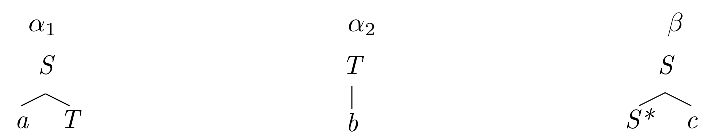
Grammar Immplementation with TAG
Homework: TAG Parsing and formal properties

Laura Kallmeyer, Benjamin Burkhardt

Question 1 (TAG CYK parsing)

Consider the TAG consisting of the following tree:



Give the trace (only successful items) of the CYK parse (the version from the course slides) of $w = abc$, i.e., a list of all successful items that get generated. Explain for each item, by which operation it is obtained and from which antecedent items.

Solution:

	Item	Rule
1.	$[\alpha_1, 1_\top, 0, -, -, 1]$	lex-scan (a)
2.	$[\alpha_2, 1_\top, 1, -, -, 2]$	lex-scan (b)
3.	$[\beta, 2_\top, 2, -, -, 3]$	lex-scan (c)
4.	$[\beta, 1_\top, 0, 0, 2, 2]$	foot-predict
5.	$[\alpha_2, \epsilon_\bot, 1, -, -, 2]$	move-unary from 2.
6.	$[\alpha_2, \epsilon_\top, 1, -, -, 2]$	null-adjoin from 5.
7.	$[\alpha_1, 2_\top, 1, -, -, 2]$	substitute 6.
8.	$[\alpha_1, \epsilon_\bot, 0, -, -, 2]$	move binary from 1. and 7.
9.	$[\alpha_1, \epsilon_\top, 0, -, -, 2]$	null-adjoin from 8.
10.	$[\beta, \epsilon_\bot, 0, 0, 2, 3]$	move-binary from 3. and 4.
11.	$[\beta, \epsilon_\top, 0, 0, 2, 3]$	null-adjoin from 10.
12.	$[\alpha_1, \epsilon_\top, 0, -, -, 3]$	adjoin 11. in 8.

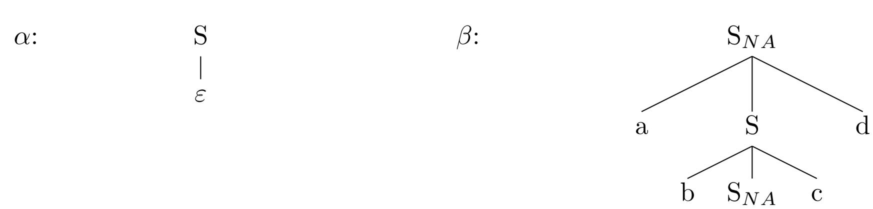
Question 2 (Pumping Lemma)

$L_4 = \{a^n b^n c^n d^n \mid n \geq 0\}$, $L_5 = \{a^n b^n c^n d^n e^n \mid n \geq 0\}$

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.

Solution:

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 $w_1, \dots w_4$, at least one of them not empty, such that they can be inserted repeatedly at four positions into w yielding a new word in L_5. At least one of the $w_1, \dots w_4$ must contain two different terminal symbols since they altogether must contain equal numbers of as, bs, cs, ds and es. Then, when doing a second insertion of the $w_1, \dots w_4$, the as, bs, cs, ds and es get mixed and the resulting word is not in L_5. Contradiction.

Question 3 (Closure Properties)

1. *Show that $L_{MIX5} = \{w \mid w \in \{a,b,c,d,e\}^*, |w|_a = |w|_b = |w|_c = |w|_d = |w|_e\}$ is not a TAL.*
   Hint: Use the closure of TALs under intersection with regular language and the result about L_5 shown above.

2. *Show that $L = \{w \mid w \in \{a,b,c,d,e,f,g\}^*, |w|_a = |w|_b = |w|_c = |w|_d = |w|_e\}$ is not a TAL.*
   Hint: Use the closure of TALs under homomorphisms and the result about L_{MIX5} shown in 1.

Solution:

1. We assume that $L_{MIX5} = \{w \mid w \in \{a,b,c,d,e\}^*, |w|_a = |w|_b = |w|_c = |w|_d = |w|_e\}$ is a TAL.
   Then $L_{MIX5} \cap L(a^*b^*c^*d^*e^*) = L_5$ must also be a TAL.
   Contradiction, consequently L_{MIX5} is not a TAL.

2. We assume that $L = \{w \mid w \in \{a,b,c,d,e,f,g\}^*, |w|_a = |w|_b = |w|_c = |w|_d = |w|_e\}$ is a TAL.
   Then its image under a homomorphism h with $h(a) = a, h(b) = b, h(c) = c, h(d) = d, h(e) = e, h(f) = \varepsilon, h(g) = \varepsilon$ must also be a TAL.
   Contradiction since $h(L) = L_{MIX5}$ is not a TAL. Consequently, L is not a TAL.