

Математика синтаксиса:
формализация закономерностей в языковых данных

[23 июля]

Марина Ермолаева

XXIII Летняя Лингвистическая Школа
Переславль-Залесский, 2023

Формализация в лингвистике





“Поиски строгих формулировок в лингвистике вызываются гораздо более серьезными мотивами, чем просто желанием соблюсти логические тонкости или упорядочить традиционные методы лингвистического анализа. [...] Выводя неприемлемые следствия из точных, но неадекватных формулировок, мы часто можем с большой точностью установить причину этой неадекватности и, таким образом, получить более глубокое представление о лингвистических данных.”

– Н. Хомский (1957),
Синтаксические структуры

Язык сбалансированных скобок: описание

- Скобки: открывающая “(” и закрывающая “)”
- Каждый открывающий символ имеет соответствующий ему закрывающий
- Скобки правильно вложены друг в друга

Сбалансированные скобки:

```
(())  
()  
(())()  
((()))  
(()((())))
```

Несбалансированные скобки:

```
(  
)()  
(((((((  
)))  
((()((
```

Язык сбалансированных скобок: формализация

Сбалансированные скобки:

```
(( ))  
( )  
( ( ) ( ) )  
((( )))  
( ( ( ) ) )
```

Несбалансированные скобки:

```
(  
) ( (  
(((( ( ( ) )  
) ) )  
( ( ) ( ( )
```

Язык сбалансированных скобок: формализация

Сбалансированные скобки:

```
(())  
()()  
(()())()  
((()))  
()((())())
```

Несбалансированные скобки:

```
(  
)()(  
(((((((  
)))  
()()()
```

- L – множество всех строк из сбалансированных скобок

Язык сбалансированных скобок: формализация

Сбалансированные скобки:

```
(())  
()()  
(()())()  
((()))  
()((())())
```

Несбалансированные скобки:

```
(  
)()  
(((((((  
)))  
()())()
```

- L – множество всех строк из сбалансированных скобок
- $\epsilon \in L$ – пустая строка принадлежит L

Язык сбалансированных скобок: формализация

Сбалансированные скобки:

```
(())  
()()  
(())()()  
(((())))  
(()((())()))
```

Несбалансированные скобки:

```
(  
)()(  
(((((((())  
)))  
(()()())
```

- L – множество всех строк из сбалансированных скобок
- $\epsilon \in L$ – пустая строка принадлежит L
- Для любой строки s : если $s \in L$, то $(s) \in L$

Язык сбалансированных скобок: формализация

Сбалансированные скобки:

```
(())  
()()  
(())()()  
((()))  
(()((())))
```

Несбалансированные скобки:

```
(  
)()(  
(((((((  
)))  
(()()()
```

- L – множество всех строк из сбалансированных скобок
- $\epsilon \in L$ – пустая строка принадлежит L
- Для любой строки s : если $s \in L$, то $(s) \in L$
- Для любых строк s и t : если $s \in L$ и $t \in L$, то $st \in L$

Языки как математические объекты

- **Алфавит:** конечное множество символов

- **Алфавит:** конечное множество символов
- **Строка:** конечная последовательность символов некоторого алфавита
 - ϵ – пустая строка (последовательность из 0 символов)

- **Алфавит:** конечное множество символов
- **Строка:** конечная последовательность символов некоторого алфавита
 - ϵ – пустая строка (последовательность из 0 символов)
- **Язык:** (конечное или бесконечное) множество строк в некотором алфавите

- **Алфавит:** конечное множество символов
- **Строка:** конечная последовательность символов некоторого алфавита
 - ϵ – пустая строка (последовательность из 0 символов)
- **Язык:** (конечное или бесконечное) множество строк в некотором алфавите
- **Грамматика:** система правил, порождающая некоторый язык

- **Элементарные синтаксические выражения**

Лексические единицы

- Элементарные **синтаксические выражения**
- Информация о лексической единице:
 - Фонологическое (или орфографическое) содержимое
 - Синтаксические **свойства и требования**

Лексические единицы

- Элементарные **синтаксические выражения**
- Информация о лексической единице:
 - Фонологическое (или орфографическое) содержимое
 - Синтаксические **свойства и требования**
- Примеры:

Лексические единицы

- Элементарные **синтаксические выражения**
- Информация о лексической единице:
 - Фонологическое (или орфографическое) содержимое
 - Синтаксические **свойства и требования**
- Примеры:
 - *girl*: **N** (существительное)

