



NLP - Foundations

Liad Magen

THE INTERNATIONAL PHONETIC ALPHABET (revised to 2020)

CONSONANTS (PULMONIC)

© 2020 IPA

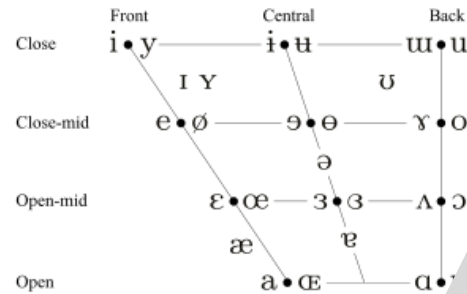
	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Symbols to the right in a cell are voiced, to the left are voiceless. Shaded areas denote articulations judged impossible.

CONSONANTS (NON-PULMONIC)

Clicks	Voiced implosives	Ejectives
◌ ɓ Bilabial	ɓ Bilabial	ʼ Examples:
◌ ɗ Dental	ɗ Dental/alveolar	pʼ Bilabial
◌ ɗ̥ (Post)alveolar	ɗ̥ Palatal	tʼ Dental/alveolar
◌ ɗ̥ Palatoalveolar	ɗ̥ Velar	kʼ Velar
◌ ɗ̥ Alveolar lateral	ɗ̥ Uvular	sʼ Alveolar fricative

VOWELS



Where symbols appear in pairs, the one to the left represents a rounded vowel.

SUPRASEGMENTALS

- ˈ Primary stress
- ˌ Secondary stress
- ː Long
- ˑ Half-long
- ◌̃ Extra-long
- ◌̥ Nasal release
- ◌̥̥ Lateral release
- ◌̥̥̥ No audible release

OTHER SYMBOLS

- ɸ Voiceless labial-velar fricative
- ɕ ʑ Alveolo-palatal fricatives
- ɰ Voiced labial-velar approximant
- ɻ Voiced alveolar lateral flap
- ɰ Voiced labial-palatal approximant
- ɰ Simultaneous ʃ and x
- ɰ Voiceless epiglottal fricative
- ɰ Voiced epiglottal fricative
- ɰ Epiglottal plosive

Affricates and double articulations can be represented by two symbols joined by a tie bar if necessary.

ts kp

DIACRITICS

◌̥ Voiceless	◌̥ ◌̥	◌̥ Breathy voiced	◌̥ ◌̥	◌̥ Dental	◌̥ ◌̥
◌̥ Voiced	◌̥ ◌̥	◌̥ Creaky voiced	◌̥ ◌̥	◌̥ Apical	◌̥ ◌̥
◌̥ Aspirated	◌̥ ◌̥	◌̥ Linguolabial	◌̥ ◌̥	◌̥ Laminal	◌̥ ◌̥
◌̥ More rounded	◌̥ ◌̥	◌̥ Labialized	◌̥ ◌̥	◌̥ Nasalized	◌̥ ◌̥
◌̥ Less rounded	◌̥ ◌̥	◌̥ Palatalized	◌̥ ◌̥	◌̥ Nasal release	◌̥ ◌̥
◌̥ Advanced	◌̥ ◌̥	◌̥ Velarized	◌̥ ◌̥	◌̥ Lateral release	◌̥ ◌̥
◌̥ Retracted	◌̥ ◌̥	◌̥ Pharyngealized	◌̥ ◌̥	◌̥ No audible release	◌̥ ◌̥
◌̥ Centralized	◌̥ ◌̥	◌̥ Velarized or pharyngealized	◌̥ ◌̥		
◌̥ Mid-centralized	◌̥ ◌̥	◌̥ Raised	◌̥ ◌̥		

hə'loʊ, wɜrld



True or False?

Read the statements and decide if they are **correct** or **not**.

<https://forms.gle/xJvLAN8GCmiM6NBfA>



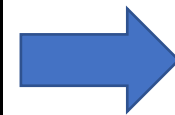
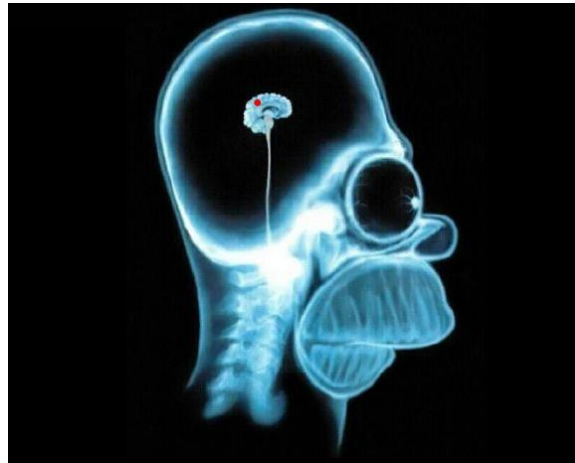
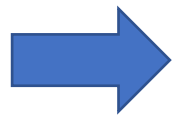


Today we'll cover:

- NLP – Motivation & Applications
- Representation Levels
- Challenges
- NLP as a classification problem
- Text Clustering

What is NLP?

