>
<td>( RO := { \square, \triangle, \bigcirc, \ldots } )</td>
</tr>
<tr>
<td><strong>Referential grid</strong></td>
<td>( RR(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 &gt; 0 } )</td>
</tr>
<tr>
<td><strong>Direction</strong></td>
<td>( D := { \vec{N} = (0,1), \vec{O} = (1,0), \vec{S} = (0,-1), \vec{W} = (-1,0) } )</td>
</tr>
<tr>
<td><strong>Object – location</strong></td>
<td>( \ell_R : RO \rightarrow RE )</td>
</tr>
</tbody>
</table>

Table 3: Collection of environmental properties.

**SCIP – Language material – entities**

<table>
<thead>
<tr>
<th><strong>Word</strong></th>
<th>the sign-object identified as vocabulary element (type) whose occurrences in (linear) sets of sign-objects (tokens) are countable</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sentence</strong></td>
<td>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</td>
</tr>
<tr>
<td><strong>Text</strong></td>
<td>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</td>
</tr>
<tr>
<td><strong>Corpus</strong></td>
<td>the (non-empty) set of texts comprising descriptions of (any or all) factually possible system-object relations within a specified systemic and environmental setting</td>
</tr>
</tbody>
</table>

Table 4: Restrictions on concepts of language material entities.
\[
T(\text{ext}) := \{ S_i \mid S_i \rightarrow S_{i+1} : \mathcal{B} \land \{ KP_1, KP_2 \} \in S_i \land \\
\{ KP_1, KP_2 \} \in S_{i+1} \land \forall KP_j \in S_i \cup S_{i+1} ; \\
j = 1, 2 ; i = 1, \ldots, I\} \\
\mathcal{B} := \{ k(0, 1), k(1, 1), k(1, 0), k(1, -1), k(0, -1), \\
k(-1, -1), k(-1, 0), k(-1, 1) : k = 1 \}
\]

\[
S_i \rightarrow NP \ VP \\
NP \rightarrow N \\
VP \rightarrow V \ PP \\
PP \rightarrow HP \ KP \\
N \rightarrow A \langle \text{triangle} \mid \text{square} \mid \text{circle} \rangle \\
V \rightarrow \text{lies} \\
HP \rightarrow \langle \text{extremely} \mid \text{very} \mid \text{rather} \rangle \langle \text{near by} \mid \text{far away} \rangle \\
KP \rightarrow \langle \text{on the left} \mid \text{on the right} \rangle \mid \langle \text{in front} \mid \text{behind} \rangle
\]

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).

<table>
<thead>
<tr>
<th>Core-predicates (KP)</th>
<th>Hedge-predicates (HP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>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</td>
<td>as distances of system-position/object-location (crisp-interpretation): in numbers of grid-points (</td>
</tr>
<tr>
<td>\begin{tabular}{lll} \text{NORTH} ( x, y ) &amp; \text{in front} &amp; \text{behind} \ \text{on the left} &amp; \text{on the right} &amp; \text{on the left} \ \text{on the right} &amp; \text{on the right} &amp; \text{on the right} \ \end{tabular}</td>
<td>\begin{tabular}{lll} \text{sp/ol-distance} &amp; \text{nearby} &amp; \text{faraway} \ \text{extremely} &amp; \text{very} &amp; \text{rather} \ \end{tabular}</td>
</tr>
<tr>
<td>\begin{tabular}{lll} \text{SOUTH} ( x, y ) &amp; \text{in front} &amp; \text{behind} \ \text{on the left} &amp; \text{on the right} &amp; \text{on the left} \ \text{on the right} &amp; \text{on the right} &amp; \text{on the right} \ \end{tabular}</td>
<td>\begin{tabular}{lll} \text{1 and 2} &amp; \text{3 and 4} &amp; \text{5} \ \text{9 and 10} &amp; \text{7 and 8} &amp; \text{6} \ \end{tabular}</td>
</tr>
</tbody>
</table>

Table 6: Semantics (referential definitions) to identify true core- and hedge-predicates (under crisp interpretation) in correct sentences as generated for fixed (unchanged) object-locations and varying (changed) system-positions.
natural language expressions. Thus, cognitive modelling of language understanding has narrowed the general (and semiotic) scope of knowledge acquisition and knowledge based processing of signs events—via re-usable results in some representational format (memory)—to the formatting of (both world and linguistic) knowledge and the application of rules of syntax and semantics. Ascribing syntactic structures to sentences whose truth-functional interpretation is determined by a referential semantics which in turn seems to allow for the definition of predicate meanings according to properties believed to be observed, recognized, indentified, and named in the world around and external to the cognitive systems, is to favour a very particularized understanding of language understanding which is hardly tenable.

