



NLP Tools

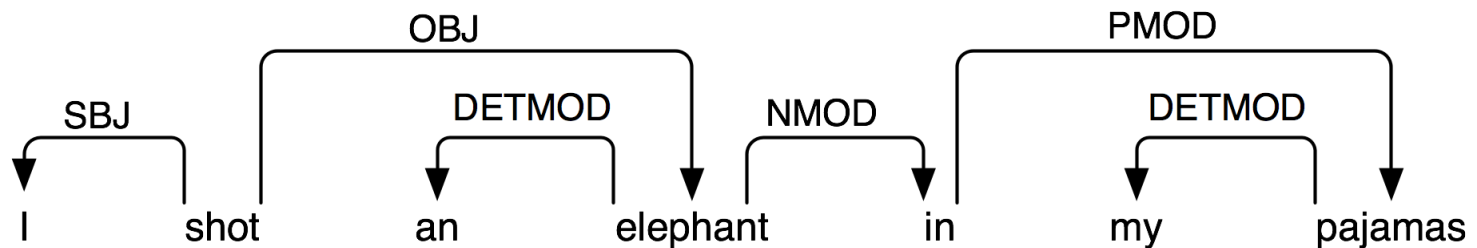
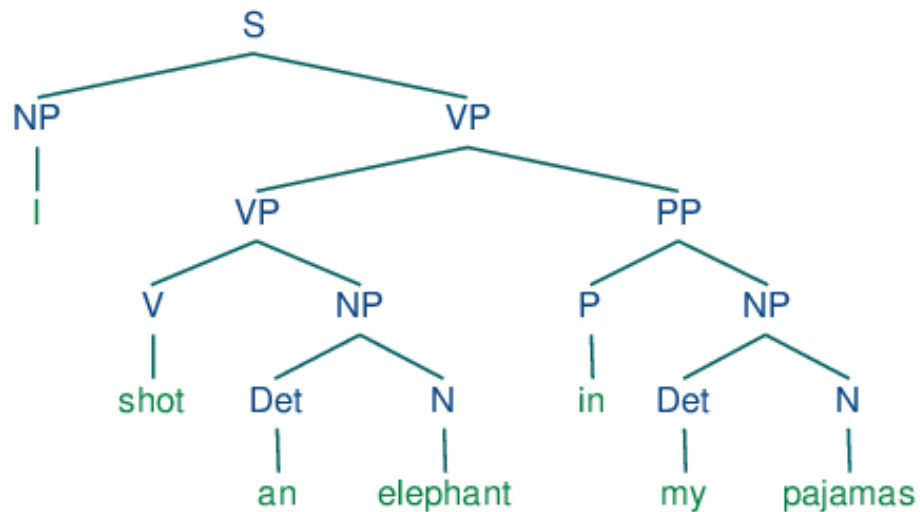
Κατσιβέλης Παναγιώτης

NLP Tools

- Syntactic Parsing
- Named-Entity Recognition
- Lexical Resources
- Language Modeling
- Sentiment Analysis
- Natural Language Generation
- General NLP Libraries

Syntactic Parsers (1)

- I shot an elephant in my pajamas.



Syntactic Parsers (2)

- Stanford Parser (Java)
 - <http://nlp.stanford.edu/software/lex-parser.shtml>
- Berkley Parser (Java)
 - <https://code.google.com/p/berkeleyparser/>

...

Part-of-Speech Taggers (1)

I shot an elephant in my pajamas.



PRP/I VBD/shot DT/an NN/elephant IN/in
PRP\$/my NN/pajamas.

Part-of-Speech Taggers (2)

- Stanford Tagger (Java)
<http://nlp.stanford.edu/software/tagger.shtml>
- Illinois POS-Tagger (Java)
http://cogcomp.cs.illinois.edu/page/software_view/3
- ...
- Greek Part-of-Speech Tager (Java)
<http://nlp.cs.aueb.gr/software.html>

Named-Entity Recognizers (1)

The origins of the order lie first in the 1978 Foreign Intelligence Surveillance Act and then in Section 215 of the Patriot Act, backed by George W Bush and passed by Congress after 9/11. In 2006 it was discovered that a similarly wide database of cellular records was being collected from customers of Verizon, AT&T and BellSouth.

