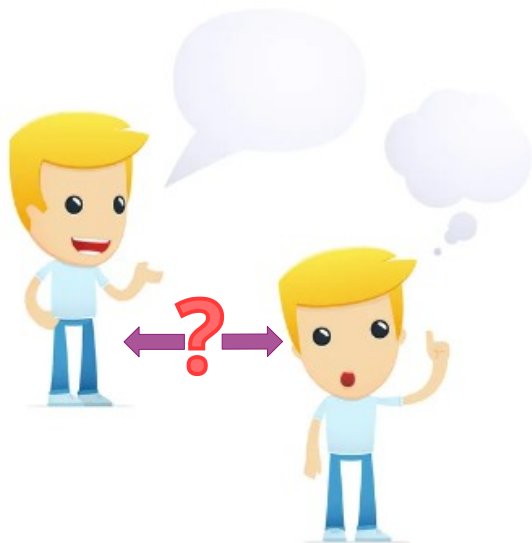


ІІТМО

Математическая ЛИНГВИСТИКА

Болгова Екатерина Владимировна, к.т.н.
преподаватель ФИКТ



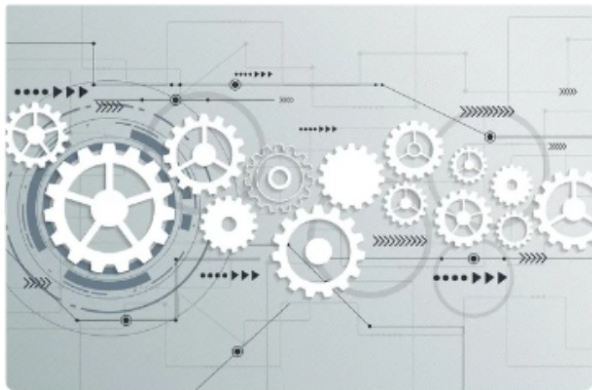


- **Аристотель:** «форма» ↔ «значение»
- слово ↔ понятие, предложение ↔ суждение, сложное предложение ↔ умозаключение
- Предложение (как языковая единица) ↔ суждение (как форма мышления)

Язык — это система знаков, служащая для хранения, переработки и передачи информации



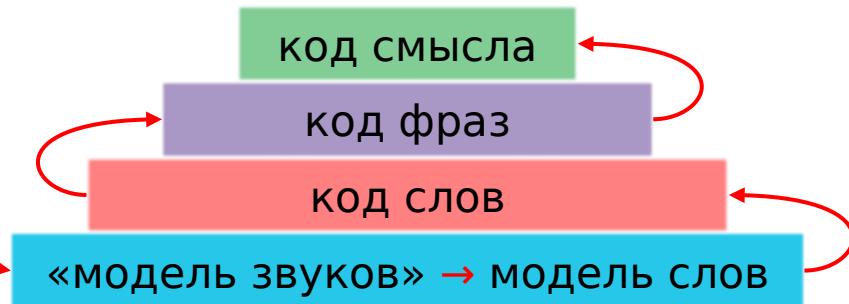
Язык - это условие для осуществления мышления и средство, позволяющее хранить мысли, уже сформировавшиеся в процессе мышления, и передавать их



Кибернетическая теория о соотношении языка и мышления:



- Система: совокупность некоторого количества одинаковых и различных элементов
- Связь: воздействие элементов системы друг на друга
- Информация: сведения, являющиеся объектом хранения, передачи, преобразования.



Язык (langue)

- Предложение — единица языка
- Социальное явление

Речь (parole)

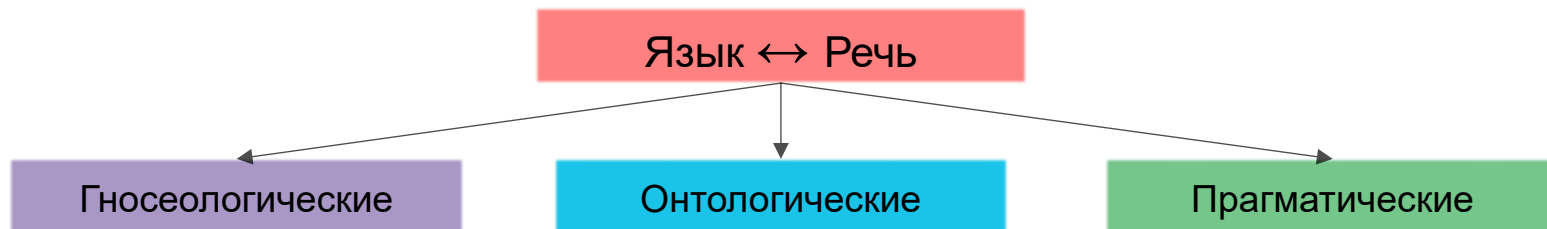
- Высказывание — случай речи
- Индивидуальное явление



