Searching for Grammar Right

Vanessa Micelli
European Media Laboratory
Schloss-Wolfsbrunnenweg 33
69118 Heidelberg, Germany
`{firstname.lastname@eml-d.villa-bosch.de}`

Abstract

This paper describes our ongoing work in and thoughts on developing a grammar learning system based on a construction grammar formalism. Necessary modules are presented and first results and challenges in formalizing the grammar are shown up. Furthermore, we point out the major reasons why we chose construction grammar as the most fitting formalism for our purposes. Then our approach and ideas of learning new linguistic phenomena, ranging from holophrastic constructions to compositional ones, is presented.

1 Introduction

Since any particular language[1] changes constantly (Cf. Hopper and Traugott, 2003; Bybee, 1998) – and even varies across domains, users, registers etc. – scalable natural language understanding systems must be able to cope with language variation and change. Moreover, due to the fact that any natural language understanding system, which is based on some formal representation of that language's grammar, will always only be able to represent a portion of what is going on in any particular language at the present time, we need to find systematic ways of endowing natural language understanding systems with means of learning new forms, new meanings and, ultimately, new form-meaning pairings, i.e. constructions.

Constructions are the basic building blocks, posited by a particular grammar framework called Construction Grammar, and are defined as follows: "C is a construction iff$_{def}$ C is a form-meaning pair <F_i, S_i> such that some aspect of F_i or some aspect of S_i is not strictly predictable from C's component parts or from other previously established constructions." (Goldberg, 1995:4).

Construction Grammar originated from earlier insights in functional and usage-based models of language mainly supposed by cognitive linguists (e.g. Lakoff, 1987; Fillmore and Kay, 1987; Kay, 2002; Talmy, 1988; etc.). It has been devised to handle actually occurring natural language, which notoriously contains non-literal, elliptic, context-dependent, metaphorical or underspecified linguistic expressions. These phenomena still present a challenge for today's natural language understanding systems. In addition to these advantages, we adhere to principles proposed by other constructivists as e.g. Tomasello (2003) that language acquisition is a usage-based phenomenon, contrasting approaches by generative grammarians who assume an innate grammar (Chomsky, 1981). Furthermore, we agree to the idea that grammatical phenomena also contribute to the semantics of a sentence which is the reason why syntax cannot be defined independently of semantics of a grammar. A more detailed outline of construction grammar and the principles we adhered to in formalizing it will be given in sections 2 and 3.

The input to the system is natural language data as found on the web, as e.g. in news tickers or blogs, initially restricted to the soccer domain. As

---

[1] This claim also holds within any solidified system of conventionalized form-meaning pairings, e.g. dialects, chronolects, sociolects, idiolects, jargons, etc.

Proceedings of the 3rd Workshop on Scalable Natural Language Understanding, pages 57–64,
New York City, June 2006. ©2006 Association for Computational Linguistics

the learning process develops the input will gradually be extended to other domains. A description of the corpus and its selection process will be given in section 4. Section 5 provides an outlook on the learning paradigm, while the last section presents some future issues and conclusions.

2 Grammar Formalism

The most crucial foundation that is needed to build a grammar learning system is a grammar formalism. Therefore, we are designing a new formalization of construction grammar called ECtoloG (Porzel et al., 2006; Micelli et al., in press).

One existing formal computational model of construction grammar is the Embodied Construction Grammar (ECG) (Chang et al., 2002; Bergen and Chang, 2002), with its main focus being on language understanding and later simulation[2]. A congruent and parallel development has led to FCG which simulates the emergence of language (Steels, 2005). FCG is mainly based on the same primitives and operators as ECG is. We decided to employ ECG in our model mainly for historical reasons (see details about its development in the following section), adhering to its main primitives and operators, but employing the state of the art in knowledge representation. We adopt insights and mechanisms of FCG where applicable.

2.1 Construction Grammar and ECG

One main difference between West Coast Grammar (Langacker, 1987; Lakoff, 1987) and East Coast Grammar (Chomsky, 1965; Katz, 1972) is the fact that construction grammar offers a vertical – not a horizontal – organisation of any knowledge concerning a language's grammar. That is, that generative grammars split form from function. Syntax, morphology, a lexicon or other formal components of the grammar constitute form, while the conventional function is defined by semantics.

