DISCOURSE UNDERSTANDING AS IMAGE GENERATION
On perception-based processing of NL texts in SCIP systems

BURGHARD B. RIEGER

FB II: Computational Linguistics, University of Trier
Universitätsring, D-54286 Trier, Germany
e-mail: rieger@uni-trier.de
URL: www.ldv.uni-trier.de:8080/rieger.html

Abstract: Semiotic Cognitive Information Processing (SCIP) is inspired by information systems theory and grounded in (natural/artificial) system-environment situations. SCIP systems’ knowledge-based natural language processing (NLP) of information makes it cognitive, their sign and symbol generation, manipulation, and understanding capabilities render it semiotic. Based upon structures whose representational status is not a presupposition to, but a result from recursive processing, SCIP algorithms initiate and modify the structures they are operating, and by simulating processes of symbol grounding they realize meaning constitution and language understanding. Whereas traditional semantics is based upon the symbolic (de)composition of propositional structures, SCIP tries to model learning and understanding dynamically by visualizing what is understood in a perception-based, sub-symbolic, multi-resolutional way of processing natural language discourse. An experimental 2-dim scenario with object locations described relative to a mobile agent’s varying positions allows to test SCIP systems’ performance against human natural language understanding in a controlled way.

Keywords: Natural language understanding, symbol grounding, fuzzy meaning constitution, semantic space, quantitative linguistics, dynamics, systems theory, visualization.

1 Cognitive Models of Meaning

It is common practice according to [BP83, p.57] in cognitive modeling and mathematical semantics to identify the real world with the (symbolic) structure that represents it. From a semiotic point-of-view, this identification is hiding rather than revealing what makes signs (and structured sign aggregates) stand for, represent, or symbolize something else.

1.1 Reality, Perception, Representations

Disciplines like language philosophy, logics, linguistic semantics, biological neuro-science, and computational connectionism, which among others focus on aspects of cognition, have outlined [PR99] that the relationship between the real world or objective reality (R) of observable entities external to a cognitive system, and the perception of such entities which constitute a system’s experienced environment or subjective actuality (A), is cognitively as well as epistemologically highly relevant and model-theoretically most decisive. Suggestions for how this relation may be mediated and (re-)constructed have resulted over the years in a number of types of models which range from simple identity as \( A = R \), and functions as \( A = f(R) \) depending on reality (R) only, or as \( A = f(R, O, C) \) being based additionally on features of the observing system (O) and its cultural and/or experiential background (C), to reach out to structurally coupled resonance phenomena of semantically closed cognitive systems as \( A_{t+1} = f(A_t, E, P) \) which relate perturbations (P) inflicted on systems and environments, the structure of a state space (E) determining a system’s possible states, and – to cope for the dynamics – the system’s actual states’ changes \( A_t \) along a time scale. In this formula, A seemingly can do altogether without R [Mat78]. This is a consequence of self-organizing, dynamic, autopoietic systems [MV80] for which the observability of entities external to a cognitive system hinges on their communicability to others which include internal results of commonly experienced external perturbations. Reality R, therefore, should be viewed more like a situational condition for the possibility of inter-subjective and social collections of experiential results rather than an independently existing realm of entities. Thus, suggesting and
finding parameters to reconstruct the background of experiential perception for the interpretation of what can be considered observable reality in this way, underscores the importance of distinguishing endo- from exo-views of reality to overcome the traditional mind/matter duality. In view of representational structures like natural language texts in discourse, the endo-exo distinction allows for a semantically more adequate approach to entities whose observable reality provides for an experiential perception which is also the precondition for their understanding (and the modeling of it).

1.2 Semantic Theory, Meaning, and Understanding

Until recently, theoretical and computational linguistics – mediated by (language) philosophy, (formal) logics, and (discrete) mathematics – have clearly dominated research and explicative theory development on how natural languages (NL), their (compositional) structures, and their (semantic) functions are to be understood and explicated as symbol manipulation and transformation systems. NL communication has long been conjectured to consist of what only recently the cognitive sciences have identified as a complex of multi-level processes that operate on (world, linguistic, situational) knowledge which has to be considered conditional for any information processing. However, the knowledge bases (KB) designed to comply with these conditions were hypothesized as physical symbol systems [New80], [Sim82] whose static conception of structure proved to be unable to adapt to changing conditions (learning). Some of the problems [Car00] that cognitive modeling along these lines encountered since, are due to the declarative (symbolic, compositional, propositional) formats employed and the (deterministic, rule-based, modular) procedures chosen in generating, forming, and manipulating linguistic concepts (morphemes, syllables, words, phrases, sentences, texts, and their meanings) as if they were clear-cut elements (aggregates, structures, relations, functions, processes, etc.) of systems of language entities whose perception is crisp and determinate, rather than variable, context dependent, fuzzy, and possibilistic in nature.

