Mildly context-sensitive grammar formalisms: TAG and related frameworks

Solutions to the exercises

Laura Kallmeyer

WS 2005/2006, Mo 11-13

Exercise 1 (Due until 7.11.05.)

$L_2 := \{a^n b^n \mid n \geq 0\}$

1. *Give a CFG for L_2 with nested dependencies, i.e., such that for each word $a_1 \dots a_n b_1 \dots b_n$ (the subscripts mark the occurrences of the as and bs respectively) a_i and b_{n+1-i} are added by the same production for all $1 \leq i \leq n$.*
2. *Show that for L_2 there is no CFG displaying cross-serial dependencies, i.e., no CFG such that for each word $a_1 \dots a_n b_1 \dots b_n$, a_i and b_i are added by the same production for all $1 \leq i \leq n$.*
   (Hint: Assume that such a CFG exists. Then show that in this case a CFG for the copy language $\{ww \mid w \in \{a,b\}^*\}$ exists. Contradiciton since this language is not context-free.)

1. $S \to \epsilon$, $S \to aSb$
2. Assume that such a CFG exists. Its productions are then all of the form $X \to \alpha a \beta b \gamma$ with $X \in N$, $\alpha, \beta, \gamma \in N^*$ such that if such a production is applied when generating a string $a_1 \dots a_n b_1 \dots b_n$, then the a and b of the production necessarily end up at positions i and $n+i$ for some $i, 1 \leq i \leq n$.

   Then replacing each of these productions $X \to \alpha a \beta b \gamma$ with $X \to \alpha a \beta a \gamma$ and $X \to \alpha b \beta b \gamma$ leads to a CFG generating the copy language. Contradiction.

□

Exercise 2 (Due until 14.11.05.)

Similar to Shieber's argument, one can apply first a homomorphim f, then intersect with some regular language, and then apply another homomorphism g in order to reduce the language of Swiss German to the copy language $\{ww \mid w \in \{a,b\}^*\}$. Find the corresponding homomorphisms and the regular language.

A first homomorphism can be as the f from Shieber. In this first homomorphism we need to keep trace of the parts that go to w, x, y and z in order to make sure that we are only considering constructions of the form as in (8).

Then intersect with the regular language $w\{a,b\}^*x\{c,d\}^*y$.

Leads to $\{wv_1xv_2y \mid v_1 \in \{a,b\}^*, v_2 \in \{c,d\}^*$ such that $|v_1| = |v_2|$ and for all i, $1 \leq i \leq |v_1|$: if the ith symbol in v_1 is an a (a b), the ith symbol in v_2 is a c (a d)$\}$.

Finally apply a second homomorphism g with $g(w) := g(x) := g(y) := \epsilon, g(a) := g(c) := a, g(b) := g(d) := b$.

Leads to the copy language.

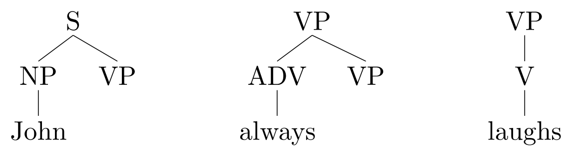
Exercise 3 (Due until 14.11.05.)

Consider the following CFG:

$S \to NP\ VP$	$NP \to John$
$VP \to ADV\ VP$	$ADV \to always$
$VP \to V$	$V \to laughs$

Find a TSG that strongly lexicalizes this grammar.
Why is this lexicalization not satisfying?

There are several possibilities. The simplest one:

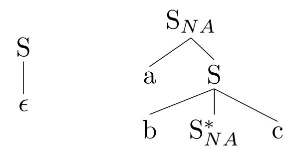


Unsatisfying since the S node comes with the lexical item *John* even though it is the maximal projection of the verb. A lexicalized TSG for the given CFG where the S node comes with the verb is not possible.

Exercise 4 (Due until 21.11.05.) $L_3 := \{a^n b^n c^n \mid n \geq 0\}$

1. *Give a TAG (with adjunction constraints) that generates* L_3.
2. *Show that TAG without adjunction constraints cannot generate* L_3.
   (Hint: Any elementary tree must contain equal numbers of a's, b's and c's. And each auxiliary tree can be adjoined at its own root.)