All constructions of a language, however, form in Langacker's terms "a structured inventory of conventional linguistic units" (Langacker, 1987:54). This inventory is network-structured, i.e. there are at least taxonomic links among the constructions (Diessel, 2004). This structure presents one of the main differences between generative and construction grammars (Croft, to appear). One of the most cited examples that evidences the necessity, that there can be no explicit separation between syntax and semantics, is Goldberg's example sentence (Goldberg, 1995:29):

(1) *he sneezed the napkin off the table.*

The whole meaning of this sentence cannot be gathered from the meanings of the discrete words. The direct object *the napkin* is not postulated by the verb *to sneeze*. This intransitive verb would have three arguments in a lexico-semantic theory: 'X causes Y to move Z by sneezing'. Goldberg states that the additional meaning of caused motion which is added to the conventional meaning of the verb *sneeze* is offered by the respective caused-motion construction. Based on this background ECG – a formal computational model of construction grammar – was developed within the Neural Theory of Language project (NTL) and the EDU project (EDU).

While other approaches consider language as completely independent from the organism which uses it, ECG claims that several characteristics of the user's sensorimotor system can influence his or her language (Gallese and Lakoff, 2005). The needed dynamic and inferential semantics in ECG is represented by embodied schemas. These schemas are known under the term of *image schemas* in traditional cognitive semantics and constitute schematic recurring patterns of sensorimotor experience (Johnson, 1987; Lakoff, 1987).

The current ASCII format of ECG is insufficient for building scalable NLU systems in the long run. Therefore, our attempt at formalizing construction grammar results in an ontological model that combines two ontological modeling frameworks endowed with a construction grammar layer, based on the main ideas behind ECG. The following section describes the resulting ontology, pointing out main challenges and advantages of that approach.

3 Formalizing Construction Grammar

The ontological frameworks mentioned above are *Descriptions & Situations* (D&S) (Gangemi and Mika, 2003) and *Ontology of Information Objects* (OIO) (Guarino, 2006), which both are extensions of the *Descriptive Ontology for Linguistic and*

---

[2] For a detailed ECG analysis of a declarative utterance, i.e. the sentence *Harry walked into the cafe*, see Bergen and Chang (2002).

Cognitive Engineering (DOLCE) (Masolo et al., 2003).

D&S is an ontology for representing a variety of reified contexts and states of affairs. In contrast to physical objects or events, the extensions of ontologies to the domain of non-physical objects pose a challenge to the ontology engineer. The reason for this lies in the fact that non-physical objects are taken to have meaning only in combination with some other *ground* entity. Accordingly, their logical representation is generally set at the level of theories or models and not at the level of concepts or relations (see Gangemi and Mika, 2003). It is, therefore, important to keep in mind that the meaning of a given linguistic expression emerges only through the combination of both linguistic and conceptual knowledge with "basic" ontological knowledge, as modeled in such ground ontologies.

Next to the support via dedicated editors and inference engines, one of the central advantages of our ensuing ontological model over the currently used ASCII-format of ECG lies in its compatibility with other ground ontologies developed within the Semantic Web framework.[3]

3.1 Modeling of Constructions

Constructions are modeled in the ECtoloG as information-objects. According to the specification of the OIO, information objects have – amongst others – the following properties: They are social objects realizable by some entity and they can express a description, which represents in this ontology the ontological equivalent of a meaning or a conceptualization. Since a construction constitutes a pairing of form and meaning according to the original theory of construction grammar, both properties are of advantage for our ontological model. To keep the construction's original structure, the form pole can be modeled with the help of the *realized-by* property[4] while the meaning pole is built via the *edns:expresses* property. Both processes are described more detailed in the following section.

Holophrastic Constructions

The class of lexical constructions is modeled as a subclass of *referringConstruction*. Since it is a subclass of the class *information-object* it inherits the *edns:expresses* property. The *referringConstruction* class has a restriction on this property that denotes, that at least one of the values of the *edns:expresses* property is of type *schema*. Modeling this restriction is done by means of the built-in *owl:someValuesFrom* constraint. The restriction counts for all constructions that *express* a *schema*. It has no effect on the whole class of *constructions*, i.e. it is possible that there exist constructions that do not *express* a single schema, as e.g. compositional ones, whose meaning is a composite of all constructions and schemas that constitute that compositional construction.