In order to understand the dynamics of how natural languages serve the communicative purposes they do, fuzzy [Zad78], [Zad99] and procedural modeling [RK99] approaches to semiotic systems [Mey95], [Rie99] and NL understanding [Rie95], [Rie00] have advanced some ideas [Zad97], [Zad01] for a computational theory of cognitive processing of fuzzy percepts. Conceived as a multi-layered process of structure identification and dynamic representation, a dynamic image generating semantics (DIGS) based on this theory will eventually be able to cope with variability and vagueness, adaptivity and learning, emergence and plasticity of knowledge and understanding in a comprehensive way. Fuzzy modeling techniques allow for (numerical, sub-symbolic, distributed, non-propositional) formats whose (parallel, pattern-based, quantitative) computation result in (the emergence of) meanings as enactment of labeled processes of choice restriction [Rie94]. Meanings are the outcome rather than the presuppositions of processing [WFKS00], whose modeling is a form of realization rather than simulation [Pat89]. It appears that a perception-based simulation of processes (of constraint detection and representation) may bring about results which realize meaning constitution and understanding (of symbolic structures) as grounded in these very processes.

2 Computational Semiotics and SCIP systems

Semiotic Cognitive Information Processing (SCIP) is inspired by information systems theory and based upon (natural or artificial) system-environment situations. A system whose processing of external, environmental data (input) is determined by its own internal structuredness, will generally gather some information (output) relative to both, internal and external conditions. As soon as the input is a flow of signals from data of signs or symbol aggregates, these have to be recognized as representations in order to be processed accordingly, i.e. interpreted as standing for something else that the perceivable signal is not.

2.1 Knowledge, Memory, and Models

Traditional models of cognitive information processing try to account for this double ontology of signs/symbols – which are physically real like data but in addition also have meaning – by providing the processing system with the necessary information via arbitrarily complex representations (sets, structures, systems) of sign-meaning correspondences, named knowledge-base. KBs extend the system’s data processing capabilities to cognitive, i.e. knowledge-based processing in generating, manipulating, and interpreting sign and symbol aggregates of different kinds. Conceived as being externally attributable to the modeled system and therefore assembled by the model designer, KBs obviously serve a function which is considered essential to the original/natural cognitive systems and their structure (i.e. knowledge and memory). In order to let models of cognitive language information processing (CLIP) systems become semiotic (SCIP),
knowledge and memory have to be conceived as procedural and internal to the systems changing their character from static determination to dynamic flexibility. Additionally, the representational format for knowledge structures and memory functions should facilitate adaptation to changing environmental and processing conditions (learning), and enable identification in changing contexts (efficiency) for a singular system concerned, as well as among a plurality of systems interacting by means of externalized sign representations (communication).

2.2 Semiotic Cognitive Information Processing

Allowing for variable, ill-defined, underdetermined data to be processed, and enabling the self-organized constitution (emergence) of vague and fuzzy entities to be represented and operated on, semiotic cognitive information processing is based on well-defined procedures which can handle imprecision in a precise way. SCIP systems’ ability comprises their performance in knowledge-based information processing and representing its results [Rie91], organizing these representations by activating others from prior processing [RT89], constituting meanings [Rie98a], allowing for (semantic) inferencing [Rie82], and planning [Rie84] by selecting from organized and represented dispositions [Rie88], and modifying them according to changing conditions, results, and states of evolving system-environment adaptedness [RT93]. Based on NL structures, SCIP performance is a form of complex, multi-resolutional information processing. As process of meaning constitution it is tied to (and even be identified with) language understanding [Rie01] or meaning acquisition. Whenever the meaning of signs is not a presupposition to but a result from algorithmic processing of (symbolic) data whose representational status (like in NL discourse) is commonly accepted, then these learning algorithms – being able to initiate and modify the structures they are operating on – may qualify as semiotic and thereby as part of computational semiotics.

3 Perception-based Text Processing

The SCIP system’s approach to natural language discourse understanding is – very much like modeling vision [Mar82] – essentially perception based. It might be considered the core of a dynamic image generating semantics (DIGS) which complements the declarative, symbolic (de)composition of propositional structures in traditional NL semantics in a
procedurally defined way of sub-symbolic, quantitative, emergent, dynamic pattern identification, representation, and manipulation.

