Категориальные грамматики

Теория формальных языков.

Степан Львович Кузнецов, Мати Рейнович Пентус, Алексей Андреевич Сорокин

МГУ им. М. В. Ломоносова, межфакультетский курс, весенний семестр 2017–2018 учебного года

Регулярные выражения

Пусть зафиксирован конечный алфавит Σ.

Регулярные выражения

Пусть зафиксирован конечный алфавит Σ.

- Базовые регулярные выражения: элементы алфавита,

Регулярные выражения

Пусть зафиксирован конечный алфавит Σ.

- Базовые регулярные выражения: элементы алфавита,
- Также есть константы 0 (пустой язык) и 1 (язык, содержащий только пустое слово ε).

Регулярные выражения

Пусть зафиксирован конечный алфавит Σ.

- Базовые регулярные выражения: элементы алфавита,
- Также есть константы 0 (пустой язык) и 1 (язык, содержащий только пустое слово ε).
- Бинарные операции: $|$ (объединение) и $\cdot$ (конкатенация или приписывание): $u \cdot v = uv$,

Регулярные выражения

Пусть зафиксирован конечный алфавит Σ.

- Базовые регулярные выражения: элементы алфавита,
- Также есть константы 0 (пустой язык) и 1 (язык, содержащий только пустое слово ε).
- Бинарные операции: $|$ (объединение) и $\cdot$ (конкатенация или приписывание): $u \cdot v = uv$,
- Унарная операция $*$ (итерация, взять любое количество раз): L^* состоит из слов вида $u_1 \dots u_r$, где $r \in \mathbb{N}$, $u_1, \dots, u_r \in L$.

Регулярные выражения

Пусть зафиксирован конечный алфавит Σ.

- Базовые регулярные выражения: элементы алфавита,
- Также есть константы 0 (пустой язык) и 1 (язык, содержащий только пустое слово ε).
- Бинарные операции: $|$ (объединение) и $\cdot$ (конкатенация или приписывание): $u \cdot v = uv$,
- Унарная операция $*$ (итерация, взять любое количество раз): L^* состоит из слов вида $u_1 \dots u_r$, где $r \in \mathbb{N}$, $u_1, \dots, u_r \in L$.
- Если α — регулярное выражение, то $L(\alpha)$ — задаваемый им язык.

Регулярные выражения

Пусть зафиксирован конечный алфавит Σ.

- Базовые регулярные выражения: элементы алфавита,
- Также есть константы 0 (пустой язык) и 1 (язык, содержащий только пустое слово ε).
- Бинарные операции: $|$ (объединение) и $\cdot$ (конкатенация или приписывание): $u \cdot v = uv$,
- Унарная операция $*$ (итерация, взять любое количество раз): L^* состоит из слов вида $u_1 \dots u_r$, где $r \in \mathbb{N}$, $u_1, \dots, u_r \in L$.
- Если α — регулярное выражение, то $L(\alpha)$ — задаваемый им язык.
- Например, $L((a|b)^*) = \{\varepsilon, a, b, aa, ab, ba, bb, \dots\}$.

Регулярные выражения

Пусть зафиксирован конечный алфавит Σ.

- Базовые регулярные выражения: элементы алфавита,
- Также есть константы 0 (пустой язык) и 1 (язык, содержащий только пустое слово ε).
- Бинарные операции: $|$ (объединение) и $\cdot$ (конкатенация или приписывание): $u \cdot v = uv$,
- Унарная операция $*$ (итерация, взять любое количество раз): L^* состоит из слов вида $u_1 \ldots u_r$, где $r \in \mathbb{N}$, $u_1, \ldots, u_r \in L$.
- Если α — регулярное выражение, то $L(\alpha)$ — задаваемый им язык.
- Например, $L((a|b)^*) = \{\varepsilon, a, b, aa, ab, ba, bb, \ldots\}$.
- Приоритет операций: итерация, конкатенация, объединение. При этом значок конкатенации можно опускать.

Регулярные выражения

Пусть зафиксирован конечный алфавит Σ.

- Базовые регулярные выражения: элементы алфавита,
- Также есть константы 0 (пустой язык) и 1 (язык, содержащий только пустое слово ε).
- Бинарные операции: $|$ (объединение) и $\cdot$ (конкатенация или приписывание): $u \cdot v = uv$,
- Унарная операция $*$ (итерация, взять любое количество раз): L^* состоит из слов вида $u_1 \dots u_r$, где $r \in \mathbb{N}$, $u_1, \dots, u_r \in L$.
- Если α — регулярное выражение, то $L(\alpha)$ — задаваемый им язык.
- Например, $L((a|b)^*) = \{\varepsilon, a, b, aa, ab, ba, bb, \dots\}$.
- Приоритет операций: итерация, конкатенация, объединение. При этом значок конкатенации можно опускать.

Язык регулярный, если он задаётся регулярным выражением.

Примеры регулярных языков

- Все слова в алфавите $\{a, b\}$: $(a|b)^*$,

Примеры регулярных языков

- Все слова в алфавите $\{a, b\}$: $(a|b)^*$,
- Слова в алфавите $\{a, b, c\}$, где предпоследняя буква — b: $(a|b|c)^* b(a|b|c)$.

Примеры регулярных языков

- Все слова в алфавите $\{a, b\}$: $(a|b)^*$,
- Слова в алфавите $\{a, b, c\}$, где предпоследняя буква — b: $(a|b|c)^*b(a|b|c)$.
- Слова в алфавите $\{a, b, c\}$, содержащие ровно 2 буквы a: $(b|c)^*a(b|c)^*a(b|c)^*$.

Примеры регулярных языков

- Все слова в алфавите $\{a, b\}$: $(a|b)^*$,
- Слова в алфавите $\{a, b, c\}$, где предпоследняя буква — b: $(a|b|c)^* b(a|b|c)$.
- Слова в алфавите $\{a, b, c\}$, содержащие ровно 2 буквы a: $(b|c)^* a(b|c)^* a(b|c)^*$.
- Слова нечётной длины в алфавите $\{a, b\}$: $(a|b)((a|b)(a|b))^*$.

