

# Introduction to NLP

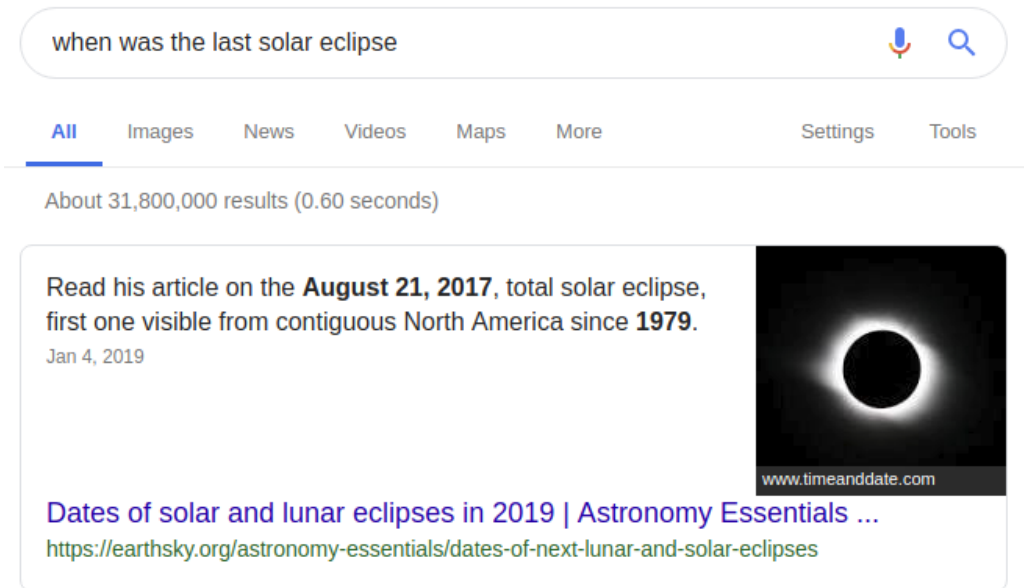
Natural Language Processing

Master in Business Analytics and Big Data

[acastellanos@faculty.ie.edu](mailto:acastellanos@faculty.ie.edu)

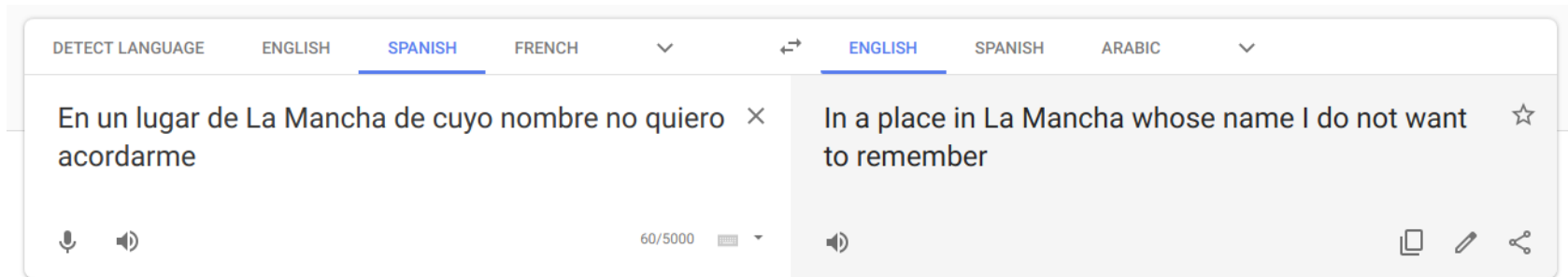
# What is NLP?

- **Automatic analysis and understanding of natural language** for the communication between computers and humans:
  - Question Answering



# What is NLP?

- **Automatic analysis and understanding of natural language** for the communication between computers and humans:
  - Question Answering
  - Machine Translation



# What is NLP?

- **Automatic analysis and understanding of natural language** for the communication between computers and humans:
  - Question Answering
  - Machine Translation
  - Language or text generation



# What is NLP?

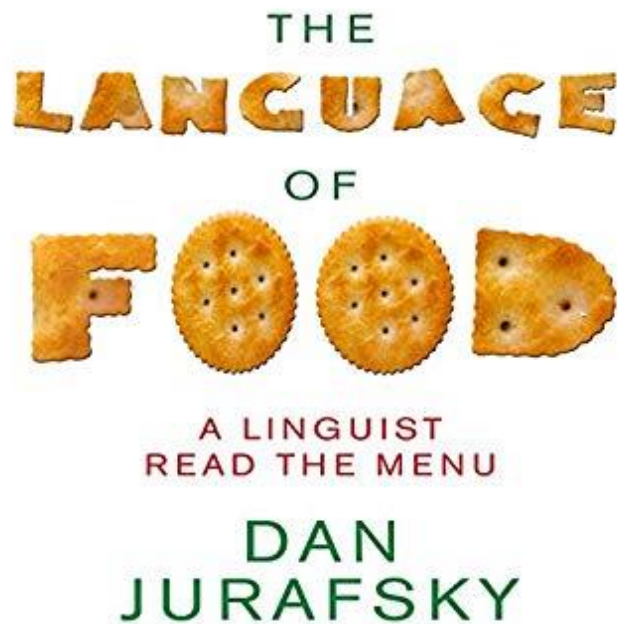
- **Automatic analysis and understanding of natural language** for the communication between computers and humans:
  - Question Answering
  - Machine Translation
  - Language or text generation
  - Fake News Detection
  - Summarization
  - Information Extraction
  - Textual classification
  - ....