Human
Language

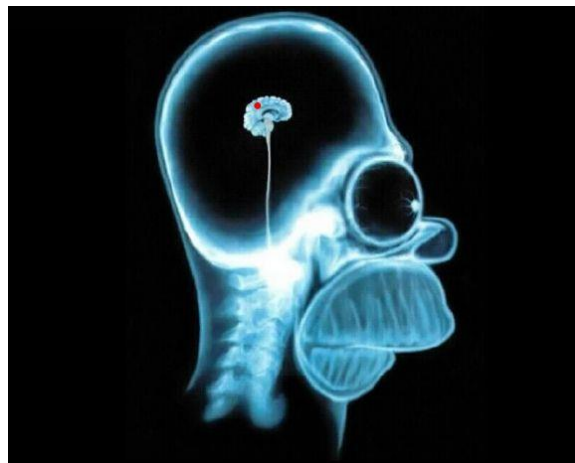


Non-trivial
useful output

Algorithms that operate on **Human Language**

What is NLP?

Human
Language



Non-trivial
useful output

Process input text in human language, in a way that suggests an intelligence was involved

What is NLP?

True NLP isn't keyword matching but rather **real understanding of language.**

AKA Computational Linguistics.

... But unfortunately, this is not yet feasible...



when did juan carlos became king of spain?

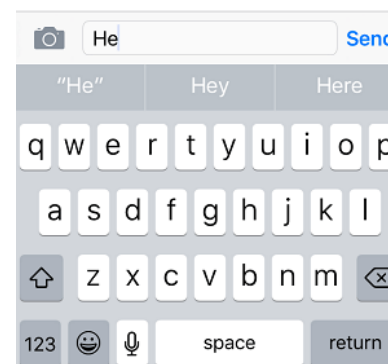
Web News Images Videos Shopping More Search tools

About 239,000 results (0.77 seconds)

Juan Carlos became King on **22 November 1975**, two days after Franco's death, the first reigning monarch since **1931**. His father did not abdicate in favor of his son until **1977**. Soon after enthronement, Juan Carlos introduced reforms to dismantle the Francoist regime and begin the Spanish transition to democracy.

Juan Carlos I of Spain - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Juan_Carlos_I_of_Spain Wikipedia

Feedback



NLP Usage



Search Engines

Exact words

Synonyms



assisting elderly



aginginplace.org > Caregiving ▾

Helping The Elderly At Home | Updated for 2020 ...

Tips for **Aging** Well – At Home · Eating · Functional Mobility (moving from one place to another while performing activities such getting in and out of bed, in and out ...

www.helpguide.org > articles > senior-housing > home-... ▾

Home Care Services for Seniors - HelpGuide.org

Help with the activities of daily living, such as dressing, bathing, or meal preparation, is called personal or custodial care. Home health aides can provide personal care services that range from a few hours a day to around-the-clock live-in care.

www.helpinghandshomecare.co.uk > home-care-services ▾

Elderly Care - Book Care & Support Of The Elderly | Helping ...

Aug 28, 2020 — Introductory senior care agencies can't enforce the training of their agency staff. However, with **Helping** Hands, each of our in-house carers are ...

What are the types of elderly care? ▾

What does care of the elderly mean? ▾

Why is elderly care important? ▾

londonlive-incare.com > what-does-a-personal-assistant... ▾

What Does a Personal Assistant For The Elderly Actually Do ...

Feb 18, 2019 — A personal **assistant** for the **elderly** helps an older person with activities related to general life, e.g. managing money, paying bills on time, making appointments, arranging transport or actually catching the transport, and tasks around the house which the person might otherwise struggle with.

en.wikipedia.org > wiki > Elderly_care ▾

Elderly care - Wikipedia

Their adult children often find it challenging to **help** their **elderly** parents make the right choices. Assisted living is one option for the **elderly** who need **assistance** ...

www.agingcare.com > ... > Home Care > Articles ▾

In-Home Services That Help Seniors Continue to Live at ...

Apr 16, 2020 — There are various sources of **assistance** for **seniors** living alone. The following list includes some common things elders need a **helping** hand ...

when was Gauss born and when did he die

All News Images Videos Maps More Settings Tools

About 10.300.000 results (2,08 seconds)

Carl Friedrich Gauss

Mathematician



Born: April 30, 1777, Brunswick, Germany

Died: February 23, 1855, Göttingen, Germany

People also search for

View 10+ more



Leonhard Euler



Bernhard Riemann



Euclid



Alexander von Humboldt



Eugene Gauss
Son



Pierre de Fermat



Isaac Newton

People also search for

Q bernhard riemann

Q eugene gauss

Q carl friedrich gauss education

Q carl friedrich gauss facts

Knowledge Graph

en.wikipedia.org › wiki › Carl_Friedrich_Gauss

Carl Friedrich Gauss - Wikipedia

Johann Carl Friedrich **Gauss** was a German mathematician and physicist who made significant ... Johann Carl Friedrich **Gauss** was **born** on 30 April 1777 in Brunswick ... On 23 February 1855, **Gauss** **died** of a heart attack in Göttingen (then ... For **Gauss**, not **he** who mumbles his creed, but **he** who lives it, is accepted.

Died: 23 February 1855 (aged 77); **Göttingen**, ... **Thesis:** Demonstratio nova... (1799)

Born: Johann Carl Friedrich Gauss; 30 April 1... **Children:** Joseph; Wilhelmina; Louis; Euge...

Heptadecagon · Sophie Germain · Disquisitiones Arithmeticae

www.britannica.com › Science › Physics › Physicists

Carl Friedrich Gauss | Biography, Discoveries, & Facts ...



in which state was yahoo founded



All

News

Images

Maps

Videos

More

Settings

Tools

About 47,300,000 results (1,04 seconds)

Yahoo! / Founded

January 1994,
Sunnyvale, California,
United States



People also search for



Google
September 4,
1998, Menlo
Park,
California,
United States



AOL
May 24, 1985,
United States



ESPN
September 7,
1979, Bristol,
Connecticut,
United States

Feedback

en.wikipedia.org › wiki › Yahoo!