Примеры регулярных языков

- Все слова в алфавите $\{a, b\}$: $(a|b)^*$,
- Слова в алфавите $\{a, b, c\}$, где предпоследняя буква — b: $(a|b|c)^*b(a|b|c)$.
- Слова в алфавите $\{a, b, c\}$, содержащие ровно 2 буквы a: $(b|c)^*a(b|c)^*a(b|c)^*$.
- Слова нечётной длины в алфавите $\{a, b\}$: $(a|b)((a|b)(a|b))^*$.
- Слова в алфавите $\{a, b, c\}$, содержащие чётное число букв a:

Примеры регулярных языков

- Все слова в алфавите $\{a, b\}$: $(a|b)^*$,
- Слова в алфавите $\{a, b, c\}$, где предпоследняя буква — b: $(a|b|c)^*b(a|b|c)$.
- Слова в алфавите $\{a, b, c\}$, содержащие ровно 2 буквы a: $(b|c)^*a(b|c)^*a(b|c)^*$.
- Слова нечётной длины в алфавите $\{a, b\}$: $(a|b)((a|b)(a|b))^*$.
- Слова в алфавите $\{a, b, c\}$, содержащие чётное число букв a: $((b|c)^*a(b|c)^*a)^*(b|c)^*$

Примеры регулярных языков

- Все слова в алфавите $\{a, b\}$: $(a|b)^*$,
- Слова в алфавите $\{a, b, c\}$, где предпоследняя буква — b: $(a|b|c)^* b(a|b|c)$.
- Слова в алфавите $\{a, b, c\}$, содержащие ровно 2 буквы a: $(b|c)^* a(b|c)^* a(b|c)^*$.
- Слова нечётной длины в алфавите $\{a, b\}$: $(a|b)((a|b)(a|b))^*$.
- Слова в алфавите $\{a, b, c\}$, содержащие чётное число букв a: $((b|c)^* a(b|c)^* a)^* (b|c)^*$.
- Слова в алфавите $\{a, b, c\}$, где перед a идёт только b:

Примеры регулярных языков

- Все слова в алфавите $\{a, b\}$: $(a|b)^*$,
- Слова в алфавите $\{a, b, c\}$, где предпоследняя буква — b: $(a|b|c)^* b(a|b|c)$.
- Слова в алфавите $\{a, b, c\}$, содержащие ровно 2 буквы a: $(b|c)^* a(b|c)^* a(b|c)^*$.
- Слова нечётной длины в алфавите $\{a, b\}$: $(a|b)((a|b)(a|b))^*$.
- Слова в алфавите $\{a, b, c\}$, содержащие чётное число букв a: $((b|c)^* a(b|c)^* a)^*(b|c)^*$.
- Слова в алфавите $\{a, b, c\}$, где перед a идёт только b: $((b|c)^* ba)^*(b|c)^*$

Примеры регулярных языков

- Все слова в алфавите $\{a, b\}$: $(a|b)^*$,
- Слова в алфавите $\{a, b, c\}$, где предпоследняя буква — b: $(a|b|c)^*b(a|b|c)$.
- Слова в алфавите $\{a, b, c\}$, содержащие ровно 2 буквы a: $(b|c)^*a(b|c)^*a(b|c)^*$.
- Слова нечётной длины в алфавите $\{a, b\}$: $(a|b)((a|b)(a|b))^*$.
- Слова в алфавите $\{a, b, c\}$, содержащие чётное число букв a: $((b|c)^*a(b|c)^*a)^*(b|c)^*$.
- Слова в алфавите $\{a, b, c\}$, где перед a идёт только b: $((b|c)^*ba)^*(b|c)^*$.
- Непустые слова в алфавите $\{a, b\}$, в которых одинаковые буквы не идут подряд:

Примеры регулярных языков

- Все слова в алфавите $\{a, b\}$: $(a|b)^*$,
- Слова в алфавите $\{a, b, c\}$, где предпоследняя буква — b: $(a|b|c)^* b(a|b|c)$.
- Слова в алфавите $\{a, b, c\}$, содержащие ровно 2 буквы a: $(b|c)^* a(b|c)^* a(b|c)^*$.
- Слова нечётной длины в алфавите $\{a, b\}$: $(a|b)((a|b)(a|b))^*$.
- Слова в алфавите $\{a, b, c\}$, содержащие чётное число букв a: $((b|c)^* a(b|c)^* a)^*(b|c)^*$.
- Слова в алфавите $\{a, b, c\}$, где перед a идёт только b: $((b|c)^* ba)^*(b|c)^*$.
- Непустые слова в алфавите $\{a, b\}$, в которых одинаковые буквы не идут подряд: $(a(ba)^*(b|1))|(b(ab)^*(a|1))$

Примеры регулярных языков

- Все слова в алфавите $\{a, b\}$: $(a|b)^*$,
- Слова в алфавите $\{a, b, c\}$, где предпоследняя буква — b: $(a|b|c)^* b(a|b|c)$.
- Слова в алфавите $\{a, b, c\}$, содержащие ровно 2 буквы a: $(b|c)^* a(b|c)^* a(b|c)^*$.
- Слова нечётной длины в алфавите $\{a, b\}$: $(a|b)((a|b)(a|b))^*$.
- Слова в алфавите $\{a, b, c\}$, содержащие чётное число букв a: $((b|c)^* a(b|c)^* a)^*(b|c)^*$.
- Слова в алфавите $\{a, b, c\}$, где перед a идёт только b: $((b|c)^* ba)^*(b|c)^*$.
- Непустые слова в алфавите $\{a, b\}$, в которых одинаковые буквы не идут подряд: $(a(ba)^*(b|1))|(b(ab)^*(a|1))$.
- Непустые слова в алфавите $\{a, b, c\}$, в которых одинаковые буквы не идут подряд:

Примеры регулярных языков

- Все слова в алфавите $\{a, b\}$: $(a|b)^*$,
- Слова в алфавите $\{a, b, c\}$, где предпоследняя буква — b: $(a|b|c)^* b(a|b|c)$.
- Слова в алфавите $\{a, b, c\}$, содержащие ровно 2 буквы a: $(b|c)^* a(b|c)^* a(b|c)^*$.
- Слова нечётной длины в алфавите $\{a, b\}$: $(a|b)((a|b)(a|b))^*$.
- Слова в алфавите $\{a, b, c\}$, содержащие чётное число букв a: $((b|c)^* a(b|c)^* a)^*(b|c)^*$.
- Слова в алфавите $\{a, b, c\}$, где перед a идёт только b: $((b|c)^* ba)^*(b|c)^*$.
- Непустые слова в алфавите $\{a, b\}$, в которых одинаковые буквы не идут подряд: $(a(ba)^*(b|1))|(b(ab)^*(a|1))$.
- Непустые слова в алфавите $\{a, b, c\}$, в которых одинаковые буквы не идут подряд: $(H|1)(cH)^*(1|c|cH)$, где H — ответ на предыдущий пункт.

