

Лекция 9: Обработка на естествен език

общуване: процес, при който един агент (**говорещ**) се стреми да направи така, че друг агент (**слушач**) да промени своето знание и(ли) цели.

Етапи на общуването за говорещия:

- намерение: решение относно информацията, която ще бъде включена в **речевия акт**. В общия случай предполага планиране (доколкото общуването е вид поведение) и разсъждения за знанието и целите на агентите;
- пораждање: превръщане на избраната информация в съобщение (низ от думи);
- синтез: извеждане на съобщението във вид, определен от средата, в която се извършва общуването (звучаща реч, видим образ (текст) и т. н.).

Етапи на общуването за слушащия:

- възприятие: превръщане на съобщението в низ от думи (разпознаване на символи, на реч и т. н.);
- анализ: установяване на информационното съдържание на съобщението на ниво
 - синтаксис: определяне на структурата на съобщението, т. е. на връзките между отделните думи; това ниво на анализа обикновено се представя като групиране на думите в йерархични единици, наречени **фрази**, откъдето говорим за **фразова структура**,
 - семантика: установяване на буквалното (речниково) значение на отделните думи и конструкции (фрази),
 - прагматика: поставяне на значението в контекста на общуването;
- усвояване: евентуално обновяване на базата знание или целите в зависимост от отношенията между агентите.

1 Многозначност

Характерно свойство на естественоезиковото общуване. Отстраняването на многозначността е един от основните проблеми на обработката на естествен език.

Нива на многозначността:

- фонетично (каламбури);

• синтактично (

я.м.	макарони	свс	сирене
------	----------	-----	--------

 , но

я.м.	макарони	с	вилница
------	----------	---	---------

 — предложната фраза *свс сирене* определя *макарони*, докато предложната фраза *с вилница*

определя *я.м.*); срв. също

Time	flies	like	an	arrow
------	-------	------	----	-------

 ‘Времето лети като

стрела’, но

Fruit	flies	like	a	banana
-------	-------	------	---	--------

 ‘Плодовете мухи обичат банан’);

- семантично (омоними);
- прагматично (*Знаете ли колко е часът?*—буквално въпрос относно знанието, с което разполага слушащият, но в контекст обикновено еквивалентен на ‘Колко е часът?’ или ‘Смятам, че вече е много късно’).

2 Синтактичен анализ

Структурата на съобщението следва правилата на синтаксиса на езика—неговата **граматика**, която определя множеството на допустимите изрази.

Фразовоструктурните граматика (**безконтекстните граматика** в теорията на формалните езици) свързват фразите (нетерминалните символи) с последователности от фрази и отделни думи (терминални символи):

$S \rightarrow NP VP$	$NP \rightarrow D N$	$D \rightarrow an$	$D \rightarrow a$
$VP \rightarrow V PP$	$NP \rightarrow N N$	$N \rightarrow time$	$N \rightarrow fruit$
$VP \rightarrow V NP$	$NP \rightarrow N$	$N \rightarrow arrow$	$N \rightarrow flies$
$PP \rightarrow P NP$		$V \rightarrow flies$	$N \rightarrow banana$
		$P \rightarrow like$	$V \rightarrow like$

Фразовата структура обикновено се изобразява графично посредством **фразовоструктурно дърво**, в чийто корен е символът на фразовата категория (напр. S 'изречение'), а в листата—думите.

Алгоритми за синтактичен анализ:

- **отгоре надолу (top-down)**—търсене в дълбочина на входния низ от думи в пространството на възможните развития на основната фразова категория;
- **отдолу нагоре (bottom-up)**—търсене на все по-големи фрази, отговарящи на части от входния низ от думи, докато се стигне до основната фразова категория.

3 Семантичен анализ

Значението на съобщението обикновено се представя чрез изрази в някаква формална система (напр. чрез езика на предикатната логика). Това означава, че за семантичния анализ е необходимо да се знае

- какво значат отделните думи (какви логически изрази – формули или части от тях – им съответствуват);
- как се свързват значенията на компонентите на синтактичните конструкции.