Fuzzy Word Meanings as Semantic Granules.
Emergent constraints for self-organizing tree structures in SCIP systems.

Burghard B. Rieger*
Department of Computational Linguistics, FB II: LDV/CL
University of Trier, D-54286 TRIER, Germany

Abstract

The notion of *semiotic cognitive information processing* (SCIP) is concerned with the situated employment of natural language expressions for communicative purposes. Natural languages (NL) provide not only linguistic structures representational for processes of understanding, but also crucial hints on the operational constitution of their processing. These allow for the decomposition of wholes into their constituents or parts (*granulation*), for the composition or integration of their parts into wholes (*organization*), and for the association of semiotic causes with effects (*meaning*). Thus, *information granulation*[1] can algorithmically be modeled and realized both, in its *crisp* as well as *fuzzy* modes of representation and processing, by exploiting the structuredness of *pragmatically homogeneous* NL text samples (PHT corpora).

1 Introduction

Based upon the notion of *Computing with Words* (CW) [13], the concept of fuzzy and/or crisp *granulation*—once their process-result ambiguity is solved—lends itself easily to a unifying view of the way structural linguists used to and still categorize (segment and classify) observable natural language phenomena (tokens like *phones, morphs, lexes, utterances*, etc.) to constitute abstract linguistic entities (types like *phonemes, morphemes, lexemes, sentences*, etc.). These may either be derived as *soft linguistic categories* or fuzzy *granules* represented as vectors (*fuzzy sets*), or they may be postulated as abstractions to form crisp *categories* representable by symbols (*signs*) whose linear compositions in well-formed strings, in turn, give rise to the notion of *correctness*. Whereas the latter may formally be characterized by *rules*, the derivation of the former can be determined procedurally by *algorithms* operating on language data [7]. Their twofold *process*-analytical and *result*-representational function render these algorithms *semiotic* [5].

*URL: `http://www.ldv.uni-trier.de:8080/rieger.html`;
E-Mail: `rieger@ldv.uni-trier.de`

[1]According to Zadeh (1997) [14], all processes of human cognition are structured by *granulation, organization* and *causation*.

2 FL and CS

In *fuzzy linguistic* (FL) models and *computational semiotic* (CS) realizations of sign processes [7], analytical procedures are derived detecting and, at the same time, operating on intrinsic (or structural) information that constitutes understanding as (intermediate) representation of the phenomena concerned. Based upon the assumption the structuredness of natural language discourse, its *organizing* functions, i.e. integration of parts into wholes (sign formation), as well as the *causative* functions, i.e. semiotic association of causes with effects (meaning constitution), is realized and accessible in PHT corpora, these may be analyzed for inherent regularities which may be explored in order to re-construct (crisp and fuzzy) semantic *granules* [9]. Tied to the empirically well founded and testable observations and rigorous mathematical description of results, entity formation in natural language discourse can be shown to constitute (different levels of) processes and/or their representational results. On word level these are viewed as enactment of universal principles which are realized in and detectable from *pragmatically homogeneous* texts (PHT) of either performed or intended communicative interaction in actual situations.

The *semantic* meaning functions have been modeled and computed earlier [11] as results of those same (semiotic) procedures by way of which (representational) structures emerge. Their actual interpretation could be simulated by analyzing the possibilistic constraints found to be imposed upon the linear ordering (*syntagmatics*) and the selective combination (*paradigmatics*) of natural language entities (word-types) in discourse [2]. In a FL/CS approach to lexical semantics this is tantamount to (re-)construct an entity's *semiotic* potential (meaning function) by a weighted graph (fuzzy *distributional pattern*) [3] representing a particular state of the modeled system's lexical state space rather than by a *single symbol* whose interpretation would have to be extrinsic to that system [8]. In this view, the emergence of semantic structure can be represented and studied as a self-organizing process based upon word usage regularities in natural language discourse. In its course, the linearly agglomerative (or *syntagmatic*) as well as the

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

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

distributionally selective (or *paradigmatic*) constraints are exploited by text analyzing algorithms which accept natural language text corpora as *input* and produce—via levels of intermediate processing and representation—a vector space structure as *output*. As *semantic hyperspace* (SHS) it may be interpreted as an internal (*endo*) representation of the SCIP system's states of adaptation to the external (*exo*) structures of its environment as mediated by the discourse processed. The degree of correspondence between these two is determined by the granularity that the texts provide in depicting an *exo*-view, and the resolution that the SCIP system is able to acquire as its *endo*-view in the course of that discourse' processing [10].