Структура именной группы в русском языке

Именная группа в русском языке состоит из:

- Существительного: (*дом*),

N

Структура именной группы в русском языке

Именная группа в русском языке состоит из:

- Существительного: (*дом*),

$$N$$

- перед которым идёт какое-то количество прилагательных (*маленький красивый дом*);

$$A^{*}N$$

Структура именной группы в русском языке

Именная группа в русском языке состоит из:

- Существительного: (*дом*),

$$N$$

- перед которым идёт какое-то количество прилагательных (*маленький красивый дом*);

$$A^* N$$

- прилагательным могут предшествовать наречия (*довольно маленький очень красивый дом*);

$$(R^* A)^* N$$

Структура именной группы в русском языке

Именная группа в русском языке состоит из:

- Существительного: (*дом*),

$$N$$

- перед которым идёт какое-то количество прилагательных (*маленький красивый дом*);

$$A^* N$$

- прилагательным могут предшествовать наречия (*довольно маленький очень красивый дом*);

$$(R^* A)^* N$$

- после существительного может стоять группа дополнения в родительном падеже (*довольно маленький очень красивый дом почти неизвестного человека*):

$$(R^* A)^* N((R^* A_g)^* N_g)?$$

- Не учтено: согласование, отрицание (*не очень красивый дом*)...

Определение конечного автомата

Пусть Σ — конечный алфавит.

Определение конечного автомата

Конечный автомат: помеченный граф $M = \langle Q, \Sigma, \Delta, q_0, F \rangle$, где

Определение конечного автомата

Пусть Σ — конечный алфавит.

Определение конечного автомата

Конечный автомат: помеченный граф $M = \langle Q, \Sigma, \Delta, q_0, F \rangle$, где

- Q — конечное множество состояний (вершин графа),

Определение конечного автомата

Пусть Σ — конечный алфавит.

Определение конечного автомата

Конечный автомат: помеченный граф $M = \langle Q, \Sigma, \Delta, q_0, F \rangle$, где

- Q — конечное множество состояний (вершин графа),
- Δ — конечное множество переходов (рёбер с метками) вида $\langle q_1, w \rangle \to q_2$, $q_1, q_2 \in Q$, $w \in \Sigma^*$.

Определение конечного автомата

Пусть Σ — конечный алфавит.

Определение конечного автомата

Конечный автомат: помеченный граф $M = \langle Q, \Sigma, \Delta, q_0, F\rangle$, где

- Q — конечное множество состояний (вершин графа),
- Δ — конечное множество переходов (рёбер с метками) вида $\langle q_1, w\rangle \to q_2$, $q_1, q_2 \in Q$, $w \in \Sigma^*$.
- $q_0 \in Q$ — стартовое состояние
- $F \subseteq Q$ — завершающие состояния.

Определение конечного автомата

Пусть Σ — конечный алфавит.

Определение конечного автомата

Конечный автомат: помеченный граф $M = \langle Q, \Sigma, \Delta, q_0, F \rangle$, где

- Q — конечное множество состояний (вершин графа),
- Δ — конечное множество переходов (рёбер с метками) вида $\langle q_1, w \rangle \to q_2$, $q_1, q_2 \in Q$, $w \in \Sigma^*$.
- $q_0 \in Q$ — стартовое состояние
- $F \subseteq Q$ — завершающие состояния.

Метка пути — конкатенация всех меток на входящих в него рёбрах.

Определение конечного автомата

Пусть Σ — конечный алфавит.

Определение конечного автомата

Конечный автомат: помеченный граф $M = \langle Q, \Sigma, \Delta, q_0, F \rangle$, где

- Q — конечное множество состояний (вершин графа),
- Δ — конечное множество переходов (рёбер с метками) вида $\langle q_1, w \rangle \to q_2$, $q_1, q_2 \in Q$, $w \in \Sigma^*$.
- $q_0 \in Q$ — стартовое состояние
- $F \subseteq Q$ — завершающие состояния.

Метка пути — конкатенация всех меток на входящих в него рёбрах.
Язык $L(M)$ — метки путей из начального состояния в завершающие.

Определение конечного автомата

Пусть Σ — конечный алфавит.

Определение конечного автомата

Конечный автомат: помеченный граф $M = \langle Q, \Sigma, \Delta, q_0, F \rangle$, где

- Q — конечное множество состояний (вершин графа),
- Δ — конечное множество переходов (рёбер с метками) вида $\langle q_1, w \rangle \to q_2$, $q_1, q_2 \in Q$, $w \in \Sigma^*$.
- $q_0 \in Q$ — стартовое состояние
- $F \subseteq Q$ — завершающие состояния.

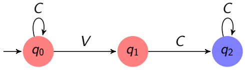
Метка пути — конкатенация всех меток на входящих в него рёбрах.
Язык $L(M)$ — метки путей из начального состояния в завершающие.
Язык — автоматный, если задаётся некоторым конечным автоматом.

Примеры конечных автоматов

- Закрытый слог

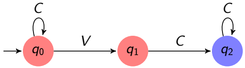
Примеры конечных автоматов

- Закрытый слог



Примеры конечных автоматов

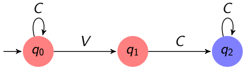
- Закрытый слог



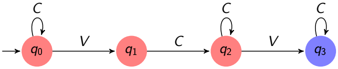
- Слово с 2 гласными, разделёнными хотя бы одним согласным:

Примеры конечных автоматов

- Закрытый слог

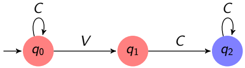


- Слово с 2 гласными, разделёнными хотя бы одним согласным:

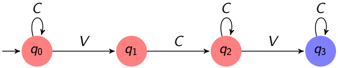


Примеры конечных автоматов

- Закрытый слог



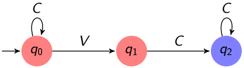
- Слово с 2 гласными, разделёнными хотя бы одним согласным:



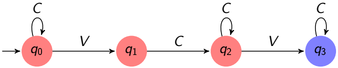
- Слова в алфавите a, b, c, где перед a обязательно идёт b:

Примеры конечных автоматов

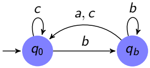
- Закрытый слог



- Слово с 2 гласными, разделёнными хотя бы одним согласным:

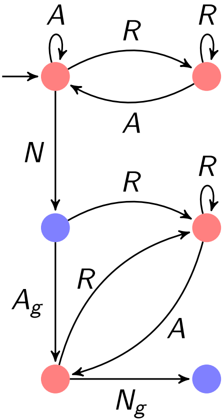


- Слова в алфавите a, b, c, где перед a обязательно идёт b:



Примеры конечных автоматов

Именная группа в русском языке $(R^*A)^*N((R^*A_g)^*N_g)?$:



Свойства автоматных языков

Теорема

Любой автоматный язык распознаётся автоматом с однобуквенными переходами.

