

From Strings to Trees to Strings to Trees ...

(Abstract)

Aravind K. Joshi

Dept. of Computer and Information Science
University of Pennsylvania, Philadelphia PA 19104

Sentences are not just strings of words (or are they?), they have some (hierarchical) structure. This much is accepted by all grammar formalisms. But how much structure is needed? The more the sentences are like strings the less the need for structure.

A certain amount of structure is necessary simply because a clause may embed another clause, or one clause may attach to another clause or parts of it. Leaving this need of structure aside, the question then is how much structure should a (minimal) clause have? Grammar formalisms can differ significantly on this issue. Minimal clauses can be just strings, or words linked by dependencies (dependency trees), or with rich phrase structure trees, or with flat (one level) phrase structure trees (almost strings) and so on. How much hierarchical structure is needed for a minimal clause is still an open question, that is being debated heatedly. How are clauses put together? Are these operations more like string manipulations (concatenation,

insertion, or wrapping, for example) or are they more like tree transformations (generalized transformations of the early transformational grammars, for example)? Curiously, the early transformational grammars, although clearly using tree transformations, actually formulated the transformations as pseudo string-like operations! More recent non-transformational grammars differ significantly with respect to their use of string rewriting or tree rewriting operations.

Grammar formalisms differ with respect to their stringiness or treeness. Also during their evolution, they have gone back and forth between string-like and tree-like representations, often combining them in different ways. These swings are a reflection of the complex interplay between aspects of language structure such as constituency, dependency, dominance, locality of predicates and their arguments, adjacency, order, and discontinuity. We will discuss these issues in an informal manner, in the context of a range of formalisms.