9 Multicomponent TAG

Weir (1988)

First introduced in Joshi et al. (1975) as *simultaneous TAG*, later redefined as *multicomponent TAG (MCTAG)*. Idea: Instead of single elementary trees, the grammar contains (finite) sets of elementary trees. In each derivation step, a new set is picked and all trees from the set are added simultaneously.

Definition 18 (MCTAG) *An MCTAG is a tuple $G = \langle N, T, I, A, O, C, \mathcal{A} \rangle$ such that:*

- $G_{TAG} := \langle N, T, I, A, O, C \rangle$ *is a TAG with adjunction constraints, and*
- $\mathcal{A} \subseteq P(I \cup A)$ *is a set of subsets of $I \cup A$, the set of elementary tree sets.*[14]

$\gamma \Rightarrow \gamma'$ *is a derivation step in G iff there is an instance $\{\gamma_1, \ldots, \gamma_n\}$ of an elementary tree set in $\mathcal{A}$ and there are pairwise different node addresses $p_1, \ldots, p_n$ such that $\gamma' = \gamma[p_1, \gamma_1] \ldots [p_n, \gamma_n]$.*

As in TAG, a derivation starts from an initial tree and in the end, in the final derived tree, all leaves must have terminal labels (or the empty word) and there must not be any OA constraints left.

MCTAGs are linguistically interesting because they extend the domain of locality.

Example: extraction out of complex NPs Kroch (1989)

(49) which painting did you see a picture of

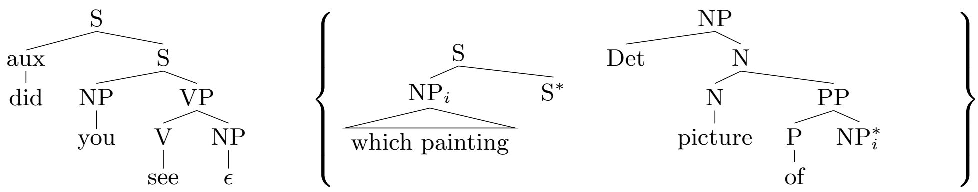


Figure 15: MCTAG elementary trees for extraction from NP

An MCTAG is called *tree-local* iff in each derivation step, the nodes the new trees attach to belong to the same elementary tree. It is called *set-local* iff in each derivation step, the nodes the new trees attach to belong to the same elementary tree set. Otherwise it is called *non-local*.

Proposition 23 *Tree-local MCTAG are strongly equivalent to TAG.*

For a given tree-local MCTAG, a strongly equivalent TAG can be easily constructed adopting corresponding adjunction constraints that enforce the simultaneous adjunctions of all elementary trees from a tree set.

In the following, "MCTAG" without specification of the locality means "set-local MCTAG".

For MCTAG *derivation trees* can be defined as follows (see Weir 1988):

- each node is an ordered elementary tree set (the initial tree the derivation starts with is considered as a unary set),
- each edge represents the simultaneous adjunctions of the trees from the daughter tree set to nodes in the trees in the mother tree set; an edge is equipped with a tuple of n node positions where n is the number of trees in the daughter set. Each node position is of the form $\langle \gamma, p \rangle$ where γ is one of the trees in the mother set and p a position in γ.

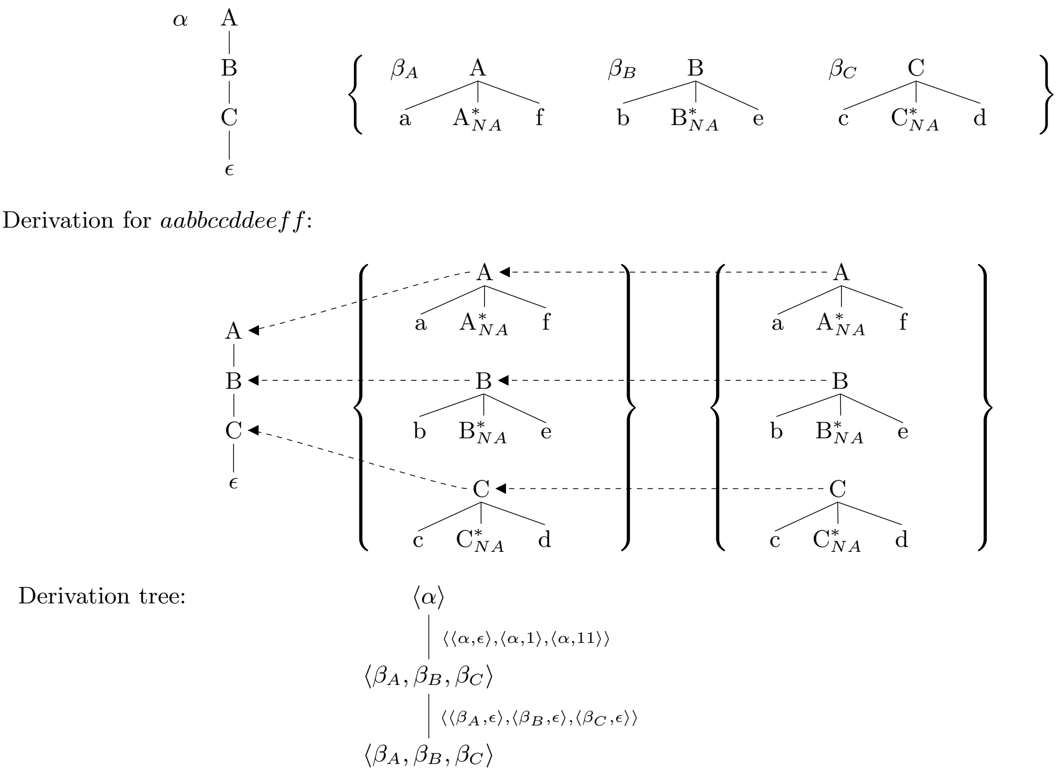


Figure 16: MCTAG for $L_6 = \{a^n b^n c^n d^n e^n f^n \mid n \geq 0\}$ with sample derivation

See for example the MCTAG in Fig. 16 with the derivation and the derivation tree for $aabbccddeeff$.