# What is NLP?

- **Automatic analysis and understanding of natural language** ~~for the communication between computers and humans~~

# What is NLP?

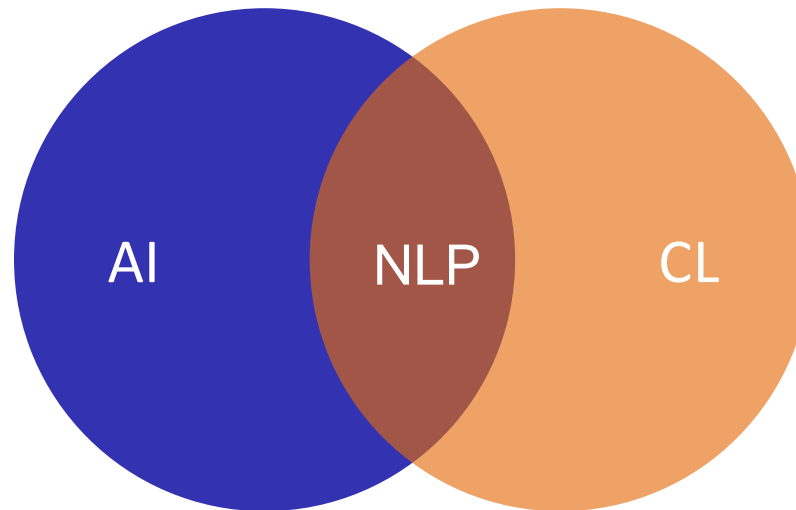
- **Automatic analysis and understanding of natural language**



- Expensive restaurants = let us be in control
  - “\$\$\$\$ restaurants have half as many dishes as cheap restaurants, are three times less likely to talk about the diner’s choice, and are seven times more likely to talk about the chef’s choice”
- Longer words in food descriptions lead to higher prices
  - Your *buckwheat orecchiette* is costing you 18 cents per additional letter per word
- **Linguistic Fillers** correlate to lower-priced food
  - Vague words (tasty!) drop prices nine percent on average.
  - Appealing adjectives (sublime!) cause a two percent drop.

# What is NLP?

- **NLP = Computer Science + Artificial Intelligence + Computational Linguistics**

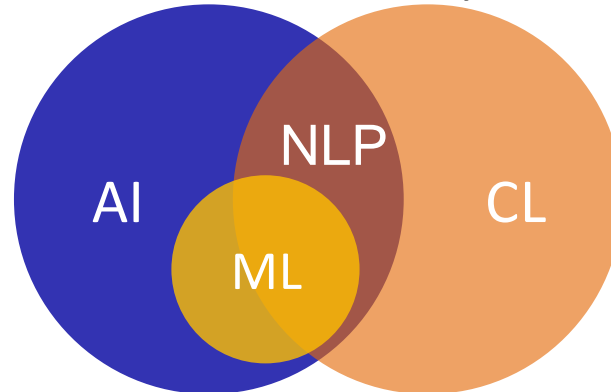




# What is NLP?

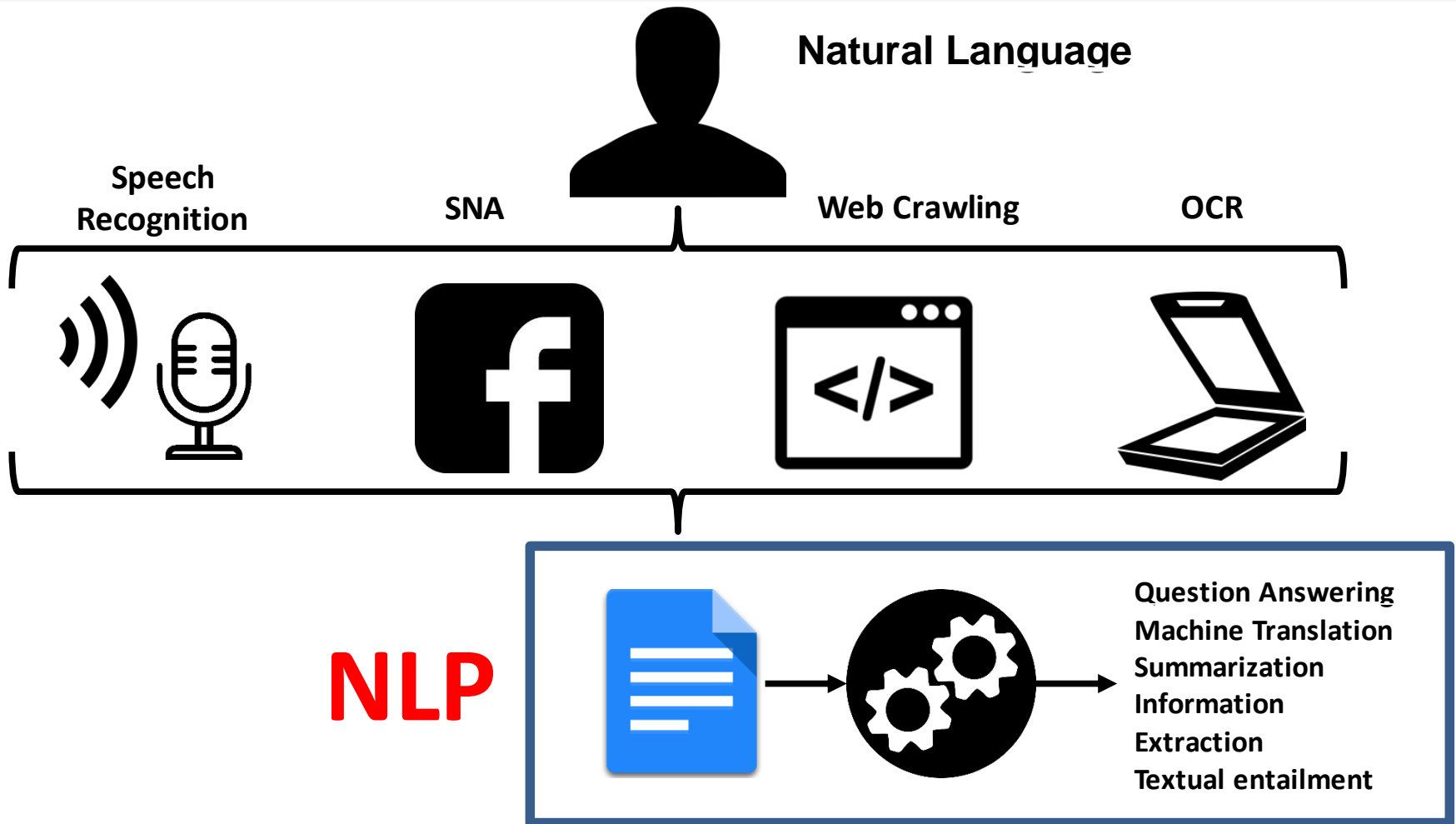
- **Statistics and ML**

- Predominant approach for NLP nowadays



- **NLP:** systems that understand language
- **ML:** systems that learn from experience
- **$\text{NLP} \cap \text{ML}$ :** systems that learn how to understand language