The form pole of each construction is modeled with the help of the *realized-by* property. This property designates that a (physical) representation – as e.g. the orthographic form of the construction – realizes a non-physical object – in this case our construction. This property is also inherited from the class *information-object,* the superclass of constructions. What fills the range of that property is the class of *edns:physical-realization*. Therefore, we define an instance of *inf:writing,* which then fills the form pole of the respective construction. This instance has once more a relation which connects it to instances of the class *inf:word* which designate the realization of the instance of the *inf:writing* class.

This way of modeling the form pole of each lexical construction enables us to automatically populate our model with new instances of constructions, as will be described more detailed in section 5.1.

Analogous to the modeling of meaning in the original ECG, the meaning pole is 'filled' with an instance of the class of *image schema*. This can be done with the help of the *edns:expresses* relation. This relation is defined, according to the specification of the D&S ontology, as a relation between information objects that are used as representations (signs) and their content, i.e. their meaning or conceptualization. In this ontology, content is reified as a description, which offered us the possibility to model image schemas as such. How image schemas are modeled will be described in section 3.2.

Compositional Constructions

Compositional constructions are constructions which are on a higher level of abstraction than holophrastic ones. This means, that there exist constructions which combine different constructions

---

[3] For more details see Porzel et al. (2006).

[4] We adhere to the convention to present both ontological properties, classes, and instances in italics.

into one unit. ECG designed a so-called constructional block, wherein several constructions are subsumed under and accessible in one more complex construction.

An example is the DetNoun construction, which combines a determiner and a noun to form one unit. There is the possibility to model different constraints both in the form pole and in the meaning pole of a construction. A form constraint applying to this exact construction is determining that the determiner comes before the noun. This understanding of *before* corresponds to Allen's definition of his interval relations (Allen, 1983), which states that they don't necessarily have to follow each other but that there could be some modifiers in between the two components of this construction.

A meaning constraint of this construction determines, that the meaning of the noun, used in this respective construction, is assigned to the meaning of the resulting complex construction.[5] To be able to represent these phenomena, we firstly defined a class *construction-parameter*, that denotes a subclass of *edns:parameter*, a subclass of *edns:concept*. There is a property restriction on the class that states that all values of the *requisite-for* property have to be of type *construction*. This determines instances of the class *construction-parameter* to be used only in constructions on a higher level of abstraction. All constructions used on level 0 of a grammar[6], i.e. lexical constructions, are at the same time instances of the class *construction-parameter* so that they can be used in more abstract constructions. The form and meaning constraints still need to be modeled in our framework. To determine which constructions are used in which more abstract construction, new properties are defined. These properties are subproperties of the *requisite-for* property. An example is the *requisite-detnoun-akk-sg* property. This property defines that the accusative singular determiner construction and the corresponding noun construction can be *requisite-for* the compositional construction that combines these two lexical constructions into one noun phrase.

3.2 Modeling of Image Schemas

Following Johnson and Lakoff (Johnson, 1987; Lakoff and Johnson, 1980; Lakoff, 1987) image schemas are schematic representations that capture recurrent patterns of sensorimotor experience. According to ECG, a schema is a description whose purpose is filling the meaning pole of a construction. It consists of a list of schematic roles that can serve as simulation parameters.

In ECG, schemas can be evoked by or can evoke other schemas, i.e. particular schematic-roles of another schema can be imported. A schema can, therefore, be defined against the background of another schema[7]. The property *evokes* and its inverse property *evoked-by* have been defined as subproperties of the *dol:generically-dependent-on* property and its inverse property *dol:generic-dependent* respectively. Generic dependence is defined in the DOLCE ontology as the dependence on an individual of a given type at some time.

The class of *image schemas* is modeled as a subclass of *edns:description* (see definition of description in 3.1), in order to enable being employed in the meaning pole of constructions.