Фердинанд де Соссюр — швейцарский лингвист

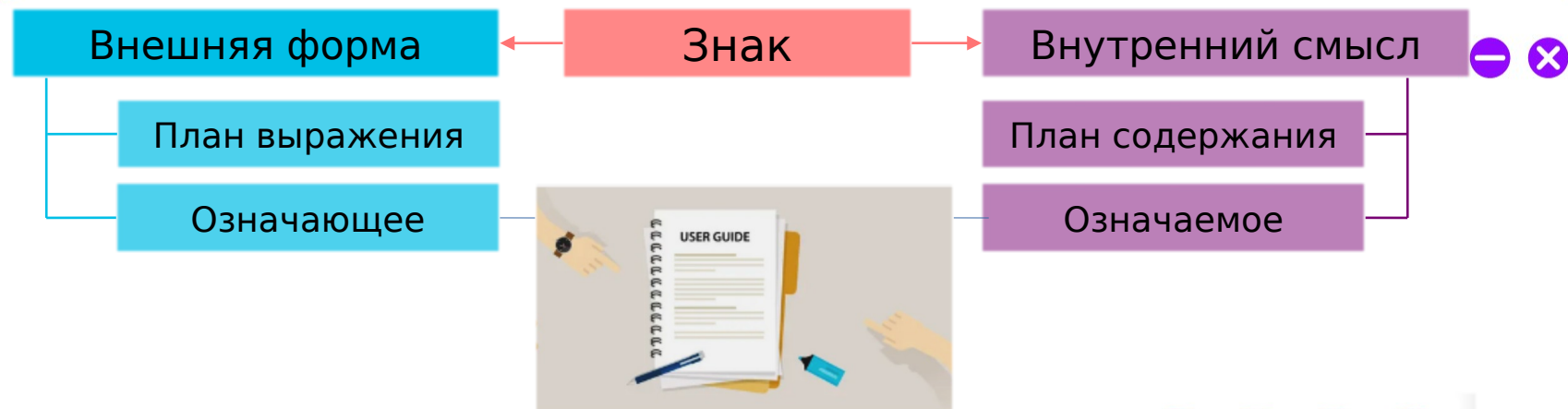
К речи Соссюр относит три ряда явлений:

- 1) звуковое пополнение, т.е. реализацию означающего при помощи голоса;
- 2) способность ассоциации звукового образа с идеей в зависимости от определенной ситуации;
- 3) способность сочетания знаков, т.е. конструирование фраз.

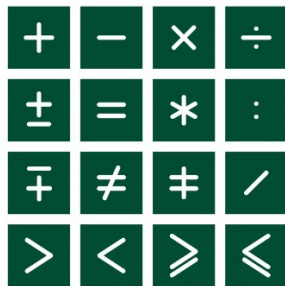


Язык как знаковая система

ИТМО

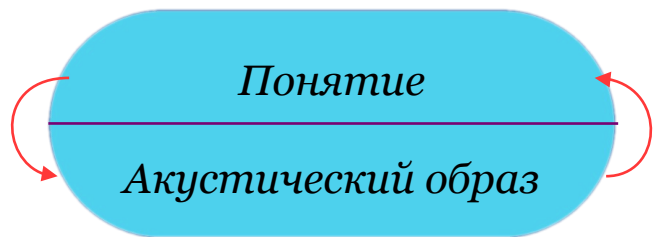


Полковник:
(последняя форма одежды)



Язык как знаковая система

ІІТМО



Все единицы функционируют следующим образом:

- **в языке** — в своем общем виде (инвариант),
- **в речи** — в виде конкретных вариантов

Единицы языка и единицы речи

- звук
- морфема
- слово
- словосочетание
- предложение

Уровни организации языка

- фонетический
- морфемный
- лексический
- морфологический
- синтаксический

Как описать язык?

