

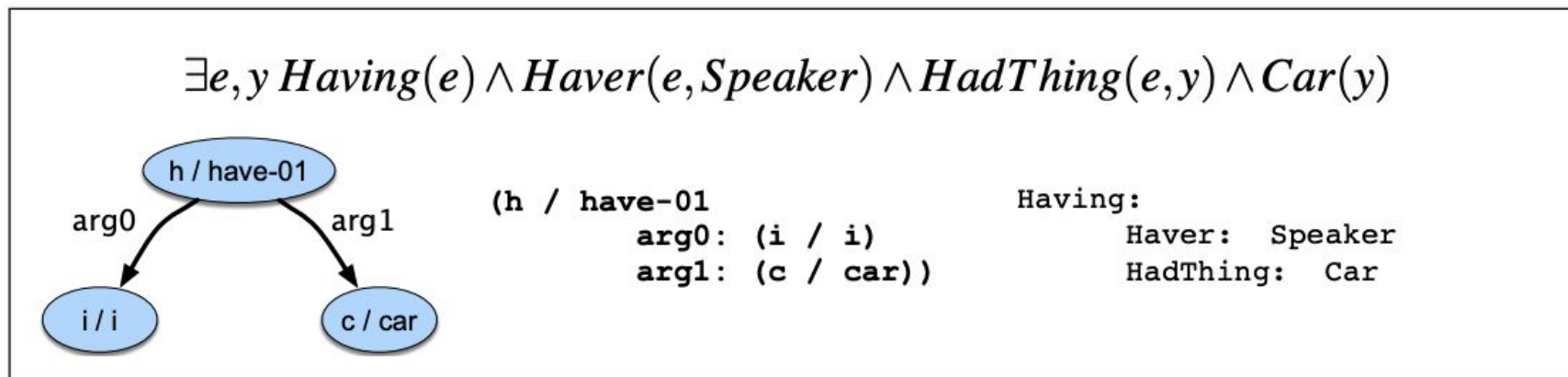
# Natural Language Processing

Лекция 6

Семантический анализ

**Семантический парсинг** (семантический анализ) - процесс нахождения формального представления смысла текста.

**Computational semantics** is the study of how to automate the process of constructing and reasoning with meaning representations of natural language expressions.



**Figure 16.1** A list of symbols, two directed graphs, and a record structure: a sampler of meaning representations for *I have a car*.

Does Maharani serve vegetarian food?

*Serves (Maharani, VegetarianFood)*

**Verifiability** (возможность осуществления проверки) - это способность системы сравнивать положение дел во входном тексте с положением дел в реальном мире, описанном в базе знаний.

# Однозначное представление

I wanna eat someplace that's close to NSU.

**Неопределённость** - понятие, при котором значение высказывания не до конца определено.

Пример: I want to eat Italian food.

# Каноническая форма

**Каноническая форма:** различные тексты, которые имеют одно и то же значение, должны иметь одинаковое представление значения.

1. Does Maharani have vegetarian dishes?
2. Do they have vegetarian food at Maharani?
3. Are vegetarian dishes served at Maharani?
4. Does Maharani serve vegetarian fare?

1. Maharani serves vegetarian dishes.
2. Vegetarian dishes are served by Maharani.

**Инференс**, вывод, заключение - это логический вывод, основанный на представлении значения входного текста и знаний о мире.

I'd like to find a restaurant where I can get vegetarian food.

*Serves(x, VegetarianFood)*

# Модельно-теоретическая семантика

**Модель** - формальная структура, в которой содержатся знания о мире.

Выражения, которые являются представлением значения, могут быть отображены на элементы модели: объекты, свойства объектов, отношения между объектами.

Словарь представления значения состоит из двух частей:

- **нелогический словарь** состоит из открытого множества названий объектов, свойств и отношений (предикаты, узлы в графах и т.д.);
- **логический словарь** состоит из закрытого множества символов, операторов и т.д.

Каждый элемент нелогического словаря должен иметь денотат в модели, т.е. каждый элемент должен соответствовать фиксированной части модели.

