Parsing Beyond CFG
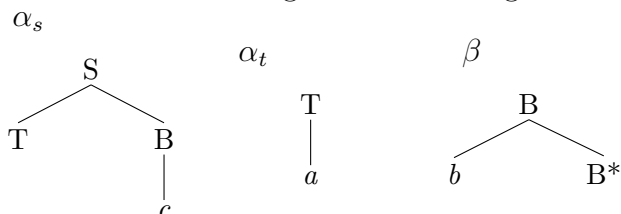# Homework 4: TAG Earley Parsing

Laura Kallmeyer, Tatiana Bladier

Sommersemester 2018

Question 1 (TAG Earley parsing)

Consider the TAG consisting of the following elementary trees α_s, α_t, and β:

Give the rest of the trace of the Earley parse (the version from the course slides) of $w = abc$ (the first 10 items are already provided in the table below). Give only those items that lead to the correct parse. Explain for each of these items, by which operation it is obtained and from which antecedent item(s).

Solution:

	Item	Dotted tree	Rule
1.	$[\alpha_s, \epsilon, la, 0, -, -, 0, 0]$		Initialize
2.	$[\alpha_s, \epsilon, lb, 0, -, -, 0, 0]$		predictNoAdj from 1.
3.	$[\alpha_s, 1, la, 0, -, -, 0, 0]$		moveDown from 2.
4.	$[\alpha_s, 1, lb, 0, -, -, 0, 0]$		predictNoAdj from 3.
5.	$[\alpha_t, \epsilon, la, 0, -, -, 0, 0]$		predictSubst from 4.
6.	$[\alpha_t, \epsilon, lb, 0, -, -, 0, 0]$		predictNoAdj from 5.
7.	$[\alpha_t, 1, la, 0, -, -, 0, 0]$		moveDown from 6.
8.	$[\alpha_t, 1, ra, 0, -, -, 1, 0]$		scanTerm from 7.
9.	$[\alpha_t, \epsilon, rb, 0, -, -, 1, 0]$		moveUp from 8.
10.	$[\alpha_t, \epsilon, ra, 0, -, -, 1, 0]$		completeNode from 9. and 7.

	Item	Dotted tree	Rule
11.	$[\alpha_s, 1, rb, 0, -, -, 1, 0]$	S → (T•, B → c)	substitute from 10.
12.	$[\alpha_s, 1, ra, 0, -, -, 1, 0]$	S → (T•, B → c)	completeNode from 11. and 3.
13.	$[\alpha_s, 2, la, 0, -, -, 1, 0]$	S → (T, •B → c)	moveRight from 12.
14.	$[\beta, \epsilon, la, 1, -, -, 1, 0]$	•B → (b, B*)	predictAdjonable from 13.
15.	$[\beta, \epsilon, lb, 1, -, -, 1, 0]$	•B → (b, B*)	predictNoAdj from 14.
16.	$[\beta, 1, la, 1, -, -, 1, 0]$	B → (•b, B*)	moveDown from 15.
17.	$[\beta, 1, ra, 1, -, -, 2, 0]$	B → (b•, B*)	scanTerm from 16.
18.	$[\beta, 2, la, 1, -, -, 2, 0]$	B → (b, •B*)	moveRight from 17.
19.	$[\beta, 2, lb, 2, -, -, 2, 0]$	B → (b, •B*)	predictNoAdj from 18.
20.	$[\alpha_s, 2, lb, 2, -, -, 2, 0]$	S → (T, •B → c)	predictAdjoined from 19.
21.	$[\alpha_s, 21, la, 2, -, -, 2, 0]$	S → (T, B → •c)	moveDown from 20.
22.	$[\alpha_s, 21, ra, 2, -, -, 3, 0]$	S → (T, B → c•)	scanTerm from 21.
23.	$[\alpha_s, 2, rb, 2, -, -, 3, 0]$	S → (T, B• → c)	moveUp from 22.
24.	$[\beta, 2, rb, 2, 2, 3, 3, 0]$	B → (b, B*•)	completeFoot from 23, 19.
25.	$[\beta, 2, ra, 1, 2, 3, 3, 0]$	B → (b, B*•)	completeNode from 18, 24.