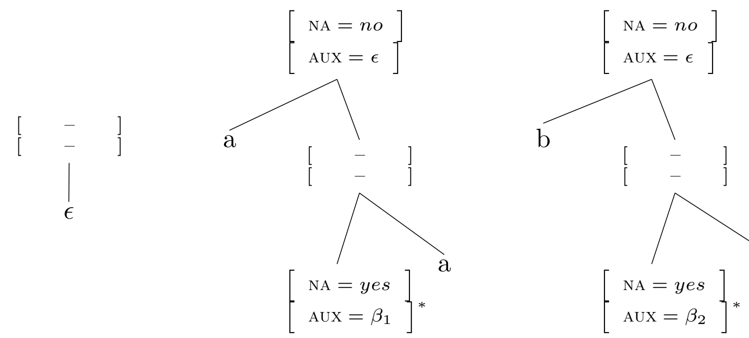
1. TAG for L_3:



2. Assume that a TAG G for L_3 without adjunction constraints exists. Assume without loss of generality that G contains no substitution nodes. G has at least one auxiliary tree β with leaves labeled with terminals. β must contain equal numbers of a's, b's and c's. (Otherwise one could derive a word with different numbers of a's, b's and c's.) One can adjoin β at its root which leads to a derived auxiliary β'. If the yield of β is $a^i b^i c^i$ ($i \geq 1$), there are the following possibilities for the foot node:
   (a) the foot node is left of all a's or right of all c's $\Rightarrow$ using β' a word with substring $a^i b^i c^i a^i b^i c^i$ can be derived. Contradiction.
   (b) the foot node is right of the kth a for some k, $1 \leq k \leq i \Rightarrow$ using β' a word with substring $a^{i-k} b^i c^i a^{i-k} b^i c^i$ can be derived. Contradiction.
   (c) the foot node is right of the kth b for some k, $1 \leq k \leq i \Rightarrow$ using β' a word with substrings $a^i b^k a^i b^k$ and $b^{i-k} c^i b^{i-k} c^i$ can be derived. Contradiction.
   (d) the foot node is left of the kth c for some k, $1 \leq k \leq i \Rightarrow$ using β' a word with substring $a^i b^i c^{k-1} a^i b^i c^{k-1}$ can be derived. Contradiction.

□

Exercise 5 (Due until 28.11.05.) *Give an FTAG for the copy language* $\{ww \mid w \in \{a,b\}^*\}$.



Exercise 6 (Due until 5.12.05.) *Show that* $\{a^n b^n c^n a^m b^m c^m \mid n, m \geq 0\}$ *is a TAL.*
Hint: The language $L_3 = \{a^n b^n c^n \mid n \geq 0\}$ *is a TAL.*

L_3 is a TAL $\Rightarrow$ with Prop. 11 $\{w_1 w_2 \mid w_1, w_2 \in L_3\} = \{a^n b^n c^n a^m b^m c^m \mid n, m \geq 0\}$ is a TAL.

□

Exercise 7 (Due until 5.12.05.) *Show that* $\{a^i b^j a^i b^j \mid i, j \geq 0\}$ *is a TAL.*
Hint: The copy language is a TAL.

The copy language is a TAL and $a^* b^* a^* b^*$ is a regular language $\Rightarrow$ with Prop. 15 $\{ww \mid w \in \{a, b\}^*\} \cap a^* b^* a^* b^* = \{a^i b^j a^i b^j \mid i, j \geq 0\}$ is a TAL.

□

Exercise 8 (Due until 5.12.05.) *Show that* $L := \{a^{2^n} \mid n \geq 0\}$ *is no TAL using the (weak) pumping lemma.*

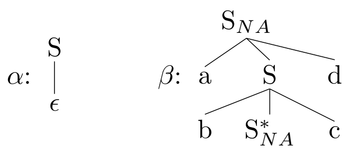
Assume that L is a TAL. Then the weak pumping lemma holds for some constant $c \Rightarrow$ For each word $w \in L$ with $|w| \geq c$ there is a $w' \in L$ with $|w'| \leq |w| + c$. Contradiction since for all $w \in L$ with $|w| > c$ this is not the case (for each $w \in L$ the $w' \in L$ next to it wrt. word length has twice its length).