3 Empirical Reconstruction

Following the procedural approach in FL/CS, the reconstruction of linguistic functions or meanings of words is based upon a fundamental analytical as well as representational formalism. It can be characterized as a two-level process of abstraction (called α- and δ-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* (Tab. 1). These may be understood to interpret semantically (by way of the meaning function) those word-types which are being instantiated by word-tokens as employed in natural language PHT corpora.

The basically descriptive statistics used to grasp these relations on the level of *words* in discourse is centered around a correlational measure (1) to specify intensities of co-occurring lexical items in texts, and a measure of similarity (or rather, dissimilarity) (2) to specify these correlation value distributions' differences. Simultaneously, these two measures may also be interpreted semiotically as providing for the set theoretical constraints or formal mappings α and δ which model the meanings of words as a function of these words' differences of usage regularities as produced in discourse and analyzed in the PHT corpus.

$$
\alpha_{i,j} = \frac{\sum_{t=1}^{T}(h_{it} - e_{it})(h_{jt} - e_{jt})}{\left(\sum_{t=1}^{T}(h_{it} - e_{it})^2 \sum_{t=1}^{T}(h_{jt} - e_{jt})^2\right)^{\frac{1}{2}}}; \qquad (1)
$$

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

The coefficient $\alpha_{i,j}$ (1) measures pairwise relatedness of word-types $(x_i, x_j) \in V \times V$ where $e_{it} = \frac{H_i}{L} l_t$ and $e_{jt} = \frac{H_j}{L} l_t$, the PHT corpus of texts $K = \{k_t\}; t = 1, \dots, T$ has the length $L = \sum_{t=1}^{T} l_t; 1 \leq l_t \leq L$ measured by the number of word-tokens per text, and a vocabulary $V = \{x_n\}; n = 1, \dots, i, j, \dots, N$ whose type frequencies are denoted by $H_i = \sum_{t=1}^{T} h_{it}; 0 \leq h_{it} \leq H_i$.

The lexical items' usage regularities detected are represented by tuples of $\alpha(x_i, x_n)$-values which—interpreted as coordinates $\alpha_i(x_n)$— 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 $x_i \in V$.

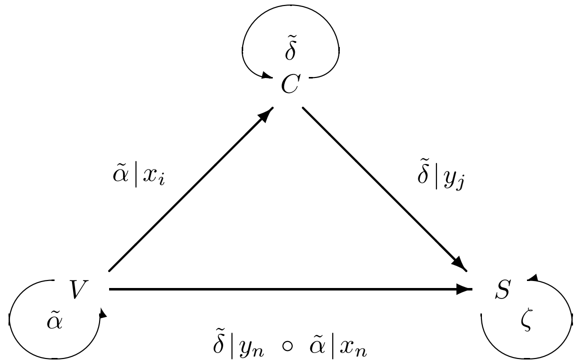

Figure 2: Fuzzy mapping relations $\tilde{\alpha}$ and $\tilde{\delta}$ between the structured sets of vocabulary items $x_n \in V$, of corpus points $y_n \in C$, and of meaning points $z_n \in S$.

Their similarities and/or dissimilarities are calculated by a distance measure δ (2) as the EUCLIDian metric on C

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

whose pairwise representations as tuples of $\delta(y_i, y_n)$-values determine—interpreted as coordinates $\delta_i(y_n)$ again—*meaning points* $z_n \in S$ or vectors in the hyperstructure or *semantic space* $\langle S, \zeta \rangle$ spanned by the number of axes corresponding to vocabulary entries $x_n \in V$ and a EUCLIDian metric ζ.