# What is NLP?



# What is NLP?

Speaker	Listener
<p>What to communicate?</p> <p>“What” <math>\Rightarrow</math> “Language”</p> <p>“Language” <math>\Rightarrow</math> Speech/text</p>	<p>Hears speech/reads text</p> <p>Process speech/text. Requires: I</p> <ul style="list-style-type: none"><li>• Recognize sequence of sounds/chars as words.</li><li>• Organize word sequence into structured entities - representations.</li><li>• Interpret structured entities for meaning. Represent meaning. Reason with it.</li></ul>

# What is NLP?

- How can a computer make sense out of a string

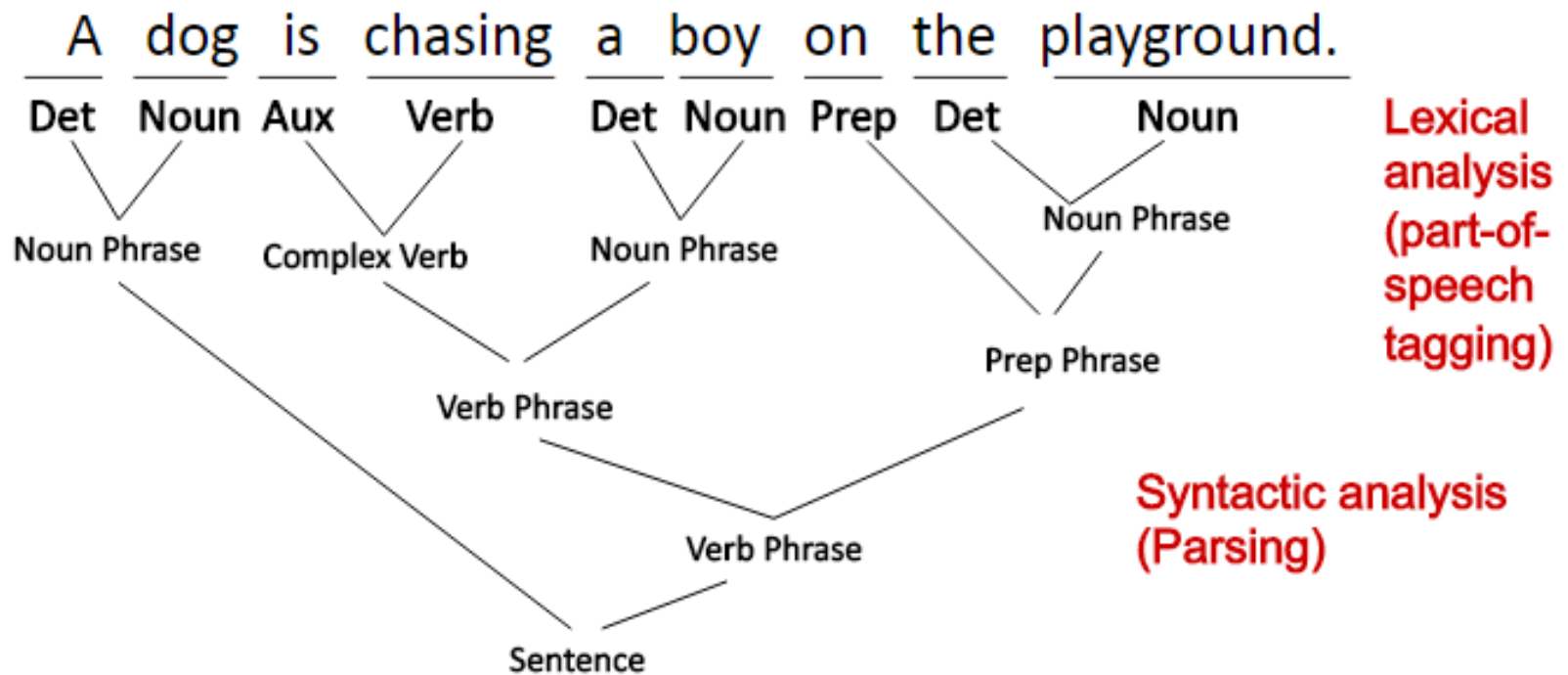
	What are the units of meaning (words)?
<b>Morphology</b>	
	What is the meaning of each Word?
<b>Syntax</b>	How are words related with each other?
<b>Semantics</b>	What is the combined meaning of words?
<b>Pragmatics</b>	What is the meta-meaning? (speech act)
<b>Inference</b>	Making Sense of everything