Лексические единицы

- Элементарные **синтаксические выражения**
- Информация о лексической единице:
 - Фонологическое (или орфографическое) содержимое
 - Синтаксические **свойства и требования**
- Примеры:
 - *girl*: **N** (существительное)
 - *the*: требует **N** справа, чтобы построить **NP** (именную группу)

Лексические единицы

- Элементарные **синтаксические выражения**
- Информация о лексической единице:
 - Фонологическое (или орфографическое) содержимое
 - Синтаксические **свойства и требования**
- Примеры:
 - *girl*: **N** (существительное)
 - *the*: требует **N** справа, чтобы построить **NP** (именную группу)
 - *John*: **NP**

Лексические единицы

- Элементарные **синтаксические выражения**
- Информация о лексической единице:
 - Фонологическое (или орфографическое) содержимое
 - Синтаксические **свойства и требования**
- Примеры:
 - *girl*: **N** (существительное)
 - *the*: требует **N** справа, чтобы построить **NP** (именную группу)
 - *John*: **NP**
 - *laughed*: требует **NP** слева, чтобы построить **S** (предложение)

Лексические единицы

- Элементарные **синтаксические выражения**
- Информация о лексической единице:
 - Фонологическое (или орфографическое) содержимое
 - Синтаксические **свойства и требования**
- Примеры:
 - *girl*: **N** (существительное)
 - *the*: требует **N** справа, чтобы построить **NP** (именную группу)
 - *John*: **NP**
 - *laughed*: требует **NP** слева, чтобы построить **S** (предложение)
 - *praised*: требует **NP** слева и справа, чтобы построить **S**

Синтаксические признаки

- Примеры:
 - *girl*: **N** (существительное)
 - *the*: требует **N** справа, чтобы построить **NP** (именную группу)
 - *John*: **NP**
 - *laughed*: требует **NP** слева, чтобы построить **S** (предложение)
 - *praised*: требует **NP** слева и справа, чтобы построить **S**
- Эту информацию можно записать в виде **синтаксических признаков**:
 - f “имеет свойство f ” – синтаксическая категория
 - $=f$ “требует выражение со свойством f справа от себя”
 - $f=$ “требует выражение со свойством f слева от себя”

Операция соединения (Merge)

- Merge:
 - видит первый признак **вершины** (главного слова) каждого выражения
 - соединяет выражения, несущие свойство и требование с одинаковым именем (например, =f и f)

Операция соединения (Merge)

- Merge:
 - видит первый признак **вершины** (главного слова) каждого выражения
 - соединяет выражения, несущие свойство и требование с одинаковым именем (например, =f и f)

$John :: np \quad + \quad laughed :: np = s \quad =$

$\begin{array}{c} \triangleright \\ \swarrow \quad \searrow \\ John \quad laughed \\ \cancel{np} \quad \cancel{np} \quad s \end{array}$

Операция соединения (Merge)

- Merge:
 - видит первый признак **вершины** (главного слова) каждого выражения
 - соединяет выражения, несущие свойство и требование с одинаковым именем (например, =f и f)

$John :: np + laughed :: np = s =$

```
graph TD; Root[">"] --- John; Root --- laughed; John --- np1["np"]; laughed --- np2["np"]; np2 --- s["s"];
```

- Использованные признаки **удаляются**
- Лексическая единица, у которой был признак вида =f или f=, становится вершиной нового выражения

Операция соединения (Merge)

- Merge:
 - видит первый признак **вершины** (главного слова) каждого выражения
 - соединяет выражения, несущие свойство и требование с одинаковым именем (например, =f и f)

$$John :: np \quad + \quad laughed :: np = s \quad =$$

```
graph TD
    Root(">") --- John("John")
    Root --- Laughed("laughed")
    John --- NP1("np")
    Laughed --- NP2("np")
    Laughed --- S("s")
```

- Использованные признаки **удаляются**
- Лексическая единица, у которой был признак вида =f или f=, становится вершиной нового выражения
- **Предложение:** выражение без признаков, кроме s у вершины

Merge: простые предложения

laughed

John

the girl

John

praised

the girl

Merge: простые предложения

laughs
laughed

John
Mary
the boy
the girl

jumps
jumped

praises
praised

John
Mary
the boy
the girl

Merge: простые предложения

John	laughs
Mary	laughed
the boy	will laugh
the girl	jumps
	jumped
	will jump

praises
praised
will praise

John
Mary
the boy
the girl