ИТМО

«Формальный»

=

Логически
последовательный

+

Однозначный

+

Абсолютно явный



Уровни
формализации

доматематический

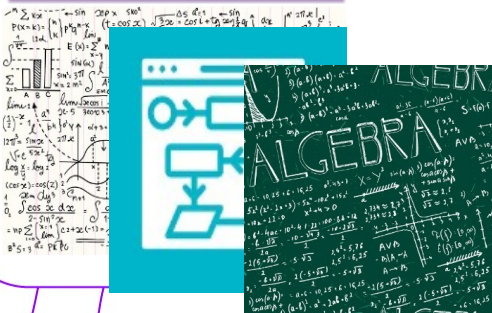
математический

Язык как функция (отображение) ИТМО



математик Вася

Функция $f(x)$



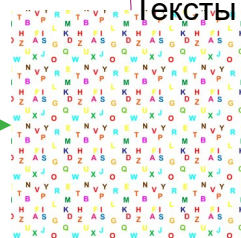
План содержания



Смыслы



Тексты

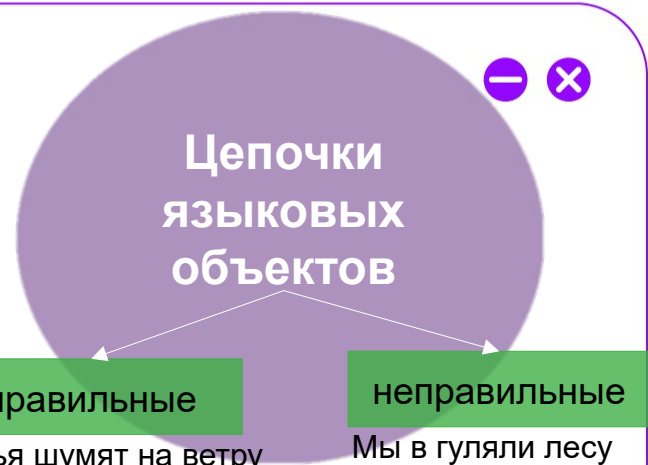
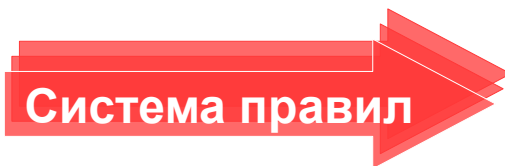


План выражения



Формальные грамматики

ІТМО



правильные

Листья шумят на ветру
Мы гуляли в лесу

неправильные

Мы в гуляли лесу
На шумят ветру листья

Формальная грамматика задает правильные цепочки, если:

- 1) либо для любой предъявленной цепочки грамматика умеет решить, является эта цепочка правильной или нет, и в случае положительного ответа дать указания о строении этой цепочки
- 2) либо грамматика умеет построить любую правильную цепочку, давая при этом указания о ее строении, и не строит ни одной неправильной цепочки

распознающие

порождающие

ФГ и способы ее описания

ІТМО

Пусть:

- **X** — некоторый *алфавит*
- **Слово** в алфавите X — последовательность символов алфавита
- **Лексика** языка — множество допустимых слов
- **Предложения** — наборы слов, построенные по определенным правилам
- **Синтаксис** языка — описание правильных предложений
- **Язык** — множество предложений
- **Грамматика** языка — набор правил синтаксиса



ФГ и способы ее описания

ІТМО

Грамматика G_0 задана следующим набором правил:

