

Интеллектуальные информационные системы

Тема 3

Синтаксический анализ

Павел Исаакович Браславский
pb@imach.uran.ru
весенний семестр 2006

План

- Что такое синтаксис?
- Модели представления синтаксических зависимостей
- Подходы к реализации СА
- Примеры



Синтаксис и СА

- Неформальное определение:
Синтаксис – правила объединения слов в корректные выражения – словосочетания и предложения (ср.: *синтаксис языка программирования*)
- Задача синтаксического анализатора (*parser*) – построить синтаксическую структуру входного предложения



Синтаксические связи

- Управление (играть роль, кататься на лыжах)
- Сogласование (заглавная роль, актер спел)
- Примыкание (хорошо петь)
- Союзное и бессоюзное подчинение и сочинение
(жить и работать; жить, чтобы работать)

Модели описания синтаксиса

- Система непосредственных составляющих
- Дерево зависимостей

Система составляющих

((Он сделал) (это) (так неловко)), ((что задел (образок (моего ангела)), (висевший (на дубовой (спинке кровати))),) и (что (убитая муха) (упала) (мне (прямо на голову)))).

Дерево зависимостей



Непроективность:

Чашку кофею я тебе горячего налью...

Аспекты реализации СА

- Словари (информация об индивидуальных единицах языка)
- Формальные правила
- Взаимодействие с соседними уровнями обработки (морфологический анализ, семантический анализ)

Контекстно-свободная грамматика

Context-free grammar (CFG):

$\langle N, T, X, R \rangle$

N – множество нетерминальных символов

T – множество терминальных символов
($N \cap T = \emptyset$)

X – аксиома ($X \in N$)

R – множество правил преобразования
(подстановки) типа $Y \rightarrow \alpha$, где $Y \in N$, α –
список терминальных и нетерминальных
символов

КСГ: пример

$N = \{S, NP, PP, V, N, D\}$

$T = \{John, a, girl, with, saw, telescope\}$

S

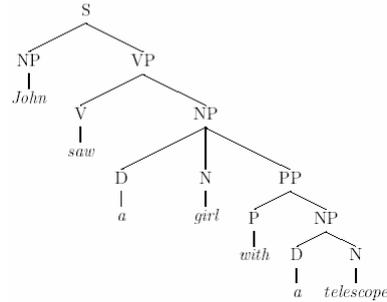
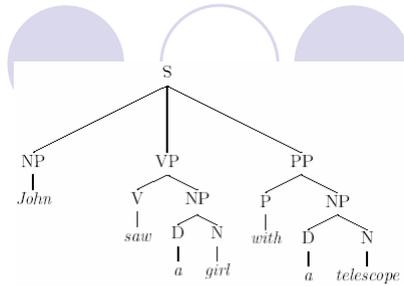
$R = \{S \rightarrow NP VP, S \rightarrow NP VP PP, NP \rightarrow D N,$
 $PP \rightarrow P NP, VP \rightarrow V NP, NP \rightarrow John,$
 $V \rightarrow saw, D \rightarrow a, N \rightarrow girl, N \rightarrow telescope,$
 $P \rightarrow with\}$

Bertrand Gaie, Guy Perrier, 2004

Неоднозначности

John saw a girl with a telescope.

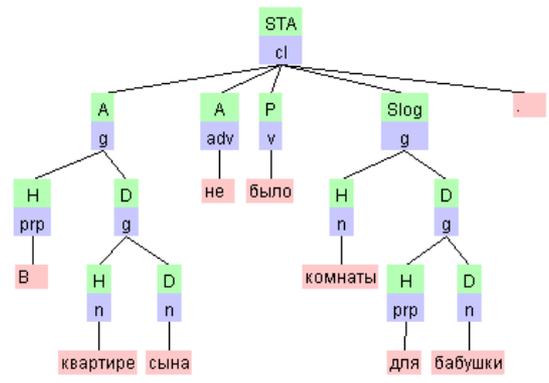
Bertrand Gaie, Guy Perrier, 2004



Пример систем

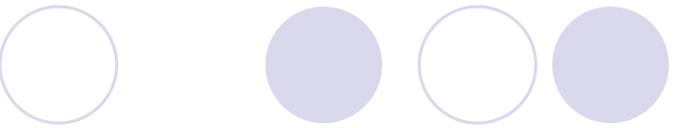
The train went on up the track out of sight, around one of the hills of burnt timber. Nick sat down on the bundle of canvas and bedding the baggage man had pitched out of the door of the baggage car.

Ernest Hemingway. Big two-hearted river



<http://visl.sdu.dk>

Пример



Часть исходного предложения:

он сделал это так неловко, что задел образец моего ангела, висевший на дубовой спинке кровати, и что убитая муха упала мне прямо на голову.

часть предложения: (* он сделал это так неловко, *)

```

---он[1](глагол)сделал[2](кого)это[3]
.....сделал[2](как)так[4]
.....сделал[2](предикатив)неловко[5]
.....сделал[2](как)неловко[5]
    
```

часть предложения: (* что задел образец моего ангела, *)

```

---образок[9](глагол)задел[8](кого)что[7]
...образок[9](какой)моего[10]
{и[19]}ангела[11](какого)моего[10]
    
```

часть предложения: (* висевший на дубовой спинке кровати, *)

```

{образок[9]}висевший[13](предлог)на[14](чём)спинке[16](какой)дубовой[15]
.....спинке[16](чего)кровати[17]
    
```

часть предложения: (* и *)

```
{образок[9]}и[19]
```

часть предложения: (* что убитая муха упала мне прямо на голову. *)

```

---муха[22](глагол)упала[23](кому)мне[24](предлог)на[26](кого)голову[27]
.....на[26](что)голову[27]
.....упала[23](предлог)на[26]
.....упала[23](как)прямо[25]
...муха[22](какая)убитая[21]
...муха[22](какая)убитая[21]
{и[19]}что[20]
---мне[24](предлог)на[26]
несвяз:он[1], муха[22],
== в предл. слов всего:24, слов несвязано:2, из них предлогов:0, время обр: 0.050 с
он[1] сделал[2] это[3] так[4] неловко[5], [6] что[7] задел[8] образец[9] моего[10] {и[19]}ангела[11] [12]
{образок[9]}висевший[13] на[14] дубовой[15] спинке[16] кровати[17], [18] {образок[9]}и[19] {и[19]}что[20]
убитая[21] муха[22] упала[23] мне[24] прямо[25] на[26] голову[27]. [28]
    
```

Parsing by chunks

«Предсинтаксис»

↔

Разбиение предложения на
непересекающиеся фразы (плоская
структура) ≠ полному разбору

(the boy (with the hat)) ↔
(the boy) with (the hat)