Yahoo! - Wikipedia

Yahoo is an American web services provider headquartered in Sunnyvale, California, and owned by Verizon Media. The original **Yahoo!** company was **founded** by Jerry Yang and David Filo in ... At its height it was one of the most popular sites in the United States. According to third-party web analytics providers Alexa and ...

Founder(s): Jerry Yang; David Filo

Revenue: \$5.17 billion

Founded: January 1994; 26 years ago

Advertising: Native

History of Yahoo! · Jerry Yang · Yahoo! Search · David Filo

en.wikipedia.org › wiki › History of Yahoo!

Yahoo!

Web services company

Yahoo! is an American web services provider headquartered in Sunnyvale, California, and owned by Verizon Media. The original Yahoo! company was founded by Jerry Yang and David Filo in January 1994 and was incorporated on March 2, 1995. Yahoo was one of the pioneers of the early Internet era in the 1990s. [Wikipedia](#)

Founded: January 1994, Sunnyvale, California, United States

Employees: 8,600 (March 2017)

Headquarters: Sunnyvale, California, United States

Parent organization: Verizon Media

Video game: Yahoo! Soccer Manager

Founders: Jerry Yang, David Filo

Profiles



Twitter



LinkedIn



Instagram



Pinterest



Tumblr

People also search for

View 10+ more



MSN



Gmail



Outlook...



Yahoo!
Search



ESPN

Advanced Knowledge Graph

Recent Advances in Machine Reading

<http://ai.stanford.edu/blog/beyond-local-pattern-matching/>

History of Yahoo!

From Wikipedia, the free encyclopedia

See also: *Timeline of Yahoo!*



This article needs to be **updated**. Please update this article to reflect recent events or newly available information. (May 2016)

Yahoo! was started at **Stanford University**. It was founded in January 1994 by **Jerry Yang** and **David Filo**, who were Electrical Engineering graduate students when they created a website named "Jerry and David's Guide to the World Wide Web". The Guide was a directory of other websites, organized in a hierarchy, as opposed to a searchable index of pages. In April 1994, Jerry and David's Guide to the World Wide Web was renamed "Yahoo!".^{[1][2]} The word "YAHOO" is a **backronym** for "**Yet Another** Hierarchically Organized Oracle"^[3] or "Yet Another Hierarchical Official Oracle."^[4] The yahoo.com domain was created on January 18, 1995.^[5]

Yahoo!

Stanford University

California

what does barack obama do now



All

Images

News

Videos

Maps

More

Settings

Tools

About 315.000.000 results (0,86 seconds)

www.townandcountrymag.com › society › politics › news ▼

What Barack Obama Is Doing Now - Barack Obama Latest ...

Jun 24, 2019 — However, the former **president did** have some kind words for Biden in general, releasing a statement following Biden's official declaration of ...

www.thesun.co.uk › news › barack-obama-doing-now-net...

What has Barack Obama done since leaving office, what is...

What has the former **President** been **doing** since leaving the leaving the White ... Mr **Obama** has reportedly told associates he **does** not intend to use the new ...

www.foxnews.com › four-surviving-ex-presidents-and-wh...

Four surviving ex-presidents and what they are doing now ...

Feb 3, 2020 — Former Presidents **Barack Obama**, George W. Bush, Bill Clinton, and Jimmy ... The former president **did** speak out, in a Facebook post in 2017, ...

en.wikipedia.org › wiki › Barack_Obama ▼

Barack Obama - Wikipedia

Barack Hussein Obama II is an American politician and attorney who served as the 44th ... He **is the only president** born outside the contiguous 48 states. ... In January 2008, **Obama** told Christianity **Today**: "I am a Christian, and I am a devout ... David, but the members of the family **do** not attend church on a regular basis.

[Early life and career](#) · [Family and personal life](#) · [Law career](#) · [Legislative career](#)

www.quora.com › What-is-Barack-Obama-doing-right-n... ▼

What is Barack Obama doing right now? - Quora

Sep 22, 2015 — Q: **What does Barack Obama do now?** Enjoy "retirement"...and wealth. This is where the Barack and Michelle Obama settled with their family after leaving the ...

