Practical AI: NLP. Syntax

Stanislav Protasov for
Harbour.Space University

Let's do it together!

Student	"Team"
Catalina Sagan	Alpha
Leonardo Fanchini	Beta
Pranav Joy	Gamma
Arjun Nair	Delta
Do Tuan Anh Tran (Bin)	Eplsion
Nikolay Shityakov	Zeta
Nathaniel Remy	Eta
Jovan Velanac	Theta
Animesh Bajpai	Iota
Anouar Meziou	Kappa
Diego Gladig	Omega
Tamino Jakobs	Gamma
Carl Philip Bluecher	Eplsion
Cornel de Vroomen	Zeta
Yousuf Labib	Omega
Olivia Engelhardt	Lambda
Nishit	Lambda

Tic-tac-time!!!

What you were supposed to do

Agenda

- **Languages**
- **Syntax**
  - Grammar
  - Part of speech
  - Syntax tree

Languages: syntax, semantics, pragmatics

~~new_var = map(lambda x: x - 2, [4, 5, 6])~~

- **Semantics**:
  - This is a valid sentence in English.
  - The worst part and clumsy looking for whoever heard light.
  - Twas brillig, and the slithy toves did gyre and gimble in the wabe.
  - Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor …
- **Grammar (syntax)**:
  - I can has cheezburger?
  - I nevr mkae tipos and erors in my sentencs.
  - I’m chuffed to bits seeing you! Do ya wanna watch some telly together, bro?
  - I’ll txt w/my ETA 2U.

Syntax of formal grammar

[Formal] grammar is

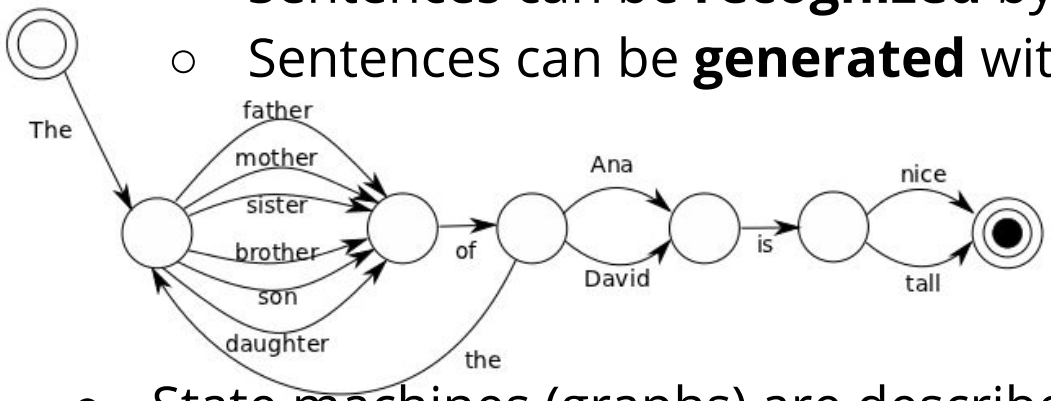
- Alphabet of terminal symbols **T (words)**
- Alphabet on non-terminals **NT (word phrases)**
- Productions **P (rules of construction and parsing)**
- Initial non-terminal symbol (sentence) **S**

Grammars are (N. Chomsky):

- **Regular**,
  - NT -> T NT
  - NT -> T
- **Context-free**,
  - NT -> *
- ...

How do we describe a grammar

- Grammar is a set of sentences
- *Regular* grammar is equivalent to finite state machine
  - Sentences can be **recognized** by walking through the graph
  - Sentences can be **generated** with random walks



- State machines (graphs) are described with listing edges or adjacency matrix (or transition matrix)
  - In regular grammars **edges** are **productions**
- For grammars there are special descriptive languages to write down productions
  - Backus-Naur form

Backus-Naur form examples and task

```
<digit> := "0" | "1" | "2" | "3" | "4" | "5" |
"6" | "7" | "8" | "9"

<integer> ::= <digit> | <integer> <digit>      (?)

<neg_integer> ::= "-" <integer>

<float> ::= _____________________              (?)

<neg_float> ::= _________________              (?)
```

Context-free grammar

```
S -> NP VP

PP -> P NP

NP -> Det N | Det N PP | 'I'

VP -> V NP | VP PP

Det -> 'an' | 'my'

N -> 'elephant' | 'pajamas'

V -> 'shot'

P -> 'in'
```

Parts of speech

Token = (lexeme, token_type)

unigram: [('This', 'DT'), ('is', 'BEZ'), ('a', 'AT'), ('sentence', 'NN'), ('that', 'CS'), ('we', 'PPSS'), ('will', 'MD'), ('use', 'VB'), ('to', 'TO'), **('test', 'NN')**, ('the', 'AT'), ...