Schematic Roles

The class of *schematic-roles* is a subclass of the *edns:concept* class. In the specification of D&S a concept is classified as a non-physical object which again is defined by a description. Its function is classifying entities from a ground ontology in order to build situations that can satisfy the description. Schematic roles are parameters that allow other schemas or constructions to refer to the schema's key variable features, e.g. the role of a trajector in a Trajector Landmark-Schema can be played by the same entity that denotes the mover in e.g. a caused-motion schema.

At the moment, they are modeled with the help of the *edns:defines* property. A schema defines its schematic roles with this property, denoting a subproperty of the *edns:component* property. According to the D&S specification, a component is a proper part with a role or function in a system or a context. It is also stated, that roles can be different for the same entity, and the evaluation of them changes according to the kind of entity. This means, that instances of the class *schema* and its

---

[5] For further information about which operators are used to model these features in ECG we refer to Bergen and Chang (2002), Chang et al. (2002) and Bryant (2004).

[6] Following Bryant's (2004) division of constructions into 5 levels of different degrees of schematicity.

[7] To clarify this claim see Langacker's hypotenuse example (Langacker, 1987:183ff.).

subclasses can have instances of the class *schematic-role* as their components. The *schematic-roles* class has to fulfil the necessary condition, that at least one of the values of the *edns:defined-by* property is of type *schema*.

The domain of the *defines* property is a *description* (which can be our schemas) and its range is set to either *concepts* or *figures* (which are our schematic roles). The problem occurring hereby is that the roles cannot be filled by complete classes which is necessary in a lot of cases, since the parameters are not always filled with atomic values but possibly with whole classes of entities. Therefore, one could think about modeling schematic roles as properties, setting the domain on the corresponding *schema* class and the range on the corresponding class whose subclasses and instances can possibly fill its range.

3.3 Linguistic Information

Since linguistic information as e.g. grammatical gender, its case, or the part-of-speech of a word is needed for analyzing natural language texts, this information has to be modeled, as well, in the ECtoloG. Therefore, we integrated the LingInfo model (Buitelaar et al., 2006) into the ECtoloG.

LingInfo constitutes an ontological model that provides other ontologies with linguistic information for different languages, momentarily for English, French, and German. Main objective of this ontology is to provide a mapping between ontological concepts and lexical items. That is, that the possibility is offered to assign linguistic information as e.g. the orthographic term, its grammatical gender, its part-of-speech, stem etc. to classes and properties. For our purposes, the LingInfo ontology had to be converted from RDFS into OWL-DL format and then integrated into the ECtoloG. For that reason, a new subclass of *owl:class* was defined: *ClassWithLingInfo*. Instances of this metaclass are linked through the *linginfo* property to *LingInfo* classes. The *LingInfo* class is used to associate a term, a language, and morphosyntactic information to classes from the ground ontology; e.g. a class *CafeConstruction*, which is an instance of *ClassWithLingInfo,* from an ontology proper, can be associated through the property *linginfo* with *Café*, an instance of the class *LingInfo*. Thus, the information that the term is *German*, its part-of-speech is *noun* and its grammatical gender *neuter* is obtained.

Following this approach, our classes of lexical constructions were defined as subclasses of *ClassWithLingInfo,* being thereby provided with all the necessary linguistic information as defined above. The central challenge resulting from this approach is, that through the definition of a metaclass the ontological format is no longer OWL-DL but goes to OWL-Full which thwarts the employment of Description Logic reasoners. Reasoning will not stay computable and decidable. Future work will address this challenge by means of intertwining the LingInfo model with the ECtoloG grammar model in such a way, that the computational and inferential properties of OWL-DL remain unchallenged.

Another possibility could be obtaining linguistic information for lexical items through an external lexicon.

4 The Web as a Corpus

The Seed Corpus *C:* The primary corpus *C* in this work is the portion of the World Wide Web confined to web pages containing natural language texts on soccer. To extract natural language texts out of web documents automatically we are using wrapper agents that fulfil this job (see Porzel et al., 2006). Our first goal is to build a grammar that can deal with all occurring language phenomena – i.e. both holophrastic and compositional ones – contained in that corpus *C*.