85 answers

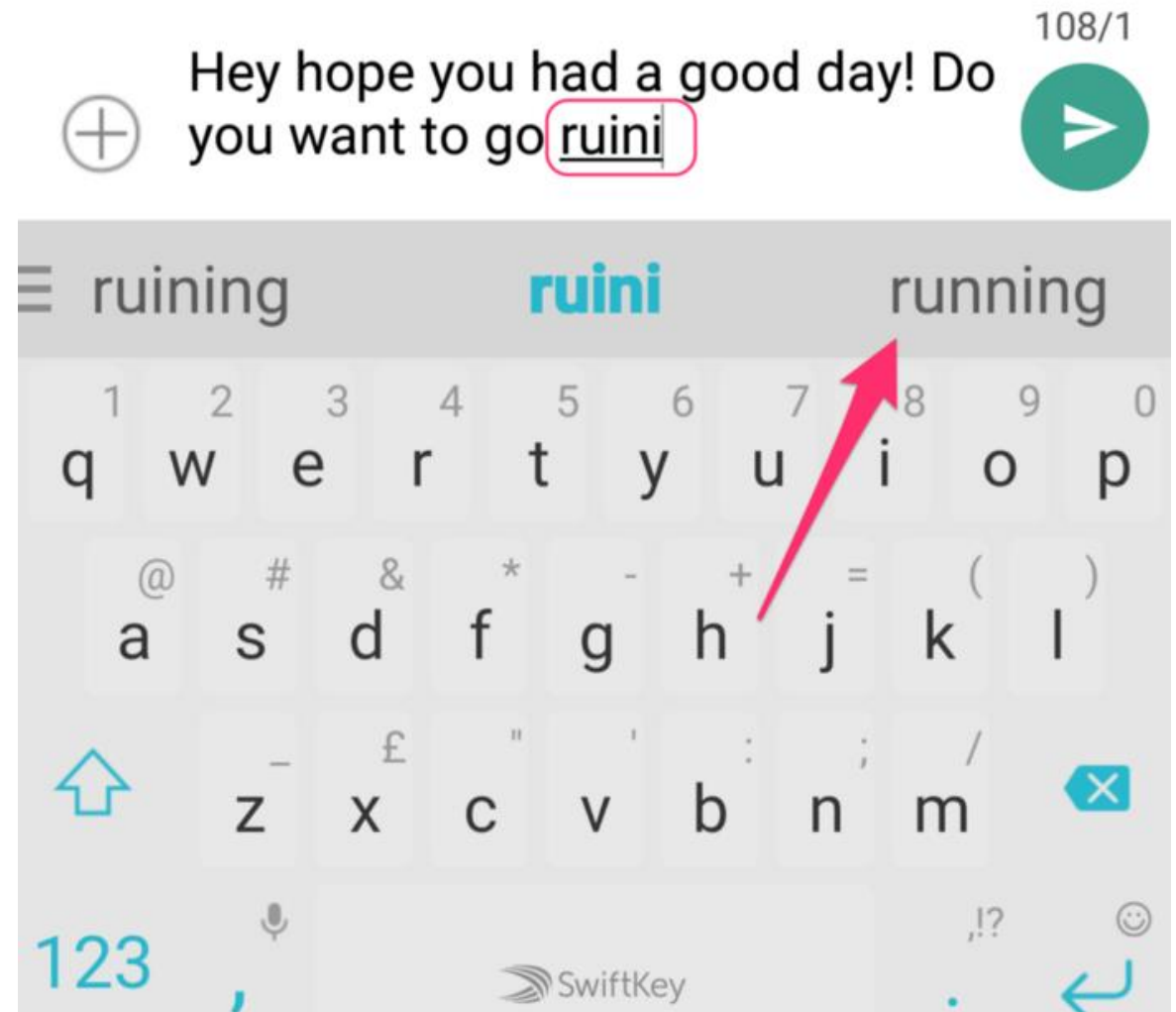
How is **Barack Obama doing today?** - Quora 9 answers 11 Mar 2019

What will **Barack Obama do** after he's done being ... 109 answers 20 Feb 2017

[More results from www.quora.com](#)

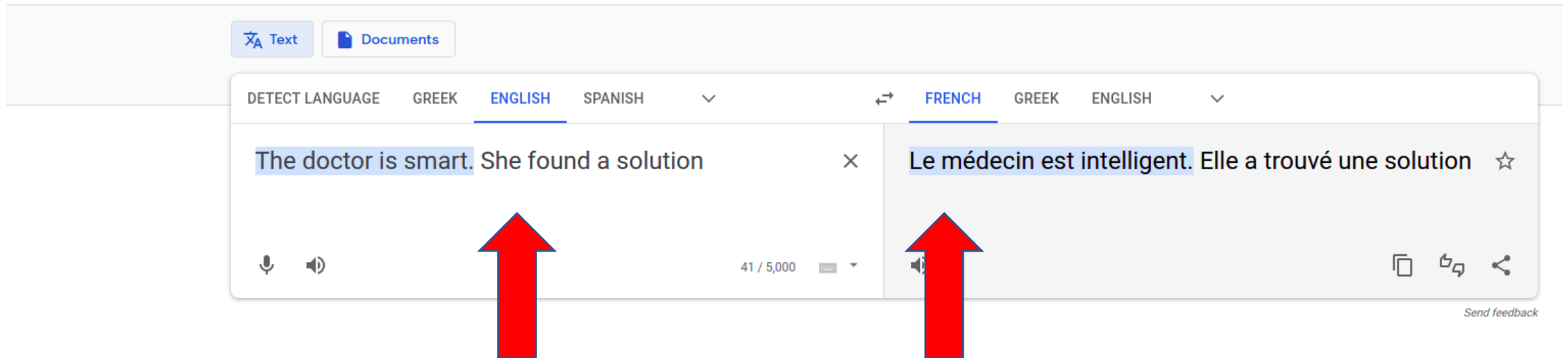
Back to normal search

Word Prediction & Spelling Correction



Machine Translation

Google Translate



... How would you translate this passage to French / German / Spanish / Italian / Portuguese / Arabic?

Hidden Model Bias

Google Translate

TextDocumentsWebsites

DETECT LANGUAGEENGLISHHEBREWGERMAN

↔GERMANHEBREWENGLISH

The nurse is smart. He found a solution.

Die Krankenschwester ist schlau. Er hat eine Lösung gefunden.

40 / 5,000

[Send feedback](#)

Hidden Model Bias

Google Translate

TextDocumentsWebsites

DETECT LANGUAGEENGLISHHEBREWGERMAN

↔SPANISHGERMANHEBREW

The nurse is smart. He found a solution.

