Problem 42

Measure the size of a CFG by summing \(|\gamma| + 1\) over all productions \(A \rightarrow \gamma\). Show that minimal CFG’s for a language are not unique.

Proof. By counter-example.

Consider the following two distinct CFG’s for the (infinite) language of all strings over the alphabet \(\Sigma = \{a\}\):

\[
\begin{align*}
S & \rightarrow aS \\
S & \rightarrow \epsilon \\
S & \rightarrow Sa \\
S & \rightarrow \epsilon
\end{align*}
\]

The following two claims hold for CFG’s that accept infinite languages:

1. A CFG for an infinite language must contain at least two production rules.
2. A CFG for an infinite language must contain a production rule with at least two symbols on the right hand side.
Proof. of 1.
Suppose not. Then we have some CFG for an infinite language that only contains one production rule. That production rule either:

1. Contains all terminals on the right hand side, in which case it cannot be infinite, as the only word in the language would be the string of terminals in the RHS of that rule. Contradiction!

2. Contains at least one nonterminal on the right hand side, in which case it cannot accept any words, as no derivation will lead to a string that contains only terminals. Contradiction!

Proof. of 2.
Suppose not. Then we have some CFG for an infinite language that only contains production rules with a single symbol on the right hand side. This means that, beginning at the start state, we either derive a terminal directly, accepting a string of length one, or derive a different single nonterminal, which can only produce strings of length 1.

Then the language is not infinite. Contradiction!

So for any infinite language, we get from the above claims that, for some \( \gamma_1 \) and \( \gamma_2 \) that are the RHS of production rules in \( P \),

\[
\sum_{A \rightarrow \gamma \in P} |\gamma| + 1 \geq |\gamma_1| + 1 + |\gamma_2| + 1 \geq 4
\]

That is, the size of the CFG for an infinite language is at least 4.

Finally, \(|S \rightarrow aS| + |S \rightarrow \epsilon| = 3 + 1 = 4\) and \(|S \rightarrow Sa| + |S \rightarrow \epsilon| = 3 + 1 = 4\), so the two distinct grammars given above must be minimal grammars for the same language.