**Домен модели** - это множество объектов, которые описывает модель.

- **Объекты** соотносятся с **элементами** домена;
- **Свойства** соотносятся с **множеством элементов** домена;
- **Отношения** соотносятся с **множеством кортежей элементов** домена.

**Интерпретация** - функция, которая соотносит элементы словаря с нужным денотатом модели.



**Domain**

Matthew, Franco, Katie and Caroline

Frasca, Med, Rio

Italian, Mexican, Eclectic

$$\mathcal{D} = \{a, b, c, d, e, f, g, h, i, j\}$$

$$a, b, c, d$$

$$e, f, g$$

$$h, i, j$$

**Properties**

*Noisy*

Frasca, Med, and Rio are noisy

$$\text{Noisy} = \{e, f, g\}$$

**Relations**

*Likes*

Matthew likes the Med

Katie likes the Med and Rio

Franco likes Frasca

Caroline likes the Med and Rio

$$\text{Likes} = \{\langle a, f \rangle, \langle c, f \rangle, \langle c, g \rangle, \langle b, e \rangle, \langle d, f \rangle, \langle d, g \rangle\}$$

*Serves*

Med serves eclectic

Rio serves Mexican

Frasca serves Italian

$$\text{Serves} = \{\langle f, j \rangle, \langle g, i \rangle, \langle e, h \rangle\}$$

**Figure 16.2** A model of the restaurant world.

Matthew likes Frasca.

**a** likes **e**

*Likes* contain **<a, e>** ?

Katie likes the Rio and Matthew likes the Med.

Katie and Caroline like the same restaurants.

Franco likes noisy, expensive restaurants.

Not everybody likes Frasca.

# Логика первого порядка

<i>Formula</i>	→	<i>AtomicFormula</i>
		<i>Formula</i> <i>Connective</i> <i>Formula</i>
		<i>Quantifier</i> <i>Variable</i> , ... <i>Formula</i>
		$\neg$ <i>Formula</i>
		( <i>Formula</i> )
<i>AtomicFormula</i>	→	<i>Predicate</i> ( <i>Term</i> , ...)
<i>Term</i>	→	<i>Function</i> ( <i>Term</i> , ...)
		<i>Constant</i>
		<i>Variable</i>
<i>Connective</i>	→	$\wedge$   $\vee$   $\implies$
<i>Quantifier</i>	→	$\forall$   $\exists$
<i>Constant</i>	→	<i>A</i>   <i>VegetarianFood</i>   <i>Maharani</i> ...
<i>Variable</i>	→	<i>x</i>   <i>y</i>   ...
<i>Predicate</i>	→	<i>Serves</i>   <i>Near</i>   ...
<i>Function</i>	→	<i>LocationOf</i>   <i>CuisineOf</i>   ...

**Figure 16.3** A context-free grammar specification of the syntax of First-Order Logic representations. Adapted from Russell and Norvig (2002).

**Терм** - обозначение объекта.

**Константа** - обозначение определённого объекта в описываемом мире.

**Переменная** - указание на конкретный объект.

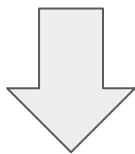
Maharani serves vegetarian food ->

Serves(Maharani, VegetarianFood)

Maharani is a restaurant -> Restaurant(Maharani)

## Логические операции (коннекторы):

I only have five dollars and I don't have a lot of time.



$\text{Have}(\text{Speaker}, \text{FiveDollars}) \wedge \neg \text{Have}(\text{Speaker}, \text{LotOfTime})$

$\forall$  - квантор всеобщности (“для всех”, “for all”) ;

$\exists$  - квантор существования (“существует”, “there exists”)

a restaurant that serves Mexican food near ICSI.:

$\exists x: \text{Restaurant}(x) \wedge \text{Serves}(x, \text{MexicanFood}) \wedge$   
 $\text{Near}(\text{LocationOf}(x), \text{LocationOf}(\text{ICSI}))$



# Вывод: modus ponens

**Modus ponens:** если левая сторона импликации является истиной, то правая сторона выводима.

`VegetarianRestaurant (Leaf)`

`$\forall x$ VegetarianRestaurant  $\Rightarrow$  Serves (x, VegetarianFood)`

---

`Serves (Leaf, VegetarianFood)`

# Теория “Смысл $\Leftrightarrow$ Текст”

И. Мельчук: нужно построить формальную модель, наиболее полно охватывающую все множество грамматических явлений конкретного языка. При этом считается, что описание процесса интерпретации текста (определение смысла) может быть получено на основе описания процесса построения текста.