# What is NLP?

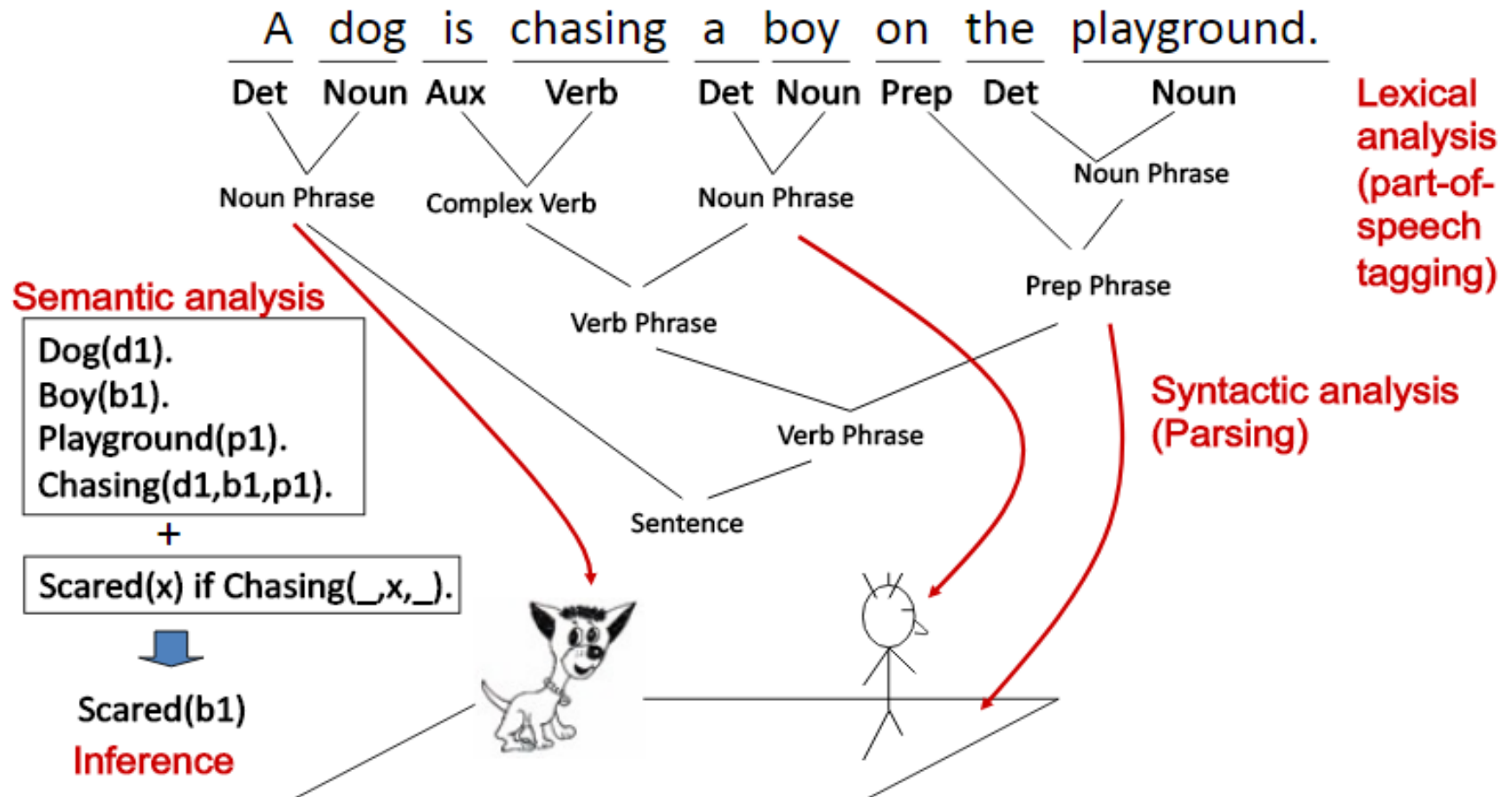
A dog is chasing a boy on the playground.  
Det Noun Aux Verb Det Noun Prep Det Noun

Lexical  
analysis  
(part-of-  
speech  
tagging)

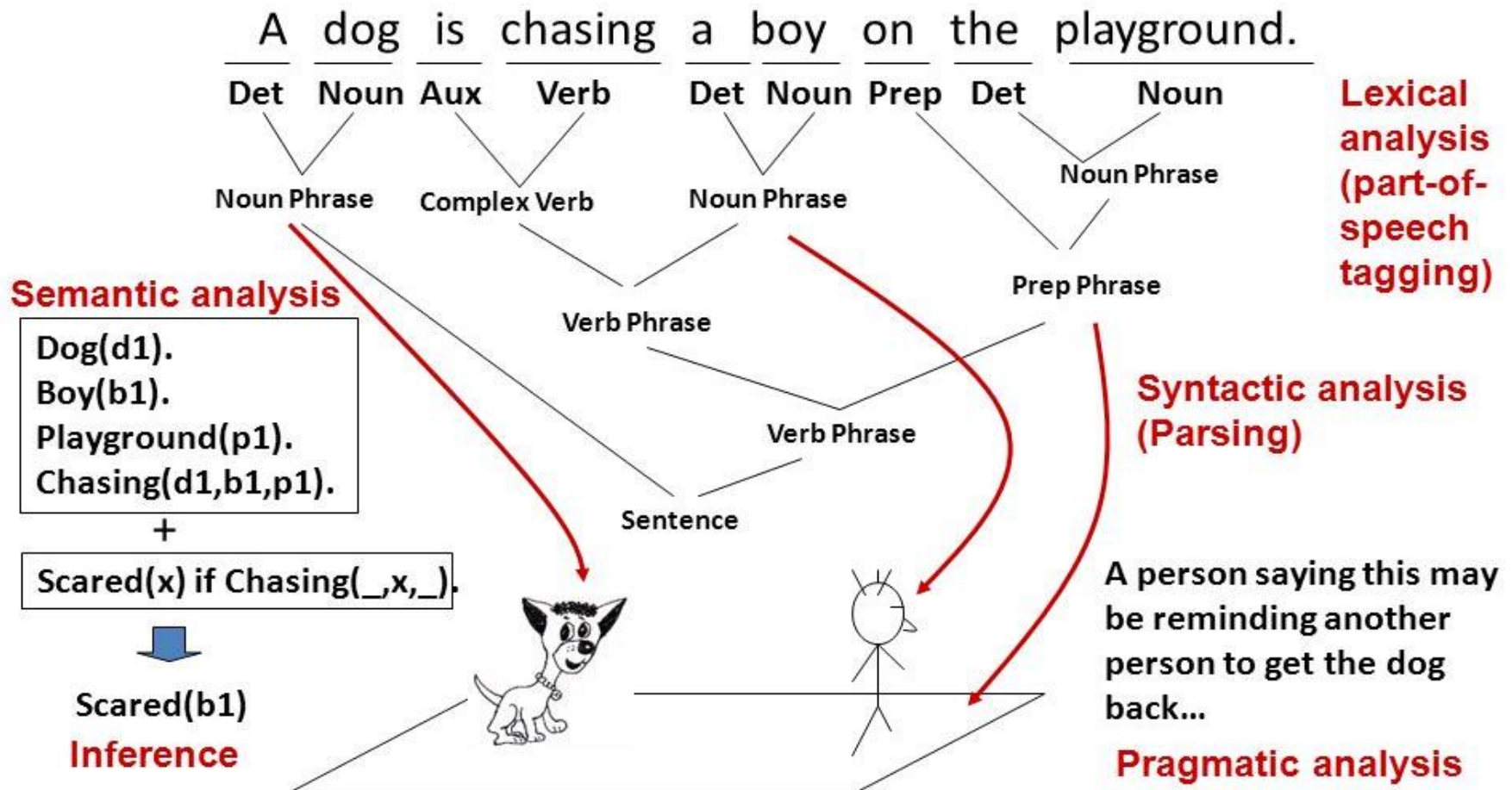
# What is NLP?