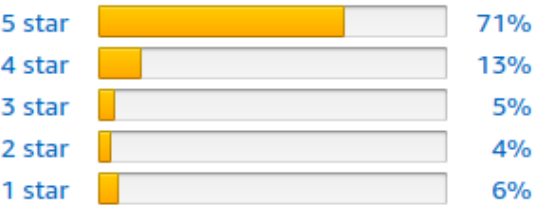
La enfermera es inteligente. Encontró una solución.

[Send feedback](#)

Customer reviews

★★★★☆ 4.4 out of 5

11,242 global ratings



How are ratings calculated?

By feature

Quality of material	★★★★☆	4.5
Battery life	★★★★☆	4.5
Ergonomic	★★★★☆	4.4

See more

Review this product

Share your thoughts with other customers

Write a customer review

Customer images



See all customer images

Read reviews that mention

battery life

stopped working

wireless mouse

great mouse

programmable buttons

thumb buttons

highly recommend

extra buttons

thumb rest

wireless gaming

buttons on the side

ever owned

Top reviews

Top reviews from the United States

Showing 1-8 of 161 reviews with "highly recommend". [Clear filter](#)



Smitty83

★★★★★ Excellent mouse for general use! Can't comment on gaming experience

Reviewed in the United States on December 9, 2019

Verified Purchase

...are all very good. I wish I had bought one of these for the office years ago. Highly recommended for \$25!
G1 - Left Mouse G2 - Right Mouse G3 - Enter G4 - Backwards G5 - Redo G6 - Paste G7 - Forward G8 - Undo
G9 - Copy G10 - DPI Shift (2200 ->... [Read more](#) >

Comments Analysis

NLP Research

- Two Dimensions:
 - Foundational Language Analysis
 - Applied NLP



WIIIFM?

Write down what you want to learn through this training?

Which objective(s) do you want to be able to do?



True NLP is AI-Complete

- Turing Test: is a computer program intelligent? (1954)
 - Would a human find out that he speaks (chats) with a computer?



Related academic fields

Linguistics

Statistical Inference
and Machine
Learning

Psycholinguistics

Computer Science

- Algorithms (e.g., indexing & search, graph-based algorithms)

Cognitive Science



Layers of Linguistic Knowledge

- Phonology: Sounds combination
- Morphology: Word Structure
- Syntax: Sentence Structure
- Semantics: Utterance Meaning
- Pragmatics: Intention / Act

A close-up profile of a woman with dark hair, shouting with her mouth wide open. A cloud of black letters and symbols (A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z, and punctuation) is being expelled from her mouth, illustrating the concept of 'leakage' in the text.



- What sounds go together
- Language differences



Morphology: Word Structure

- Word Structure:

Independently: in | depend | ent | ly

Lebensversicherungsgesellschaftsangestellter



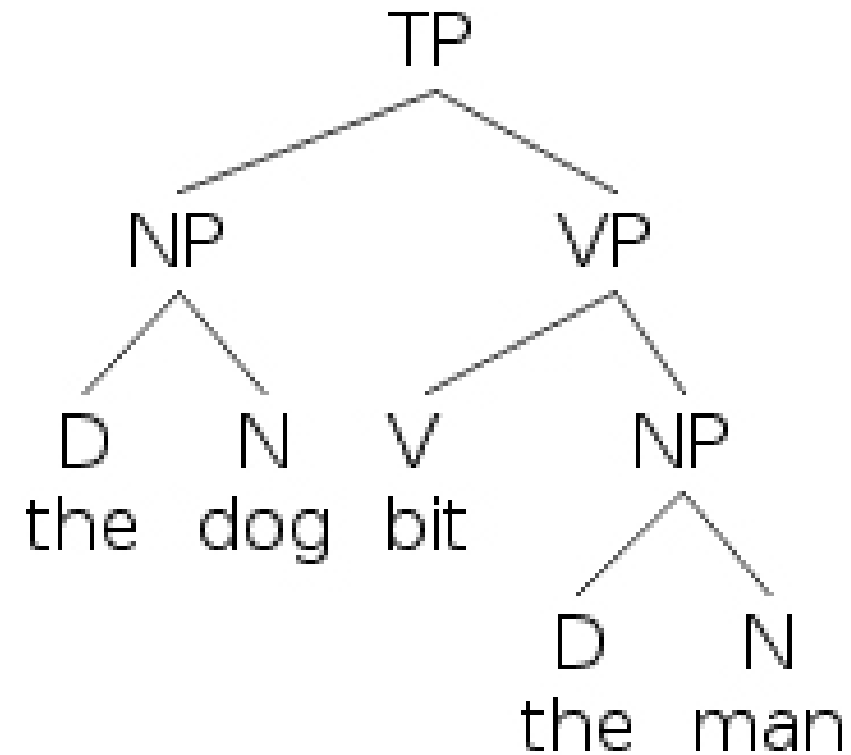
Syntax: Sentence Structure

- Categorization - Part of Speech (POS)
- Phrases (more than one word)
- Sentence Hierarchy

Good for:

- Question Answering
- Entity & Information extraction

The dog bit the man



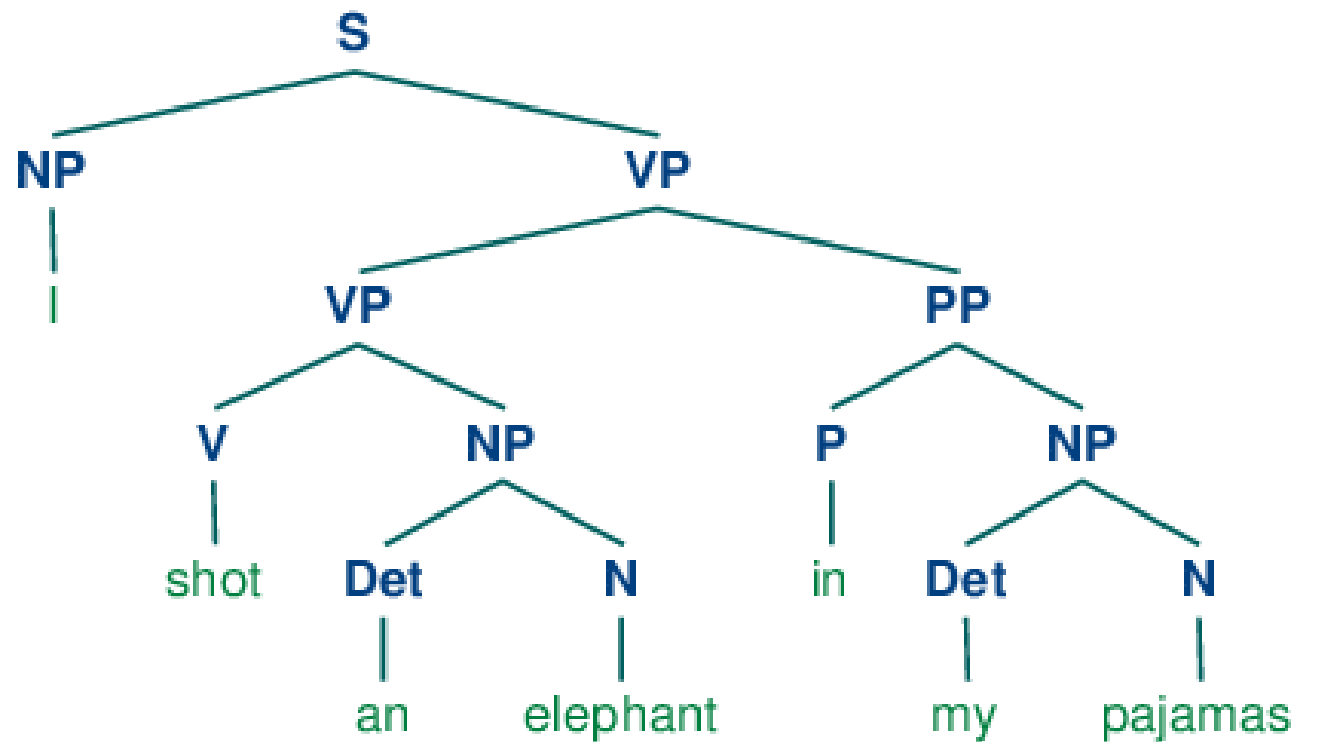
Syntax: Sentence Structure

- Categorization - Part of Speech (POS)
- Phrases (more than one word)
- Sentence Hierarchy

Good for:

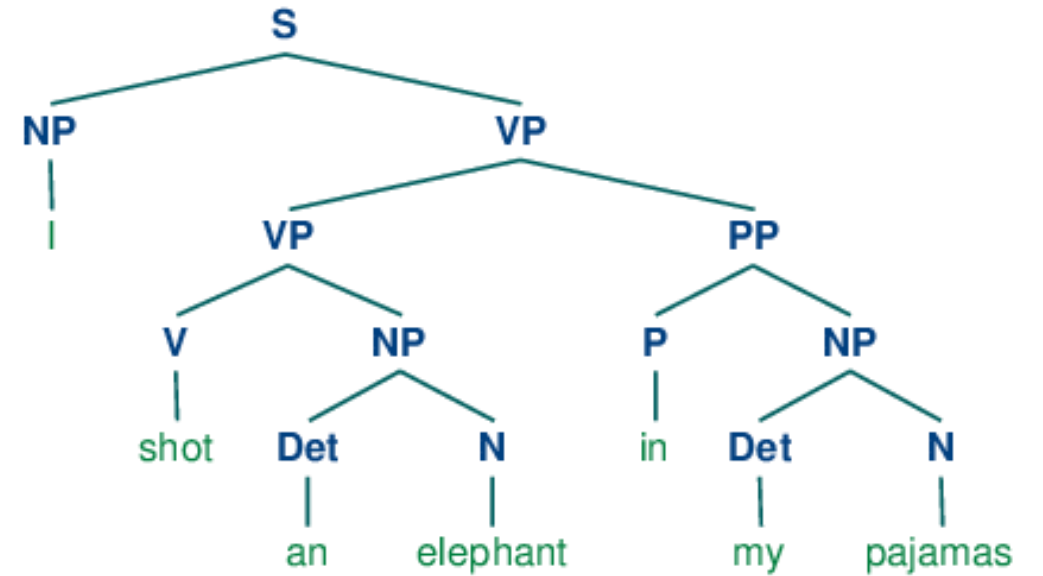
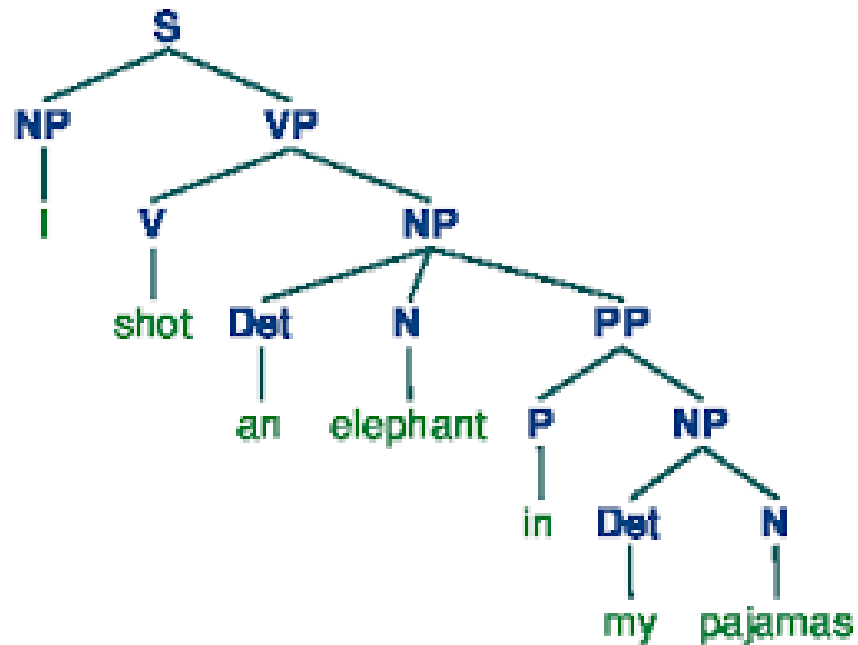
- Question Answering
- Entity & Information extraction

