Lexicalized Tree Adjoining Grammar

Valerii Maliulin

Abstract

Tree-adjoining grammar (TAG) is a grammar formalism defined by Aravind Joshi. The elementary objects manipulated by a tree-adjoining grammar are trees. The properties of TAGs relate directly to the strong generative capacity (structural description) which is more relevant to linguistic descriptions than the weak generative capacity (set of strings). The tree sets of TAGs are not recognizable sets but are equivalent to the tree sets of linear indexed languages. Hence, tree-adjoining grammars generate some context-sensitive languages. However, tree-adjoining languages are strictly contained in the class of indexed languages.

The lexicalization of grammar formalisms is of linguistic and formal interest. There's an idea that rules should not be separated totally from their lexical realization. In this "lexicalized" approach, each elementary structure is systematically associated with a lexical anchor. These structures specify extended domains of locality (as compared to Context Free Grammars) over which constraints can be stated.

The process of lexicalization of context-free rules forces us to use operations for combining structures that make the formalism fall in the class of mildly context sensitive languages. Substitution and adjunction give us the freedom to lexicalize CFGs. Elementary structures of extended domain of locality, when they are combined with substitution and adjunction, yield Lexicalized TAGs. TAGs were so far introduced as an independent formal system. They derive from the lexicalization process of context-free grammars. Also, TAGs are closed under lexicalization.

It is still an open problem whether or not adjoining is the 'minimal' operation needed for lexicalizing CFGs, i.e., whether there exists a tree-gluing operation say Φ such that substitution and Φ can lexicalize any CFG, and such that the tree sets of the tree system with substitution and Φ are properly contained in the tree sets of TAG.