Parsing

Homework 5 (CYK), due 17 May 2021

Laura Kallmeyer

SS 2021, Heinrich-Heine-Universität Düsseldorf

Question 1 (CYK recognition – general version)

Consider the CFG with non-terminals S, A, C, terminals a, b, start symbol S and productions

$$S \to AS \,|\, ASA \,|\, A \quad A \to a \,|\, AC \,|\, \varepsilon \quad C \to c$$

Give the chart (the $(n+1) \times (n+1)$-table) that results from the general CYK algorithm for the input aaac.

Solution:

4	S				
3	S	S			
2	S	S	A, S		
1	a, A, S	a, A, S	a, A, S	c, C, A, S	
0	A, S	A, S	A, S	A, S	A, S
	1	2	3	4	5

Question 2 (CYK parsing for CNF grammars)

Consider the CFG with non-terminals $\{A, S\}$, terminal a, start symbol S and productions

$$S \to SS \,|\, AS \,|\, a \quad A \to a$$

.

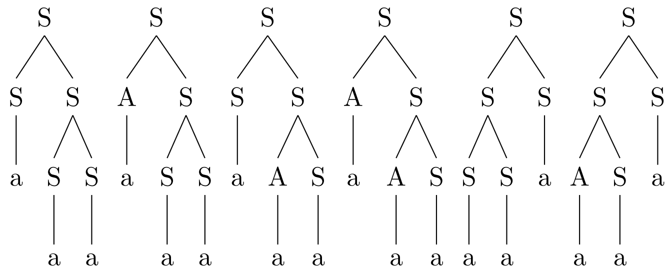
This grammar is in Chomsky Normal Form.

1. *Give the chart (the $n \times n$-table) that results from the CYK parsing algorithm (for CNF) for the input aaaa. The chart should include not only the non-terminals that we find but the entire productions with, in the rhs, the indices of the antecedent chart items in the complete rule that has been applied.*
2. *Now consider an input aaa. The chart entries for $i \leq 3$ and $l \leq 3$ from the chart obtained in 1. give the complete chart for this input. Read off all parse trees for aaa.*

Solution:

1. Chart:

l					
4	$S \to S_{1,1}S_{2,3}$, $S \to S_{1,2}S_{3,2}$, $S \to S_{1,3}S_{4,1}$, $S \to A_{1,1}S_{2,3}$				
3	$S \to S_{1,1}S_{2,2}$, $S \to S_{1,2}S_{3,1}$, $S \to A_{1,1}S_{2,2}$	$S \to S_{2,1}S_{3,2}$, $S \to S_{2,2}S_{4,1}$, $S \to A_{2,1}S_{3,2}$			
2	$S \to S_{1,1}S_{2,1}$, $S \to A_{1,1}S_{2,1}$	$S \to S_{2,1}S_{3,1}$, $S \to A_{2,1}S_{3,1}$	$S \to S_{3,1}S_{4,1}$ $S \to A_{3,1}S_{4,1}$		
1	$S \to a,\ A \to a$	$S \to a,\ A \to a$	$S \to a,\ A \to a$	$S \to a,\ A \to a$	
	1	2	3	4	i

2. 

Question 3 (CYK – completeness) *Consider the CNF recognizer from the course slides. We have already discussed how to prove the soundness of the algorithm.*

Now prove in addition completeness, i.e., prove that the following claim holds for a given input $w_1 \dots w_n$*, and for any* $1 \leq i \leq n$ *and* $1 \leq l \leq n - i + 1$*:*

If $A \stackrel{*}{\Rightarrow} w_i \dots w_{i+l-1}$*, then* $[A, i, l]$*.*

Show this via induction over l*, i.e.,*

1. *first prove that the claim holds for* $l = 1$*.*

   (To this end, assume that $A \stackrel{*}{\Rightarrow} w_i$ *for some* $1 \leq i \leq n$ *and show that then we can apply the scan rule to obtain* $[A, i, 1]$*.)*

2. *then show that, if the claim holds for* l*, it also holds for* $l + 1$*.*

   (To this end, assume that $A \stackrel{*}{\Rightarrow} w_i \dots w_{i+l}$ *(length* $l + 1$*). Then you can reason about the first step done in this derivation, which leads to subsequent subderivations for strings of smaller length, for which we already know that our claim holds ...)*

Solution:

To show: If $A \stackrel{*}{\Rightarrow} w_i \dots w_{i+l-1}$, then $[A, i, l]$ for every $1 \leq i, l \leq n$.

$l = 1$ Trivially, if $A \stackrel{*}{\Rightarrow} w_i$ for some $1 \leq i \leq n$, then $A \to w_i \in P$, and then (because of the scan rule) $[A, i, 1]$.

$l \Rightarrow l + 1$ We assume that our claim holds for any length up to l. And we assume that $A \stackrel{*}{\Rightarrow} w_i \dots w_{i+l}$ (length $l + 1$). To show: $[A, i, l + 1]$. Since $l > 1$, the first step in the derivation must be the application of some production of the form $A \to BC$ such that $A \Rightarrow BC$, $B \stackrel{*}{\Rightarrow} w_i \dots w_{i+l_1-1}$ and $C \stackrel{*}{\Rightarrow} w_{i+l_1} \dots w_{i+l}$ for some l_1 with $1 \leq l_1 \leq l$. Then, since our induction claim holds for all lengths $\leq l$, we have $[B, i, l_1]$ and $[C, i + l_1, l - l_1 + 1]$. Consequently, with the complete rule, we also have $[A, i, l + 1]$.

Consequently, our completeness claim holds for every derivation of some substring of the input of length l for any $l \geq 1$.