<предложение> → <подлежащее> <сказуемое>

<подлежащее> → <имя существительное>

<подлежащее> → <местоимение>

<имя существительное> → КОТ

<имя существительное> → ПЕС

<местоимение> → ОН

<сказуемое> → <глагольная форма>

<глагольная форма> → ИДЕТ

<глагольная форма> → ЛЕЖИТ

<Пр> → <П> <С>

<П> → <ИС>

<П> → <М>

<ИС> → КОТ

<ИС> → ПЕС

<М> → ОН

<С> → <ГФ>

<ГФ> → ИДЕТ

<ГФ> → ЛЕЖИТ



ФГ и способы ее описания

- **R** — множество правил грамматики
 - **V** — словарь — символы, входящие в правила
 - **V^*** — Множество всех конечных последовательностей символов (строк) из словаря V , включая пустую строку
- $V = \{<Пр>, <П>, <С>, <ИС>, <М>, <ГФ>, КОТ, ПЕС, ОН, ИДЕТ, ЛЕЖИТ\}$
- V делят на 2 подмножества:
 - **T** — символы, которые входят только в правые части правил R ($T = \{КОТ, ПЕС, ОН, ИДЕТ, ЛЕЖИТ\}$) — *Терминальные символы* или *терминалы*
 - **$N = V \setminus T$** ($N = \{<Пр>, <П>, <С>, <ИС>, <М>, <ГФ>\}$) — *Нетерминальные символы* или *переменные*
 - **S** — *начальный символ* грамматики или *аксиома* $S = \{<Пр>\}$

$G = \{V, T, R, S\}$, где элементы множества R : $x \rightarrow y, (x \neq y, x \in (V \setminus T)^*, y \in V^*)$

Порождающая грамматика

- **Грамматика в виде ориентированного графа:**
- Символ левой части правила изображают выше символа правой части и соединяют их стрелкой
- Если в правую часть входят несколько символов, то их объединяют узлом, отмеченным знаком «+»
- Узлы, помеченные одинаковыми терминалами, объединяются в один узел.
- Для изображения правил с одинаковыми левыми частями используют узел, отмеченный знаком «v»



ФГ и способы ее описания

ІТМО

$\langle \text{Пр} \rangle \rightarrow \langle \text{П} \rangle \langle \text{С} \rangle$

$\langle \text{П} \rangle \rightarrow \langle \text{ИС} \rangle$

$\langle \text{П} \rangle \rightarrow \langle \text{М} \rangle$

$\langle \text{ИС} \rangle \rightarrow \text{КОТ}$

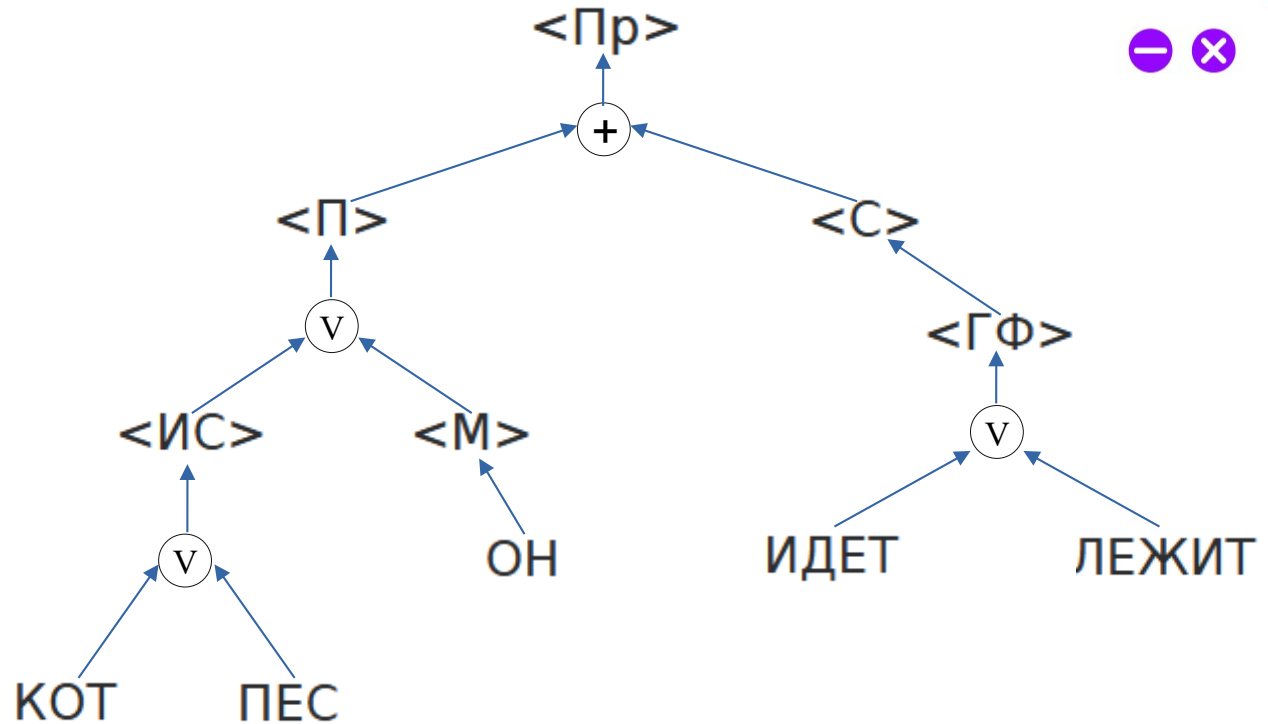
$\langle \text{ИС} \rangle \rightarrow \text{ПЕС}$