Свойства автоматных языков

Теорема

Любой автоматный язык распознаётся автоматом с однобуквенными переходами.

Определение

Автомат с однобуквенными переходами — детерминированный, если ни из какого состояния нет двух переходов по одной и той же букве.

Теорема

Любой автоматный язык распознаётся детерминированным конечным автоматом.

Свойства автоматных языков

Теорема

Любой автоматный язык распознаётся автоматом с однобуквенными переходами.

Определение

Автомат с однобуквенными переходами — детерминированный, если ни из какого состояния нет двух переходов по одной и той же букве.

Теорема

Любой автоматный язык распознаётся детерминированным конечным автоматом.

Теорема (Клини, 1956)

Классы автоматных и регулярных языков совпадают.

Доказательства неавтоматности языков

Теорема

Пусть найдётся бесконечное множество слов $u_1, v_1, \ldots, u_n, v_n, \ldots$ *таких, что* $\forall i (u_i v_i \in L)$, *если* $i \neq j$, *то* $u_i v_j \notin L$. *Тогда язык* L *— не автоматный.*

Доказательства неавтоматности языков

Теорема

Пусть найдётся бесконечное множество слов $u_1, v_1, \ldots, u_n, v_n, \ldots$ *таких, что* $\forall i (u_i v_i \in L)$, *если* $i \neq j$, *то* $u_i v_j \notin L$. *Тогда язык* L *— не автоматный.*

Пример

Язык $L = \{a^n b^n \mid n \in \mathbb{N}\}$ *— не автоматный.*

Доказательства неавтоматности языков

Теорема

Пусть найдётся бесконечное множество слов $u_1, v_1, \ldots, u_n, v_n, \ldots$ *таких, что* $\forall i (u_i v_i \in L)$, *если* $i \neq j$, *то* $u_i v_j \notin L$. *Тогда язык* L *— не автоматный.*

Пример

Язык $L = \{a^n b^n \mid n \in \mathbb{N}\}$ *— не автоматный.*

Теорема (Лемма о разрастании)

Для всякого автоматного языка L *найдётся число* p, *такое что для всякого* $w \in L$, *такого что* $|w| \geqslant p$, *верно что* $\exists x, y, z\ (w = xyz,\ |xy| \leqslant p, |y| > 0)$ *и* $\forall k \in \mathbb{N} (xy^k z \in L)$.

Доказательства неавтоматности языков

Теорема

Пусть найдётся бесконечное множество слов $u_1, v_1, \ldots, u_n, v_n, \ldots$ *таких, что* $\forall i (u_i v_i \in L)$, *если* $i \neq j$, *то* $u_i v_j \notin L$. *Тогда язык* L *— не автоматный.*

Пример

Язык $L = \{a^n b^n \mid n \in \mathbb{N}\}$ *— не автоматный.*

Теорема (Лемма о разрастании)

Для всякого автоматного языка L *найдётся число* p, *такое что для всякого* $w \in L$, *такого что* $|w| \geqslant p$, *верно что* $\exists x, y, z\ (w = xyz,\ |xy| \leqslant p, |y| > 0)$ *и* $\forall k \in \mathbb{N} (xy^k z \in L)$.

Упражнение

Доказать, что $L = \{a^n b^n \mid n \in \mathbb{N}\}$ *— не автоматный язык, с помощью леммы о разрастании.*

Неавтоматность естественных языков

Замечание: естественные языки неавтоматны.

Неавтоматность естественных языков

Замечание: естественные языки неавтоматны.

- The dog chases the cat.

Неавтоматность естественных языков

Замечание: естественные языки неавтоматны.

- The dog chases the cat.
- The dog whom a dog chases chases the cat.

Неавтоматность естественных языков

Замечание: естественные языки неавтоматны.

- The dog chases the cat.
- The dog whom a dog chases chases the cat.
- The dog whom a dog whom a dog chases chases the cat.

Неавтоматность естественных языков

Замечание: естественные языки неавтоматны.

- The dog chases the cat.
- The dog whom a dog chases chases the cat.
- The dog whom a dog whom a dog chases chases the cat.
- ...
- The dog (whom a dog)^m chases^{m+1} the cat.

Неавтоматность естественных языков

Замечание: естественные языки неавтоматны.

- The dog chases the cat.
- The dog whom a dog chases chases the cat.
- The dog whom a dog whom a dog chases chases the cat.
- ...
- The dog (whom a dog)^m chases^{m+1} the cat.
- Тот же шаблон, что в $a^n b^n$.

Неавтоматность естественных языков

Замечание: естественные языки неавтоматны.

- The dog chases the cat.
- The dog whom a dog chases chases the cat.
- The dog whom a dog whom a dog chases chases the cat.
- ...
- The dog (whom a dog)^m chases^{m+1} the cat.

- Тот же шаблон, что в $a^n b^n$.
- На самом деле регулярных языков не хватит даже для полного описания морфологии (редупликация):

Пример (Редупликация в языке бамбара, семья манде)

wulu	*"собака"*
wulu-o-wulu	*"любая собака"*
wulunyinina	*"видящий собаку"*
wulunyinina-o-wulunyinina	*"любой, кто видит собаку"*

Контекстно-свободные грамматики

Определение

Контекстно-свободная грамматика: $G = \langle N, \Sigma, P, S \rangle$, *где*

- Σ — *алфавит,*
- N — *множество вспомогательных символов,*
- P — *множество правил* $A \to \alpha$, $A \in N, \alpha \in (\Sigma \cup N)^*$,
- $S \in N$ — *стартовый символ.*

Контекстно-свободные грамматики

Определение

Контекстно-свободная грамматика: $G = \langle N, \Sigma, P, S \rangle$, *где*

- Σ — *алфавит,*
- N — *множество вспомогательных символов,*
- P — *множество правил* $A \to \alpha$, $A \in N, \alpha \in (\Sigma \cup N)^*$,
- $S \in N$ — *стартовый символ.*

