

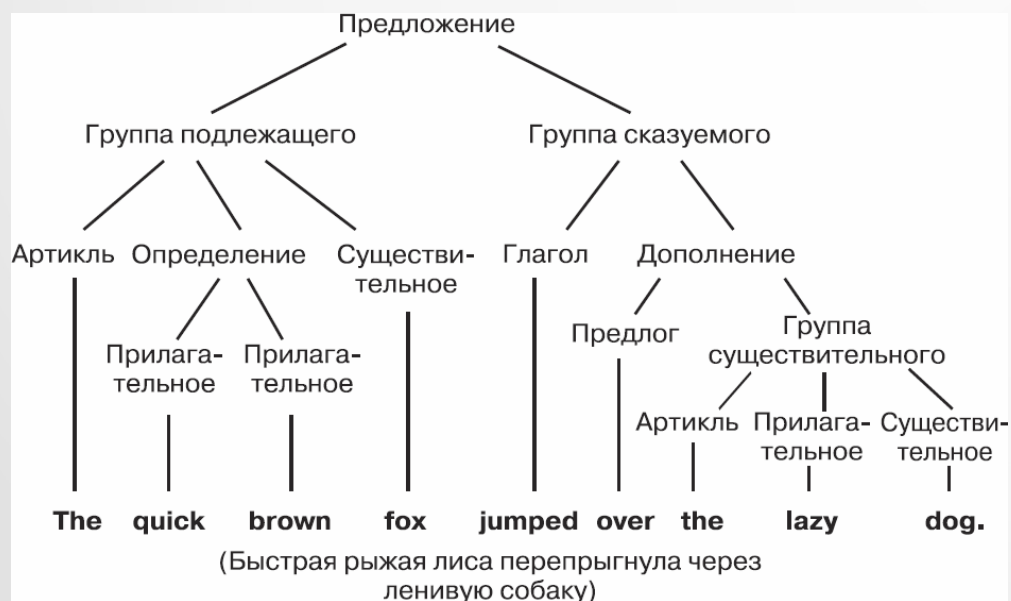


Грамматика зависимостей

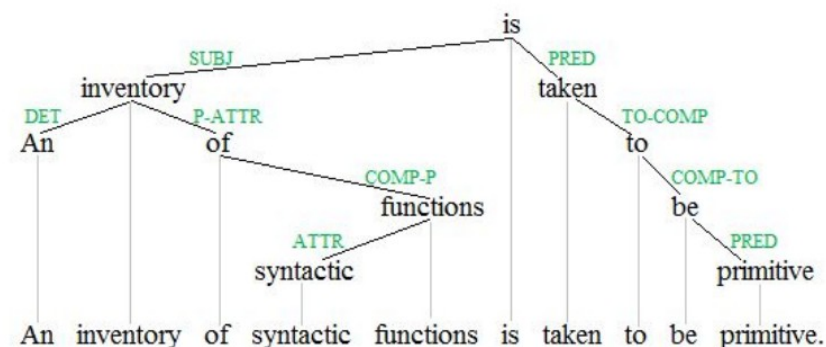
Евгений Борисов

Синтаксический анализ

грамматика составляющих



грамматика зависимостей



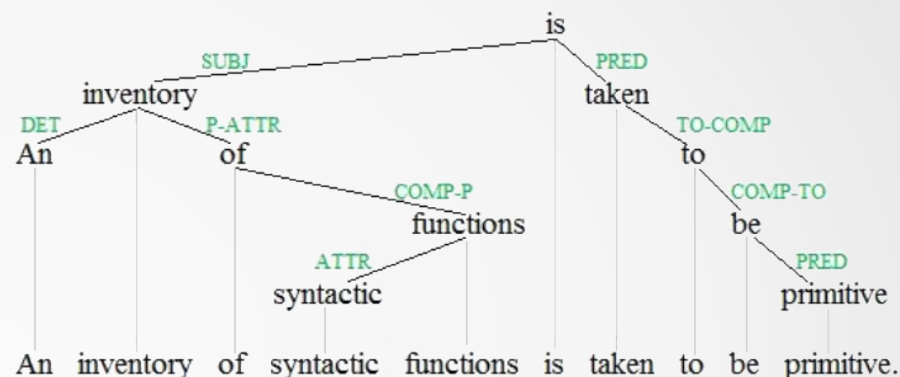
Синтаксический анализ

Грамматика зависимостей (Dependency grammar)

Главные члены предложения

Подлежащее — предмет. кто? что?

Сказуемое — что делать? что сделать? каков?