□

Exercise 9 (Due until 5.12.05.) $L_4 := \{a^n b^n c^n d^n \mid n \geq 0\}$, $L_5 := \{a^n b^n c^n d^n e^n \mid n \geq 0\}$

1. *Give a TAG generating* L_4.
2. *Show that* L_5 *is not a TAL using the weak pumping lemma.*
   Hint: Consider the word $w = a^{c+1} b^{c+1} c^{c+1} d^{c+1} e^{c+1}$ *with* c *being the constant from the pumping lemma.*

1. TAG for L_4:



2. Assume that L_5 is a TAL and satisfies the weak pumping lemma with some constant c. Take $w = a^{c+1} b^{c+1} c^{c+1} d^{c+1} e^{c+1}$. According to the pumping lemma one can find $w_1, \ldots w_4$, at least one of them not empty, such that they can be inserted repeatedly at 4 positions into w yielding a new word in L_5. At least one of the $w_1, \ldots w_4$ must contain two different terminals symbols since they altogether must contain equal numbers of a's, b's, c's, d's and e's. Then, when doing a second insertion of the $w_1, \ldots w_4$, the a's, b's, c's, d's and e's get mixed and the resulting word is not in L_5. Contradiction.

□

Exercise 10 (Due until 12.12.05.)

Give an indexed grammar for $L_4 = \{a^n b^n c^n d^n \mid n \geq 0\}$.

See above for the TAG for L_4.

Indexed grammar for L_4: Start symbol is $\langle \alpha, S \rangle$.

Productions:

1. Traversal of α without adjunction:

$$\langle \alpha, S \rangle \to \epsilon$$

2. Adjunction at root of α:

$$\langle \alpha, S \rangle \to \langle \beta, S_1 \rangle [\overline{\langle \alpha, S \rangle}]$$

3. Traversal of β without adjunctions:

$$\langle \beta, S_1 \rangle [\ldots] \to a \langle \beta, S_2 \rangle [\ldots] d$$

$$\langle \beta, S_2 \rangle [\ldots] \to b \langle \beta, S_3 \rangle [\ldots] c$$

4. Adjunction at β:

$$\langle \beta, S_2 \rangle [\ldots] \to \langle \beta, S_1 \rangle [\overline{\langle \beta, S_2 \rangle} \ldots]$$

5. Going back from the foot node of β to the tree where β was adjoined:

$$\langle \beta, S_3 \rangle [\overline{\langle \alpha, S \rangle}] \to \epsilon$$

$$\langle \beta, S_3 \rangle [\overline{\langle \beta, S_2 \rangle} \ldots] \to b \langle \beta, S_3 \rangle [\ldots] c$$

Exercise 11 (Due until 19.12.05.)

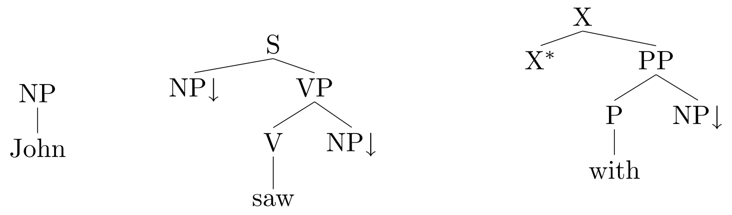
Propose elementary trees for the following sentences:

(1) John saw a man with a telescope

(2) John buys the house that Bill lives in

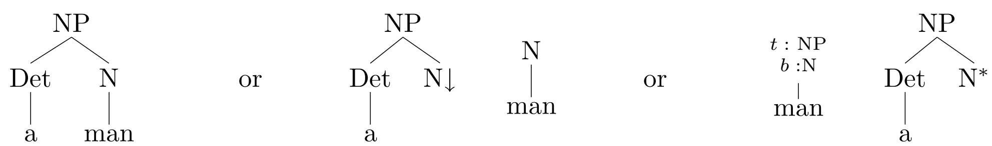
(3) Mary took a decision

(1) John saw a man with a telescope



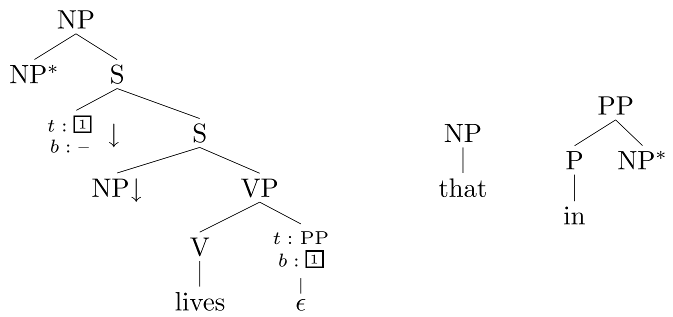
X=N or X=VP

Concerning the determiners, there are three possibilities:

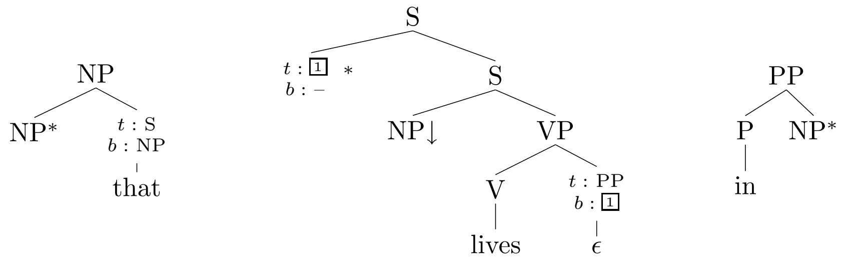


(2) John buys the house that Bill lives in

elementary trees for *John*, *buys* and *the house*: see (1).

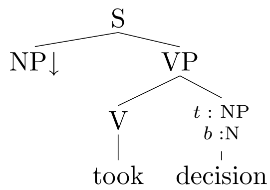


Other possibility: Add relative pronoun first:



(There are probably other possibilities as well.)

(3) Mary took a decision

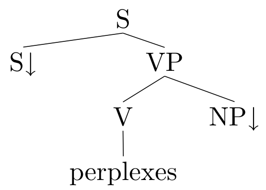


Exercise 12 (Due until 19.12.05.) *Consider sentential subjects as in*

(4) That John wins perplexes Bill

Do you prefer adding them by substitution or adding the matrix verb to the sentential subject by adjunction (similar to sentential complements)?
(Note that extraction out of sentential subjects is not allowed.)
Give the elementary tree for perplexes *that you would choose.*

Since extraction out of sentential subjects is not allowed, it makes sense to add them by substitution. Otherwise one could adjoin them to a sentence with wh-extracted elements. (Of course such an adjunction could also be prevented within the feature structures.)

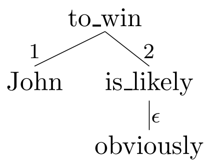


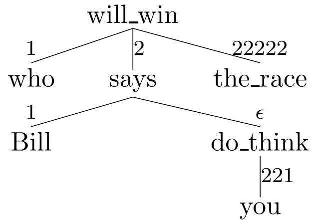
Exercise 13 (Due until 9.1.06.)

Give the derivation trees for

(5) John obviously is likely to win

(6) Who do you think Bill says will win the race?





Exercise 14 (Due until 16.1.06.) *The simplest verb tree family in XTAG is Tnx0V, the tree family for intransitive verbs such as* laugh. *Find sample sentences containing uses of forms of* laugh *in different constructions that should be included in this tree family.*

(7) a. John laughs

b. Who laughs all the time?

c. Laugh!

d. John's laughing is irritating.

e. Do you approve of John laughing?

f. the man who laughs all the time

g. the man that laughs all the time

h. the place where everybody was laughing all the time

i. the day he laughs about this

j. the laughing man

k. he promised me not to laugh

l. he prevented me from laughing

Exercise 15 (Due until 23.1.06.)

1. *Show that the copy language $\{ww \mid w \in T^*\}$ for some alphabet T is semilinear using the Parikh-Theorem.*
2. *Show that $\{a^{2^n} \mid n \geq 0\}$ is not semilinear.*

   Hint: if the language was semilinear it would satisfy the constant growth property. Show that this is not the case.

1. The copy language $L := \{ww \mid w \in T^*\}$ is letter equivalent to $L' := \{wr(w) \mid w \in T^*$ and $r(w)$ is w in reverse order$\}$ which is a CFL: it is generated by the CFG with productions $S \to \epsilon$ and $S \to xSx$ for all $x \in T$. Consequently (with Parikh's theorem) L' and also L are semilinear.