Отношение перехода $\vdash'$ на множестве $(\Sigma \cup N)^*$ — наименьшее отношение, такое что

$$(A \to \alpha \in P) \Leftrightarrow \forall \gamma, \eta \in (\Sigma \cup N)^* \ (\gamma A \eta \vdash' \gamma \alpha \eta)$$

Контекстно-свободные грамматики

Определение

Контекстно-свободная грамматика: $G = \langle N, \Sigma, P, S \rangle$, *где*

- Σ — *алфавит,*
- N — *множество вспомогательных символов,*
- P — *множество правил* $A \to \alpha$, $A \in N, \alpha \in (\Sigma \cup N)^*$,
- $S \in N$ — *стартовый символ.*

Отношение перехода $\vdash'$ на множестве $(\Sigma \cup N)^*$ — наименьшее отношение, такое что

$$(A \to \alpha \in P) \Leftrightarrow \forall \gamma, \eta \in (\Sigma \cup N)^* \ (\gamma A \eta \vdash' \gamma \alpha \eta)$$

$$(\alpha \vdash \beta) \leftrightarrow (\exists \alpha_1, \ldots, \alpha_m \ (\alpha \vdash' \alpha_1 \ldots \alpha_2 \ldots \alpha_m \vdash' \beta))$$

Контекстно-свободные грамматики

Определение

Контекстно-свободная грамматика: $G = \langle N, \Sigma, P, S \rangle$, *где*

- Σ — *алфавит,*
- N — *множество вспомогательных символов,*
- P — *множество правил* $A \to \alpha,\ A \in N, \alpha \in (\Sigma \cup N)^*$,
- $S \in N$ — *стартовый символ.*

Отношение перехода $\vdash'$ на множестве $(\Sigma \cup N)^*$ — наименьшее отношение, такое что

$$(A \to \alpha \in P) \Leftrightarrow \forall \gamma, \eta \in (\Sigma \cup N)^* \ (\gamma A \eta \vdash' \gamma \alpha \eta)$$

$$(\alpha \vdash \beta) \leftrightarrow (\exists \alpha_1, \ldots, \alpha_m \ (\alpha \vdash' \alpha_1 \ldots \alpha_2 \ldots \alpha_m \vdash' \beta))$$

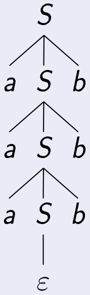
$L(G) = \{w \in \Sigma^* \mid S \vdash w\}$ — язык, порождаемый грамматикой.

Примеры контекстно-свободных грамматик

Пример

$L(G) = \{a^n b^n \mid n \in \mathbb{N}\}$: $S \to aSb$, $S \to \varepsilon$.

Пример (Дерево вывода)



Примеры контекстно-свободных грамматик

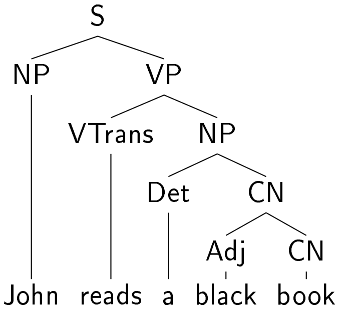
Пример

$$
\begin{array}{lll}
S \to NP\ VP & NP \to Det\ CN & NP \to John \\
VP \to VTrans\ NP & VTrans \to reads & Det \to a \\
CN \to Adj\ CN & Adj \to black & CN \to book
\end{array}
$$

Примеры контекстно-свободных грамматик

Пример

$S \to NP\ VP$	$NP \to Det\ CN$	$NP \to John$
$VP \to VTrans\ NP$	$VTrans \to reads$	$Det \to a$
$CN \to Adj\ CN$	$Adj \to black$	$CN \to book$

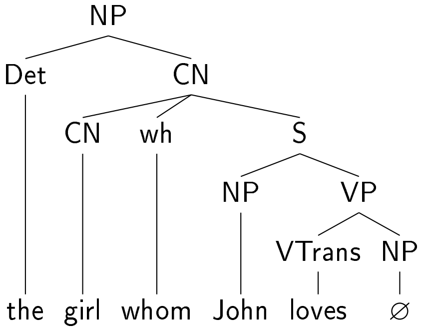


Примеры контекстно-свободных грамматик

The girl whom John loves

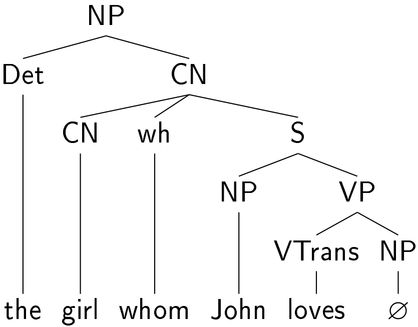
Примеры контекстно-свободных грамматик

The girl whom John loves



Примеры контекстно-свободных грамматик

The girl whom John loves



Однако мы не можем разрешить переход NP в пустое слово!

Примеры контекстно-свободных грамматик

Пример

$NP \to Det\ CN$
$CN \to CN\ wh\ S/NP$
$S/NP \to NP\ VP/NP$

$Det \to the$
$wh \to whom$
$VP/NP \to VTrans\ \cancel{NP}$

$NP \to John$
$CN \to girl$
$VTrans \to loves$

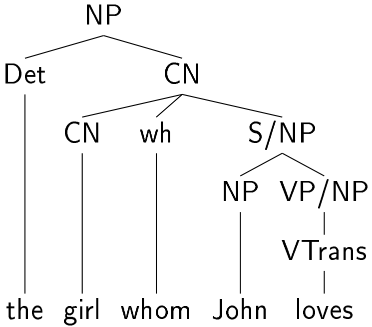
Примеры контекстно-свободных грамматик

Пример

$NP \to Det\ CN$ $\qquad$ $Det \to the$ $\qquad$ $NP \to John$

$CN \to CN\ wh\ S/NP$ $\qquad$ $wh \to whom$ $\qquad$ $CN \to girl$

$S/NP \to NP\ VP/NP$ $\qquad$ $VP/NP \to VTrans\ \cancel{NP}$ $\qquad$ $VTrans \to loves$



Свойства контекстно-свободных языков

- Всякий автоматный язык является контекстно-свободным.

Свойства контекстно-свободных языков

- Всякий автоматный язык является контекстно-свободным.

Идея доказательства

Положить $N = Q$, $S = q_0$, заменить переходы $\langle q_1, a \rangle \to q_2$ на правила $q_1 \to a q_2$, а для всех завершающих состояний $q \in F$ добавить правила $q \to \varepsilon$.