# Лексическая функция

**Лексическая функция** – функция, аргументами которой являются некоторые слова или словосочетания данного языка, а значениями – множества слов и словосочетаний этого же языка.

Значения одной лексической функции от разных аргументов могут полностью или частично совпадать; могут совпадать и значения разных функций от одного аргумента:

$$\exists f_1 \neq f_2 \exists x: f_1(x) = f_2(x);$$

$$\exists f \exists x_1 \neq x_2: f(x_1) = f(x_2)$$

# Лексическая функция

Примеры:

- Значение предиката  $Syn(x, y)$  истинно, если  $x, y$  - синонимы; иначе - ложно;
- Значение предиката  $Anti(x, y)$  истинно, если  $x, y$  - антонимы; иначе - ложно;
- $Gener(x, y)$ :  $y$  – обобщающее понятие по отношению к понятию, обозначенному  $x$  ( $x$ =клубника,  $y$ =ягода).

# Ситуация

**Ситуация** – определенное лексическое отражение (в данном языке) некоторой части действительности.

Семантические актанты: A, B, C, D.

Глубинно синтаксические актанты: 1, 2, 3, 4.

$Loc(x, y)$ ,  $y$  is a preposition of localization (spatial, temporal or abstract).

$Loc_{in}(x, y)$ ,  $y$  is a “statistic” localization (*in Moscow*).

$Loc_{ad}(x, y)$ ,  $y$  is a preposition of direction (*to Moscow*).

$Loc_{ab}(x, y)$ ,  $y$  is a preposition of distance (*from Moscow*).

$Func_1(x, y)$ ,  $y$  is a verb, linking a situation as a subject with actants as objects (*Sun is shining through the window.*).

$Labor_{12}(x, y)$ : “*The president awarded him the order*”.

$Perf(x, y)$ ,  $y$  is a completed action ( $x = \text{читать}$ ,  $y = \text{прочитать}$ ).

$Destr(x, y)$ ,  $y$  is an “aggressive” action ( $x = \text{bee}$ ,  $y = \text{bite}$ ).

$Cap(x, y)$ ,  $y$  is a “chef” ( $x = \text{department}$ ,  $y = \text{dean}$ ).

# Валентность

**Валентность слова** - это способность слова вступать в связи с другими словами.

Валентностями обладают слова, которые задают ситуацию.

Это все глаголы, некоторые существительные (отглагольные), прилагательные (обозначающие сравнение: больше, меньше, выше, ниже), некоторые предлоги и наречия.

# Валентность

**Семантическая валентность** определяется лексическим анализом ситуации, выраженной определённым словом.

*Арендовать:  $P(x_1, x_2, x_3, x_4, x_5)$ , где  $x_1$  - “кто”,  $x_2$  - “что”,  $x_3$  - “у кого”,  $x_4$  - “цена”,  $x_5$  - “период”.*

**Синтаксическая валентность** определяется количеством синтаксических актантов, которые присутствуют в тексте, и зависит от контекста.



# Толково-комбинаторный словарь

Языковая модель, предложенная И. А. Мельчуком, представляет язык как совокупность словарных статей с огромным количеством разнообразной информации; грамматические правила при таком словаре играют скорее второстепенную роль. Толково-комбинаторный словарь отражает прежде всего нетривиальную сочетаемость лексем.

Статья толково-комбинаторного словаря содержит:

- основное слово;
- лексические предикаты (характеризующие сочетаемость слов);
- валентность и для каждого актанта указание, в каких падежах и с какими предлогами используются слова, соответствующие данному актанту

Thank you for your attention!