# What is NLP?



# What is NLP?





# What do we need for NLP?

- Tokenization.
- Word representation.
- Morphological segmentation/chunking.
- Part of Speech (PoS) tagging.
- Phrase/sentence/document representation.
- Shallow and deep parsing.
- Anaphora resolution.
- Meaning disambiguation.
- Representation of meaning.

# A short history of Natural Language Processing

- Machine translation started with the first computers in the 1950s
  - deeply influenced by the Chomskyian cognitive revolution
- Until the 1990s NLP centred around:
  - implementing linguistic theories of syntax
  - with parsers based on symbolic AI theorem-proving methods
- The statistical revolution started in speech recognition
  - Hidden Markov Models worked better than rule-based systems
  - in general, probabilistic approaches work better than rule-based ones
- We are at the start of a new deep learning neural network revolution

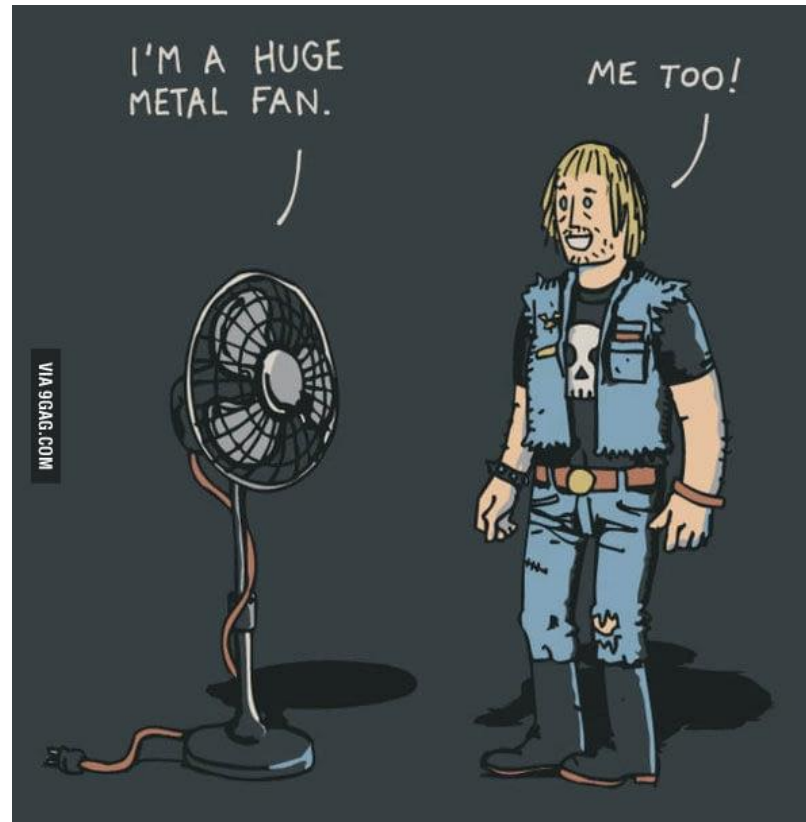
# NLP is difficult, very difficult (sorry ☹️)

## Natural Language is designed to be efficient

- We omit a lot of “**common sense**” knowledge
  - “Help me with the window”
- We have limitless ability to **say new things**
  - Language Productivity leads to combinatorial explosion
- We keep a lot of **ambiguities**
  - Ambiguity is a killer!

# NLP is difficult, very difficult (sorry ☹)

## Ambiguity



# NLP is difficult, very difficult (sorry 😞)

## Ambiguity

- Violinist Linked to JAL Crash Blossoms
- Teacher Strikes Idle Kids
- Red Tape Holds Up New Bridges
- Hospitals Are Sued by 7 Foot Doctors
- Juvenile Court to Try Shooting Defendant
- Local High School Dropouts Cut in Half

# NLP is difficult, very difficult (sorry ☹)

## non-standard English

Great job @justinbieber! Were SOO PROUD of what youve accomplished! U taught us 2 #neversaynever & you yourself should never give up either♥

## segmentation issues

the New York-New Haven Railroad  
the New York-New Haven Railroad

## idioms

dark horse  
get cold feet  
lose face  
throw in the towel

## neologisms

unfriend  
Retweet  
bromance

## world knowledge

Mary and Sue are sisters.  
Mary and Sue are mothers.

## tricky entity names

Where *is A Bug's Life* playing ...  
*Let It Be* was recorded ...  
... a mutation on the *for* gene ...

# Where are we?

## mostly solved

### Spam detection

Let's go to Agra!



Buy V1AGRA ...



### Part-of-speech (POS) tagging

ADJ ADJ NOUN VERB ADV

Colorless green ideas sleep furiously.

### Named entity recognition (NER)

PERSON ORG LOC

Einstein met with UN officials in Princeton

## making good progress

### Sentiment analysis

Best roast chicken in San Francisco!



The waiter ignored us for 20 minutes.



### Coreference resolution

Carter told Mubarak he shouldn't run again.

### Word sense disambiguation (WSD)

I need new batteries for my *mouse*.



### Parsing

I can see Alcatraz from the window!

### Machine translation (MT)

第13届上海国际电影节开幕...



The 13<sup>th</sup> Shanghai International Film Festival...

### Information extraction (IE)

You're invited to our dinner party, Friday May 27 at 8:30



Party  
May 27  
add

## still really hard

### Question answering (QA)

Q. How effective is ibuprofen in reducing fever in patients with acute febrile illness?

### Paraphrase

XYZ acquired ABC yesterday

ABC has been taken over by XYZ

### Summarization

The Dow Jones is up

The S&P500 jumped

Housing prices rose



Economy is good

### Dialog

Where is Citizen Kane playing in SF?



Castro Theatre at 7:30. Do you want a ticket?