As it appears by no means convincing why some (well understood) formalisms representing certain functional results on different levels of (more or less arbitrary) abstraction should also provide the only levels or moduls to study and investigate the (yet enigmatic) processes related to (or even underlying) these (or other cognitive) functions in language processing, this hypothesis common to models of language understanding in cognitive linguistics is questioned.

To enable an intersubjective scrutiny, it is suggested here that 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. To achieve this, it has to be guaranteed that

- a corpus of pragmatically homogeneous texts providing the system’s environmental data is compiled as a collection of (natural language) expressions of true propositions denoting the system-object-relations according to some specified syntax and semantics (representing the exo-view or described situations), and
- the system’s internal picture of its surroundigs (representing the endo-view or discourse situations) which will (and can) be derived from this textual language environment other than by way of propositional reconstruction, i.e. without syntactic parsing and semantic interpretation of sentence structures.

Therefore, the exo-knowledge allowing the designers of the experimental setting to control the propositional encoding and decoding of environmental information in texts which the system in a specified environment would process, had to be kept strictly apart from and was essentially not to be included in the SCIP system’s endo-capacities. Thus, the system’s own non-propositional processing would allow to come up with some results which cannot be interpreted as mere repetitious reproductions of knowledge structures which the system had been endowed with externally, but which instead may be considered the system’s internal representation quite comparable (however different in format) to the exo-view of the environment.

4.2 Simulations: semiotic cognitive information processing

To start with, let the system’s environment be thought of as a referential plane with some (fixed) objects (Fig. 2) and let the SCIP system be conceived as a mobile abstract entity

---

16 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.

17 v. p. 15
moving (linearly) about that plane in a (limited) number of directions and at a certain pace. Then, for any point that the system may occupy in that referential plane there is (at least) one true proposition denoting the relation between the system’s position and the object’s location. Natural language expressions of such true propositions denoting factual position-location-relations (which will vary with the system moving) can easily be generated given a limited vocabulary, a simple syntax and semantics.

Figure 3: 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 formally specify and control the system’s environment or language input and to allow for a semantic interpretation and visible image of the internal structure of the system’s acquired knowledge.
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).

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 generation of referentially descriptive language material and its processing within the experimental SCIP setting. Whereas the given definitions of word and corpus suffice to specify the elementary and global dimensions of the language material for its non-propositional processing, the intermediate dimensions of sentence and text want further specification. A (very simple) phrase-structure grammar (Tab. 5) and a reference semantic (Tab. 6) with different (crisp or possibly fuzzy) interpretations serve the purpose. Together, they determine the generation of situational adequate texts that consist of true and correct sentences describing the location of an object relative to the position the system might take. Localities are denoted by simple local predicate expressions composed of cores like "on the left" or "on the right" and "in front" or "behind" together with hedges like "extremely" or "very" or "rather" and "nearby" or "far away" to form sentences like e.g. "A triangle lies rather nearby on the left." – "A square lies very far away in front." – etc. This leads to the schematic diagram (Fig. 3) of the general set up, the SCIP system, and its environment. It illustrates the situational components conditional for the external view of the environment or exo-reality—as specified (by Tab. 3)—mediated by referentially true and situational adequate textual descriptions (language training material) generated according to the syntax and semantics (Tab. 5 and Tab. 6) for the SCIP system—as specified (by Tab. 2)18—whose (language) perception, (cognitive) processing, and (internal) structuring will result in the dynamic built-up of its own or endo-view of the environment. How can it be tested and evaluated to correspond—at least partially—with the exo-view which also happens to be ours but may certainly not necessarily be identified with it?