3.1 Dynamic Image Generating Semantics

The dynamics of DIGS depends essentially on the SCIP system’s format of non-symbolic, distributed representations whose processing allow new representations to emerge. These are tying the system to those segments of the real world which the language expressions are a part of and – when processed properly – convey information about as their meanings. They do so both, according to their grammaticality and propositional contents external to the system in a formally specifiable sense, and according to the system’s own or internal understanding based upon the non-propositional, syntagmatic and paradigmatic regularities in textual structures which can be visualized accordingly. This is achieved by formalizing these ties not as functions abstracted from grammatical rules that are represented symbolically, but as a class of restrictions that are typified by non-symbolic understanding. Although the semantic contents conveyed cannot always be represented in a language independent way, operations and/or processes may exist whose procedures may be found even without being understood prior to their algorithmized enactment resulting in some observable (re)presentation. The difficulties of controlled production, test, and evaluation of results of non-symbolic understanding is why traditional cognitive approaches easily accept linguistic analyses of propositional language structure as ready made model and explicative theory of understanding, and why linguistic semantics in turn appeals to formal logics as an available format for the representation of NL expressions’ propositional functioning. 

---

Figure 2: Situated test cycle to compare the system’s (unknown) internal-view (endo-reality) resulting from the modeled SCIP system’s (well-known) processing, against the observer’s (well known) external-view (exo-reality) which traditional, symbol based, computational linguistic models identify prematurely with the (unknown) processes underlying natural language understanding. However, referential semantics and propositional text grammar allow to generate PHT corpora of true NL descriptions of (real world) SPOL relations. Their subsymbolic, two-level processing results in the SCIP system’s semantic space structure. Its algorithmic visualization (Fig.4) based on clustering allows for a comparison with what traditional models describe by grammatically correct and semantically true propositions (Fig.3) encoded as referential meaning or informational content.
(soft) constraints, modeled as procedures which produce (fuzzy) relations represented as (word type/numerical value) distributions. These are not just another instance of transformed data representation but – as they result from non-symbolic, numerical computation – a new type of structural representation associating emergent entities (concepts) with observable entities (objects/signs) to realize what may be named understanding.

### 3.2 Describing and Understanding: Morphisms

Being grounded in system-environment situations, DIGS may formally be characterized by morphisms\(^2\) \([\text{DDS dependency graphs} \text{ \cite{Rie98b}}]\) of abstract objects \((\text{meaning points} p \in M \subseteq I)\). These emerge in multi-layered vector space mappings \((\text{corpus space } C, \text{ semantic space } I)\) from computation of aggregational \(\text{syn (syntagmatic)}\) and selective \(\text{par (paradigmatic)}\) patterns of constraints on language signs \(z \in T \subseteq V\) in very large corpora of texts which describe the real world situation \(x \in X \subseteq U\) to be understood.

**B. 1)** The process of describing entities in the universe of discourse \(dsc : X \rightarrow T\) (in Fig.1) can theoretically be specified and algorithmically determined by formal expressions \(e \in E \subseteq G\) of grammatical adequacy as provided by computational linguistics. The morphisms \(\text{syn} \) and \(\text{sem}\) define a notion of constrained syntactic correctness and semantic truth of propositional structures. These are dynamically generated to describe real world entities \(z \in X \subseteq U\) in a controlled way as NL texts \(z \in T \subseteq V\). Assembled into collections of increasing size, this language material \(T_n \subseteq V\) forms PHT-corpora (of pragmatically homogeneous texts) whose semantic contents \((\text{meaning})\) are the described situations these texts refer to.

**B. 2)** The process of understanding the reference relation as morphism \(\text{ref} : T \rightarrow X\) (in Fig.1) is (re)constructed by implemented semiotic algorithms for the recursive computation of the combinatorial constraints \(\text{syn} \) and \(\text{par}\) and their multi-layered, multi-resolutional representation \(y \in C\) in (patterns) of distributions of (observable and emergent) entities \(p \in M \subseteq I\). Thus, morphisms characterize a very general type of relatedness that allows to specify a procedural notion of \textit{semioticy} which realizes Peirce’s conception of \textit{semiosis}\(^3\) for operational application in a SCIP setting.

**B. 3)** In order to demonstrate the suggested DIGS potential as modeled by a SCIP system’s discourse understanding capability, it will be made to constitute meaning (i.e. realization) internally by performing (i.e. simulation) some perception-based signal processing whose computed results (end-view) ground the denotation morphism \(\text{den} : M \rightarrow X\) (Fig. 1). It can be visualized according to systemic \(\text{sys}\) and environmental \(\text{env}\) constraints of system-environment relatedness or situation types \(S\) (Fig. 1), and evaluated against the – externally observable \(x \in X \subseteq U\) (exo-view) – the true and correct descriptions of which are given by the NL discourse processed.

