Mildly Context-Sensitive Grammar Formalisms:

Multicomponent TAG

Laura Kallmeyer
Heinrich-Heine-Universität Düsseldorf

Sommersemester 2011

Overview

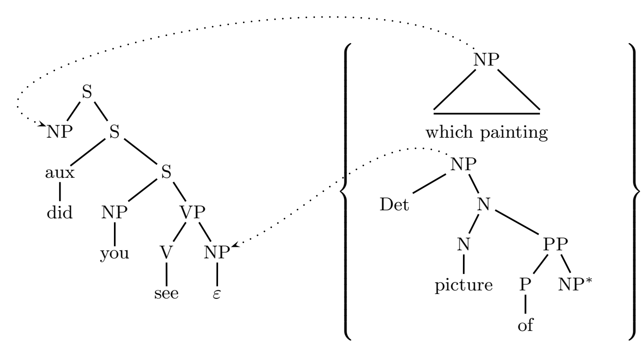
1. Motivation
2. Different types of MCTAG
3. Semilinearity of MCTALs

Motivation (1)

- Multicomponent Tree Adjoining Grammars (MCTAGs) were first introduced in [Joshi et al., 1975] as *simultaneous TAGs*, later redefined as *multicomponent TAGs (MCTAGs)* in [Weir, 1988, Joshi, 1985].
- The underlying linguistic motivation is the idea to separate the contribution of a lexical item into several components.
- In each derivation step, a new set is picked and all trees from the set are added simultaneously, i.e., they are attached (by substitution or adjunction) to different nodes in the already derived tree.

Motivation (2)

(1) which painting$_i$ did you see a picture of t_i



Motivation (3)

Constructions that require multicomponents:

- Extraction out of complex NPs [Kroch, 1989], stranding phenomena, in particular "picture-NPs":

  (2) which castle did you paint a picture of?

- Subject-aux inversion in raising questions [Frank, 2008]

  (3) Does John seem to annoy you?

- Scrambling in German [Rambow, 1994]

  (4) dass den Kühlschrank niemand zu reparieren versprochen hat

Different types of MCTAG (1)

Definition 1 (MCTAG) *An MCTAG is a tuple* $G = \langle N, T, S, I, A, f_{OA}, f_{SA}, \mathcal{A}\rangle$ *such that:*

- $G_{TAG} := \langle N, T, S, I, A, f_{OA}, f_{SA}\rangle$ *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.*[a]

Without loss of generality, we can assume that $\mathcal{A}$ is a partition of $I \cup A$.

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

Different types of MCTAG (2)

Definition 2 (MCTAG derivation) $\gamma \Rightarrow \gamma'$ *is a derivation step in* G *iff there is an instance* $\{\gamma_1, \dots, \gamma_n\}$ *of an elementary tree set in* $\mathcal{A}$ *and there are pairwise different nodes* $v_1, \dots, v_n$ *in* γ *such that* $\gamma' = \gamma[v_1, \gamma_1] \dots [v_n, \gamma_n]$.

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.

Different types of MCTAG (3)

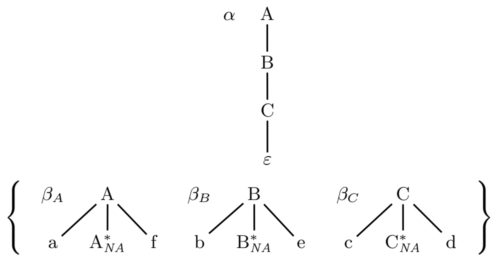
An MCTAG is called

- *tree-local* iff in each derivation step, the nodes the new trees attach to belong to the same elementary tree.
- *set-local* iff in each derivation step, the nodes the new trees attach to belong to the same elementary tree set.
- *non-local* otherwise.

Usually, the term "MCTAG" without specification of the locality means "set-local MCTAG".

