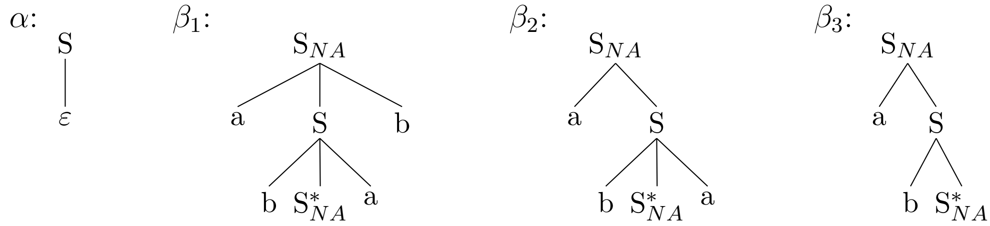
Parsing Beyond CFG
Homework 3: TAG (Formal properties)

Laura Kallmeyer

Question 1 *Give a TAG (with adjunction constraints) for the following language:*

$$\{a^n b^n a^m b^k \mid n \geq 0, n \geq m \geq k\}$$

Solution



Question 2 *Show that $L_5 = \{a^n b^n c^n d^n e^n \mid n \geq 0\}$ is not a Tree Adjoining language.*

Solution: 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, therefore our initial assumption does not hold, L_5 is not a TAL.

Question 3 *Show that the following languages are not Tree Adjoining languages:*

1. $L_6 = \{a^n b^n c^n d^n e^n f^n \mid n \geq 0\}$
   Hint: You can make use of the fact the language L_5 above is not a TAL.
2. $L_{wwww} = \{wwww \mid w \in \{a, b\}^*\}$

Solution:

1. We assume that L_6 is a TAL. Then its image under the homomorphism g with $g(a) = a, g(b) = b, g(c) = c, g(d) = d, g(e) = e, g(f) = \varepsilon$ is also a TAL. This image is L_5, a language for which we already know that it is not a TAL. Contradiction, therefore our initial assumption does not hold, L_6 is not a TAL.

2. We assume that L_{wwww} is a TAL. Then

   $$L = L_{wwww} \cap L(a^+b^+a^+b^+a^+b^+a^+b^+) = \{a^n b^m a^n b^m a^n b^m a^n b^m \mid n, m \geq 1\}$$

   is also a TAL.

   Consequently, L satisfies the weak pumping lemma with some constant c. Take $w = a^{c+1}b^{c+1}a^{c+1}b^{c+1}a^{c+1}b^{c+1}a^{c+1}b^{c+1}$. 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. Furthermore, between w_1 and w_2 and also between w_3 and w_4 there must not be more than c terminals. Consequently, we cannot choose to add equal numbers of as in the four a^{c+1} parts in each iteration or equal numbers of bs in the four b^{c+1} parts. But any other choice for w_1, w_2, w_3, w_4 and the positions in w where they are added leads to a word outside L when iterating. Contradiction.