Burghard B. Rieger

From Computational Linguistics to Computing with Words.

For the past decades, the concept of symbolic representation together with the computer metaphor appeared to offer an adequate framework to deal with cognitive processes scientifically. Formally grounded by logical calculi and implemented as algorithms operating on representational structures, cognition is considered a form of information processing in the *cognitive sciences*. Thus, *computational linguistics* (CL) as part of *cognitive theory* identified the complex of language understanding as a modular system of subsystems of information processing which could be modeled accordingly. The alliance of logics and linguistics, mediated mainly by (language) philosophy in the past, and by (discrete) mathematics since the first half of the last century, has long been (and partly still is) dominating in what way and terms natural languages and their functioning should be explicated and how their processing could be modeled. In replicating (and in parts also supplementing) semiotically motivated strata of systematic sign description and analysis, different levels of modular aggregation of information – external and/or internal to a processing system – have been distinguished in cognitive models of language understanding. They partly correspond to and partly cut across the *syntactics-semantics-pragmatics* distinction in the *semiotic* relatedness of signs, the *utterance-discourse-corpus* levels of performative language analysis, and the hierarchy of *morpho-phonological*, *syntax-sentencial* and *lexico-semantic* descriptions in *structural* models of linguistics. It is ironic, however, that the dramatic increase of computational power and symbol manipulation means has changed the fundamentals of many scientific disciplines, creating even new ones, but has left linguistically oriented disciplines, even new ones, adhere to the lore of seemingly well grounded and traditionally dignified concepts in describing natural language structures and their functions.

1. Cognition

Cognitive approaches tend to model mind/brain processes based upon the evaluation of (in parts linguistic) data generated in more or less sophisticated experiments of human thought/thinking and understanding. For a linguist, however, more immediate results of cognitively most relevant mind/brain activities – not the experimentally reduced segments of them – come to mind as being easily accessible in form and structure of natural language discourse which is abundantly available now in machine readable form. Other than what premises of theoretical linguistics and main stream computational linguistics suggest, some processes of language understanding require and might well be modeled along observable but as yet unexploited traces of meaning constitution in natural language text corpora which speakers/writers and hearers/readers have enacted in situations of communicative language *use*[1]. This enactment is tied to representational or semiotic functions, based upon regularities of entity *usages*[2] which not only generate observable structures, but also serve in turn to allow these functions being activated to modify these structures simultaneously. Thus, the complexities and dynamics of natural languages themselves may be taken as a salient paradigm for information granulation both, in its *fuzzy* as well as *crisp* modes of structural representation and functional processing. It appears that the conception of fuzzy and/or crisp *granularity* – once the process-result ambiguity and its cognitive-linguistic ambivalence is solved as addressed by a facet of the symbol-matter problem [Pattee89] – lends itself easily

[1] In the sense of employment and act of being used under certain conditions and to a particular communicative purpose.
[2] In the sense of customary practice and manner of using which may establish and modify rules and standards.

to a unifying view of how natural language understanding or *meaning constitution* may be arrived at as a computational process on structural entities adequately identified.

2. Computational Linguistics

The way structural linguists used to and still categorize (*segment* and *classify*) observable natural language phenomena as *tokens* (like *phones, morphs, utterances*, etc.) to constitute abstract linguistic entities as *types* (like *phonemes, morphemes, phrase*, etc.) can be shown to be based on the very processes of granular meaning constitution, however imperfectly. What may procedurally be derived either as *soft linguistic categories* or fuzzy *granules* represented by vectors, distributions, or *fuzzy sets* for (numerical) computation, has so far been (over-) generalized and abstracted to form crisp *categories* represented by signs (*symbols*) for string manipulation. Certainly, linear aggregation of these symbols serve to understand and control one type of observable natural language phenomena as part of aggregational *string formatio*n or *formal grammar*. Its core concepts of well-formedness (*syntax*) and truth-function (*semantics*) were made explicit by way of specifying conditions of formal *correctness* and derivational *compositionality*. Their symbolic representations in the form of productions or rewrite *rules* – allowing for recursive application and generative string formation – not only constituted a wealth of symbol aggregation systems (*formal languages*) but were also employed to model comparable properties (in processes) of natural language string formation.

However, whereas formal rules whose application would allow to specify generative properties and truth-functional constraints in formal constructs like *sentence* and *proposition*, other properties of natural language expressions which are communicationally more relevant, like e.g. making *sense* by having specific *meaning* in situational *contexts* which are to be *understood*, tended to be abstracted away. The process of understanding natural language discourse below and above formal sentence reconstruction is, by and large, still in want of principled analysis, formal representation, computational simulation, and procedural realization. This is, were CL is to head in near future in order to assist people not only in producing huge amounts of texts, but also in skim the of verbal information worldwide without the need to read the texts concerned.