Образец разбора предложения

Где? Какие? Какие?

В саду расцвели красные и белые розы.



Это предложение – **повествовательное, невосклицательное**. Основа предложения – **розы** (подлежащее) **расцвели** (сказуемое). В предложении есть второстепенные члены, поэтому оно **распространённое**. **Розы** (какие?) **красные и белые** – однородные определения, произносятся с интонацией перечисления. **Расцвели** (где?) **в саду** – обстоятельство.

Второстепенные члены предложения

Определение — признак предмета.
какой? чей? который?

Обстоятельство — время, место, способ действия.
где? когда? куда? откуда? почему? зачем? как?

Дополнение — предмет. кого? чего? кому? чему?
кого? что? кем? чем? о ком? о чём?

Синтаксический анализ

Разбор в грамматику зависимостей

- Dependency parser
 - Malt parser (2006)
 - Stanford Neural Network Dependency Parser (2014)

строим ориентированный граф зависимостей
на упорядоченном множестве слов

Синтаксический анализ

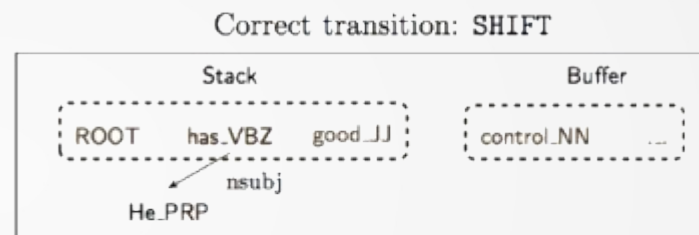
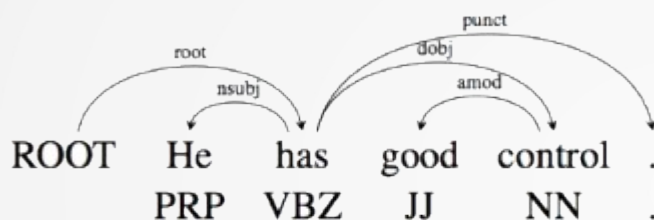
Разбор в грамматику зависимостей

- Итеративный алгоритм разбора предложения w_1, w_2, \dots, w_n
- Состояние парсера $c = (s, b, A)$:
 - стек $s = [Root]$
 - буфер $b = [w_1, w_2, \dots, w_n]$
 - множество дуг зависимостей $A = \emptyset$
- Остановка
 - стек содержит один узел $ROOT$
 - и буфер пуст
- На каждой итерации происходит выбор одного из трех правил (для выбора используется классификатор)
 - *LEFT-ARC(label)*: добавление дуги $s_1 \rightarrow s_2$ с меткой $label$ и удаление s_2 из стека. Предусловие: $|s| \geq 2$
 - *RIGHT-ARC(label)*: добавление дуги $s_2 \rightarrow s_1$ с меткой $label$ и удаление s_1 из стека. Предусловие: $|s| \geq 2$
 - *SHIFT*: перенос b_1 из буфера в стек. $|b| \geq 1$

используем ML классификатор для выбора правила на каждом шаге

Синтаксический анализ

Пример



Transition	Stack	Buffer	A
	[ROOT]	[He has good control .]	\emptyset
SHIFT	[ROOT He]	[has good control .]	
SHIFT	[ROOT He has]	[good control .]	
LEFT-ARC (nsubj)	[ROOT has]	[good control .]	$A \cup \text{nsubj}(\text{has}, \text{He})$
SHIFT	[ROOT has good]	[control .]	
SHIFT	[ROOT has good control]	[.]	
LEFT-ARC (amod)	[ROOT has control]	[.]	$A \cup \text{amod}(\text{control}, \text{good})$
RIGHT-ARC (dobj)	[ROOT has]	[.]	$A \cup \text{dobj}(\text{has}, \text{control})$
...
RIGHT-ARC (root)	[ROOT]	[]	$A \cup \text{root}(\text{ROOT}, \text{has})$

Синтаксический анализ

Литература

git clone https://github.com/mechanoid5/ml_nlp.git

Турдаков Д.Ю.

Основы обработки текстов. лекция 8.

Статистические методы синтаксического анализа. ИСП РАН, 2017

<https://www.youtube.com/watch?v=Cq9E4Fa2Jsk>

Steven Bird, Ewan Klein, and Edward Loper

Analyzing Text with the Natural Language Toolkit

<https://www.nltk.org/book/>

D.Jurafsky, J.H.Martin Speech and Language Processing. third edition, 2020