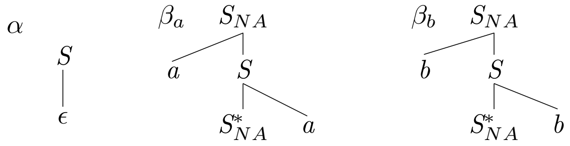
□

2. Assume that $\{a^{2^n} \mid n \geq 0\}$ satisfies the constant growth property with c_0 and C. Then take a $w = a^{2^m}$ with $|w| = 2^m > max(\{c_0\} \cup C)$. Then, according to constant growth, for $w' = a^{2^{m+1}}$ there must be a $w'' = a^{2^k}$ with $|w'| = |w''| + c$ for some $c \in C$. I.e., $2^{m+1} = 2^k + c$. Consequently (since $k \leq m$) $c \geq 2^m$. Contradiction.

□

Exercise 16 (Due until 30.1.06.)

Give the corresponding LCFRS for the TAG for the copy language:

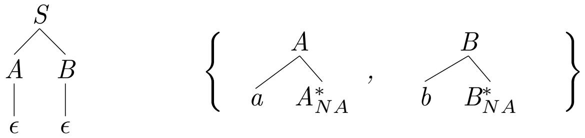


Productions with denotations and yields:

- No adjunctions:
  $\alpha \rightarrow f_\alpha()$, $[\![f_\alpha()]\!] := \alpha$, $\phi(f_\alpha()) := \langle\epsilon\rangle$.
  $\beta_a \rightarrow f_{\beta_a}()$, $[\![f_{\beta_a}()]\!] := \beta_a$, $\phi(f_{\beta_a}()) := \langle a, a\rangle$.
  $\beta_b \rightarrow f_{\beta_b}()$, $[\![f_{\beta_b}()]\!] := \beta_b$, $\phi(f_{\beta_b}()) := \langle b, b\rangle$.
- Adjunctions to α:
  $\alpha \rightarrow f_{\alpha:\epsilon}(\beta_a)$, $\alpha \rightarrow f_{\alpha:\epsilon}(\beta_b)$,
  $[\![f_{\alpha:\epsilon}(t)]\!] := \alpha[\epsilon, [\![t]\!]]$,
  $\phi(f_{\alpha:\epsilon}(t)) := \langle w_1 w_2\rangle$ where $\phi(t) = \langle w_1, w_2\rangle$.
- Adjunctions to β_a:
  $\beta_a \rightarrow f_{\beta_a:1}(\beta_a)$, $\beta_a \rightarrow f_{\beta_a:1}(\beta_b)$,
  $[\![f_{\beta_a:1}(t)]\!] := \beta_a[1, [\![t]\!]]$,
  $\phi(f_{\beta_a:1}(t)) := \langle a w_1, a w_2\rangle$ where $\phi(t) = \langle w_1, w_2\rangle$.
- Adjunctions to β_b:
  $\beta_b \rightarrow f_{\beta_b:1}(\beta_a)$, $\beta_b \rightarrow f_{\beta_b:1}(\beta_b)$,
  $[\![f_{\beta_b:1}(t)]\!] := \beta_b[1, [\![t]\!]]$,
  $\phi(f_{\beta_b:1}(t)) := \langle b w_1, b w_2\rangle$ where $\phi(t) = \langle w_1, w_2\rangle$.

Exercise 17 (Due until 6.2.06.)

Consider the following MCTAG:



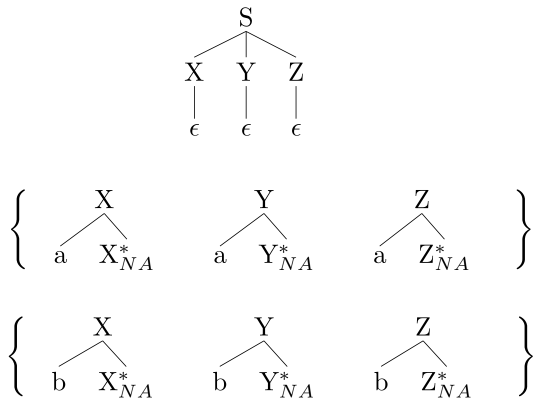
Which string language does this MCTAG generate if the derivations are a) tree-local? b) set-local?

a) $\{\epsilon, ab\}$

b) $\{a^n b^n \mid n \geq 0\}$

Exercise 18 (Due until 6.2.06.)

