# Review

Mark all correct items.

Unless stated otherwise, $G = (N, T, R, S)$ is a context-free grammar without useless letters.

(a) If a grammar is not $LL(10)$ then there exist two distinct words in $L(G)$ whose prefixes consisting of the first 10 letters of each are identical.

(b) If $L(G)$ contains two distinct words whose prefixes consisting of the first 10 letters of each are identical, then $G$ is not an $LL(10)$ grammar.

(c) Suppose $G_i = (N, T, R_i, S)$ for $i = 1, 2$, where $R_1 \subseteq R_2$. If $G_2$ is $LL(k)$ for some $k$ then so is $G_1$.

(d) Suppose $N = \{S, A\}$ and $T = \{a, b\}$. Then there exists an $LL(2)$ grammar with $|R| = 12$.

(e) If $|N| = |T| = 5$ and $|R| = 100$, then $G$ is not $LL(1)$.

## Solution

(a) Consider the grammar defined by the rules

$$S \to A \mid a^{10}\,,$$
$$A \to a^{10}\,.$$

Clearly, $L(G) = \{a^{10}\}$, so that $L(G)$ does not even contain two distinct words. However, the grammar is not $LL(10)$ (and is in fact even ambiguous) since the word $a^{10}$ can be produced in two distinct ways, namely $S \Longrightarrow a^{10}$ and $S \Longrightarrow A \Longrightarrow a^{10}$.

(b) The grammar defined by the rules

$$S \to aS \mid bS \mid \varepsilon$$

is clearly $LL(10)$ (and even $LL(1)$), yet $L(G) = \{a,b\}^*$ contains numerous pairs of distinct words with identical prefixes of length 10.

(c) Every parse tree for $G_1$ is clearly a parse tree for $G_2$. The fact that $G_2$ is $LL(k)$ means that, at each point of constructing the parse tree top-down, the next $k$ letters of the input give at most one possible continuation. Since every production of $G_1$ is also a production of $G_2$, we certainly have at most one possible continuation for $G_1$ corresponding to these next $k$ letters.

(d) The grammar defined by the rules

$$S \to aaA \mid abS \mid baS \mid bbA \mid a \mid b\,,$$
$$A \to aaS \mid abS \mid baA \mid bbS \mid a \mid b\,,$$

is $LL(2)$. In fact, when parsing a word, at each stage, if we still have to produce a word of length exceeding 1, the following two letters of the input word give exactly one possible production to apply (one of the first four productions for the relevant non-terminal). If the remaining word is of length 1 we again have exactly one possible continuation (one of the last two productions), while if the remaining word is empty then we are stuck. (Notice, by the way, that $L(G)$ consists of all words of odd length over $\{a,b\}$.)

(e) For a grammar to be $LL(1)$, the FIRST sets of the right-hand sides of the rules corresponding to each non-terminal need to be pairwise disjoint. Thus, if $|T| = 5$, then for each non-terminal there are at most 5 rules with non-empty FIRST sets. Recall that, in an $LL(1)$ grammar, for each non-terminal $A$ there may be at most at most one rule $A \to \alpha$ with $\mathsf{Nullable}(\alpha)$. Hence, in our case, each non-terminal may also admit one rule $A \to \alpha$ such that the only word in $T^*$ that may be produced from $\alpha$ is $\varepsilon$. Altogether, for each non-terminal we have at most 6 rules, so that $|R| \leq 5 \cdot 6 = 30$. We mention in passing that the following grammar with $|R| = 30$ is in fact $LL(1)$:

$$
\begin{aligned}
&S \to aS \mid bS \mid cA \mid dC \mid eA \mid \varepsilon\,,\\
&A \to aS \mid bD \mid cD \mid dB \mid eC \mid \varepsilon\,,\\
&B \to aD \mid bB \mid cS \mid dC \mid eD \mid \varepsilon\,,\\
&C \to aB \mid bB \mid cS \mid dB \mid eS \mid \varepsilon\,,\\
&D \to aC \mid bA \mid cA \mid dB \mid eB \mid \varepsilon\,.
\end{aligned}
$$

Thus, (c), (d) and (e) are correct.