Corpus *C'*: Next step is the development of a corpus *C'*, where $C' = C + \varepsilon$ and ε is constituted by a set of new documents. This new corpus is not designed in an arbitrary manner. We search similar pages, adding add them to our original corpus *C*, as we expect the likelihood of still pretty good coverage together with some new constructions to be maximal, thereby enabling our incremental learning approach. The question emerging hereby is: what constitutes a similar web page? What, therefore, has to be explored are various similarity metrics, defining similarity in a concrete way and evaluate the results against human annotations (see Papineni et al., 2002).

4.1 Similarity Metric

To be able to answer the question which texts are actually similar, similarity needs to be defined precisely. Different approaches could be employed, i.e. regarding similarity in terms of syntactic or semantic phenomena or a combination of both. Since construction grammar makes no separation between syntax and semantics, phenomena that should be counted are both constructions and image schemas. As for holophrastic constructions this presents less of a challenge, we rather expect counting compositional ones being a ‘tough cookie’.

To detect image schemas in natural text automatically, we seek to employ different methodologies, e.g. LSA (Kintsch and van Dijk, 1978), using synonym sets (Fellbaum, 1998) or other ontologies, which could assist in discovering the semantics of an unknown word with its corresponding schematic roles and the appropriate fillers. This or a similar methodology will be applied in the automatic acquisition process as well.

Another important point is that some terms, or some constructions, need to get a higher relevance factor than others, which will highly depend on context. Such a relevance factor can rank terms or constructions according to their importance in the respective text. Ranking functions that can be examined are, e.g., the TF/IDF function (e.g. Salton, 1989) or other so called *bag of words* approaches.

Term statistics in general is often used to determine a scalable measure of similarity between documents so it is said to be a good measure for topical closeness. Also part-of-speech statistics could be partly helpful in defining similarity of documents based on the ensuing type/token ratio.

The following five steps need to be executed in determining the similarity of two documents:

Step 1: Processing of the document D; analyzing the text and creating a list of all occurring words, constructions and/or image schemas. We assume that the best choice is counting constructions and corresponding image schemas, since they represent the semantics of the given text.

Step 2: Weighing of schemas and constructions

Step 3: Processing of the document D+1; executing of step 1 and 2 for this document.

Step 4: Comparing the documents; possibly adding synonyms of sources as e.g. WordNet (Fellbaum, 1998).

Step 5: Calculating the documents’ similarity; defining a threshold up to which documents are considered as being similar. If a document is said to be similar, it is added to the corpus, which becomes the new corpus *C’*.

Analysis of the New Corpus C’: The new corpus *C’* is analyzed, whereby the coverage results in coverage *A* of *C’* where:

$$A = 100\% - (\delta_h + \delta_c)$$

δ_h denotes all the holophrastic phenomena and δ_c all compositional phenomena not observed in *C.*

5 Grammar Learning

To generate a grammar that covers this new corpus *C’* different strategies have to be applied for holophrastic items δ_h which are lexical constructions in our approach and for compositional ones δ_c – meaning constructions on a higher level of abstraction as e.g. constructions that capture grammatical phenomena such as noun phrases or even whole sentences.

5.1 Learning Lexical Constructions

Analogous to the fast mapping process (Carey, 1978) of learning new words based on exposure without additional training or feedback on the correctness of its meaning, we are employing a method of filling our ontology with whole paradigms of new terms[8], enabled through the modeling of constructions described in 3.1. First step herein is employing a tool – Morphy (Lezius, 2002) – that enables morphological analysis and synthesis. The analysis of a term yields information about its stem, its part-of-speech, its case, its number, and its grammatical gender. This information can then easily be integrated automatically into the ECtoloG.

As already mentioned in section 4.3, we are not only trying to automatically acquire the form pole of the constructions, but also its image schematic meaning, that means the network of the schemas that hierarchically form the meaning pole of such a term, applying ontology learning mechanisms (e.g. Loos, 2006) and methods similar to those described in section 4.3. Additionally, investigations are underway to connect the grammar learning framework proposed herein to a computer vision system that provides supplementary feedback con-

---

[8] We are aware of the fact that fast mapping in humans is limited to color terms, shapes or texture terms, but are employing the method on other kinds of terms, nevertheless, since the grammar learning paradigm in our approach is still in its baby shoes.

cerning the hypothesized semantics of individual forms in the case of multi-media information.