# Where are we?

A dog is chasing a boy on the playground

Det Noun Aux Verb Det Noun Prep Det Noun

Noun Phrase Complex Verb Noun Phrase Noun Phrase

Verb Phrase Prep Phrase

Verb Phrase

Sentence

**Semantics: some aspects**

- Entity/relation extraction
- Word sense disambiguation
- Anaphora resolution

**Inference: ???**



**Speech act analysis: ???**

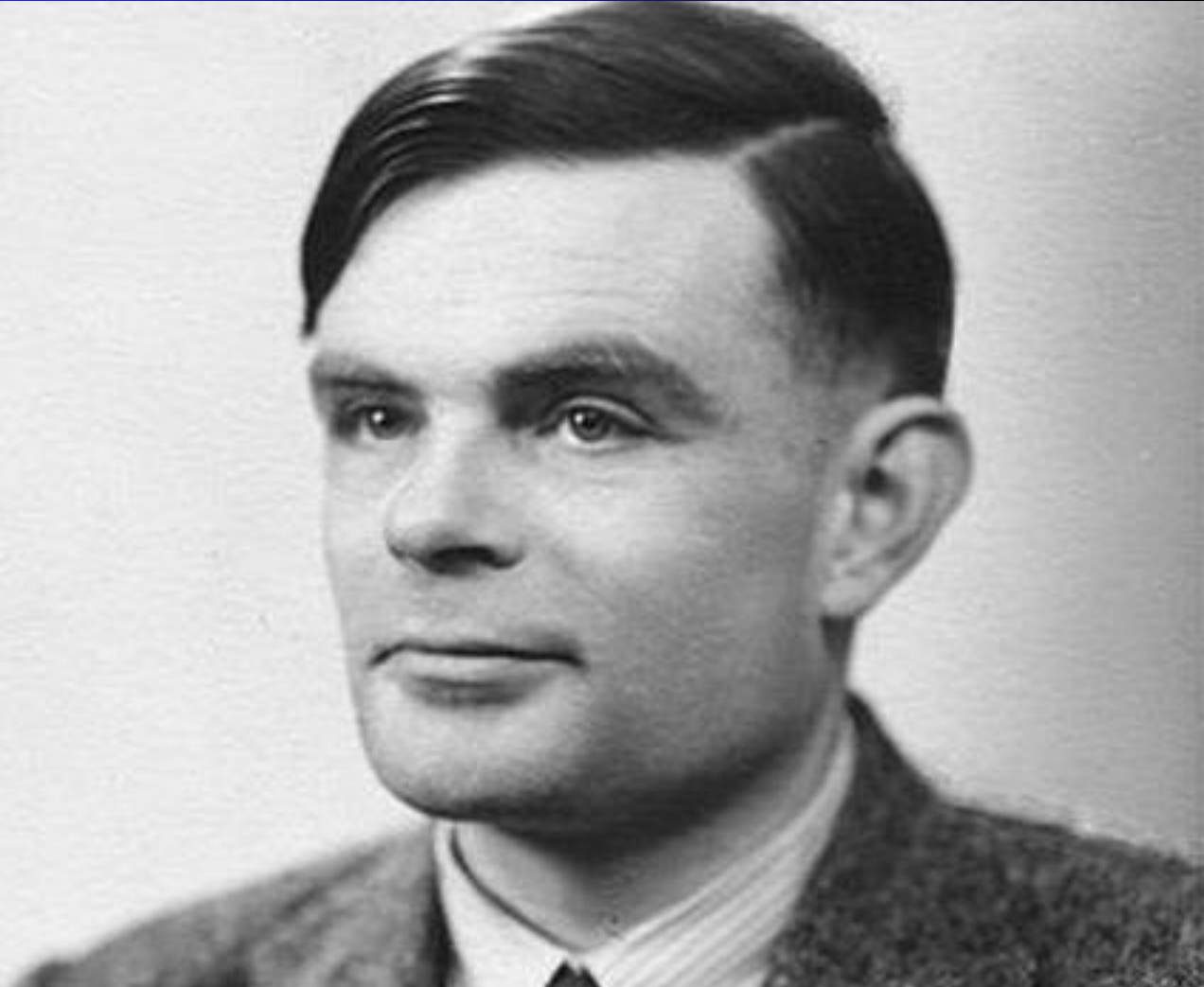
**POS  
Tagging:  
97%**

**Parsing: partial >90%**



What is NLP  
What do we need for NLP?  
NLP is difficult  
Where are we?  
Why NLP and why now?

# Where do we want to arrive?



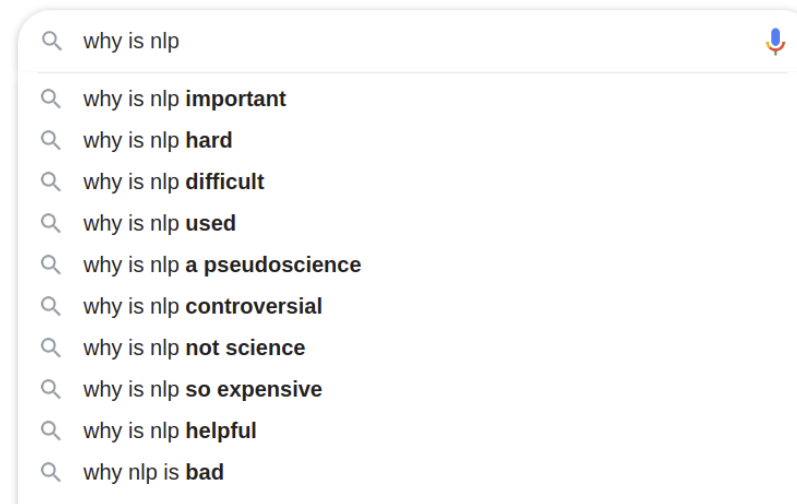
# Why NLP and why now?



# Why NLP and why now?

## What for?

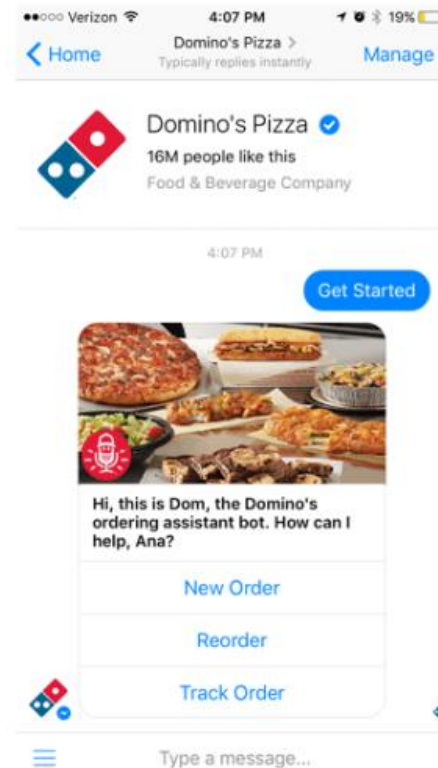
- Improve user experience



# Why NLP and why now?

## What for?

- Automate support



# Why NLP and why now?

## What for?

- Monitor and analyze feedback

### Read reviews that mention



Showing 1-8 of 21 reviews with "ios" [Clear filter](#)

★★★★☆ Great Phone But...

