

# Базовые категориальные грамматики с однозначным присвоением типов

Вишникин Максим Евгеньевич

МГУ, механико-математический факультет

29 июня 2023

## Пример

Пусть  $G = \langle \{(, )\}, \{S\}, P \rangle$ , где  $P$  состоит из следующих правил:

$$S \rightarrow (S) \quad S \rightarrow SS \quad S \rightarrow \epsilon$$

порождает язык правильных скобочных последовательностей.

## Теорема

Любую контекстно свободную грамматику можно привести к нормальной форме Грейбах, т.е. к форме где содержатся только правила следующих видов:

$$P \rightarrow \epsilon \quad P \rightarrow a \quad P \rightarrow aQ \quad P \rightarrow aQR$$

## Пример

Любой регулярный язык без пустого слова может быть задан как язык КС грамматики только с правилами:

$$P \rightarrow a \quad P \rightarrow aQ$$

## Определение

Базовая категориальная грамматика есть тройка  $\langle \Sigma, N, R, \triangleright \rangle$ , где :

- $\Sigma$  и  $N$  — произвольные непересекающиеся множество
- $R(\in N)$  — произвольный нетерминальный символ
- $\triangleright$  — произвольное конечное бинарное отношение  $\triangleright \subset \Sigma \times Tr(\backslash, /)$ , где  $Tr(\backslash, /)$  индуктивно определяющийся набор категорий (например:  $p/r$  или  $p \backslash (p/q)$ )

Язык, порождаемый базовой категориальной грамматикой, определяется как множество всех непустых слов  $a_1 \dots a_n$  в алфавите  $\Sigma$ , для которых существует такие типы  $p_1, \dots, p_n$ , что для любого  $i \leq n$  выполняется  $p_i \triangleright a_i$  и секвенция  $p_1 \dots p_n \rightarrow R$  выводится в исчислении со схемой аксиом  $p \rightarrow p$ , где  $p \in N$  и правилами:

$$\frac{\Pi \rightarrow A \quad \Gamma B \Delta \rightarrow C}{\Gamma \Pi (A \backslash B) \Delta \rightarrow C} \quad \frac{\Gamma B \Delta \rightarrow C \quad \Pi \rightarrow A}{\Gamma (B/A) \Pi \Delta \rightarrow C}$$

## Пример

В данном исчислении выводится  $p/p, p, p, p \setminus (p \setminus p) \rightarrow p$ :

$$\frac{\frac{\frac{p \rightarrow p \quad p \rightarrow p}{p \rightarrow p \quad p, p \setminus p \rightarrow p}}{p, p, p \setminus (p \setminus p) \rightarrow p} \quad p \rightarrow p}{p/p, p, p, p \setminus (p \setminus p) \rightarrow p}$$

## Теорема

Для любого контекстно-свободного языка (без пустого слова) существует базовая категориальная грамматика задающая данный язык.

## Доказательство.

Возьмём грамматику в форме Грейбах  $G$  и для любой буквы  $a$  добавим в  $\triangleright$ :

- $(a, p)$  если правило  $P \rightarrow a$  было в  $G$
- $(a, p/q)$  если правило  $P \rightarrow aQ$  было в  $G$
- $(a, (p/q)/s)$  если правило  $P \rightarrow aSQ$  было в  $G$



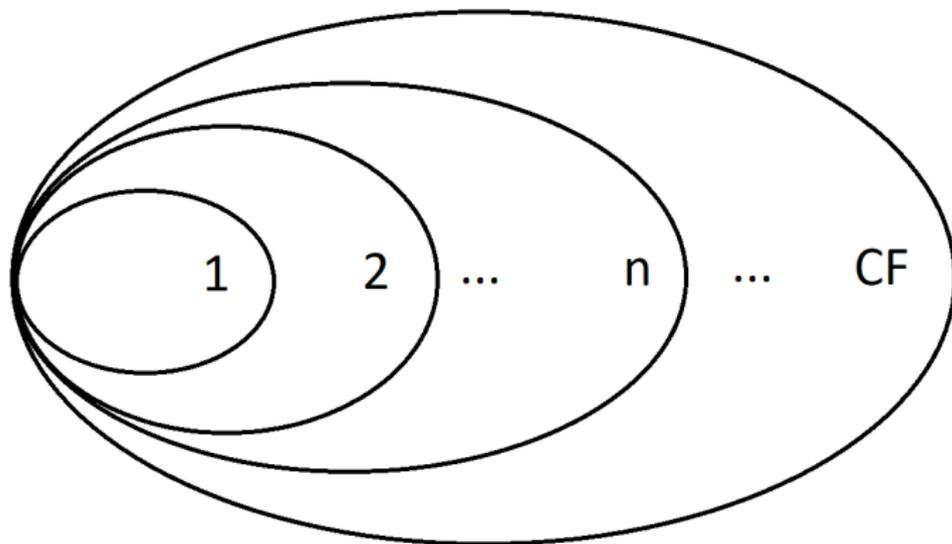
## Определение

Будем называть базовую категориальную грамматику  $G = \langle \Sigma, N, R, \triangleright \rangle$  базовой категориальной грамматикой с однозначным присвоением типов, если отношение  $\triangleright$  функционально.

## Определение

Будем называть базовую категориальную грамматику  $G = \langle \Sigma, N, R, \triangleright \rangle$  базовой категориальной грамматикой с  $\leq k$  присвоением типов, если любой букве присвоено не более  $k$  категорий.

- Любой контекстно-свободный язык является образом гомоморфизма некоторого языка, заданного базовой категориальной грамматикой с однозначным присвоением типов. Достаточно "склеить" необходимые буквы  $\Gamma_N$ .
- Более того, глубина типа в такой базовой категориальной грамматике может быть ограничена (форма Грейбах).
- Базовые категориальные грамматики с однозначным присвоением типов имеют одну "степень свободы" помимо количества букв.



