1 Решение задачи 1

Пусть $\mathcal{L}$ — язык сентенциальных форм, порождаемых грамматикой

I $S \to aSSbS$,

II $S \to bSb$,

III $S \to a$,

и содержащих одинаковое число встречающихся в них термов. Рассмотрим слово $w_0 = a^3b(aS^2b)^2bSb$. Заметим, что $w_0 \in \mathcal{L}$. Действительно, $|w_0|_a = |w_0|_b = |w_0|_S = 5$, а само w_0 получается применением к стартовому нетерминалу S трёх правил грамматики I, одного правила II и двух правил III.

Если применить к w_0 четыре правила I, одно правило II и два правила III, получим слово $w_1 = (a^3b)^2(aS^2b)^5b^2Sb^2 \in \mathcal{L}$, $|w_1|_a = |w_1|_b = |w_1|_S = 11$, причём w_1 — «наименьшее» слово языка, получаемое из w_0.

Аналогично, применение четырёх правил I, одного правила II и двух правил III к w_1 порождает $w_2 = (a^3b)^3(aS^2b)^8b^3Sb^3 \in \mathcal{L}$, $|w_2|_a = |w_2|_b = |w_2|_S = 17$, и w_2 — «наименьшее» слово языка, получаемое из w_1.

В общем случае получаем $w_i = (a^3b)^{i+1}(aS^2b)^{3i+2}b^{i+1}Sb^{i+1}$, где $|w_i|_a = |w_i|_b = |w_i|_S = 6i+5$. Покажем, что язык $\mathcal{L} \notin \texttt{CFG}$, применяя лемму о накачке для CFL к слову w_i. Будем предполагать выполненным пересечение с регулярной аппроксимацией $\mathcal{R} = (a^3b)^+(aS^2b)^+b^+Sb^+$.

Пусть n — длина накачки, и пусть $w = (a^3b)^{n+1}(aS^2b)^{3n+2}b^{n+1}Sb^{n+1}$. Рассмотрим всевозможные разбиения $w = w_0w_1w_2w_3w_4$, $|w_1w_3| > 0$, $|w_1w_2w_3| \leq n$, и покажем, что w не накачивается.

1. $w_1w_3 = (a^3b)^i$ для некоторого i (заметим, что при ином выборе фрагментов накачек в $(a^3b)^{n+1}$ слово выпадает из языка в силу пересечения с $\mathcal{R}$). Тогда при отрицательной накачке в слове уменьшится число букв a, b, но число букв S останется неизменным — слово окажется не в языке;

2. $w_1 = (a^3b)^i$, $w_3 = (aS^2b)^j$ для некоторых i, j. Аналогично, при выборе других фрагментов накачек в $(a^3b)^{n+1}$ и $(aS^2b)^{3n+2}$ соответственно слово оказывается не в языке. Отрицательной накачкой удаляется большее число букв a, чем b и c;

3. $w_1w_3 = (aS^2b)^i$ для некоторого i. Вновь отрицательная накачка удаляет из слова букв S больше, чем a, b;

4. $w_1 = (aS^2b)^i$, $w_3 = b^j$ для некоторых i, j. С отрицательной накачкой в слове остаётся различное число букв;

5. $w_1w_3 = b^i$ для некоторого i. Отрицательная накачка удаляет в слове только буквы b;

6. случай $w_1 = b^i$, $w_3 = b^j$ для некоторых i, j аналогичен предыдущему.

Таким образом, при любом разбиении слово оказывается не в языке (даже не за счёт невозможности вывести данную сентенциальную форму, а просто из-за дисбаланса термов — всё благодаря пересечению с $\mathcal{R}$). Язык $\mathcal{L} \notin \mathtt{CFL}$.

2 Решение задачи 2

Пусть $\mathcal{L} = \{c^i a^n b^k a^j \mid (k > n) \vee (i = j \,\&\, n > 2)\}$. Язык $\mathcal{L} \in \mathtt{CFL}$, поскольку $\mathcal{L} = \mathcal{L}_1 \cup \mathcal{L}_2$, где

$$\mathcal{L}_1 = \{c^i a^n b^k a^j \mid k > n\} \in \mathtt{CFL},$$
$$\mathcal{L}_2 = \{c^i a^n b^k a^j \mid i = j \,\&\, n > 2\} \in \mathtt{CFL}.$$

Язык $\mathcal{L}$ недетерминирован. Докажем это с помощью леммы о накачке для $\mathtt{DCFL}$. Пусть n — длина накачки. Рассмотрим слова

$$w_1 = c^n a^{n+2} \in \mathcal{L},$$
$$w_2 = c^n a^{n+2} b a^n \in \mathcal{L}.$$

У них общий префикс $x = c^n a^{n+1}$, $|x| > n$, и различные суффиксы $y = a$ и $z = aba^n$ соответственно, причём $y[0] = z[0]$. Будем предполагать выполненным пересечение с регулярной аппроксимацией $c^* a^2 a^* b? a^*$.

Пусть накачивается только префикс x, т.е. существует разбиение $x = x_0 x_1 x_2 x_3 x_4$, $|x_1 x_3| > 0$, $|x_1 x_2 x_3| \leq n$, такое, что $(\forall i \in \mathbb{N})\ x_0 x_1^i x_2 x_3^i x_4 y \in \mathcal{L}$ и $x_0 x_1^i x_2 x_3^i x_4 z \in \mathcal{L}$. Рассмотрим разбиения префикса x.

- $x_1 x_3 = c^i$ для некоторого i. Отрицательная накачка рассинхронизирует число букв c и a в слове w_1;
- $x_1 = c^i$, $x_3 = a^j$ для некоторых i, j. При отрицательной накачке наблюдаем рассинхронизацию числа букв c и a уже в слове w_2;
- $x_1 x_3 = a^i$ для некоторого i. Вновь отрицательная накачка рассинхронизирует число букв c и a в слове w_1.

Пусть теперь префикс x и суффиксы y, z накачиваются синхронно, т.е. существуют разбиения $x = x_0 x_1 x_2$, $y = y_0 y_1 y_2$, $z = z_0 z_1 z_2$, где $|x_1 x_2| \leq n$, $|x_1| > 0$, такие, что $(\forall i \in \mathbb{N})\ x_0 x_1^i x_2 y_0 y_1^i y_2 \in \mathcal{L}$ и $x_0 x_1^i x_2 z_0 z_1^i z_2 \in \mathcal{L}$. Заметим, что $x_1 = a^i$ для некоторого i. Какое бы мы ни выбрали разбиение y ($y_1 = a$ или $y_1 = \varepsilon$), при отрицательной накачке слово w_1 выходит из языка из-за рассинхронизации числа букв a и c.

Таким образом, $\mathcal{L} \notin \mathtt{DCFL}$.