By [harsh gaur](#) on 16 July 2018

Colour: Space Grey | Size: 256GB | [Verified Purchase](#)

...chip that came significantly later. A11 chip coupled with greatly optimized **iOS** makes this phone faster than almost 95% smartphones available. The facelid with Ir dot projection is most secure than any other. 2. The latest released **iOS 11** comes packed with features and is faster than ever. It has the most simplistic and natural user interface on a smartphone. 3. It has currently the 3rd best camera present on any smartphone out there after the samsung galaxy s9+ and HUawei p20 pro. But unlike the first two, it provides more details, similar quality in a smaller image size meaning it takes up less storage space. 4. I am just in love with the design of this phone. The phone is just gorgeous, I mean just beautiful. Depending on who you ask, the notch is debatable but I don't mind it because it gives more screen space. But with more and more phones coming with the notch, you just have to accept that it's going to stay. 5. The biggest selling point of apple devices is the after sales service that apple provides, being it software updates, hardware issues or just any customer support whatsoever. Apple service is best in class and nobody can argue that. 6. [Read more](#)

18 people found this helpful

[Helpful](#) [Not Helpful](#) [Comment](#) [Report abuse](#)

★★★★★ seamless and best OS ever in Mobile history - **iOS**

By [harshad](#) on 5 June 2018

Colour: Space Grey | Size: 64GB | [Verified Purchase](#)

Masterclass from the house of Apple. Used for month and no complaint. Robust, secure, seamless and best OS ever in Mobile history - **iOS**.

4 people found this helpful

[Helpful](#) [Not Helpful](#) [Comment](#) [Report abuse](#)

# Why NLP and why now?



I believe that the really interesting challenge in NLP, which will be the key to actual "**natural language understanding**", is the design of learning algorithms that will be able to learn to represent meaning.

**Gartner**®

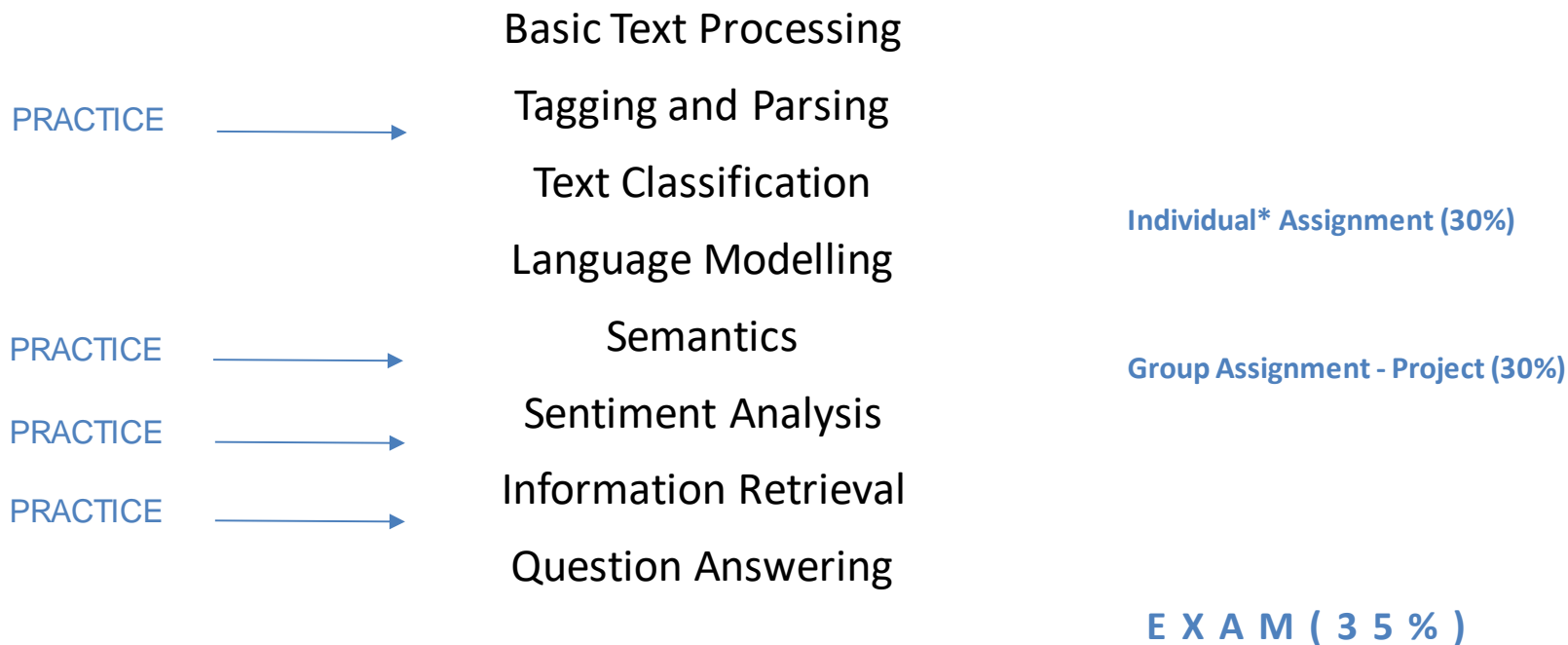
By 2020 **50% of analytics queries will be generated by NLP, NLG**, search and automated Insights

# Do I need to be a master programmer?

## Skills you will need

- Simple linear algebra (vectors, matrices)
- Basic probability theory
- Python programming
- Some Machine Learning (ML2 😊)