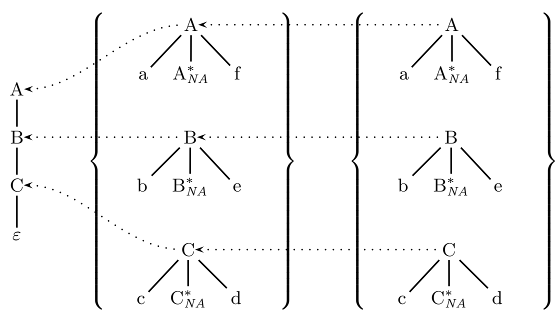
Different types of MCTAG (4)

Set-local MCTAG for $L_6 = \{a^n b^n c^n d^n e^n f^n \mid n \geq 0\}$:



Different types of MCTAG (5)

Derivation for $aabbccddeeff$:



Different types of MCTAG (6)

Tree-local MCTAG and TAG are equivalent since we can precompile the possible adjunctions and substitutions in an elementary tree:

Proposition 1 *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.

But: the number of elementary trees in the grammar can increase in an exponential way in this construction (⇒ rather a bad strategy for tree-local MCTAG parsing).

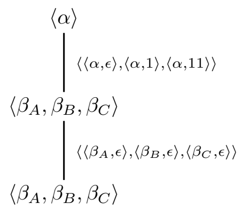
Semilinearity of MCTALs (1)

For tree-local and set-local MCTAG, *derivation trees* can be defined as follows [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.

Semilinearity of MCTALs (2)

Derivation tree of the previous sample derivation



Semilinearity of MCTALs (3)

Construction of a letter-equivalent CFG for a given set-local MCTAG: The non-terminals are a start symbol S and the tuples indicating the elementary tree sets.

- For every unary elementary tree set tuple Γ containing an initial tree with root label S where Γ contains terminals $a_1 \dots a_m$, we add
$$S \to \Gamma a_1 \dots a_m$$
- Whenever a tuple Γ allows us to attach the tuples $\Gamma_1, \dots \Gamma_k$ and the new trees add the terminals $a_1 \dots a_m$, we add
$$\Gamma \to \Gamma_1, \dots \Gamma_k a_1 \dots a_m$$
- For every tuple Γ containing neither OA-nodes nor substitution nodes, we add
$$\Gamma \to \varepsilon$$

Semilinearity of MCTALs (4)

Letter-equivalent CFG for our sample MCTAG:

$S \to \langle\alpha\rangle$

$\langle\alpha\rangle \to \varepsilon \qquad \langle\alpha\rangle \to \langle\beta_A, \beta_B, \beta_C\rangle abcdef$

$\langle\beta_A, \beta_B, \beta_C\rangle \to \varepsilon \qquad \langle\beta_A, \beta_B, \beta_C\rangle \to \langle\beta_A, \beta_B, \beta_C\rangle abcdef$

We will see later that MCTALs are also in PTIME and consequently mildly context-sensitive.

References

[Frank, 2008] Frank, R. (2008). Syntax and ltag. Slides of a tutorial at TAG+9, Tübingen.

[Joshi, 1985] Joshi, A. K. (1985). Tree adjoining grammars: How much contextsensitivity is required to provide reasonable structural descriptions? In Dowty, D., Karttunen, L., and Zwicky, A., editors, *Natural Language Parsing*, pages 206–250. Cambridge University Press.

[Joshi et al., 1975] Joshi, A. K., Levy, L. S., and Takahashi, M. (1975). Tree Adjunct Grammars. *Journal of Computer and System Science*, 10:136–163.

[Kroch, 1989] Kroch, A. (1989). Asymmetries in long-distance extraction in a Tree Adjoining Grammar. In Baltin and Kroch, editors, *Alternative Conceptions of Phrase Structure*. University of Chicago.

[Rambow, 1994] Rambow, O. (1994). *Formal and Computational Aspects of Natural Language Syntax*. PhD thesis, University of Pennsylvania.

[Weir, 1988] Weir, D. J. (1988). *Characterizing Mildly Context-Sensitive Grammar Formalisms*. PhD thesis, University of Pennsylvania.