(Such derivation trees exist of course only for tree-local and set-local MCTAG, not for non-local MCTAG.)

MCTAG can be defined as LCFRS.

Example: LCFRS for the MCTAG in Fig. 16.

- GCFG: $\alpha \to f_\alpha(), \alpha \to g_\alpha(\beta_{A,B,C}), \beta_{A,B,C} \to f_{A,B,C}(), \beta_{A,B,C} \to g_{A,B,C}(\beta_{A,B,C})$.
- Denotation: $[\![f_\alpha()]\!] := \alpha$, $[\![f_{A,B,C}()]\!] := \langle \beta_A, \beta_B, \beta_C \rangle$,
  $[\![g_\alpha(X)]\!] := \alpha[\epsilon, \beta_1][1, \beta_2][11, \beta_3]$ where $[\![X]\!] = \langle \beta_1, \beta_2, \beta_3 \rangle$,
  $[\![g_{A,B,C}(X)]\!] := \langle \beta_A[\epsilon, \beta_1], \beta_B[\epsilon, \beta_2], \beta_C[\epsilon, \beta_3] \rangle$ where $[\![X]\!] = \langle \beta_1, \beta_2, \beta_3 \rangle$,
- Yield: $\phi(f_\alpha()) := \langle \epsilon \rangle$, $\phi(f_{A,B,C}()) := \langle a, b, c, d, e, f \rangle$,
  $\phi(g_\alpha(t)) := \langle w_1 w_2 w_3 w_4 w_5 w_6 \rangle$ where $\langle w_1, w_2, w_3, w_4, w_5, w_6 \rangle = \phi(t)$,
  $\phi(g_{A,B,C}(t)) := \langle w_1 a, w_2 b, w_3 c, d w_4, e w_5, f w_6 \rangle$ where $\langle w_1, w_2, w_3, w_4, w_5, w_6 \rangle = \phi(t)$

The construction of a strongly equivalent LCFRS for a given MCTAG is possible in general. The yield of an elementary set with n auxiliary trees has $2n$ components $\Rightarrow$ an MCTAG with n being the maximal number of elementary trees per tree set is in LCFRS($2n$).

Proposition 24 *For each MCTAG there is a strongly equivalent LCFRS.*

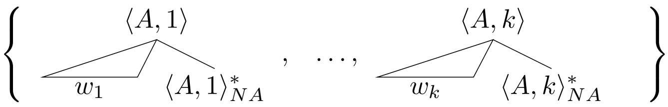
On the other hand, for each LCFRS, one can construct a weakly equivalent MCTAG:

Proposition 25 *For each LCFRS, there is a weakly equivalent MCTAG.*

[14] $P(X)$ is the set of subsets of some set X.

The construction goes as follows: Assume without loss of generality that all yields are k-tuples for some $k > 0$.

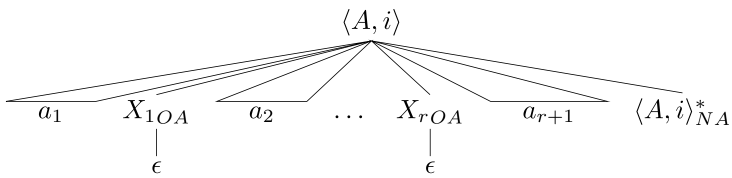
- For each production of the form $A \to f()$ in the LCFRS with $\phi(f()) = \langle w_1, \dots, w_k \rangle$ there is an elementary tree set of the form

  (the letters of the w_i are immediate daughters of the root nodes)

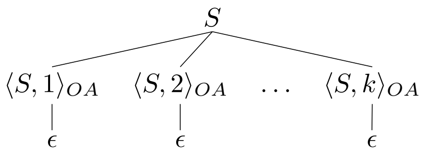
- For each production of the form $A \to f(A_1, \dots, A_n)$ in the LCFRS with $\phi(f(A_1, \dots, A_n)) = \langle y_1, \dots, y_k \rangle$ there is an elementary tree set containing k auxiliary trees of the following form:

  For each $1 \leq i \leq k$: assume that $y_i = a_1 x_1 a_2 \dots x_r a_{r+1}$ where $x_1, \dots, x_r$ are components from the yields of the $A_1, \dots, A_n$ and $a_1, \dots, a_{r+1}$ are additional terminal words. The ith auxiliary tree has the form

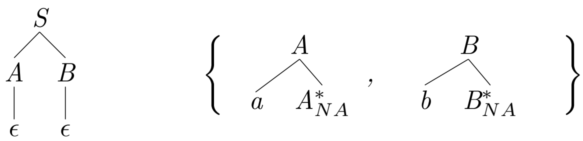
  Where X_j is the following label: if x_j is the q-th component of the yield of A_p, then $X_j := \langle A_p, q \rangle$.

- There is one initial tree of the form

$\Rightarrow$ we have shown $CFL \subsetneq TAL \subsetneq MCTAL = LCFRL \subseteq mildly CSL$

Exercise 17 *Consider the following MCTAG:*



Which string language does this MCTAG generate if the derivations are a) tree-local? b) set-local?

Exercise 18 *Give an MCTAG for the double copy language $\{www \mid w \in \{a, b\}^*\}$.*