### 5 Enacting the game

To give a start with an example situation, the reference plane is one with two object-locations as introduced (Fig. 2) above19 which can easily be identified to have the approximate coordinates $m \approx 2.5, n \approx 7.5$ for the triangle and $m \approx 6.5, n \approx 2.5$ for the square in a $10^2$ referential grid $R_R$ to be used here. These object locations have (automatically) been described in a corpus of language expressions comprising some 12 432 word

---

18 cf. p. 16  
19 v. p. 15
Figure 4: Dendrogram of internal SHS structure of hedged core predicate adjacencies as analysed from textual object location descriptions relative to system positions oriented south (a = extremely, s = very, z = rather, w = far away, n = nearby, Li = on the left, Re = on the right, Vo = in front, Hi = behind, Dreieck = triangle, Quadrat = square, liegt = lies).

tokens of 26 word types in 2,483 sentences and 684 texts generated according to the specified syntax and semantics\(^\text{20}\) for all possible system-positions and one orientation. 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, i.e. according to $W, \mathcal{F}$ and $\mathcal{K}$ as specified (by Tab. 2) and defined (by Eqns. 1–4, Eqns. 5, 8, 11 and Eqns. 6, 9) above\(^\text{21}\). 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. In the course of the processing, the two-level consecutive mappings (Tab. 1 and Fig. 1)\(^\text{22}\)

\(^{20}\)cf. Tab. 5 and Tab. 6, p.17

\(^{21}\)cf. p. 16 and pp. 10–12

\(^{22}\)v. p. 13 above
result in the semantic hyper space (SHS) structure (as declared by Eqns. 10 and 11)\(^2\) whose vector space intrinsic data structure can be made visible in a three stage process:

1. **first**, this comprises methods of Kohonen-maps (Kohonen 1989) or—with comparable results—cluster analysis (Fig. 4),
2. **second**, an intermediate representation of the system’s own oriented view (Tab. 8), and
3. **third**, a mapping which gives an orientation independent representation of the system’s view of its environment (Tab. 9) that can further be visualized.

<table>
<thead>
<tr>
<th>Crisp 1.0</th>
<th>DR</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP</td>
<td>1</td>
</tr>
<tr>
<td>extremely nearby</td>
<td>1.0</td>
</tr>
<tr>
<td>very nearby</td>
<td>0</td>
</tr>
<tr>
<td>rather nearby</td>
<td>0</td>
</tr>
<tr>
<td>rather far away</td>
<td>0</td>
</tr>
<tr>
<td>very far away</td>
<td>0</td>
</tr>
<tr>
<td>extremely far away</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 7: *Crisp* semantic definition of hedge predicates (HP) in membership-values of distance ranges (DR) measured by reference plane grid-points.

### 5.1 Structures of understanding (internal): semantic clusters

Operating on the SHS data or vector space matrix, well known cluster- or mapping-procedures will assemble most similar meaning points in agglomerative binary dendrograms or two-dimensional maps respectively for any one orientation. The linguistic labels of such pairs of pairs . . . of pairs of meaning points agglomerated separately in the dendrogram according to the cluster criterion of average linkage (Fig. 4) now apparently correspond to and depict the vectorial or distributional similarities as structural adjacencies detected by the algorithm. Thus, cutting accross the dendrogram at any similarity level will produce hedged core predicate label pairings like e.g. "rather far behind" and "very far behind" clustering with "extremely nearby on the right" and "very nearby on the right", never, however, will structural adjacency clusters be found that would pair orientationally impossible\(^2\) core predicate labels like "...on the right" with "...on the left" or "...in front" with "...behind". Even referentially implausible pairs of hedged core predicate labels like e.g. "extremely nearby on the left" and "rather faraway on the left" do not occur as structurally adjacent.

In a second step, pairs of hedged predicate clusters as produced by the dendrogram cuts may referentially be decoded and numerically be specified for an intermediate representational frame or working structure $Endo1$. With this end, the same (crisp) interpretations of hedged core predicate labels are used (Tab. 7) which also served to encode the denotational meaning in the propositional texts of the language training material. However,

\(^2\) cf. p. 12 above

\(^2\) This impossibility is derived from the formal definitions specified by the semantics (Tab. 6), not by our own knowledge or intuitions.
whereas the encoding was derived from the exo-viewed (or factual) object locations \((n, m)\) relative to changing system positions \((x, y)\) at any point in the referential plane (Fig. 2) yielding numerical distances \(|m - x|, |n - y|\) for some choice of referential grid, \(R_R(m, n)\); \(m=x=10, n=y=10\) according to the defined semantics of hedged core predicates, the decoding is now derived from the system’s own or endo-expectations visualized as a plane or matrix \(Endo_{1,i,j}\); \(i=20, j=20\) with the system fixed in central position relative to its directionally determined orientations of potential object-locations (Tab. 8).

