A TAG-grammar for "Swiss-style" cross-serial dependencies

Jan Strunk, January 21, 2004

1 Motivation

Shieber(1985) proved that Swiss-German subordinate clauses exhibit cross-serial dependencies of the form $wa^m b^n x c^m d^n y$. The string set of Swiss-German can thus not be generated by a context-free phrase-structure grammar. However, there exists a class of so-called *mildly context-free languages* which are a proper subset of the context-sensitive languages and contain the context-free languages as a proper subset. One example is the so-called Tree Adjoining Grammar (TAG) described in (Joshi & Schabes 1997). As Joshi & Schabes (1997) show that TAG can model cross-serial dependencies for a language $a^n e b^n$, I want to show here — as an exercise — that a TAG can also model cross-serial dependencies of the form needed for Swiss-German[1].

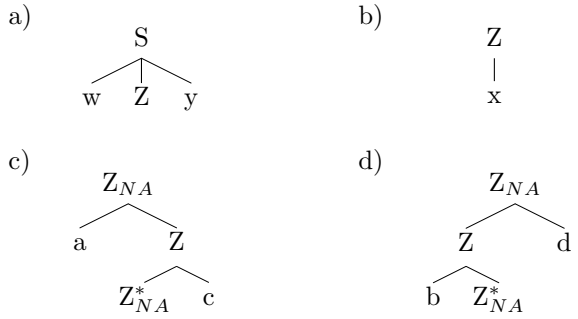
2 Short Introduction to TAG

Tree adjoining grammars manipulate trees as their basic objects. There exist two kinds of so-called *elementary trees*: *initial trees* which consist of interior non-terminal symbols and terminal and/or non-terminal symbols on the frontier; *auxiliary trees* are basically the same except for the fact that they must have a non-terminal node on the frontier, the so-called *foot node* (marked by a *) which must have the same label as its root node. The *substitution* operation can replace a non-terminal node on the frontier of an initial tree with an initial tree that has a root of the same label. *Adjunction* can be used to splice trees and insert an auxiliary tree at a node which may be an internal node and which has the same label as the root and foot node of the auxiliary tree provided the node at which the tree is spliced is not marked as unavailable for adjunction (with the symbol NA).

3 A simple model grammar

In this section, I will propose a small model grammar which generates the language $L = \{wa^m b^n x c^m d^n y | m, n \in \mathbf{N}\}$. I assume the following trees:

[1] This fact is of course already known. However, there is no proof of this in the paper by Joshi & Schabes (1997).



The derivation has to start with tree a) as it is the only one with the root node S so that the symbols w and y are always generated once on the absolute left and right side respectively. Next tree b) has to be substituted for the Z node on tree a) because it cannot later replace the Z on the frontier of the non-initial trees c) or d). If n and m are both zero, no further operation is need to yield the string wxy. If m or n or both are ≥ 1, the tree constructed so far can be spliced at the internal node Z and either tree c) or tree d) can be adjoined. After this operation, the top most Z node is unavailable for further adjunction because the root nodes of both c) and d) which are inserted are marked as NA. Further adjunctions are now only possible on the internal Z node of the auxiliary tree that has just been inserted. x will always be in the center of the string because the two auxiliary trees c) and d) which can be adjoined always introduce two non-terminals; one on both sides of x. As tree c) always introduces a and c at the same time, the number of a and c will always be equal. The same is true for b and d because of tree d). Moreover, because of the linear precedence encoded in the auxiliary trees, the sequence will always be $a \prec b \prec c \prec d$. It can thus be shown that the grammar above generates the language $L = \{wa^m b^n x c^m d^n y | m, n \in \mathbf{N}\}$. The example string *waaabbxcccddy* can be produced by first substituting tree b) into tree a), then adjoining tree c) three times and tree d) two times.

4 Linguistic relevance

One of the main advantages of TAG is that dependencies which are treated as non-local in other formalism are treated as local dependencies in TAG. Suppose a in the toy language above stands for a verb that has a dative object and c for a dative noun phrase, then the relation of government between them can easily by modeled by the TAG even though material that is not an argument of the verb a occurs between it and its dative object.

Joshi, A.K. and Schabes, Y. (1997): Tree-Adjoining Grammars. In: Rozenberg, G. and Salomaa, A. (eds.): Handbook of Formal Languages, Vol. 3, Springer, Berlin, New York, pp. 69-124.