Parsing

Homework 4 (Deduction-based Parsing), due 10 May 2022

Laura Kallmeyer

SS 2022, Heinrich-Heine-Universität Düsseldorf

Question 1 (Top-down Parsing with deduction rules)

Consider again the following CFG: G with $N = \{S, X\}, T = \{a\}$, start symbol S and productions

$$S \to aSX \mid aX \mid a, X \to aX \mid a$$

and the input $w = aaa$.

We use a constrained version of top-down parsing that is possible for CFGs in Greibach Normal form (items are defined as on the slides for top-down parsing):

Whenever the topmost stack symbol is A, there is an A-production $A \to a\gamma$, and the next input symbol is a, we can perform a combined predict-scan:

$$\textit{PredictScan:} \quad \frac{[A\alpha, i]}{[\gamma\alpha, i+1]} \quad A \to w_{i+1}\gamma \in P, |\gamma\alpha| \leq n - i - 1$$

This is the only rule we use. Axiom and goal items are defined as in the standard version.

1. *Give **all** items the parser generates for the input w, using this special top-down parsing schema. Note, however, that items $[\alpha, i]$ with the length of α being greater than $n - i$ are not allowed, i.e., the rules block them (as specified on slide 18).*

   For every item, indicate the production that was used to deduce this item and indicate the antecedent item.

2. *How does the parser know whether $w = aaa$ is in the language generated by the grammar?*

Question 2 (Unger deduction rules for GNF and for CNF)

1. *Consider the following version of the Unger deduction rules for CFGs in Greibach Normal Form:*

   $$\textit{Predict\_scan:} \quad \frac{[\bullet A, i_0 - 1, j]}{[a\bullet, i_0 - 1, i_0], [\bullet A_1, i_0, i_1], \ldots, [\bullet A_k, i_{k-1}, i_k]} \quad \begin{array}{l} A \to aA_1 \ldots A_k \in P \\ w_{i_0} = a, j = i_k, i_m < i_{m+1} \text{ for } 0 \leq m < k \end{array}$$

   This replaces the original scan and predict. Complete remains as the original rule, and the items have the same form, and the axiom and the goal item remain the same.

   *Perform again a parsing of aaa with the grammar from Question 1. Which items arise (list **all** items)?*

2. *Now assume that we are dealing only with grammars in Chomsky Normal Form. The rules of the Unger parsing are then as follows:*

   $$\textit{Predict:} \quad \frac{[\bullet A, i, k]}{[\bullet B, i, j], [\bullet C, j, k]} \quad A \to BC \in P, i < j < k$$

   $$\textit{Scan:} \quad \frac{[\bullet A, i, i+1]}{[A\bullet, i, i+1]} \quad A \to w_{i+1} \in P$$

   $$\textit{Complete:} \quad \frac{[\bullet A, i, k], [B\bullet, i, j], [C\bullet, j, k]}{[A\bullet, i, k]} \quad A \to BC \in P$$

   What is the time complexity of this algorithm in the length of the input string, i.e., of the fixed recognition problem? Explain your answer by estimating the maximal number of different rule applications possible for a given CFG in CNF.

3. *The grammar from Question 1, transformed into CNF, is as follows:*

   G with $N = \{S, X, Y, A\}, T = \{a\}$*, start symbol* S *and productions*

$$S \to AY \mid AX \mid a, X \to AX \mid a, A \to a, Y \to SX$$

   Perform an Unger parsing with the CNF deduction rules from 2. and this grammar for the input word aaa. Give again all the items that arise, together with information about the respective deduction rule and antecedent items.