Review

Mark all correct items.

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

(a) If $LC(S) \supseteq LC(S) \cdot \{S\}$, then $LC(S) \supseteq LC(S) \cdot L(G)$.

(b) Denote (for the purposes of this question):

$$LC'(A) = \{\beta \in (N \cup T)^* : S' \underset{r}{\overset{*}{\Longrightarrow}} \beta A \gamma \text{ for some } \gamma \in (\mathrm{N} \cup \mathrm{T})^*\}, \qquad \mathrm{A} \in \mathrm{N}.$$

Then $LC'(A)$ is regular for every $A \in N$.

(c) Suppose $A, B \in N$ are such that $A \to \alpha \in R$ if and only if $B \to \alpha \in R$. Then $LC(A) = LC(B)$.

(d) The grammar defined by the rules

$S \to abcdeS \mid abcedS \mid \ldots \mid edcbaS \mid f$

(namely, there are 121 rules, the right-hand sides of the first 120 of which are the 5! permutations of the word *abcde*, all followed by S, and that of the last one is f) is $LR(0)$.

(e) The grammar defined by the rules

$S \to a^{10}b^{20}Sa^{30} \mid a^{20}b^{30}Sa^{40} \mid \varepsilon$

is not $LR(20)$.

Solution

(a) It is true that the condition $LC(S) \supseteq LC(S) \cdot \{S\}$ implies that, for every word $w \in L(G)$, there exists a word of the form $wS\alpha$ which can be produced from S'. Indeed, since $LC(S)$ includes the word ε, the above condition implies that it includes the word S, so we can produce from S' by rightmost derivations some word of the form $SS\alpha$, and then, operating with the first occurrence of S, all the words $wS\alpha$ with $w \in L(G)$. However, these latter derivations are not rightmost. In fact, consider the grammar defined by the rules:

$S \to SS \mid a.$

It is easy to verify that $LC(S) = \{S\}^*$ while $L(G) = \{a\}^+$.

(b) We claim that $LC'(A) = LC(A)$, and in particular $LC'(A)$ is regular. In fact, we obviously have $LC'(A) \supseteq LC(A)$. Now let $\beta \in LC'(A)$. Take $\gamma \in (N \cup T)^*$ such that $S' \underset{r}{\overset{*}{\Longrightarrow}} \beta A \gamma$. Applying to $\beta A \gamma$ a suitable sequence of rightmost derivations, we can produce from it a string of the form $\beta A w$ with $w \in T^*$. Hence $\beta \in LC(A)$, which implies the inclusion $LC'(A) \subseteq LC(A)$.

(c) The relationship between possible derivations of A and of B has little implication on rules having these letters on their right-hand side. Thus, for example, for the grammar defined by the rules

$S \to AB,$

$A \to a,$

$B \to a,$

the condition in question is clearly satisfied, yet one checks easily that $LC(A) = \{\varepsilon\}$ while $LC(B) = \{A\}$.

(d) A word in $L(G)$ consists of a concatenation of words, each of which is some permutation of *abcde*, and a single f at the end. When parsing the word, the first opportunity for reducing is when one arrives at the f. After reducing this f to S, it is necessary each time to reduce the block consisting of the last 6 letters (some permutation of *abcde* with S at the end) to S. Hence the grammar is $LR(0)$.

Alternatively, one checks that

$$LC(S) = \{abcde, abced, \ldots, edcba\}^*.$$

The various $LR(0)$-C sets are the concatenations of this set with all strings consisting of some permutation of *abcde* and an additional S at the end, as well as with f. It is easy to see that no word in one of these sets is a prefix of any word in another, which again yields the same conclusion.

(e) Note first that $L(G)$ consists of concatenations of the two words $a^{10}b^{20}$ and $a^{20}b^{30}$, in any number and order, and then a block of a's of length depending on the number of times each of the two words above has been used before. Thus, when parsing a word bottom-up, one should reduce the ε just after the last b at the word to S, and then shift each time either 30 or 40 times, depending on whether the block of consecutive b's preceding the S is of length 20 or 30, respectively. After each of these shifts one can reduce either $a^{10}b^{20}Sa^{30}$ or $a^{20}b^{30}Sa^{40}$ to S. Thus, the only time it is necessary to shift beyond the reduction place is at the first reduction, of ε to S. For example, suppose the input is $a^{20}b^{30}a^{40}$. After reading $a^{20}b^{30}a^{20}$ we still do not know if the input is $a^{20}b^{30}a^{40}$, so that we had to reduce already 20 steps earlier, or $a^{20}b^{30}a^{20}b^{30}a^{80}$ (or another possibility out of an infinite variety of possibilities), so that we will have to reduce only in the future. It follows that the grammar is not $LR(20)$. Notice that it follows from these considerations that the grammar is $LR(21)$.

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