Thus, the two-stage mapping corresponds to a category-type morphism or composition $\tilde{\delta} | y_n \circ \tilde{\alpha} | x_n$ (Fig. 2), re-

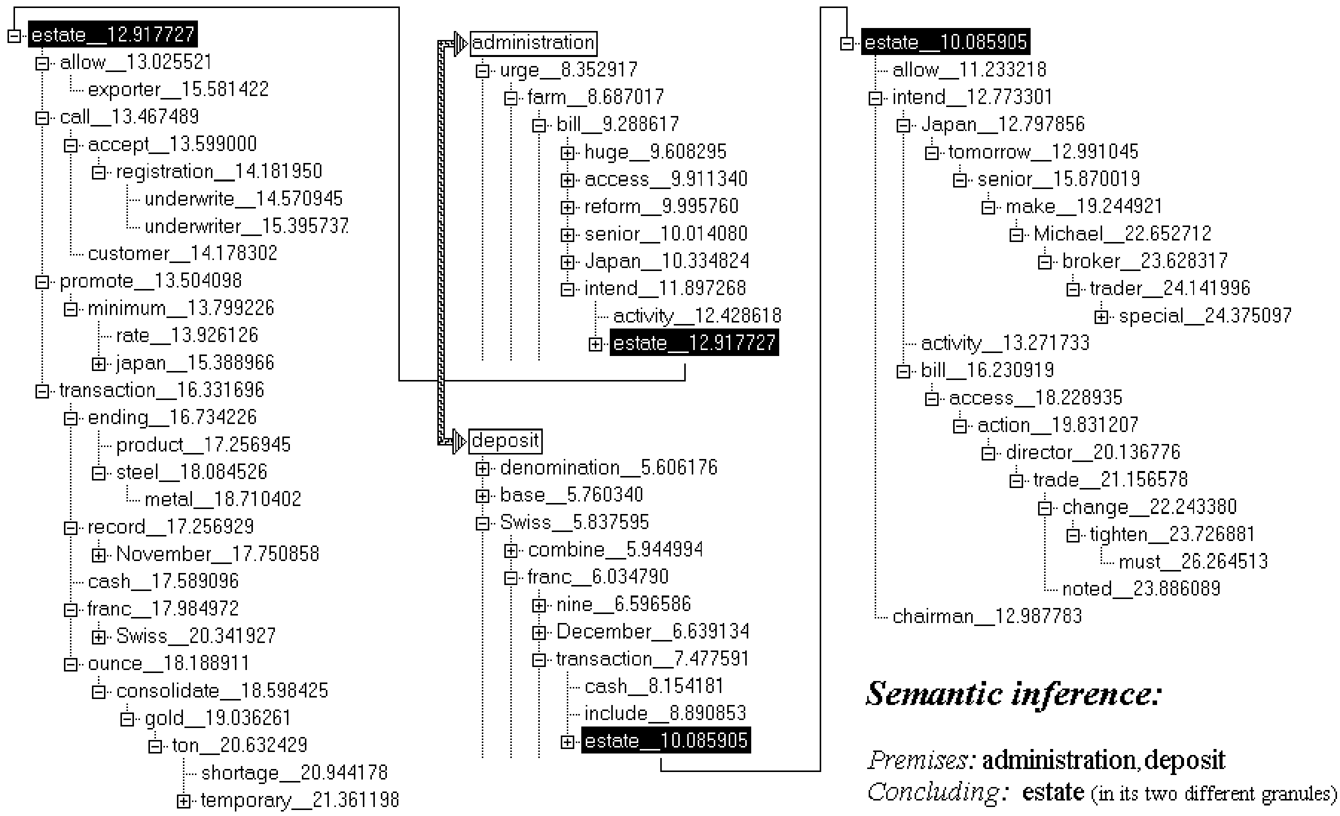

Figure 1: The *semantic inference* procedure is a parallel process activated from start nodes (*premises*) generating DDS graphs and stopped by first node common to all (*conclusion*). Subtrees constitute perspectively determined *information granules* of differing connotative, resolutional, and dependency structure.

sulting in the format of *semantic hyperspace* (SHS) which constitutes a system of *meaning points* as an empirically founded and functionally derived representation of a lexically labeled knowledge structure.

