Parsing

Homework 6 (CYK), due 17 May 2022

Laura Kallmeyer

SS 2022, Heinrich-Heine-Universität Düsseldorf

Question 1 (CYK recognition – general version)

Consider the CFG with non-terminals S, terminals a, start symbol S and productions $S \rightarrow SS \mid a \mid \epsilon$.

1. *Give the chart (the $(n+1) \times (n+1)$-table) that results from the general CYK algorithm for the input aa.*
2. *Now assume that, instead of writing just the non-terminal or terminal into a chart field, we replace any non-terminal with the entire production that we have used, together with indices on the rhs elements. For instance $S \rightarrow S_{1,1}S_{2,0}$ for having used the first S production with antecedent items $[S,1,1]$ and $[S,2,0]$. (For productions with righthand sides consisting of terminals or ε, we just enter the production without indices.)*

   Give the chart that we obtain in that case.
3. *How many parse trees do we obtain?*

Question 2 (CYK parsing for CNF grammars)

Consider the CFG with non-terminal S, terminals a, b, start symbol S and productions:

$S \rightarrow T\ B \mid A\ A \mid U\ U$,
$U \rightarrow U\ B \mid S\ B$, $T \rightarrow A\ X$, $X \rightarrow A\ B$,
$A \rightarrow a$, $B \rightarrow b$

This grammar is in Chomsky Normal Form.

1. *Do a strictly agenda-based chart parsing of the input aabb with this grammar. Assume that in each step the element is taken from the agenda next that has entered the agenda before all others. Put differently: the agenda is a first-in-first-out list, i.e., you can consider it as a list where new elements are added at the end while, in each step, the first element is removed.*

   Give the trace in the form of a table with two columns, the first one for the chart, the second one for the agenda. List in the first one only newly added elements, but give in the second in each step the entire content of the agenda.

chart (newly added items)	*Agenda (entire content)*
$[A,1,1]$, $[A,2,1]$, $[B,3,1]$, $[B,4,1]$	$[A,1,1]$, $[A,2,1]$, $[B,3,1]$, $[B,4,1]$
$[S,1,2]$	$[A,2,1]$, $[B,3,1]$, $[B,4,1]$, $[S,1,2]$
...	...

2. *Is this order of items the on-line or off-line order we had on the slides? In order to justify your answer, give the sequence of lengths and the sequence of end positions of spans that we obtain from the items (when ordered by their order of creation, as can be obtained from the chart column).*