Dependency Representations, Grammars, Folded Structures, among Other Things![*]

Abstract of invited talk

Aravind K. Joshi
Department of Computer and Information Science
and Institute for Research in Cognitive Science
University of Pennsylvania
Philadelphia PA USA
joshi@seas.upenn.edu

In a dependency grammar (DG) dependency representations (trees) directly express the dependency relations between words. The hierarchical structure emerges out of the representation. There are no labels other than the words themselves. In a phrase structure type of representation words are associated with some category labels and then the dependencies between the words emerge indirectly in terms of the phrase structure, the non-terminal labels, and possibly some indices associated with the labels. Behind the scene there is a phrase structure grammar (PSG) that builds the hierarchical structure. In a categorical type of grammar (CG), words are associated with labels that encode the combinatory potential of each word. Then the hierarchical structure (tree structure) emerges out of a set of operations such as application, function composition, type raising, among others. In a tree-adjoining grammar (TAG), each word is associated with an elementary tree that encodes both the hierarchical and the dependency structure associated with the lexical anchor and the tree(s) associated with a word. The elementary trees are then composed with the operations of substitution and adjoining. In a way, the dependency potential of a word is localized within the elementary tree (trees) associated with a word. Already TAG and TAG like grammars are able to represent dependencies that go beyond those that can be represented by context-free grammars, but in a controlled way. With this perspective and with the availability of larger dependency annotated corpora (e.g. the Prague Dependency Treebank) one is able to assess how far one can cover the dependencies that actually appear in the corpora. This approach has the potential of carrying out an 'empirical' investigation of the power of representations and the associated grammars. Here by 'empirical' we do not mean 'statistical or distributional' but rather in the sense of covering as much as possible the actual data in annotated corpora!

If time permits, I will talk about how dependencies are represented in nature. For example, grammars have been used to describe the folded structure of RNA biomolecules. The folded structure here describes the dependencies between the amino acids as they appear in an RNA biomolecule. One can then ask the question: Can we represent a sentence structure as a folded structure, where the fold captures both the dependencies and the structure, without any additional labels?

[*] Part of this work is in cooperation with Joan Chen Main, University of Pennsylvania, Philadelphia PA and Johns Hopkins University Baltimore, MD.

Proceedings of the Second International Conference on Dependency Linguistics (DepLing 2013), page 12,
Prague, August 27–30, 2013.