- 1. $\mathcal{L}_1 = \{ v_1 z v_2 z \mid |z| \ge 2 \& z, v_2 \in \{a, b, c\}^* \& v_1 \in \{a, b\}^* \}$
- 2. Множество троичных чисел, кратных 5.
- 3. Множество трасс грамматик с правилами вида $N_i \to \gamma N_j, \, N_i \to N_j \gamma, \, N_i \to \epsilon.$
- 4. Множество трасс грамматики G_k с правилами вида $S \to S_1S_1, S_1 \to S_2S_2\dots, S_k \to \mathfrak{a}, S_k \to SS.$

Решение задачи І

«Интуитивно» язык не регулярный — есть требование вхождения одинаковых подслов, которые «по идее» можно как-то отделить от аморфных значений v_1 и v_2 . Попробуем провести это отделение более последовательно.

Два момента в данном языке обращают на себя внимание:

- словарь (язык) v_1 отличается от словаря смежного с ним слова z. Это значит, что если взять первую букву z из разности языка z и языка v_1 , то никакое значение v_1 не сможет её поглотить.
- словарь (язык) v_2 точно такой же, как у смежных вхождений z. Поэтому в принципе v_2 может поглотить любой суффикс первого вхождения z, а также любой префикс второго вхождения.

Чтобы слово v_2 не могло поглотить префикс второго вхождения z, надо сделать так, чтобы его положение определялось однозначно. Мы уже знаем, что выгодно взять z начинающимся c буквы c (это исключит поглощение префикса z значением v_1). Значит, если в слове будет всего две буквы c, то ровно одна из них должна начинать первое вхождение z, и ровно одна должна начинать второе вхождение. Осталось дополнить z достаточно длинными суффиксами, исключающими возможность «накачки» их одновременно.

Кандидат на контрпример — серия слов с \mathfrak{a}^{n+2} с \mathfrak{a}^{n+2} , где \mathfrak{n} — длина накачки.

Теперь можно доказать нерегулярность языка посредством теоремы Майхилла— Нероуда. Действительно, слова вида с $\alpha^m c \alpha^{m+k+1}$ языку не принадлежат, а с $\alpha^{m+k} c \alpha^m$ — точно принадлежат (при этом $z = c \alpha^m, v_2 = \alpha^k$), что порождает нижнетреугольную матрицу принадлежности: наименования строк здесь — это префиксы слов, а столбцов — соответствующие суффиксы.

Также можно использовать этот контрпример для построения короткого доказательства нерегулярности \mathcal{L}_1 , если пересечь его с языком (регулярным) $R = c \alpha^* c \alpha^*$. В слове $c \alpha^{n+2} c \alpha^{n+2}$, принадлежащем пересечению $R \cap \mathcal{L}_1$, можно накачивать лишь фрагмент, состоящий только из букв α , чтобы не выйти из языка R. Пусть такой фрагмент имеет длину k (где k>0). Тогда при отрицательной накачке получится слово $c \alpha^{n-k+2} c \alpha^{n+2}$, которое не входит в \mathcal{L}_1 .

Наиболее неприятный путь — прямое применение леммы о накачке к слову $ca^{n+2}ca^{n+2}$ без использования свойств замыканий. На этом пути придётся разобрать два случая.

- Фрагмент накачки имеет вид a^k . Этот случай аналогичен уже рассмотренному в решении с пересечением.
- Фрагмент накачки имеет вид ca^k . Отметим, что k < n. Тогда при положительной накачке в одну итерацию получим слово $ca^kca^{n+2}ca^{n+2}$. Поскольку это слово начинается c c, то значение z должно начинаться c c, а значит, должно заканчиваться на a^{n+2} (потому что первое c конца вхождение c уж точно будет относиться k e). Но это значит, что e должна содержать также и фрагмент e0 (иначе первое вхождение e0 не сможет заканчиваться на e1), а он в этом слове не повторяется.

В этой задаче у большинства возникла одна из двух проблем:

- Или взято значение z с префиксом в языке $\{a,b\}$, из-за чего оно смешалось со значением v_1 .
- Или взято значение z, равное c^n (очевидно, вы заметили, что в противном случае анализу мешает v_1). Тогда мы можем весь суффикс z, кроме двух первых букв, положить в v_2 . Вообще, в этой задаче не стоит брать значения z, значение префикс-функция у которых больше, чем 0 (и осторожнее с такими словами в других задачах).