### 4 Tests and Future Work

For the structurally coupled system-environment relation whose situated processing is enacted as being

\[^3\]By \textit{semiosis} I mean […] an action, or influence, which is, or involves, a cooperation of 
three subjects, such as sign \([z \in T \subseteq V]\), its object \([x \in X \subseteq U]\), and its \[cognitive: \ p \in M \subseteq I\] or \[logical: \ e \in E \subseteq G\] interpretant, this tri-relative influence not being in any way resolvable into actions between pairs.” \cite[Pei06, p.282]
Figure 4: 2-dim visualizations of potential object locations (isoreferentials) showing the simulative results of the SCIP system’s incremental meaning constitution or learning process (without any semantic and syntactic knowledge of grammar) entirely based upon the sub-symbolic, numerical computation of textual (syntagmatic and paradigmatic) constraints in growing sets of (10 to 300) texts which describe a randomly walking system’s positions relative to stationary objects’ locations (SPOL relations) in a formally controlled way.

based on NL descriptions, an experimental scenario was devised whose simplifications would hopefully not trivialize the issues to be tested (Fig.2). Confining the discourse material to (syntactically correct, semantically true) natural language descriptions\(^4\) of an external observer’s view (exo-reality) first, these descriptions would then be submitted to the perception based, sub-symbolic, cognitive processing and structuring according to the defined DIGS formalisms implemented as SCIP algorithms. These will result in some mappings and/or representations which form the semantic space structure whose clustering and visualization reveals it being part of the system’s internal view of its environment (endo-reality) constituting its understanding. As the computational visualization of the endo-view is independent from all symbolic processing provided by computational linguistic (CL) techniques, its imaging results allow for an inter-subjectively controlled, repeatable, and experimental testing of the artificial SCIP system’s capacity to understand the referential meaning in NL text material processed, against the externally observable situational reality as represented and described by that discourse.

4.1 Experimental Setting

The 2-dim real world scenario (Fig. 3) is a reference plane with two stationary objects $\triangle, \square \in X \subset U$ (environment), and an oriented mobile agent $A \in U$ (SCIP system). System and environment are structurally coupled [VTR91] by a text corpus $T$ of (true and correct) natural language (NL) expressions\(^4\) $z \in T$ of possible system-position/object-location (SPOL) relations. The perception-based,
non-symbolic processing \( \text{par} \circ \text{syn} = \text{des} : T \to M \) (Fig. 1) of these text corpora (see Fig. 2) yield vectorial representations of meaning points \( p \in M \subseteq I \) in semantic space. Its over-all structure (see Fig. 2) may computationally be visualized \( \text{env} \circ \text{sys} = \text{den} : M \to X \) (Fig. 1) which – according to the incremental processing of growing numbers \( 1 \leq n \leq N \) of texts in larger corpora \( T_n \) – will produce images (Fig. 4) which depict regions of potential object locations \( X_n \subset U \) by profile lines of common likelihood (isoreferentials). Their development – from 25 to 400 texts – shows increasingly distinct maxima that identify object locations computationally from the texts which describe them, demonstrating the SCIP system’s understanding capability as performed by its non-symbolic, perception-based, and grammar-independent processing.

A software prototype of the SCIP system-environment has been implemented as a testbed for the modeled processes of description \( \text{dsc} : X \to T_n \) and understanding \( \text{ref} : T_n \to X_n \), covering variable system-environment situations and their comparison \( X \Leftrightarrow X_n \) some preliminary versions of which have been discussed and presented earlier [Rie02]. The testbed will also be accessible via internet soon [RFJ03] to illustrate the performance of a perception based, procedural approach to the dynamics of semiotically grounded (natural language) meaning constitution for referential expressions as part of dynamic image generating semantics (DIGS).

4.2 Outlook

Future research will primarily be directed towards discourse dialog situations allowing for (two and more) agents. These will be concerned with NL descriptions generated as above. However, the text corpora being derived from one system’s SPOL related views of its environment will serve as input for the other agent(s), and vice versa. Their mutual processing should add structural information for their object and/or system identification respectively. The exo-view distinction of mobile/objective system positions (SP) from stationary/fixed object locations (OL) will have to be translated to the endo-view level in order to address (and hopefully solve) the problem of how the variations in a changing and/or stable environment may be (re)cognized and understood by mobile agents, i.e. relative to and against their own (space-time) movements for which there is no independent representation (yet), apart from the text corpora of NL descriptions that mediate them indirectly. It might be suspected that additional sensory channels (e.g. vision) will have to be allowed to enhance (and differentiate the semiotic) cognitive information processing capacities of the SCIP systems modeled so far.

References


L.A. Zadeh: From computing with numbers to computing with words – from manipulation of measurement to manipulation of perceptions. In: [Wan01], pp. 35–68.