5.2 Learning Compositional Constructions

Learning of compositional constructions still presents an issue which has not been accounted for, yet. What has already been proposed (Narayanan, inter alia) is that we have to assume a strong inductive bias and different learning algorithms, as e.g. some form of Bayesian learning or model merging (Stolcke, 1994) or reinforcement learning (Sutton and Barto, 1998).

Another important step that has to be employed is the (re)organization of the so-called constructicon, i.e. our inventory of constructions and schemas. These need to be merged, split or maybe thrown out again, depending on their utility, similarity etc.

5.3 Ambiguity

Currently the problem of ambiguity is solved by endowing the analyzer with a chart and employing the semantic density algorithm described in (Bryant, 2004). In the future probabilistic reasoning frameworks as proposed by (Narayanan and Jurafsky, 2005) in combination with ontology-based coherence measures as proposed by (Loos and Porzel, 2004) constitute promising approaches for handling problems of construal, whether it be on a pragmatic, semantic, syntactic or phonological level.

6 Concluding Remarks

In this paper we described our ongoing work in and thoughts on developing a grammar learning system based on a construction grammar formalism used in a question-answering system. We described necessary modules and presented first results and challenges in formalizing construction grammar. Furthermore, we pointed out our motivation for choosing construction grammar and the, therefore, resulting advantages. Then our approach and ideas of learning new linguistic phenomena, ranging from holophrastic constructions to compositional ones, were presented. What should be kept in mind is that our grammar model has to be strongly adaptable to language phenomena, as e.g. language variation and change, maps, metaphors, or mental spaces.

Evaluations in the light of the precision/coverage trade-off still present an enormous challenge (as with all adaptive and learning systems). In the future we will examine the feasibility of adapting ontology evaluating frameworks, as e.g. proposed by Porzel and Malaka (2005) for the task of grammar learning. We hope that future evaluations will show that our resulting system and, therefore, its grammar will be robust and adaptable enough to be worth being called ‘Grammar Right’.

References

Allen, J.F. 1983. Maintaining Knowledge about Temporal Intervals. *Communications of the ACM, 26(11).*

Bergen, B. and Chang, N. 2002. Simulation-Based Language Understanding in Embodied Construction Grammar. ICSI TR-02-004, Berkeley, CA, USA.

Bryant, J. 2004. Recovering coherent interpretations using semantic integration of partial parses. *Proceedings of the 3rd ROMAND workshop*, Geneva, Switzerland.

Buitelaar, P., Declerck, T., Frank, A., Racioppa, S., Kiesel, M., Sintek, M., Engel, M., Romanelli, M., Sonntag, D., Loos, B., Micelli, V., Porzel, R. and Cimiano, P. 2006. LingInfo: Design and Applications of a Model for the Integration of Linguistic Information in Ontologies. *Proceedings of OntoLex 2006*. In Press. Genoa, Italy.

Bybee, J. 1998. A functionalist approach to grammar and its evolution. *Evolution of Communication 2(2).*

Carey, S. 1978. The child as word-learner. *Linguistic theory and psychological reality*. MIT Press, Cambridge, MA.

Chang, N., Feldman, J., Porzel, R. and Sanders, K. 2002. Scaling Cognitive Linguistics: Formalisms for Language Understanding. *Proceedings of the 1st ScaNaLU Workshop*, Heidelberg, Germany.

Chomsky, N. 1965. *Aspects of the theory of syntax*. MIT Press. Cambridge, Mass.

Croft, W. To appear. Logical and typological arguments for Radical Construction Grammar. *Construction Grammar(s): Cognitive and cross-language dimensions (Constructional approaches to Language, 1).* John Benjamins, Amsterdam.

Diessel, H. 2004. The Acquisition of Complex Sentences. *Cambridge Studies in Linguistics (105).* Cambridge University Press, Cambridge.

EDU: *Even Deeper Understanding* http://www.eml-development.de/english/research/edu/index.php (last access: 31/03/06).

Fellbaum, C. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, Mass.