$\langle \text{М} \rangle \rightarrow \text{ОН}$

$\langle \text{С} \rangle \rightarrow \langle \text{ГФ} \rangle$

$\langle \text{ГФ} \rangle \rightarrow \text{ИДЕТ}$

$\langle \text{ГФ} \rangle \rightarrow \text{ЛЕЖИТ}$



ФГ и способы ее описания

Порождающая грамматика для арифметических выражений, использующих операции сложения и умножения, скобки и идентификаторы:

$\langle \text{Выражение} \rangle \rightarrow \langle \text{Терм} \rangle$

$\langle \text{Выражение} \rangle \rightarrow \langle \text{Выражение} \rangle + \langle \text{Терм} \rangle$

$\langle \text{Терм} \rangle \rightarrow \langle \text{Множитель} \rangle$

$\langle \text{Терм} \rangle \rightarrow \langle \text{Терм} \rangle \times \langle \text{Множитель} \rangle$

$\langle \text{Множитель} \rangle \rightarrow \langle \text{Идентификатор} \rangle$

$\langle \text{Множитель} \rangle \rightarrow (\langle \text{Выражение} \rangle)$



ФГ: определения

Пустая цепочка – цепочка, не содержащая ни одного символа.



- **Конкатенацией** непустых цепочек v и w в мн-ве V называется цепочка vw
- Если для некоторых цепочек v, w, x, y в V выполняется равенство $v = xwy$, то такое представление цепочки v будем называть **вхождением цепочки w в цепочку v**
- Если существует вхождение цепочки w в цепочку v , то w называют **подцепочкой** цепочки v .
- Количество повторений некоторой цепочки n раз подряд записывают в виде степени, например, вместо $aaaaa$ пишется a^5 , вместо $avcavcavc$ пишется $(avc)^3$

ФГ: определения

Пусть для некоторой грамматики $G = \langle V, T, S, R \rangle$ цепочки w и v имеют вид:

$$w = xzy \quad v = xqy,$$

а в множестве правил R есть правило $z \rightarrow q$

В этом случае говорят, что цепочка v **непосредственно выводима** из цепочки w в грамматике G (путём применения к выделенному вхождению z правила $z \rightarrow q$)



Последовательность цепочек $D = (\omega_0, \omega_1, \dots, \omega_n)$, где $n \geq 1$, называют **выводом** ω_n из ω_0 в грамматике G , если для каждого i ($1 \leq i \leq n$) цепочка ω_i непосредственно выводима из ω_{i-1} в G .

**Спасибо
за внимание!**

it^{'s}**MO** *re than a*
UNIVERSITY

E-mail: ekaterina_bolgova@itmo.ru
telegram: @Katerina_Bolgova

Практическая работа №2

ІІТМО

1. Выберите 4 поэтов:

- Творчество каждых 2 из них должно принадлежать близким временным отрезкам
- Творчество пар — далеким временным отрезкам

2. У каждого поэта отберите не менее 9 стихотворений (примерно одной длины)



Практическая работа №2

ІІТМО

3. Пользуясь количественными методами оцените:

- Среднее количество строк в стихотворении
 - Среднее количество строф
 - Среднее количество строк в строфе
 - Среднее количество каждой из частей речи
 - Среднее количество слов в строке
 - Среднее количество слов в предложении
- 4. Для каждого стихотворения определить размер и такие тропы, как эпитеты, метафоры, олицетворения и сравнения. Проанализировать их.



Практическая работа №2

ІІТМО

5. Пользуясь получившейся статистикой проанализировать и выявить особенности творчества поэтов внутри группы и между группами



6. Построить графики, сделать отчет