4 SHS

The *semantic hyperspace* (SHS) structure resulting from the *performance oriented* approach allows to reconstruct formally 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 *language games* (or cotexts[2]) and of *situations* (or contexts) [5]. As this function corresponds to the two-level actualization of cognitive processes in language understanding, SHS provides the format and structural information for an intermediate representational (tree-graph) structures to be generated as *semantic* or *dispositional dependencies* (DDS) introduced elsewhere [2]. Their property of situational (co- and contextual) sensitivity gave rise to the algorithmic derivation and diagrammatical emulation of (perspective and relevance driven) *information granulation* and *semantic inferencing* which operate on DDS-structures [8].

4.1 Granular structure and constraints

Dispositional dependency structures (DDS) (*Fig.* 1) can be viewed as an alternative procedural format of *fuzzy information granulation* which extends the rule-based frame as introduced by the concept of *generalized constraint* [12] and exemplified in [14] as *unconditional constraints*.

According to Zadeh's (1997) *theory of fuzzy information granulation* (TFIG), a generalized constraint on values of X is expressed as X *isr* R, where X is a variable which takes values in a universe of discourse U, *isr* is a parametric copula with r being a discrete variable whose values define the way in which R constrains X, and R is the constraining relation. For r different values may be defined as *equality, possibility, verity, probability, random set,* and *fuzzy graph*, and their (definitional, operational, procedural, computational) interpretations can be given.

From our perspective it is important to observe that r is a means to extend the copula's interpretations in a controlled and operationally defined way which relates to R in a predicative sense, i.e. specifying the interpretation of R (generally a distribution of grades of membership) as being *possibilities, truth values, probabilities* or compos-

[2] The text-linguistic term refers to the language environment (*cotext*) of an expression embedded in its discourse situation (*context*).

ites thereof. As these functional types of r needed to be specified for rule-based mechanisms in order to determine their different interpretations of R, this necessity may be relaxed or even become obsolete when the rule-based inference mechanism is replaced by an algorithmic procedure, operating on a well-defined structure like SHS as specified numerically by the value distributions which constitute the meaning points' interpretations.

4.2 Deriving semantic granules

Taking the concept of generalized constraints being applicable likewise for sentences (propositions) as well as for words (DDS), then the TFIG notational format translates to $X \simeq \{x_n\}$ where X is a variable which takes values—via α- and δ-abstraction—of $z_n \in \langle S \rangle$ with $S \subseteq U$. A semiotically generalized constraint on values of X is expressed by $X\ dds_i\ S$ where dds relates x_i via z_i to S by restricting SHS procedurally in generating the tree structure from meaning point z_i as its root, and z_n as its discrete variables whose values determine different structures (*dependency paths*) which constrain the topology of S in a semantically perspective way.

Thus, *dependency path*is a structural representation for a dynamic concept of *granular word meaning* which induces a reflexive, symmetric, and weakly transitive relation between relevant meaning points as its components, allowing for the procedural definition and computational enactment of *semantic* inferencing on the *word* level [8], very much like the *rule-based* models of inferencing in granular *fuzzy information* processing based on *fuzzy rules*, or the *syntagmatically* defined propositional formats of symbolic processing in (cognitive linguistic) sentence semantics based on *crisp* logic calculi.