I shot an elephant in my pajamas



Syntax: Sentence Structure

I shot an elephant in my pajamas



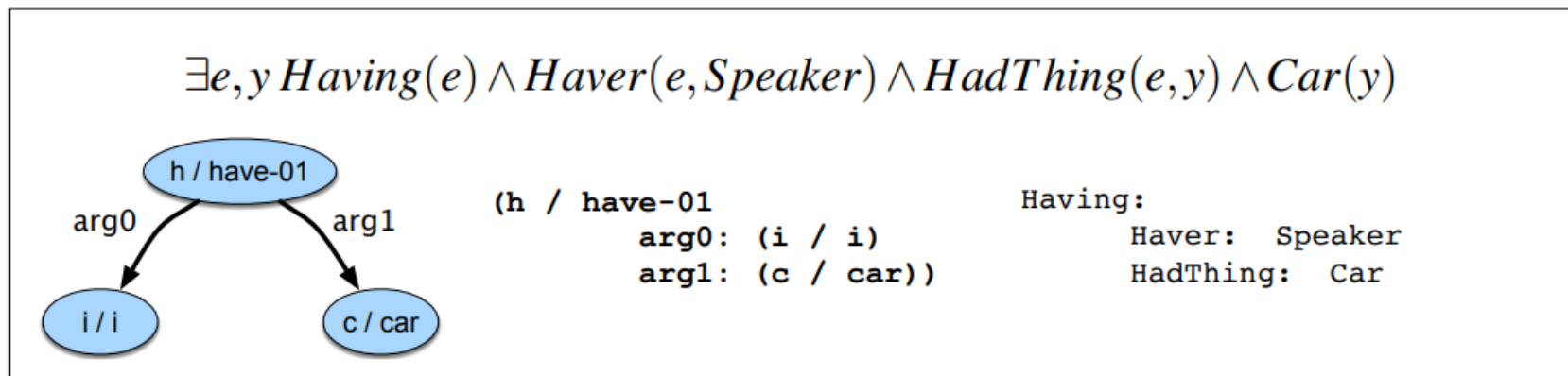


Figure 16.1 A list of symbols, two directed graphs, and a record structure: a sampler of meaning representations for *I have a car*.



Semantics: Meaning

- Situation representation
 - Graph
 - Logics
 - Vector ...

Pragmatics: Intention / Act

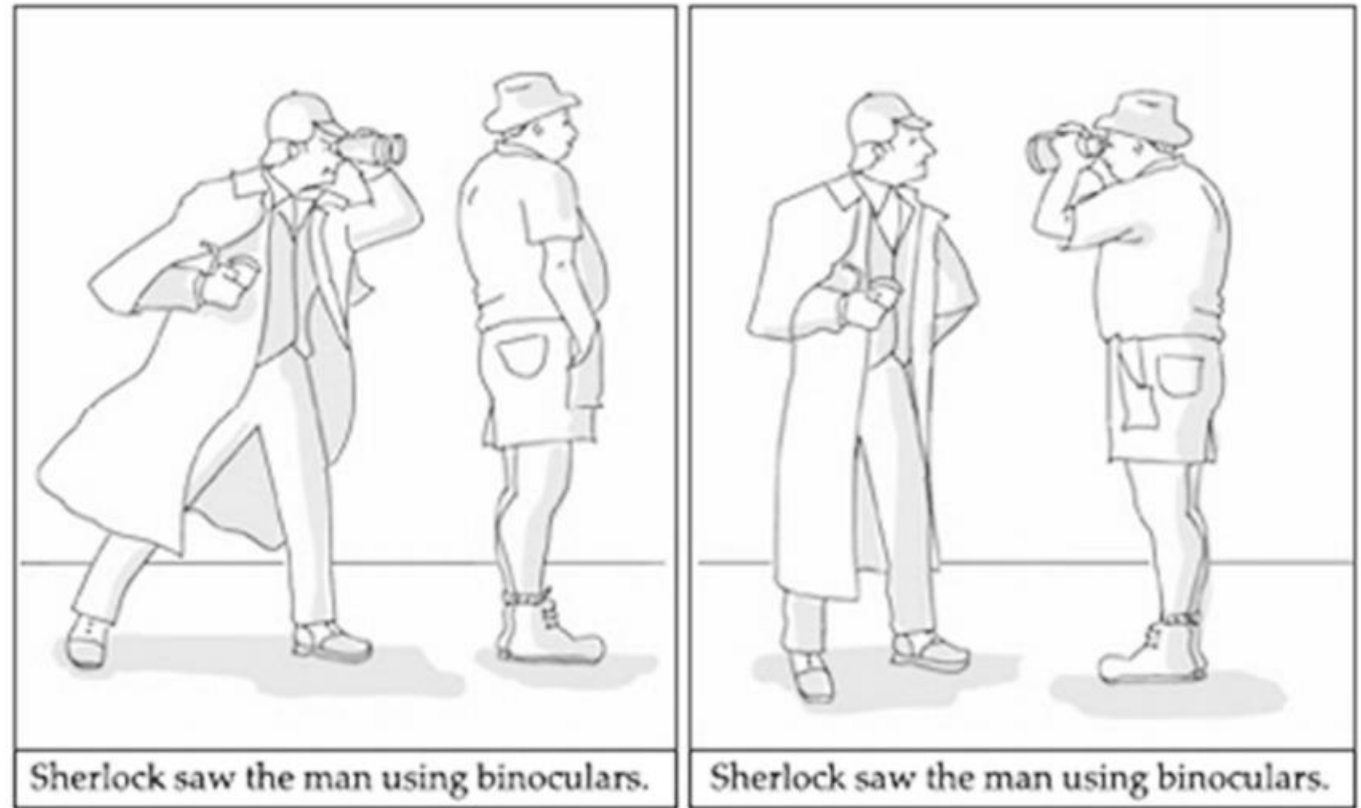
- Aspects beyond the text
- Social relationship / Context
- Time & Place

