

(Еще раз) о формальных грамматиках

Марина Ермолаева

Математические методы в лингвистических исследованиях II
Формальные грамматики и синтаксис

6 сентября 2022

- Электронная почта
 - mail@mermolaeva.com
 - Отвечаю на письма в течение 24 часов
- Dropbox
 - Общая папка: [ссылка](#)
 - Материалы к курсу
- Анонимные вопросы и комментарии
 - Google Forms: [ссылка](#)

Алфавиты

- **Алфавит:** конечное множество символов; обычно обозначается Σ
- Примеры:
 - $\{a, б, в, \dots, я\}$ – алфавит
 - $\{0, 1, +, -, =\}$ – алфавит
 - $\{\text{б}\}$ – алфавит
 - $\{и, в, не, на, я\}$, пять самых частотных слов русского языка по версии Викисловаря – алфавит
 - $\mathbb{N} = \{0, 1, 2, 3, \dots\}$, множество натуральных чисел – **не** алфавит
 - Множество всех грамматичных предложений русского языка – **не** алфавит

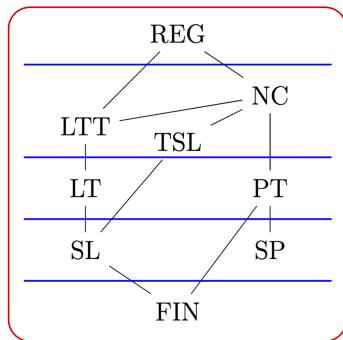
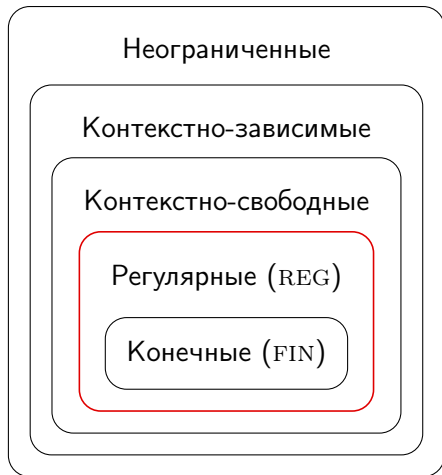
- **Строка** (или **слово**) в алфавите Σ :
конечная последовательность символов из Σ
- Пустая строка обозначается ϵ
- Σ^* : множество всех строк в алфавите Σ
Если $\Sigma = \{a, b\}$, то $\Sigma^* = \{\epsilon, a, b, aa, ab, ba, bb, aaa, \dots\}$
- Σ^+ : множество всех непустых строк в алфавите Σ
- Если x – строка и $n \in \mathbb{N}$, то x^n – слово x , повторенное n раз
- Примеры:
 - $a^3 = aaa$
 - $ab^2 = abb$
 - $(ab)^2 = abab$
 - $b^0 = \epsilon$

(Формальные) языки

(Формальные) языки

- Язык над алфавитом Σ :
(конечное или бесконечное) подмножество Σ^*
- Примеры языков над алфавитом $\Sigma = \{a, b\}$:
 - \emptyset
 - $\{\epsilon\}$
 - Σ
 - Σ^*
 - все последовательности из a и b короче 10 символов
 - все последовательности из a четной длины

Иерархия формальных языков



(J.Heinz (2018), *The computational nature of phonological generalizations*)

- **Строго локальные языки (SL)**
 - локальные зависимости
 - пример: оглушение шумных согласных на конце слова
- **Строго кусочные языки (SP)**
 - нелокальные зависимости без блокирования
 - пример: ограничение на тоны (*HLH) в луганда
- **Ярусные строго локальные языки (TSL)**
 - локальные зависимости
 - нелокальные зависимости с блокированием
 - пример: гармония гласных в турецком
- **Регулярные языки (Reg)**
 - всё вышеперечисленное и не только

- **Грамматика:** конечное множество правил, описывающих некоторый формальный язык
- Принято говорить, что грамматика **порождает** язык
- Элементы грамматики:
 - Алфавит **терминальных** символов Σ
 - Алфавит **нетерминальных** символов N такой, что $\Sigma \cap N = \emptyset$
 - Начальный символ $S \in N$
 - Множество правил вида $\alpha \rightarrow \beta$, где α и β – строки в $(\Sigma \cup N)^*$

- Классы иерархии формальных языков имеют различные ограничения на допустимые правила
- **Праволинейные грамматики** включают только следующие виды правил:
 - $A \rightarrow a$, где $A \in N$, $a \in \Sigma^*$
 - $A \rightarrow aB$, где $A, B \in N$; $a \in \Sigma^*$

Как порождаются строки

- Начать со строки, содержащей только стартовый символ S
- Применение правил вида $\alpha \rightarrow \beta$:
найти подстроку α и заменить ее на β
- Продолжать, пока в строке есть нетерминальные символы

“Дом, который построил Джек”

This is the maiden all forlorn
That milked the cow with the crumpled
horn
That tossed the dog that worried the cat
That killed the rat that ate the malt
That lay in the house that Jack built.