It is this relation that now translates the structural adjacencies of meaning points identified by pairs of hedged core predicate labels. Their numerical hedge interpretation yields the distance values and their directional core interpretation determines the regions of object locations. The representational frame selected for these endo-viewed regions is to be found in overlapping referential locations allowing to denote everything that is ”on the left”, ”in front”, ”on the right” of the system or ”behind” it. These directional overlaps combine to form a coordinate system of four quadrants each with the referential grid’s cardinality, and the SCIP system in the center.

To illustrate this translation, the matrix \(Endo_{1,i,j}\); \(i=20, j=20\) (Tab. 8) contains the number of marks per grid point which are identified according to the corresponding distance values of clustered hedge predicate pairs, and which are distributed around the system’s central position \((i=10, j=10)\) for the southern orientation \((O\)-value of Tab. 2\). The profile of \(Endo_{1,i,j}\) (Fig. 5) allows quite clearly to see that it is but a 3-dim-translation of the 2-dim-structure of the cluster dendrogram (Fig. 4) with the base being formed by the core predicate adjacencies to denote potential referential areas from a centrally positioned system’s view.

The \(Endo_{1,i,j}\) data (Tab. 8) serves as base for the following third step of a line- and column-wise transform which results in a new mapping \(Endo_{2,m,n}\) (Tab. 9) according to the summation equation

\[
Endo_{2,m,n} = \sum_{i=m}^{m+10} \sum_{j=n}^{n+10} Endo_{1,i,j} \tag{12}
\]

The matrix \(Endo_{2,m,n}\) (Tab. 9) contains the data for the external observer’s representation of the endo-view that the system has computed from the training corpus of texts describing object locations relative to system positions in the reference plane (Fig. 2). This becomes evident when looking for highest \(m,n\)-values in the matrix which reveals 295 at \((3, 8)\) and 291 at \((6, 3)\) to be exactly or very near the object locations which were encoded in the training corpus of texts as generated to form the system’s language environment. A (two-dimensional) 2-dim-scattergram of \(Endo_{2}\) (Fig. 6) gives an overall picture as a pattern of polygons which connect points of even referential likelihood or so-called isoreferentials denoting potential object locations quite clearly, however fuzzy. The 3-dim-scattergram of the same data (Fig. 7) pictures the differences in reference potentials in a more obvious way.

5.2 Images of knowledge (external): isoreferentials

The fuzziness of this image is quite remarkable insofar as it does not concern the object locations themselves but rather the referential space around them allowing for their differentiation. This sort of holistic and indirect way of specification—as opposed to the
Table 8: **Endo1*<sub>i,j</sub>* showing possible regional object locations from system position (oriented south ▽) by sums of marks that grid points received according to value pairs of (*crisp*ly interpreted) hedged core predicate adjacencies.