Fillmore, C. and Kay, P. 1987. The goals of Construction Grammar. *Berkeley Cognitive Science Program TR 50*. University of California, Berkeley.

Gallese, V. and Lakoff, G. 2005. The brain´s concepts: the role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology 21/2005*.

Gangemi A., Mika P. 2003. Understanding the Semantic Web through Descriptions and Situations. *Proceedings of ODBASE03 Conference*, Springer.

Goldberg, A. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. University of Chicago Press. Chicago.

Guarino, N. 2006. Ontology Library. WonderWeb Deliverable D202, I STC-CNR, Padova, Italy. www.loa-cnr.it/Papers/Deliverable%202.pdf (last access: 31/03/2006).

Hopper, P. and Traugott, E. 2003. *Grammaticalization*. Cambridge University Press. Cambridge, UK.

Johnson, M. 1987. *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. University of Chicago Press. Chicago.

Katz, J. J. 1972. *Semantic theory*. Harper & Row. New York.

Kay, P. 2002. An Informal Sketch of a Formal Architecture for Construction Grammar. *Grammars 1/5*.

Kintsch, W. and van Dijk, T. 1978. Toward a model of text comprehension and production. *Psychological Review, 85 (5)*.

Lakoff, G. and Johnson, M. 1980. *Metaphors We Live By*. Chicago University Press. London.

Lakoff, G. 1987. *Women, Fire, and Dangerous Things*. University of Chicago Press. Chicago and London.

Langacker, R. 1987. *Foundations of Cognitive Grammar, Vol. 1*. University Press. Stanford.

Lezius, W. 2000. Morphy - German Morphology, Part-of-Speech Tagging and Applications. *Proceedings of the 9th EURALEX International Congress*, Stuttgart, Germany.

Loos, B. 2006. Scaling natural language understanding via user-driven ontology learning. *This volume*.

Loos, B. and Porzel, R. 2004. Resolution of Lexical Ambiguities in Spoken Dialogue System. *Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue*, Cambridge, Massachusetts, USA.

Masolo, C., Borgo, S., Gangemi, A., Guarino, N., and Oltramari, A. 2003. Ontology Library. WonderWeb Deliverable D18, I STC-CNR, Padova, Italy. http://wonderweb.semanticweb.org/deliverables/documents/D18.pdf.

Micelli, V., Porzel, R., and Gangemi, A. In Press. ECtoloG: Construction Grammar meets the Semantic Web. *Proceedings of ICCG4*, Tokyo, Japan.

Narayanan, S. and Jurafsky, D. 2005. *A Bayesian Model of Human Sentence Processing* (in preparation).

Narayanan, S. 2006. Lecture Series given at Interdisciplinary College, Guenne, Lake Moehne, March 10th – 17th.

NTL: *Neural Theory of Language* http://www.icsi.berkeley.edu/NTL/

Papineni, K., Roukos, S., Ward, T. and Zhu, W. 2002. Bleu: a Method for Automatic Evaluation of Machine Translation. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, Philadelphia.

Porzel, R. and Malaka, R. 2005. A Task-based Framework for Ontology Learning, Population and Evaluation. Ontology Learning from Text: Methods, Evaluation and Applications Frontiers. *Artificial Intelligence and Applications Series, Vol. 123,* IOS Press.

Porzel, R., Micelli, V., Aras, H., and Zorn, H.-P. 2006. Tying the Knot: Ground Entities, Descriptions and Information Objects for Construction-based Information Extraction. *Proceedings of OntoLex 2006*. In Press. Genoa, Italy.

Salton, G. 1989. *Automatic Text Processing: the Transformation, Analysis, and Retrieval of Information by Computer*. Addison-Wesley. Reading, MA.

Steels, L. 2005. The Role of Construction Grammar in Fluid Language Grounding. Submitted. Elsevier Science.

Stolcke, A. 1994. Bayesian Learning of Probabilistic Language Models. *Ph.D. Thesis*, Computer Science Division, University of California at Berkeley.

Sutton, R. and Barto, A. 1998. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge.

Talmy, L. 1988. Force dynamics in language and cognition. *Cognitive Science, 12*.

Tomasello, M. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press.