In *Fig.* 1 the semantic hyperspace $\langle S, \zeta \rangle$ was computed from a corpus of Reuters 1987 newswire articles[3]. Two vocabulary items x_i = `administration`, x_j = `deposit`, corresponding to meaning points z_i, z_j were chosen as *premises* for the semantic inference process. It restricts $\langle S \rangle$ simultaneously by generating the graphs DDS_i, DDS_j in parallel. The inferred *conclusion* is the first common node z_k = `estate` whose different *dependency paths* $dep_i(z_k), dep_j(z_k)$ are given (center column). Depending on the semantic perspectives, however, as determined by the root node z_i, z_j respectively, the subtrees or *information granules* $ig_i(k), ig_j(k)$, headed by z_k = `estate` (left and right column) demonstrate the i and j induced differences both, in connotative meaning and in semantic resolution of these fuzzy information granules.

5 Conclusion

The dynamics of *semiotic* knowledge structures and the processes operating on them essentially consist in their recursively applied mappings of multi-level [2] representations resulting in a multi-resolutional [1] granularity of fuzzy word meanings which emerge from and are modified by such text processing. Computational tests and experiments with different PHT corpora have produced promising evidence on the SCIP system's granular meaning acquisition and language understanding capacity without any explicit initial morphological, lexical, syntactic, or semantic knowledge.

[3]Reuters-21578 (1.0) Text Categorization Test Collection, prepared by D.D.Lewis (AT&T Labs) and thankfully acknowledged here (`www.research.att.com/~lewis/reuters21578.html`).

References

[1] A. Meystel: *Semiotic Modeling and Situation Analysis.* Bala Cynwyd, PA (AdRem Inc), 1995.

[2] B. Rieger: *Unscharfe Semantik.* Frankfurt/ Bern/ Paris (Lang), 1989.

[3] B. B. Rieger: Distributed Semantic Representation of Word Meanings. In: Becker/Eisele/Mündemann (eds): *Parallelism, Learning, Evolution.* Berlin/ Heidelberg/ New York (Springer), 1991, pp. 243–273.

[4] B. B. Rieger: Meaning Acquisition by SCIPS. In: Ayyub (ed): *ISUMA/NAFIPS-95-Proceedings*, Los Alamitos, CA (IEEE), 1995, pp. 390–395.

[5] B. B. Rieger: Situations, Language Games, and SCIPS. In: Meystel/Nerode (eds): *Semiotic Modeling and Situation Analysis: 10th IEEE Symposium on Intelligent Control.* Bala Cynwyd, PA (AdRem), 1995, pp. 130–138.

[6] B. B. Rieger: Situation Semantics and Computational Linguistics: towards Informational Ecology. In: Kornwachs/Jacoby (eds): *Information. New Questions to a Multidisciplinary Concept*, Berlin (Akademie), 1996, pp. 285–315.

[7] B. B. Rieger: Computational Semiotics and Fuzzy Linguistics. In: Meystel (ed): *A Learning Perspective, ISAS-97-Proceedings*, Washington, DC (US Gov. Printing), 1997, pp. 541–551.

[8] B. B. Rieger: Tree-like Dispositional Dependency Structures for non-propositional Semantic Inferencing. In: Bouchon-Meunier/Yager (eds): *IPMU-98-Proceedings*, Paris (EKD), 1998, pp. 351–358.

[9] B. B. Rieger: Semiotics and Computational Linguistics. In: Zadeh/Kacprzyk (ed): *Computing with Words in Information/ Intelligent Systems I*, Heidelberg/ New York (Physica), 1999, pp. 93–118.

[10] B. B. Rieger: Computing Granular Word Meanings. In: Wang/Meystel/Albus (eds): *Computing with Words*, New York, NY (Wiley), 2000, [in print].

[11] B. B. Rieger/C. Thiopoulos: Semiotic Dynamics: a self-organizing lexical system in hypertext. In: Köhler/Rieger (eds): *QUALICO-91-Proceedings*, Dordrecht (Kluwer), 1993, pp. 67–78.

[12] L. A. Zadeh: Outline of a computational approach to meaning and knowledge representation. In: Thoma/Wyner (eds): *AI and Man-Machine Systems*, Heidelberg (Springer), 1986, pp. 198–211.

[13] L. A. Zadeh: Fuzzy logic = Computing with words. *IEEE-Transactions on Fuzzy Systems*, 4: 103–111, 1996.

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