# Syllabus, Goals, Practices and Assignments





# Environment



**Notebook**



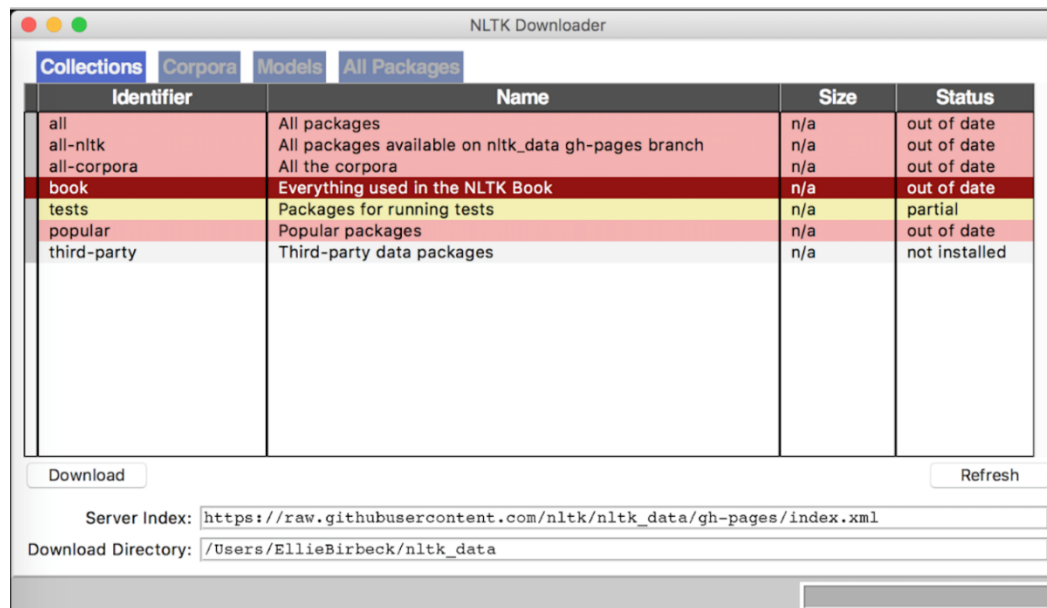
# Libraries



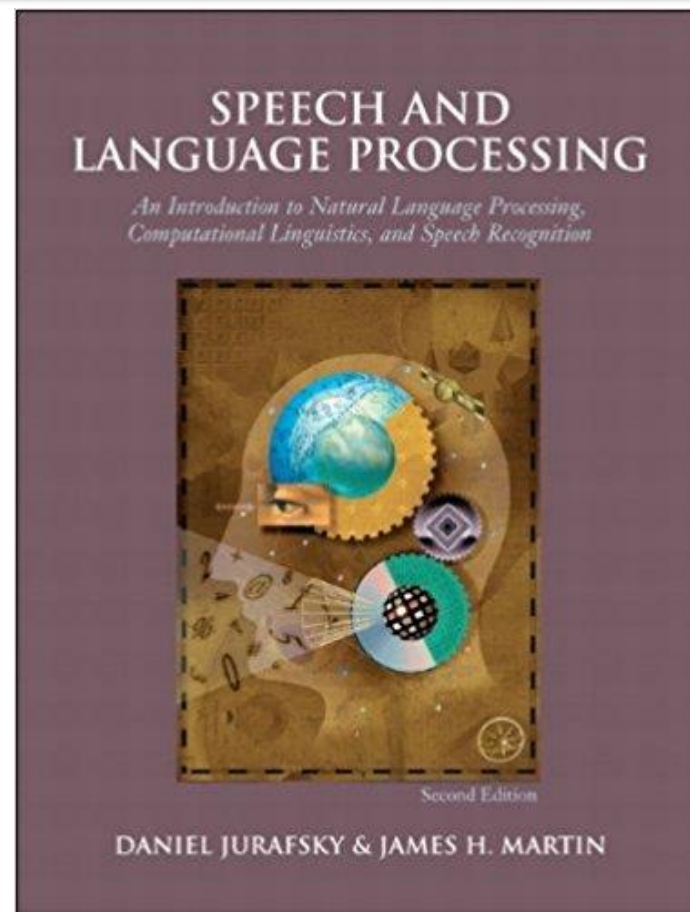
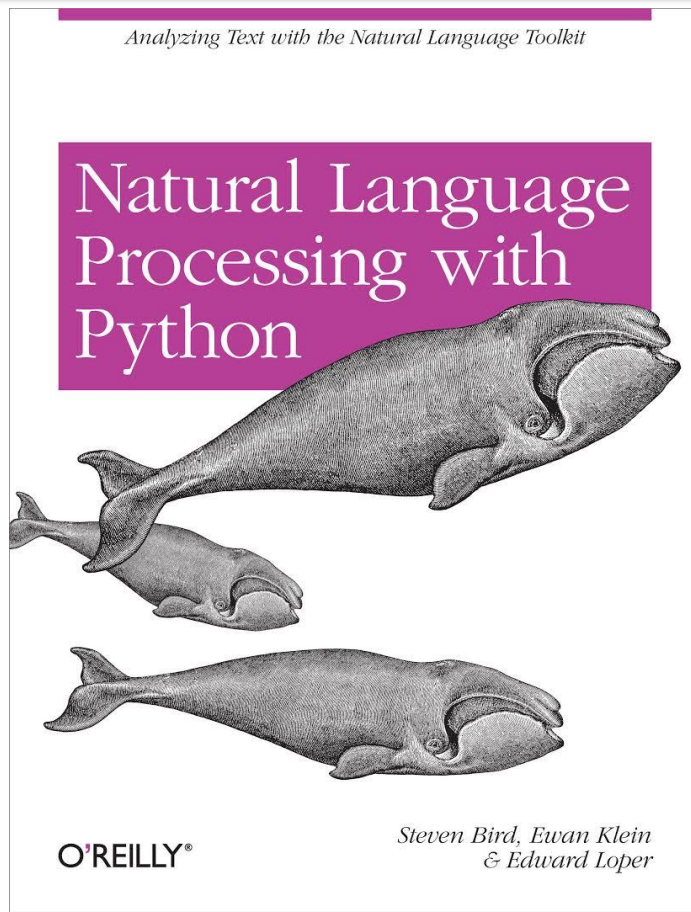
Natural Language Analysis  
with Python NLTK

# NLTK

- Installing NLTK: <https://www.nltk.org/install.html>
- Installing NLTK Data: <https://www.nltk.org/data.html>



# Books



# Need more help?