## Теорема

Для любой контекстно-свободного языка  $L (\subseteq \Sigma^*)$  существует категориальная грамматика с двумя присвоениями  $G$  с терминалами из  $\Sigma \times \mathbb{N}$  и гомоморфизм  $h : \Sigma \rightarrow (\Sigma \times \mathbb{N})^*$  такое, что  $h(a) = a_1 \dots a_n$ :

- $h^{-1}(L(G)) = L$

## Пример

Рассмотрим грамматику:

$$a \triangleright p \quad a \triangleright p/p \quad a \triangleright p \setminus p$$

Тогда грамматика :

$$\begin{array}{ll} a \triangleright (p/r)/q & a \triangleright t \\ b \triangleright t \setminus ((p/p)/r) & b \triangleright q \\ c \triangleright q \setminus (t \setminus (p \setminus p)) & c \triangleright r \end{array}$$

задаёт язык  $(abc)^*$ .

Для КС грамматик неразрешимы следующие алгоритмические проблемы:

- $L(G) = L(G')$
- $L(G) \cap L(G') = \emptyset$
- $L(G) = \Sigma^*$
- $L(G)$  - регулярный

## Теорема

*Не существует алгоритма, позволяющего по произвольной контекстно-свободной грамматике узнать, существует ли базовая категориальная грамматика эквивалентная  $G$ .*

## Теорема

*Не существует алгоритма, позволяющего по произвольной категориальной грамматике с двумя присвоениями типов узнать, существует ли базовая категориальная грамматика с однозначным присвоением типов эквивалентная  $G$ .*

# Что происходит в других исчисления эквивалентных КС грамматикам

## Теорема

*Класс языков, порождаемых грамматиками Ламбека, совпадает с классом всех контекстно-свободных языков без пустого слова.*

## Теорема

*Для всякой контекстно-свободной грамматики без  $\epsilon$ -правила существует эквивалентная категориальная грамматика Ламбека, в которой каждой букве алфавита присвоен ровно один тип.*

# Представление базовых категориальных грамматик в виде графа

Оказывается удобно рассматривать грамматику в виде ориентированного графа:

## Определение

Вершины графа – множество примитивных типов  $V = Pr$ , существуют два типа рёбер  $E = L \cup R$ :  $pLq \Leftrightarrow \exists a \in \Sigma : p \rightarrow AqBaC$ , где  $A, B, C \in Pr^*$ ,  $pRq \Leftrightarrow \exists a \in \Sigma : p \rightarrow AaBpC$ , где  $A, B, C \in Pr^*$ .

Глубиной ребра  $pLq$ , заданного правилом  $p \rightarrow AqBaC$ , будем называть длину  $B$ .

## Теорема

Язык  $L \subseteq \Sigma^*$  является автоматным тогда и только тогда, когда множество  $\{C_L^{(r)}(y) \mid y \in \Sigma^*\}$  конечно.

## Пример

Язык  $a^n b^n$  не является регулярным.

## Теорема

*При условии отсутствия бесполезных типов. Если в графе базовой категориальной грамматики  $G$  существует:*

- *Цикл с переходом глубины  $\geq 2$ .*
- *Простой цикл, меняющий ориентацию (есть и  $L$ , и  $R$  рёбра).*

*Тогда язык  $L(G)$  не является автоматным.*

- Язык с присвоениями  $a \triangleright (R \setminus (R \setminus R))$ ,  $b \triangleright R$  - не является автоматным.
- Язык с присвоениями  $a \triangleright (R \setminus p)$ ,  $b \triangleright (R/p)$ ,  $c \triangleright p$  - не является автоматным.
- Данная теорема не является критерием как видно из примера:  $a \triangleright (p/p)$ ,  $b \triangleright (R/R)$ ,  $c \triangleright (p \setminus R)$ ,  $d \triangleright p$ .

Условия регулярности для БКГОПТ:

- Рассмотрим множество слов  $\{b^n\}$  и заметим, что множества правых контекстов данных слов отличны.
- Рассмотрим множество слов  $\{a^n\}$  и заметим, что множества правых контекстов данных слов отличны.
- Рассмотрим множество слов  $\{(ab)^n\}$  и заметим, что множества правых контекстов данных слов отличны.

## Теорема

*Для любой БКГОПТ с одним делением разрешима задача регулярности.  
Более того достаточно быстро.*

**Задача:** Пусть дан класс грамматик. И пусть где-то есть какая-то грамматика данного языка, мы её не знаем, но постепенно шаг за шагом получаем слова из языка данной грамматики.

**Вопрос:** Можем ли мы понять какой язык мы получаем?

## Пример

Предположим, что класс грамматик  $\mathcal{G}$  состоит только из правил  $S \rightarrow \alpha$ . Очевидно, что множество языков заданных данными грамматиками содержит произвольные конечные языки и только их. Положим  $\phi : \bigcup_{k \geq 1} (\Sigma^*)^k \rightarrow (\mathcal{G})$  такая, что  $\phi(\alpha_1, \dots, \alpha_n) = G$ , где  $G$  - грамматика с правилами  $S \rightarrow \alpha_1 \dots S \rightarrow \alpha_n$ . Заметим, что функция очевидно вычислима. Более того заметим следующее свойство:

- Если мы зафиксируем произвольную грамматику из описанного класса и в любом порядке будем подавать слова из данной грамматики, то значение  $\phi$  будет отличаться от выбранной грамматики только на конечном количестве входных наборов слов.

Спасибо за внимание!