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



Exercise 19 (Due until 13.2.06.)

What are the sentences generated by the non-local MCTAG with dominance links from Fig. 17, p. 42?

The string language L is as follows:

- the three NP trees are intial trees, i.e., they are in the tree language:

  niemand $\in L$

  den Kühlschrank $\in L$

  ihm $\in L$

- without using *zu versprechen*, the three *reparieren* trees must necessarily be attached to the three VP node in the *bereit* tree:

  den Kühlschrank niemand zu reparieren bereit ist $\in L$

- one adjunction of *versprechen* to *bereit*, then in the resulting tree three VPs with dominance relations among them must be found. Leads to

  den Kühlschrank ihm zu reparieren niemand zu versprechen bereit ist $\in L$

  den Kühlschrank ihm niemand zu reparieren zu versprechen bereit ist $\in L$

  ihm den Kühlschrank niemand zu reparieren zu versprechen bereit ist $\in L$

  ihm niemand den Kühlschrank zu reparieren zu versprechen bereit ist $\in L$

- a second adjunction of *versprechen* before adjoining *reparieren* leads to

  den Kühlschrank ihm zu reparieren ihm zu versprechen niemand zu versprechen bereit ist $\in L$

  den Kühlschrank ihm ihm zu reparieren zu versprechen niemand zu versprechen bereit ist $\in L$

  ihm den Kühlschrank ihm zu reparieren zu versprechen niemand zu versprechen bereit ist $\in L$

  ihm ihm den Kühlschrank zu reparieren zu versprechen niemand zu versprechen bereit ist $\in L$

  and also

  den Kühlschrank ihm zu reparieren ihm niemand zu versprechen zu versprechen bereit ist $\in L$

  den Kühlschrank ihm ihm zu reparieren niemand zu versprechen zu versprechen bereit ist $\in L$

  den Kühlschrank ihm ihm niemand zu reparieren zu versprechen zu versprechen bereit ist $\in L$

  ihm den Kühlschrank ihm zu reparieren niemand zu versprechen zu versprechen bereit ist $\in L$

  ihm den Kühlschrank ihm niemand zu reparieren zu versprechen zu versprechen bereit ist $\in L$

ihm ihm den Kühlschrank niemand zu reparierenzu versprechen zu versprechen bereit ist $\in L$
ihm ihm niemand den Kühlschrank zu reparieren zu versprechen zu versprechen bereit ist $\in L$
and
den Kühlschrank ihm zu reparieren niemand ihm zu versprechen zu versprechen bereit ist $\in L$
den Kühlschrank ihm niemand zu reparieren ihm zu versprechen zu versprechen bereit ist $\in L$
den Kühlschrank ihm niemand ihm zu reparieren zu versprechen zu versprechen bereit ist $\in L$
ihm den Kühlschrank niemand zu reparieren ihm zu versprechen zu versprechen bereit ist $\in L$
ihm den Kühlschrank niemand ihm zu reparieren zu versprechen zu versprechen bereit ist $\in L$
ihm niemand den Kühlschrank ihm zu reparierenzu versprechen zu versprechen bereit ist $\in L$
ihm niemand ihm den Kühlschrank zu reparieren zu versprechen zu versprechen bereit ist $\in L$

In general, the strings are such that

- all arguments (VPs and NPs) are left of the verbs they depend on,
- *niemand* is left of the rightmost *versprechen*
- the rightmost *versprechen* immediately precedes *bereit*
- immediately left of an NP there cannot be a verb and its NP argument together
- at least three NPs are left of the leftmost *versprechen*

Exercise 20 (Due until 13.2.06.)

What are the sentences generated by the MCTAG with dominance links from Fig. 17 if this MCTAG is considered being a V-TAG and if, additionally, all foot nodes carry NA constraints?

The language contains the following strings:

- *niemand* $\in L$
  den Kühlschrank $\in L$
  ihm $\in L$
- *niemand den Kühlschrankzu reparieren bereit ist* $\in L$
  den Kühlschrank niemand zu reparieren bereit ist $\in L$
  den Kühlschrank zu reparieren niemand bereit ist $\in L$
- iteratively using *versprechen*, all scrambling order are possible as well: only condition: all arguments must precede their verbs because of the dominance links