	Item	Dotted tree	Rule
26.	$[\beta, \epsilon, rb, 1, 2, 3, 3, 0]$	B$_\bullet$ with children b, B*	moveUp from 25.
27.	$[\beta, \epsilon, ra, 1, 2, 3, 3, 0]$	B$^\bullet$ with children b, B*	completeNode from 26, 14
28.	$[\alpha_s, 2, rb, 1, -, -, 3, 1]$	S with children T, B$_\bullet$ (B over c)	adjoin 23, 27.
29.	$[\alpha_s, 2, ra, 0, -, -, 3, 0]$	S with children T, B$^\bullet$ (B over c)	completeNode from 13, 28.
30.	$[\alpha_s, \epsilon, rb, 0, -, -, 3, 0]$	S$_\bullet$ with children T, B (B over c)	moveUp from 23.
31.	$[\alpha_s, \epsilon, ra, 0, -, -, 3, 0]$	S$^\bullet$ with children T, B (B over c)	completeNode from 1, 30.

Question 2 (TAG induction)

a) Consider the following sentence from the Penn Treebank and the percolation table for identification of the modifiers (the table should be read as follows: if the parent node with the label given in the left column has children nodes with labels given in the right column, mark these children as modifiers):
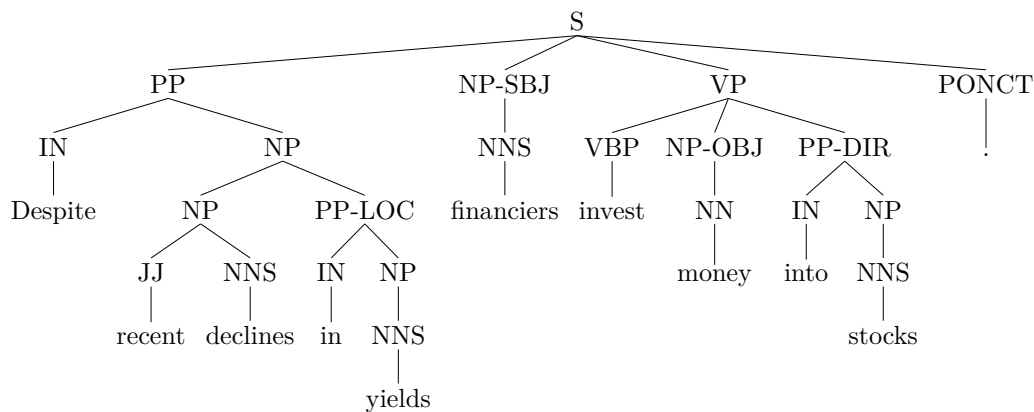

label of the parent node	modifier nodes
S	PP JJ PP-LOC PONCT
NP	JJ PONCT PP-LOC PP-DIR
NP-SBJ	JJ PONCT PP-DIR
NP-OBJ	JJ PONCT PP-DIR
VP	JJ PONCT

b) Complete the last three lines of the percolation table for identification of the constituent heads using labels from the example sentence:

label of the parent node	search direction	head candidates
S	right-to-left	VP
PP	left-to-right	IN
NP	right-to-left	NP NNS
VP	right-to-left	VBP
NP-SBJ	right-to-left	NNS NP NN
NP-OBJ	right-to-left	NNS NP NN
PP-DIR	left-to-right	IN P PP

c) Highlight all head nodes in the example sentence (for example using different colour or through encircling the label of the head node).

Solution:
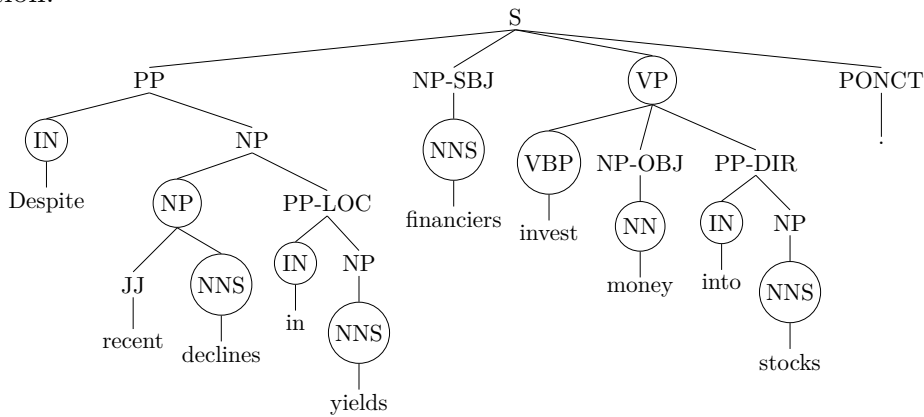
d) Extract and give **only the auxiliary trees** from the example sentence (only one lexical anchor per elementary tree is allowed; please mark the substitution nodes with ↓).

Hint: there should be five auxiliary trees.