<table>
<thead>
<tr>
<th>B E H I N D</th>
<th>L E F T</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 0 0 0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 0 2 2 2 2 4 4 4 4</td>
<td>1 1 1 1 1 1 0 0 0 0</td>
</tr>
<tr>
<td>0 0 2 2 2 2 4 4 4 4</td>
<td>1 1 1 1 1 1 0 0 0 0</td>
</tr>
<tr>
<td>0 0 2 2 2 2 4 4 4 4</td>
<td>1 1 1 1 1 1 0 0 0 0</td>
</tr>
<tr>
<td>0 0 5 5 5 5 2 2 2 2</td>
<td>1 1 1 1 1 1 0 0 0 0</td>
</tr>
<tr>
<td>0 0 5 5 5 5 2 2 2 2</td>
<td>1 1 1 1 1 1 0 0 0 0</td>
</tr>
<tr>
<td>0 0 5 5 5 5 2 2 2 2</td>
<td>1 1 1 1 1 1 0 0 0 0</td>
</tr>
<tr>
<td>0 0 5 5 5 5 2 2 2 2</td>
<td>1 1 1 1 1 1 0 0 0 0</td>
</tr>
<tr>
<td>0 0 5 5 5 5 2 2 2 2</td>
<td>1 1 1 1 1 1 0 0 0 0</td>
</tr>
<tr>
<td>0 0 5 5 5 5 2 2 2 2</td>
<td>1 1 1 1 1 1 0 0 0 0</td>
</tr>
<tr>
<td>0 0 2 2 2 2 4 4 4 4</td>
<td>1 1 1 1 1 1 0 0 0 0</td>
</tr>
<tr>
<td>0 0 2 2 2 2 4 4 4 4</td>
<td>1 1 1 1 1 1 0 0 0 0</td>
</tr>
<tr>
<td>0 0 2 2 2 2 4 4 4 4</td>
<td>1 1 1 1 1 1 0 0 0 0</td>
</tr>
<tr>
<td>0 0 2 2 2 2 4 4 4 4</td>
<td>1 1 1 1 1 1 0 0 0 0</td>
</tr>
<tr>
<td>0 0 5 5 5 5 2 2 2 2</td>
<td>1 1 1 1 1 1 0 0 0 0</td>
</tr>
<tr>
<td>0 0 5 5 5 5 2 2 2 2</td>
<td>1 1 1 1 1 1 0 0 0 0</td>
</tr>
<tr>
<td>0 0 5 5 5 5 2 2 2 2</td>
<td>1 1 1 1 1 1 0 0 0 0</td>
</tr>
<tr>
<td>0 0 5 5 5 5 2 2 2 2</td>
<td>1 1 1 1 1 1 0 0 0 0</td>
</tr>
<tr>
<td>0 0 5 5 5 5 2 2 2 2</td>
<td>1 1 1 1 1 1 0 0 0 0</td>
</tr>
<tr>
<td>0 0 0 0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 0 0 0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 0 0 0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 0 0 0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0 0 0 0</td>
</tr>
</tbody>
</table>

Table 9: **Endo2*<sub>m,n</sub>* showing regions of object location likelyhood computed for each grid-point *m,n* by superimposing locality patterns from **Endo1*<sub>i,j</sub>* value.

<table>
<thead>
<tr>
<th>N O R T H</th>
</tr>
</thead>
<tbody>
<tr>
<td>226 240 251 232 213 194 164 141 118 95</td>
</tr>
<tr>
<td>240 260 274 257 240 223 192 168 144 120</td>
</tr>
<tr>
<td>251 274 295 284 271 258 226 201 176 151</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>W</th>
<th>237 262 280 285 277 269 238 216 194 172</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>223 250 280 276 272 242 223 204 185</td>
</tr>
<tr>
<td>S</td>
<td>209 238 271 275 275 246 230 214 198</td>
</tr>
<tr>
<td>T</td>
<td>191 222 258 269 276 283 258 243 228 213</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>S O U T H</th>
</tr>
</thead>
<tbody>
<tr>
<td>173 206 245 263 277 291 270 256 242 228</td>
</tr>
<tr>
<td>144 176 214 236 254 272 256 244 232 220</td>
</tr>
<tr>
<td>119 150 187 212 233 254 242 232 222 212</td>
</tr>
</tbody>
</table>
Figure 5: 3-dim-profile of Endol matrix values from Tab. 8.

Figure 6: External observer’s visualisation (2-dim-image) from Tab.9 data of the SCIP system’s endo-view of its reality showing regions of potential object locations (isoreferentials) under crisp hedge interpretation.
Figure 7: 3-dim-profile according to Tab. 9 data 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.

direct by stating coordinate values to determine a location—is self-including and organized around the entities to be specified. It does not, therefore, need (or rely on) any categorial presuppositions exterior to the self-organizing process whose emergent results structure space to become (potentially) referential.

<table>
<thead>
<tr>
<th>Fuzzy 1.1</th>
<th>DR</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP</td>
<td>1</td>
</tr>
<tr>
<td>extremely nearby</td>
<td>1.0</td>
</tr>
<tr>
<td>very nearby</td>
<td>0.2</td>
</tr>
<tr>
<td>rather nearby</td>
<td>0</td>
</tr>
<tr>
<td>rather far away</td>
<td>0</td>
</tr>
<tr>
<td>very far away</td>
<td>0</td>
</tr>
<tr>
<td>extremely far away</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 10: Fuzzy1.1 semantic definition of hedge-predicates (HP) in membership-values of distance ranges (DR) measured by reference-plane grid-points.