Named-Entity Recognizers (2)

- Stanford NER (Java)

<http://nlp.stanford.edu/software/CRF-NER.shtml>

- Illinois Named-Entity Tagger (Java)

http://cogcomp.cs.illinois.edu/page/software_view/NETagger

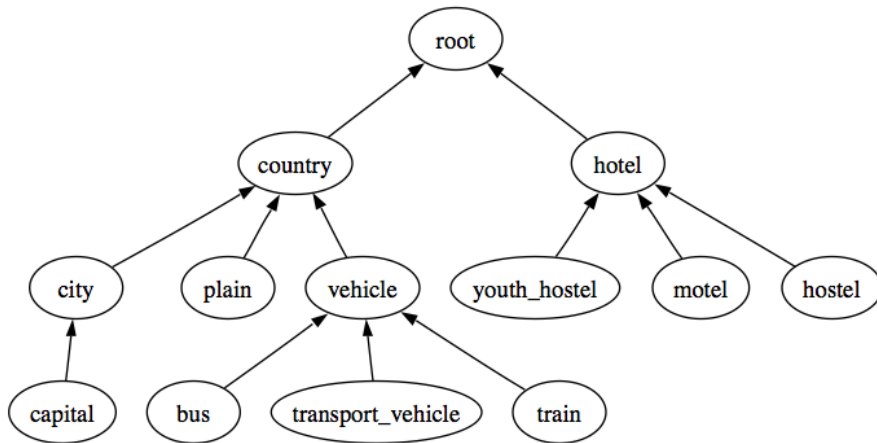
...

- Greek NER (Java)

<http://nlp.cs.aueb.gr/software.html>

Lexical Resources

WordNet



Try it online:

<http://wordnetweb.princeton.edu/perl/webwn>

Download:

<http://wordnet.princeton.edu/wordnet/download/>

APIs:

JAWS (Java): <http://lyle.smu.edu/~tspell/jaws/>

JWNL (Java): <http://sourceforge.net/projects/jwordnet/>

Wordnet-blast (C++):

<https://code.google.com/p/wordnet-blast/>

FrameNet

[Cook Matilde] fried [Food the catfish] [Heating instrument in a heavy iron skillet].

[Item Colgate's stock] rose [Difference \$3.64] [Final value to \$49.94].

APIs:

FrameNet API (Java): <http://www.cl.uni-heidelberg.de/trac/FrameNetAPI>

Stanford FrameNet II Reader (Java):

<http://nlp.stanford.edu/software/frame-net.shtml>

FrameNet Project:

Data Hub: <http://datahub.io/dataset/framenet>

Language Modeling (1)

- Assign probabilities to sentences
 - Machine Translation
 - $P(\text{Alexander the Great}) > P(\text{Alexander the Big})$
 - Spell Correction
 - $P(\text{too many people}) > P(\text{too many pepole})$
 - Speech Recognition
 - $P(\text{I believe}) > P(\text{Eye be live})$
 - Natural Language Generation, Summarization etc.
- $P(\text{"my house on the hill"}) = P(\text{my}) \times P(\text{house} \mid \text{my}) \times P(\text{on} \mid \text{my house}) \times P(\text{the} \mid \text{my house on}) \times P(\text{hill} \mid \text{my house on the})$

Language Modeling (2)

- SRILM (C++) <http://www.speech.sri.com/projects/srilm/download.html>
 - Language Modeling
 - Smoothing Techniques
 - Data Testing
 - Word Prediction
 - ...
- BerkleyLM (Java)
<https://code.google.com/p/berkeleylm/>

Sentiment Analysis

- Calculate word/sentence/text polarity (positive/negative/objective).
 - SentiWordNet (<http://sentiwordnet.isti.cnr.it/>)
 - Alchemy API (<http://www.alchemyapi.com/>)
 - Text-Processing (<http://text-processing.com/docs/sentiment.html>)
 - ReputaState (<https://www.reputastate.com/>)
 - SocialMention (<http://www.socialmention.com/>)
 - NLTK
 - LingPipe
 - Custom solutions

Natural Language Generation (1)

- **Content Determination**
 - Databases
 - RDF
 - ...
 - Information Extraction
- **Sentence Planning**
 - Templates
 - 'Standard' NLG
- **Surface Realization**
 - Realization Engines
 - Readability, grammaticality

Natural Language Generation (2)

- SimpleNLG

<https://code.google.com/p/simplenlg/wiki/Section1>

```
SPhraseSpec p = nlgFactory.createClause();  
p.setSubject("Mary");  
p.setVerb("chase");  
p.setObject("the monkey");
```

Mary chases the monkey.

General NLP Libraries

- Stanford CoreNLP (Java + Python Interface)

<http://nlp.stanford.edu/software/corenlp.shtml>

- LingPipe (Java)

<http://alias-i.com/lingpipe/>

- Apache OpenNLP (Java)

<http://opennlp.apache.org/>

- NLTK 2.0 (Python)

<http://nltk.org/>

General NLP Libraries

	CoreNLP	LingPipe	OpenNLP	NLTK
POS-Tagging	✓	✓	✓	✓
Parsing	✓	✓	✓	✓
Lexical Resources				✓
Named Entity Recognition	✓	✓	✓	✓
Text Classification		✓		✓
Sentiment Analysis		✓		