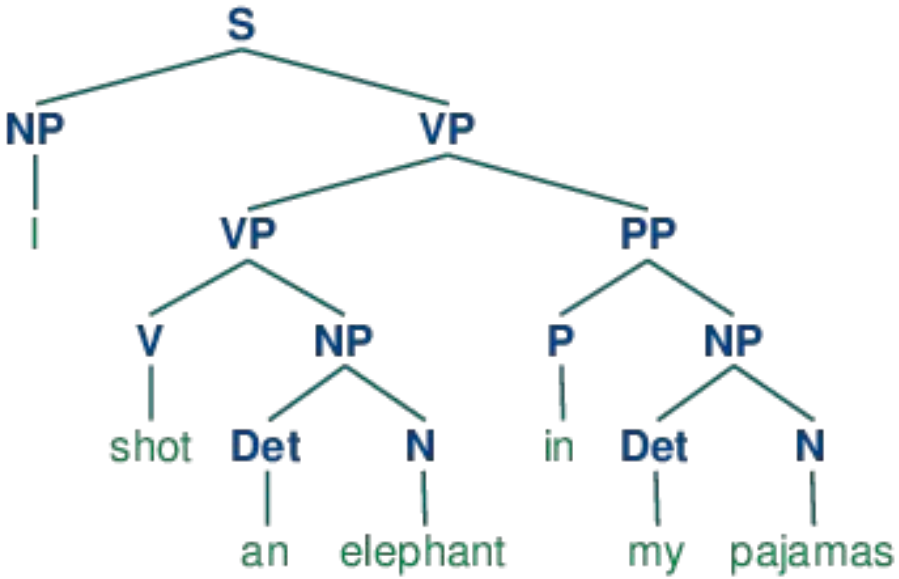
bigram: [('This', 'DT'), ('is', 'BEZ'), ('a', 'AT'), ('sentence', 'NN'), ('that', 'CS'), ('we', 'PPSS'), ('will', 'MD'), ('use', 'VB'), ('to', 'TO'), **('test', 'VB')**, ('the', 'AT'), ...

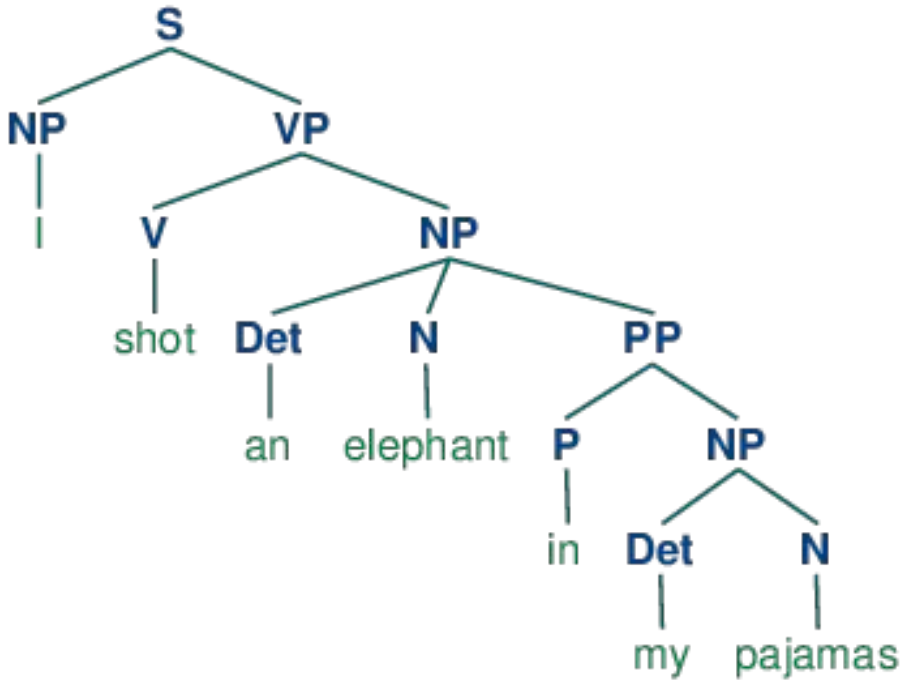
Lab #1. PoS

Study example.
https://github.com/str-anger/hsu.ai/blob/master/code/04.%20NLTK%20PoS%20example.ipynb

Discuss what is PoS, lemmas, tokens, ...

Syntax tree


S
NP
VP
I
VP
PP
V
NP
P
NP
shot
Det
N
in
Det
N
an
elephant
my
pajamas


S
NP
VP
I
V
NP
shot
Det
N
PP
an
elephant
P
NP
in
Det
N
my
pajamas

Lab #2. Syntax

Study example. Address docs or ask questions if need.

https://github.com/str-anger/hsu.ai/blob/master/code/04.%20NLTK%20Grammar%20example.ipynb

Lab #3: your own programming language

Implement formal language, that will be used to give orders to a bot on a board: turn left/right an go forward.

- Write **grammar** for your language.
  - Sentence is an operator, e.g. TURN LEFT
- Write a function, that **executes parsed operators**
- Each operator has return value (e.g. True/False)
- (*) Support GOTO and IF operators.

Write a **program in your language**, that will bring a bot from one corner of the 4x4 map to another. Run it.

- With just a list of commands on an empty map
- (*) Using GOTO and IF operators
- (**) what if map has obstacles?

Other difficulties

- Anaphoras
  - *Susan dropped the plate. It shattered loudly*
- Typos
- New words
- Homographs
- …

Lab #4: understanding NL commands

Write a tool that will from a natural sentence extract:

1. Subject (Who)
2. Action (What)
3. Object (With that)

Write few test productions that will fire with predefined action-object facts.

Test your tool using these sentences:

- Hey, [Siri], please evaluate two plus three.
- Call my mom.
- Can you play “Thriller” by Michael Jackson?
- Tell/Read me about [Elon Musk].
- Let’s play tic-tac-toe|matches

Hometask

Read provided links chapter. For a given text. Answer the following questions:

- Find top 20 most popular words and top 20 word bigrams **(mandatory)**
- (*) What are the most popular subjects (noun phrase) and actions (verbs) of sentences?

Reading for tomorrow

Please, refresh

- Matrix, Vector
- Dot product
- Matrix multiplication