It is to be noted here, too, that the initial visualization chosen to be a two-dimensional plane spanned by orthogonal coordinates is not a situational necessity of the space concept but only the most conventionalized frame for representing definite locations abstracting from their situational embedding. As soon as gradation (or fuzziness) is included as a consequence of such co- and contextual situatedness—i.e. by way of textual language descriptions of object locations relative to system positions (described situations) and their actualization or understanding by the system concerned (discourse situations)—the

26
representational frame is immediately extended by structural or pragmatic information\textsuperscript{25} (Weiszäcker 1974) which may conventionally be represented only by adding another dimension. What may be observed on the representational level as a transition from the initial reference plane (Fig. 2) to the 2-dim-image of the system’s endo-view of it (Fig. 6), whose inherent structural information gives rise to represent the isoreferentials in another (third) dimension (Fig. 7), may find an analogue correspondence on the procedural level of the processes which have been dealt with in view of their semiotic potential so far. As may be gathered from what has been outlined so far, the non-propositional processing of a set (corpus) of sets (texts) of correct and true language expressions (sentences) of propositions describing (fixed) object locations relative to (changing) system positions resulted in an internal memory of labelled meaning points in a vector space (SHS) format whose intrinsic structure was made visible by three consecutive stages of representations. These visualizations were based on crisp interpretations (Tab. 7) of the hedges. Using fuzzy definitions instead (Tab. 10) to interpret the adjacencies of hedged core predicate labels from the cluster analysis (Fig. 4)\textsuperscript{26} will—as is to be expected—produce comparable images of emerging structuredness which can be visualized again in the form of isoreferentials (Fig. 8). What is surprising though in doing so is the fact that these new exo-representations of the system’s endo-view appear to be a more immediate image of the referential space structure under fuzzy\textsubscript{1,1} interpretations of hedges, as they are derived without an intermediate transform like Endo1-Endo2 above (Eqn. 12).

\textsuperscript{25}For a detailed discussion confer Gernert (1995) and Kornwachs (1995) in this volume.

\textsuperscript{26}v. p. 21.
A still very tentative explanation may be offered by assuming that the fuzzy interpretation of hedged core predicate adjacencies on this (representational) level appears to have a similar effect as changing positions of the system would produce on the situational level to generate more language material. Described by varying but definite relations of (object) locations and (system) positions under crisp interpretation of hedges, such variations seem to become expressed also by the fuzzy \(_{1,1}\) definitions of hedge predicates. This might be the reason—pending thorough testing under varying object locations—for an overall structure being found to emerge from non-propositional processing of these situational descriptions that does not need transformation in order to reveal most likely object locations, but anticipates them by an interpretational substitute of material description, i.e. the fuzzy denotational definitions.

6 Concluding the set

Where do we stand at this juncture and what can be said of this venture’s results so far? It was hypothesized that the notion of reference and denotation as specified by realistic semantic theories favoured in competence linguistics might be reconstructed also within structural semantics based on theories of linguistic performance and an ecological understanding of informational systems. Drawing on the procedural modelling and numerical instantiation of processes that simulate the constitution of meanings and their distribu-
tional or fuzzy representations in vector-space formats, an abstract semiotic cognitive information processing system (SCIPS) was introduced that operates in a well defined, experimental language environment as a meaning acquisition and understanding device. For the purpose of evaluating the understanding of meanings which the system might acquire in the course of its language processing, particular pains were taken to allow the system’s endo-view of its environment to differ observably from the exo-view of that same environment. In order to show that a semiotic, essentially non-propositional processing of language expressions can detect very much the same referentials which a propositional and truth-functional processing of these expressions would decode, the exo-view of the environment was formalized by constraints explicitly stated. These constraints—as given by the syntax and semantics specifying correct and true descriptions of object locations relative to system positions—do not, therefore, presuppose the existence of an objective reality external to the system or the system-environment designers (observers), but introduce a formal model to specify the conditions under which the results of propositional vs. non-propositional processing can be compared and evaluated either to the negative or to the positive.

The author is quite certain to have produced some evidence for the latter.

References