- всякий контекстно-свободный язык порождается грамматикой только с правилами вида $A \to BC$, $A \to a$ и $S \to \varepsilon$, причём S не содержится в правой части правил (нормальная форма Хомского).

Свойства контекстно-свободных языков

- Всякий автоматный язык является контекстно-свободным.

Идея доказательства

Положить $N = Q$, $S = q_0$, заменить переходы $\langle q_1, a\rangle \to q_2$ на правила $q_1 \to aq_2$, а для всех завершающих состояний $q \in F$ добавить правила $q \to \varepsilon$.

- всякий контекстно-свободный язык порождается грамматикой только с правилами вида $A \to BC$, $A \to a$ и $S \to \varepsilon$, причём S не содержится в правой части правил (нормальная форма Хомского).
- Пересечение контекстно-свободного языка с автоматным является контекстно-свободным.

Свойства контекстно-свободных языков

- Всякий автоматный язык является контекстно-свободным.

Идея доказательства

Положить $N = Q$, $S = q_0$, заменить переходы $\langle q_1, a\rangle \to q_2$ на правила $q_1 \to aq_2$, а для всех завершающих состояний $q \in F$ добавить правила $q \to \varepsilon$.

- всякий контекстно-свободный язык порождается грамматикой только с правилами вида $A \to BC$, $A \to a$ и $S \to \varepsilon$, причём S не содержится в правой части правил (нормальная форма Хомского).
- Пересечение контекстно-свободного языка с автоматным является контекстно-свободным.
- Если отображение ϕ заменяет букву a на некоторое фиксированное слово w_a, то из того, что L — контекстно-свободный язык, следует, что $\phi(L)$ – тоже контектсно-свободный язык.

Лемма о разрастании для контекстно-свободных языков

Теорема (Лемма о разрастании)

Для всякого контекстно-свободного языка L найдётся число p, такое что для всякого $w \in L$, такого что $|w| \geqslant p$, верно что $\exists x, u, y, v, z \; (w = xuyvz, \; |uyv| \leqslant p, |uv| > 0)$ и $\forall k \in \mathbb{N} (xu^k y v^k z \in L)$.

Идея доказательства

Лемма о разрастании для контекстно-свободных языков

Теорема (Лемма о разрастании)

Для всякого контекстно-свободного языка L найдётся число p, такое что для всякого $w \in L$, такого что $|w| \geqslant p$, верно что $\exists x, u, y, v, z$ $(w = xuyvz,\ |uyv| \leqslant p, |uv| > 0)$ и $\forall k \in \mathbb{N}(xu^k yv^k z \in L)$.

Идея доказательства

- Рассмотрим грамматику в нормальной форме Хомского.

Лемма о разрастании для контекстно-свободных языков

Теорема (Лемма о разрастании)

Для всякого контекстно-свободного языка L найдётся число p, такое что для всякого $w \in L$, такого что $|w| \geqslant p$, верно что $\exists x, u, y, v, z$ $(w = xuyvz,\ |uyv| \leqslant p, |uv| > 0)$ и $\forall k \in \mathbb{N}(xu^k yv^k z \in L)$.

Идея доказательства

- Рассмотрим грамматику в нормальной форме Хомского.
- Тогда глубина дерева вывода для w не меньше $\log_2 |w| + 1$.

Лемма о разрастании для контекстно-свободных языков

Теорема (Лемма о разрастании)

Для всякого контекстно-свободного языка L найдётся число p, такое что для всякого $w \in L$, такого что $|w| \geqslant p$, верно что $\exists x, u, y, v, z$ $(w = xuyvz,\ |uyv| \leqslant p, |uv| > 0)$ и $\forall k \in \mathbb{N}(xu^kyv^kz \in L)$.

Идея доказательства

- Рассмотрим грамматику в нормальной форме Хомского.
- Тогда глубина дерева вывода для w не меньше $\log_2 |w| + 1$.
- Значит, если взять слово с $|w| \geqslant 2^{|N|}$, то в самой длинной ветви есть повторящиеся вспомогательные символы.

Лемма о разрастании для контекстно-свободных языков

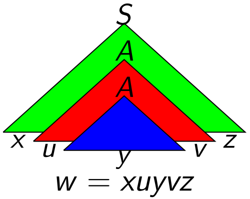
Теорема (Лемма о разрастании)

Для всякого контекстно-свободного языка L найдётся число p, такое что для всякого $w \in L$, такого что $|w| \geqslant p$, верно что $\exists x, u, y, v, z$ $(w = xuyvz,\ |uyv| \leqslant p, |uv| > 0)$ и $\forall k \in \mathbb{N}(xu^k yv^k z \in L)$.

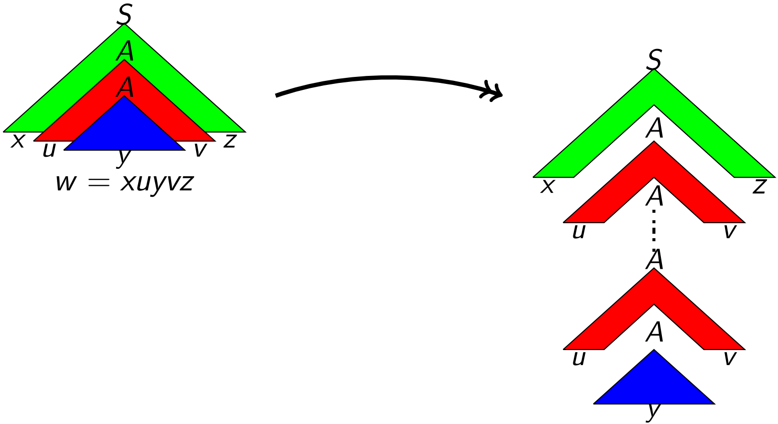
Идея доказательства

- Рассмотрим грамматику в нормальной форме Хомского.
- Тогда глубина дерева вывода для w не меньше $\log_2 |w| + 1$.
- Значит, если взять слово с $|w| \geqslant 2^{|N|}$, то в самой длинной ветви есть повторящиеся вспомогательные символы.
- Можно неограниченно дублировать поддерево между ними.

