Evaluating a semantic algebra on a large grammar of English

Dan Flickinger

Stanford University (ret'd), USA, danflick@alumni.stanford.edu

Keywords: Computational linguistics, semantic algebra

One of the ambitious goals of computational linguists is to develop implementations of the grammars of human languages, explicitly modeling for each particular language the lexicon and the inventory of syntactic rules used to compose the well-formed phrases and sentences of that language. The grammar of a language can be taken to define a mapping from form (the observable sequence of words in a sentence) to meaning (the semantic content communicated by the sentence), and thus a grammar implementation should also aim to make explicit the rules of semantic composition into some target logical formalism. Ideally, this method for constructing the meaning of each phrase from its parts would pair each syntactic rule of a language-specific grammar with one or more composition operations drawn from a small library of formally precise and language-independent operations. Such a semantic algebra was proposed by Copestake et al. (2001) for a formalism called Minimal Recursion Semantics (MRS: Copestake et al. 2005) which has been used for grammar implementations of a variety of languages, and its basic operations were illustrated using an early version of the English Resource Grammar (ERG: Flickinger 2000, 2011).

In this talk I will present ongoing collaborative work with Stephan Oepen and Emily M. Bender (see Flickinger et al. 2014) to extend and revise this semantic algebra to accommodate most of the wide syntactic and semantic coverage in MRS of the current ERG, and to identify those constructions in the grammar which cannot be readily shown to be compliant with the algebra. We have also implemented a language-independent model of this revised algebra in order to evaluate the compliance of the ERG with the algebra over a representative set of example sentences illustrating the broad range of syntactic phenomena and their corresponding MRS logical forms computed using the grammar. Our goal is to gain a better understanding of the degree to which a small set of composition operations suffices to construct the meanings of the many construction types in a language, and to focus attention on non-compliant constructions defined in the ERG in order to see whether it is the algebra which should be revised, or the grammar's analysis of those constructions.

Bibliography

Copestake, Ann, Dan Flickinger, Carl Pollard and Ivan A. Sag (2005) "Minimal Recursion Semantics. An Introduction," in the Journal of Research on Language and Computation 3:4, pp. 281-332.

Copestake, Ann, Alex Lascarides and Dan Flickinger (2001) "An Algebra for Semantic Construction in Constraint-based Grammars," in Proceedings of the 39th Conference of the Association for Computational Linguistics, pp. 140-147. Toulouse, France.

Flickinger, Dan (2000) "On Building a More Efficient Grammar by Exploiting Types," in Natural Language Engineering 6:1, pp. 15-28.

Flickinger, Dan (2011) "Accuracy vs. Robustness in Grammar Engineering," in E.M. Bender and J.E. Arnold, eds., Language from a Cognitive Perspective: Grammar, Usage, and Processing, pp. 31-50.

Flickinger, Dan, Emily M. Bender and Stephan Oepen (2014) "Towards an Encyclopedia of Compositional Semantics. Documenting the Interface of the English Resource Grammar," in Proceedings of the 9th International Conference on Language Resources and Evaluation, pp. 875-881. Reykjavik, Iceland.