(английское народное стихотворение)

А это старушка, седая и строгая,
Которая доит корову безрогую,
Лягнувшую старого пса без хвоста,
Который за шиворот треплет кота,
Который пугает и ловит синицу,
Которая часто ворует пшеницу,
Которая в тёмном чулане хранится
В доме,
Который построил Джек.

(перевод С. Маршака)

Относительные предложения в английском

- $\Sigma = \{\text{this, is, the, cat, dog, frog, mouse, bit, caught, scared, that}\}$

Относительные предложения в английском

- $\Sigma = \{\text{this, is, the, cat, dog, frog, mouse, bit, caught, scared, that}\}$
- this is...

Относительные предложения в английском

- $\Sigma = \{\text{this, is, the, cat, dog, frog, mouse, bit, caught, scared, that}\}$
- this is...
... the mouse

Относительные предложения в английском

- $\Sigma = \{\text{this, is, the, cat, dog, frog, mouse, bit, caught, scared, that}\}$

- this is...

... the mouse

... the cat that bit the mouse

Относительные предложения в английском

- $\Sigma = \{\text{this, is, the, cat, dog, frog, mouse, bit, caught, scared, that}\}$

- this is...

... the mouse

... the cat that bit the mouse

... the dog that caught the cat that bit the mouse

Относительные предложения в английском

- $\Sigma = \{\text{this, is, the, cat, dog, frog, mouse, bit, caught, scared, that}\}$
- this is...
 - ... the mouse
 - ... the cat that bit the mouse
 - ... the dog that caught the cat that bit the mouse
 - ... the frog that scared the dog that caught the cat that bit the mouse

Относительные предложения в английском

- $\Sigma = \{\text{this, is, the, cat, dog, frog, mouse, bit, caught, scared, that}\}$
- this is...
 - ... the mouse
 - ... the cat that bit the mouse
 - ... the dog that caught the cat that bit the mouse
 - ... the frog that scared the dog that caught the cat that bit the mouse
 - ... the mouse that the cat bit

Относительные предложения в английском

- $\Sigma = \{\text{this, is, the, cat, dog, frog, mouse, bit, caught, scared, that}\}$
- this is...
 - ... the mouse
 - ... the cat that bit the mouse
 - ... the dog that caught the cat that bit the mouse
 - ... the frog that scared the dog that caught the cat that bit the mouse
 - ... the mouse that the cat bit
 - ... the mouse that the cat that the dog caught bit

Относительные предложения в английском

- $\Sigma = \{\text{this, is, the, cat, dog, frog, mouse, bit, caught, scared, that}\}$
- this is...
 - ... the mouse
 - ... the cat that bit the mouse
 - ... the dog that caught the cat that bit the mouse
 - ... the frog that scared the dog that caught the cat that bit the mouse
 - ... the mouse that the cat bit
 - ... the mouse that the cat that the dog caught bit
 - ... the mouse that the cat that the dog that the frog scared caught bit

Синтаксические зависимости в английском

the frog that scared the dog that caught the cat that bit the mouse

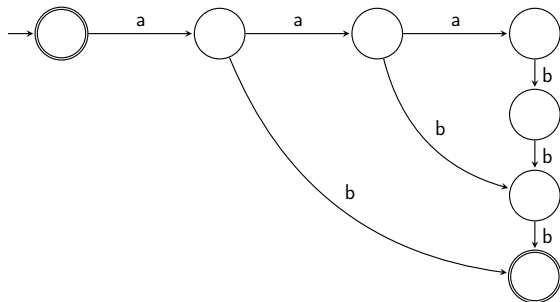
the mouse that the cat that the dog that the frog scared caught bit

– вложенные зависимости (nested dependencies)

Чего не умеют конечные автоматы

- Существует ли конечный автомат, распознающий эти языки?

$\{a^n b^n : n \leq 3\}$ – да



$\{a^n b^n : n \leq 1000000\}$ – да (но большой и непрактичный)

$\{a^n b^n\}$ – нет

В следующий раз

- Контекстно-свободные грамматики