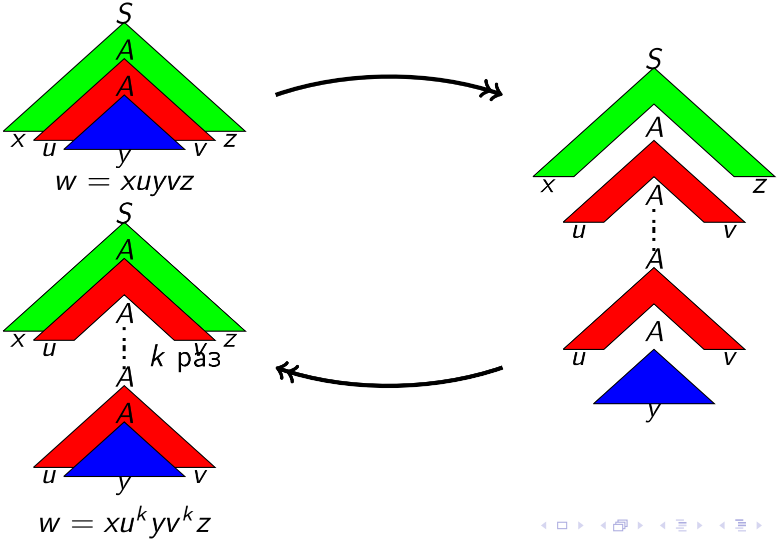
Лемма о разрастании для контекстно-свободных языков



Лемма о разрастании для контекстно-свободных языков



Лемма о разрастании для контекстно-свободных языков



Примеры неконтекстно-свободных языков

Пример

Язык $L = \{a^m b^n a^m b^n \mid m, n \in \mathbb{N}\}$ *не является контекстно-свободным.*

Примеры неконтекстно-свободных языков

Пример

Язык $L = \{a^m b^n a^m b^n \mid m, n \in \mathbb{N}\}$ *не является контекстно-свободным.*

Схема доказательства

- Возьмём $w = a^p b^p a^p b^p$ и представим его в виде $w = xuyvz$ из леммы о разрастании.

Примеры неконтекстно-свободных языков

Пример

Язык $L = \{a^m b^n a^m b^n \mid m, n \in \mathbb{N}\}$ *не является контекстно-свободным.*

Схема доказательства

- Возьмём $w = a^p b^p a^p b^p$ и представим его в виде $w = xuyvz$ из леммы о разрастании.
- Без ограничения общности u содержит букву a, пусть она из первого сегмента.

Примеры неконтекстно-свободных языков

Пример

Язык $L = \{a^m b^n a^m b^n \mid m, n \in \mathbb{N}\}$ *не является контекстно-свободным.*

Схема доказательства

- Возьмём $w = a^p b^p a^p b^p$ и представим его в виде $w = xuyvz$ из леммы о разрастании.
- Без ограничения общности u содержит букву a, пусть она из первого сегмента.
- Чтобы слово xu^2yv^2z имело требуемый вид, нужно, чтобы u или v содержало a из второго сегмента.

Примеры неконтекстно-свободных языков

Пример

Язык $L = \{a^m b^n a^m b^n \mid m, n \in \mathbb{N}\}$ *не является контекстно-свободным.*

Схема доказательства

- Возьмём $w = a^p b^p a^p b^p$ и представим его в виде $w = xuyvz$ из леммы о разрастании.
- Без ограничения общности u содержит букву a, пусть она из первого сегмента.
- Чтобы слово xu^2yv^2z имело требуемый вид, нужно, чтобы u или v содержало a из второго сегмента.
- Но тогда $|uyv| \geqslant p + 2$. Противоречие.

Примеры неконтекстно-свободных языков

Пример

Язык $L = \{a^m b^n a^m b^n \mid m, n \in \mathbb{N}\}$ *не является контекстно-свободным.*

Схема доказательства

- Возьмём $w = a^p b^p a^p b^p$ и представим его в виде $w = xuyvz$ из леммы о разрастании.
- Без ограничения общности u содержит букву a, пусть она из первого сегмента.
- Чтобы слово xu^2yv^2z имело требуемый вид, нужно, чтобы u или v содержало a из второго сегмента.
- Но тогда $|uyv| \geqslant p + 2$. Противоречие.

Пример

Язык $L = \{ww \mid w \in \Sigma^*\}$ *не является контекстно-свободным для любого алфавита* Σ, *содержащего хотя бы 2 буквы.*

Контекстно-свободны ли естественные языки

- Достаточно ли контекстно-свободных языков для моделирования синтаксиса и семантики?

Контекстно-свободны ли естественные языки

- Достаточно ли контекстно-свободных языков для моделирования синтаксиса и семантики?
- Возможно, недостаточно даже для морфологии:

Пример (Редупликация в языке бамбара, семья манде)

wulu	*“собака”*
wulu-o-wulu	*“любая собака”*
wulunyinina	*“видящий собаку”*
wulunyinina-o-wulunyinina	*“любой, кто видит собаку”*

Контекстно-свободны ли естественные языки

- Достаточно ли контекстно-свободных языков для моделирования синтаксиса и семантики?
- Возможно, недостаточно даже для морфологии:

Пример (Редупликация в языке бамбара, семья манде)

wulu	*“собака”*
wulu-o-wulu	*“любая собака”*
wulunyinina	*“видящий собаку”*
wulunyinina-o-wulunyinina	*“любой, кто видит собаку”*

- Для синтаксиса заведомо недостаточно.

Контекстно-свободны ли естественные языки

- Рассмотрим предложения вида
  Маша, Петя, Вася пришли к финишу первой, седьмым, четвёртым, соответственно.

Контекстно-свободны ли естественные языки

- Рассмотрим предложения вида
  Маша, Петя, Вася пришли к финишу первой, седьмым, четвёртым, соответственно.
- Они все подходят под регулярный шаблон R
  $\mathbf{Name}^{+}$ *пришли к финишу* $\mathbf{Num}^{+}$ *соответственно.*
  Здесь **Name** – множество имён, **Num** – множество порядковых числительных в единственном числе и творительном падеже.

Контекстно-свободны ли естественные языки

- Рассмотрим предложения вида
  Маша, Петя, Вася пришли к финишу первой, седьмым, четвёртым, соответственно.
- Они все подходят под регулярный шаблон R
  Name$^+$ *пришли к финишу* **Num**$^+$ *соответственно*.
  Здесь **Name** – множество имён, **Num** – множество порядковых числительных в единственном числе и творительном падеже.
- Если множество корректных предложений L контекстно-свободно, то и $L \cap L(R)$ – контекстно-свободный язык.