In order to follow this line of structural, functional, simulative, and enactive modeling of machine understanding, traits of traditional approaches to cognition and natural language processing as forwarded by *linguistics proper* (LP), *computational linguistics* (CL), and language processing in *artificial intelligence* (AI) research will be reviewed in order to identify points of departure from which to advance our understanding of how natural languages function the way they do. Some presuppositions will have to be revised to understand how communicative employment of languages is not only based upon (*use*) but also establishes (*usage*) structural constraints which may be made explicit by assumptions motivated by systems theory and/or by empirically testable hypotheses derived thereof [Rieger01]. It will be argued that the introspective assessment and judgment of any speakers' own language faculty on linguistic functions and the correctness of singular sentences or phrasal structures as conceived by an *ideal* speaker's/hearer's internal language (IL) is not at all sufficient, let alone superior to modern means of empirical investigation of masses of natural language discourse or *external language* (EL) [3] as being produced by *real* speakers/writers in situations of intended/successful communication. Instead, some of the

[3] Borrowing the term not "to refer to any other notion of language … never characterized in any coherent way" (Chomsky, personal communication), but understood – and diverging from Chomsky – to cover all phenomena of observable language performance.

inadequacies of CL models of natural language processing that *competence oriented* linguistics have inspired so far, will hopefully be revealed to be due to unwarranted abstractions from relevant characteristics (e.g. *contextuality, vagueness, variability, adaptivity, openness* etc., to name only the most salient) of processes of natural language communication. Other than these *idealizations* which purportedly allow immediate access to cognitively relevant entities, we shall argue for an empirically controlled understanding of functional sign constitution which does not readily abstract from the emergent structures which models of a more *semiotic* cognitive information processing (SCIP) may bring about. It is hoped to collect and produce some evidence that the traces of such processing can not only be identified, but that these identification procedures may also be employed to systematically (re)construct *fuzzy information granulation* procedurally.

3. Computing with Words

The notion of *computing with words* (CW) hinges crucially on the employment of natural language expressions. These are considered to provide not only the representational structures of what can semantically be meant but also the operational means of what can cognitively be understood by processing these structures. They allow for decomposition of wholes into their constituents or parts (*granulation*), or conversely, for composition and integration of parts into wholes (*organization*), and for the association of signs with meaningss (*causation*). According to Zadeh's early introduction of the notion of *granularity* [Zadeh79] and his recent elaboration of that concept [Zadeh97] as *theory of fuzzy information granulation* (TFIG), human cognition may be understood as based upon and structured by processes of g*ranulation, organization* and *causation*. These are meant to specify different types of mind/brain activities which can be characterized as being computational in nature and hence to be modeled mathematically and/or procedurally. Although this characterization suggests different modes of these mind/brain activities to be distinguished sharply both, as the enactment of processes and as the results which these processes produce – a distinction that will have to be drawn and obeyed more clearly – there is the need for yet another discrimination to be made in order to clarify what CL will have to deal with in future. It is tied to the first one and concerns the way mind/brain activities are made accessible by techniques, models, or disciplines in different ways.

In *fuzzy linguistic* (FL) models of *computational semiotics* (CS), the situatedness of natural language communication is considered conditional. This requirement is met by corpora of *pragmatically homogeneous texts* (PHT) which assemble language material which is situationally constrained by a number of variables (like e.g. communicative media or domain, register, topic, author, etc.) whose values (like newspaper, report, economy, etc.) define the PHT profile the incorporated texts will satisfy. There is good reason to assume that such PHT collections realize the *organized* structure of natural language discourse, i.e. integrating parts into wholes (sign formation), as well as the *causative* functions, i.e. associating causes with effects (meaning constitution). The inherent structuredness of a PHT corpus gives rise to the multi-resolutional representations of meaning functions which may be explored in order to model (crisp and fuzzy) SCIP *granules*. Based upon the empirically well founded observation and rigorous mathematical description of universal regularities in natural language discourse, these regularities can be shown to structure and constitute (different levels of) processes and/or their representational results when made to operate on pragmatically homogeneous texts of either performed or intended communicative interaction in actual situations. Only such a *performance oriented* semiotic approach will give a chance to formally reconstruct and model 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 actualization of cognitive processes in language understanding [Rieger02]
.

References

Pattee, H. (1989): Simulations, Realizations, and Theories of Life. In: Langton (ed): *Artificial Life*, [SFI Studies in the Science of Complexity VI], Reading, MA (Addison Wesley), pp. 63-77.

Rieger, B. B. (2001): Computing Granular Word Meanings. A fuzzy linguistic approach in Computational Semiotics. In: Wang (ed): *Computing with Words*, [Wiley Series on Intelligent Systems 3]. New York (J. Wiley & Sons), pp.147-208.

Rieger, B. B. (2002): Semiotic Cognitive Information Processing: Learning to Understand Discourse. A systemic model of meaning constitution. In: Stamatescu et.al. (eds): *Perspectives on Adaptivity and Learning.* Heidelberg/ Berlin/ New York (Springer), [in print].

Zadeh, L. A. (1979): Fuzzy Sets and Information Granularity. In: Gupta/Ragade/Yager (eds): *Advances in Fuzzy Set Theory and its Application.* Amsterdam (North Holland), pp. 3-18.

Zadeh, L. A. (1997): Toward a Theory of Fuzzy Information Granulation and ist Centrality in Human Reasoning and Fuzzy Logic. *Fuzzy Sets and Systems*, 90(3): 111-127.