Изкуствен интелект—Зимен семестър 2001/2002. Преподавател: Иван А. Держански.

Лекция 9: Обработка на естествен език

общуване: процес, при който един агент (**говорещ**) се стреми да направи така, че друг агент (**слушащ**) 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;
- пораждане: 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 ($\boxed{\boxed{\textit{ям}}\ \boxed{\boxed{\textit{макарони}}\ \boxed{\boxed{\textit{със}}\ \boxed{\textit{сирене}}}}}$, но $\boxed{\boxed{\textit{ям}}\ \boxed{\textit{макарони}}\ \boxed{\boxed{\textit{с}}\ \boxed{\textit{вилица}}}}$ — предложната фраза *със сирене* определя *макарони*, докато предложната фраза *с вилица* определя *ям*); срв. също $\boxed{\boxed{\textit{Time}}\quad \boxed{\boxed{\textit{flies}}\ \boxed{\boxed{\textit{like}}\ \boxed{\boxed{\textit{an}}\ \boxed{\textit{arrow}}}}}}$ ‘Времето лети като стрела’, но $\boxed{\boxed{\boxed{\textit{Fruit}}\ \boxed{\textit{flies}}}\quad \boxed{\boxed{\textit{like}}\ \boxed{\boxed{\textit{a}}\ \boxed{\textit{banana}}}}}$ ‘Плодовите 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 граматики (**безконтекстните граматики** в теорията на формалните езици) свързват фразите (нетерминалните символи) с последователности от фрази и отделни думи (терминални символи):

		$\mathsf{D} \to an$	$\mathsf{D} \to a$
$\mathsf{S} \to \mathsf{NP}\ \mathsf{VP}$	$\mathsf{NP} \to \mathsf{D}\ \mathsf{N}$	$\mathsf{N} \to time$	$\mathsf{N} \to fruit$
$\mathsf{VP} \to \mathsf{V}\ \mathsf{PP}$	$\mathsf{NP} \to \mathsf{N}\ \mathsf{N}$	$\mathsf{N} \to arrow$	$\mathsf{N} \to flies$
$\mathsf{VP} \to \mathsf{V}\ \mathsf{NP}$	$\mathsf{NP} \to \mathsf{N}$	$\mathsf{V} \to flies$	$\mathsf{N} \to banana$
$\mathsf{PP} \to \mathsf{P}\ \mathsf{NP}$		$\mathsf{P} \to like$	$\mathsf{V} \to like$

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 съответстват);
- как се свързват значенията на компонентите на синтактичните конструкции.