Контекстно-свободны ли естественные языки

- Рассмотрим предложения вида
  Маша, Петя, Вася пришли к финишу первой, седьмым, четвёртым, соответственно.
- Они все подходят под регулярный шаблон R
  $\mathbf{Name}^{+}$ *пришли к финишу* $\mathbf{Num}^{+}$ *соответственно*.
  Здесь **Name** – множество имён, **Num** – множество порядковых числительных в единственном числе и творительном падеже.
- Если множество корректных предложений L контекстно-свободно, то и $L \cap L(R)$ – контекстно-свободный язык.
- Так ли это?

Контекстно-свободны ли естественные языки

- Рассмотрим предложения вида
  Маша, Петя, Вася пришли к финишу первой, седьмым, четвёртым, соответственно.
- Они все подходят под регулярный шаблон R
  $\mathbf{Name}^+$ *пришли к финишу* $\mathbf{Num}^+$ *соответственно*.
  Здесь **Name** – множество имён, **Num** – множество порядковых числительных в единственном числе и творительном падеже.
- Если множество корректных предложений L контекстно-свободно, то и $L \cap L(R)$ – контекстно-свободный язык.
- Так ли это?
- В предложениях указанного вида числительные согласуются с существительными в роде...

Контекстно-свободны ли естественные языки

- Чтобы предложение вида
  Маша, Петя, Вася, ... пришли к финишу первой, седьмым, четвёртым, ..., соответственно.
  было корректно, требуется выполнение следующих условий:

Контекстно-свободны ли естественные языки

- Чтобы предложение вида
  Маша, Петя, Вася, ... пришли к финишу первой, седьмым, четвёртым, ..., соответственно.
  было корректно, требуется выполнение следующих условий:
  - Число существительных и числительных совпадает.

Контекстно-свободны ли естественные языки

- Чтобы предложение вида
  Маша, Петя, Вася, … пришли к финишу первой, седьмым, четвёртым, …, соответственно.
  было корректно, требуется выполнение следующих условий:
  - Число существительных и числительных совпадает.
  - Род i-го слева существительного совпадает с родом i-го слева числительного.

Контекстно-свободны ли естественные языки

- Чтобы предложение вида
  Маша, Петя, Вася, ... пришли к финишу первой, седьмым, четвёртым, ..., соответственно.
  было корректно, требуется выполнение следующих условий:
  - Число существительных и числительных совпадает.
  - Род i-го слева существительного совпадает с родом i-го слева числительного.
- Построим следующее отображение ϕ:

Контекстно-свободны ли естественные языки

- Чтобы предложение вида
  Маша, Петя, Вася, … пришли к финишу первой, седьмым, четвёртым, …, соответственно.
  было корректно, требуется выполнение следующих условий:
  - Число существительных и числительных совпадает.
  - Род i-го слева существительного совпадает с родом i-го слева числительного.
- Построим следующее отображение ϕ:
  - $\phi(\text{мужское имя}) = a$, $\phi(\text{женское имя}) = b$,

Контекстно-свободны ли естественные языки

- Чтобы предложение вида
  Маша, Петя, Вася, ... пришли к финишу первой, седьмым, четвёртым, ..., соответственно.
  было корректно, требуется выполнение следующих условий:
  - Число существительных и числительных совпадает.
  - Род i-го слева существительного совпадает с родом i-го слева числительного.
- Построим следующее отображение ϕ:
  - $\phi(\text{мужское имя}) = a$, $\phi(\text{женское имя}) = b$,
  - $\phi(\text{числительное мужского рода}) = a$,
    $\phi(\text{числительное женского рода}) = b$,

Контекстно-свободны ли естественные языки

- Чтобы предложение вида
  Маша, Петя, Вася, ... пришли к финишу первой, седьмым, четвёртым, ..., соответственно.
  было корректно, требуется выполнение следующих условий:
  - Число существительных и числительных совпадает.
  - Род i-го слева существительного совпадает с родом i-го слева числительного.
- Построим следующее отображение ϕ:
  - $\phi(\text{мужское имя}) = a$, $\phi(\text{женское имя}) = b$,
  - $\phi(\text{числительное мужского рода}) = a$,
    $\phi(\text{числительное женского рода}) = b$,
  - $\phi(\textit{пришли к финишу}) = \phi(\textit{соответственно}) = \varepsilon$.
- Тогда $\phi(L \cap L(R)) = \{ww \mid w \in \{a, b\}^+\}$ — не контекстно-свободный.

Контекстно-свободны ли естественные языки

- Чтобы предложение вида
  Маша, Петя, Вася, ... пришли к финишу первой, седьмым, четвёртым, ..., соответственно.
  было корректно, требуется выполнение следующих условий:
  - Число существительных и числительных совпадает.
  - Род i-го слева существительного совпадает с родом i-го слева числительного.
- Построим следующее отображение ϕ:
  - $\phi(\text{мужское имя}) = a$, $\phi(\text{женское имя}) = b$,
  - $\phi(\text{числительное мужского рода}) = a$,
    $\phi(\text{числительное женского рода}) = b$,
  - $\phi(\textit{пришли к финишу}) = \phi(\textit{соответственно}) = \varepsilon$.
- Тогда $\phi(L \cap L(R)) = \{ww \mid w \in \{a, b\}^+\}$ — не контекстно-свободный.
- Значит, и сами языки $L \cap L(R)$ не контекстно-свободны.

КС-языки и естественные языки

- По крайней мере, русский язык не контекстно-свободный.

КС-языки и естественные языки

- По крайней мере, русский язык не контекстно-свободный.
- Аналогичные конструкции есть в голландском, немецком и т.д.

КС-языки и естественные языки

- По крайней мере, русский язык не контекстно-свободный.
- Аналогичные конструкции есть в голландском, немецком и т.д.
- Даже если множество слов оказывается контекстно-свободным, КС-грамматики неадекватно моделируют процесс порождения.

КС-языки и естественные языки

- По крайней мере, русский язык не контекстно-свободный.
- Аналогичные конструкции есть в голландском, немецком и т.д.
- Даже если множество слов оказывается контекстно-свободным, КС-грамматики неадекватно моделируют процесс порождения.
- Кроме того, возникают проблемы с семантикой.

КС-языки и естественные языки

- По крайней мере, русский язык не контекстно-свободный.
- Аналогичные конструкции есть в голландском, немецком и т.д.
- Даже если множество слов оказывается контекстно-свободным, КС-грамматики неадекватно моделируют процесс порождения.
- Кроме того, возникают проблемы с семантикой.
- Кстати, а как её вообще моделировать?