If red and no traffic, then allowed to go
It's cold in here...



Pragmatics

Sherlock saw the man using binoculars [with binoculars]

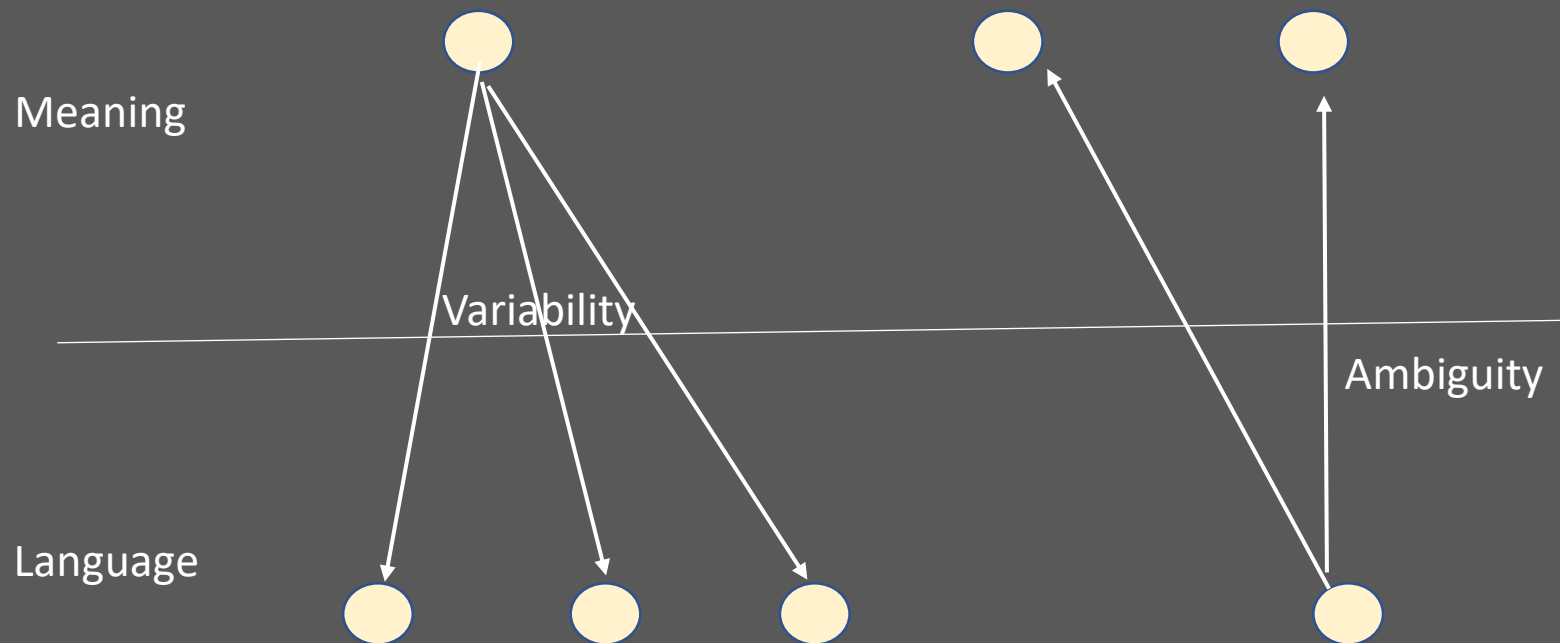


$see^1_{using\ binoculars}$

$see^2_{using\ binoculars}$

Fig. 4 A modification of a predicate on different argument places

Why is NLP hard?



Variability

Semantical / Pragmatical

He acquired it
He purchased it
He bought it
It was bought by him
It was sold to him
She sold it to him
She sold him that

Lexical *

I loooooove it!
everytime → every time
oscar nom'd doc → Oscar-nominated
documentary
is bananas → is great

* Wei Xu:
<https://cocoxu.github.io>



Negativity
and
Restrictivity

Martin is a Data Scientist

Martin is **not** a data Scientist

Martin is a great Data Scientist

Martin is **not** a great data scientist

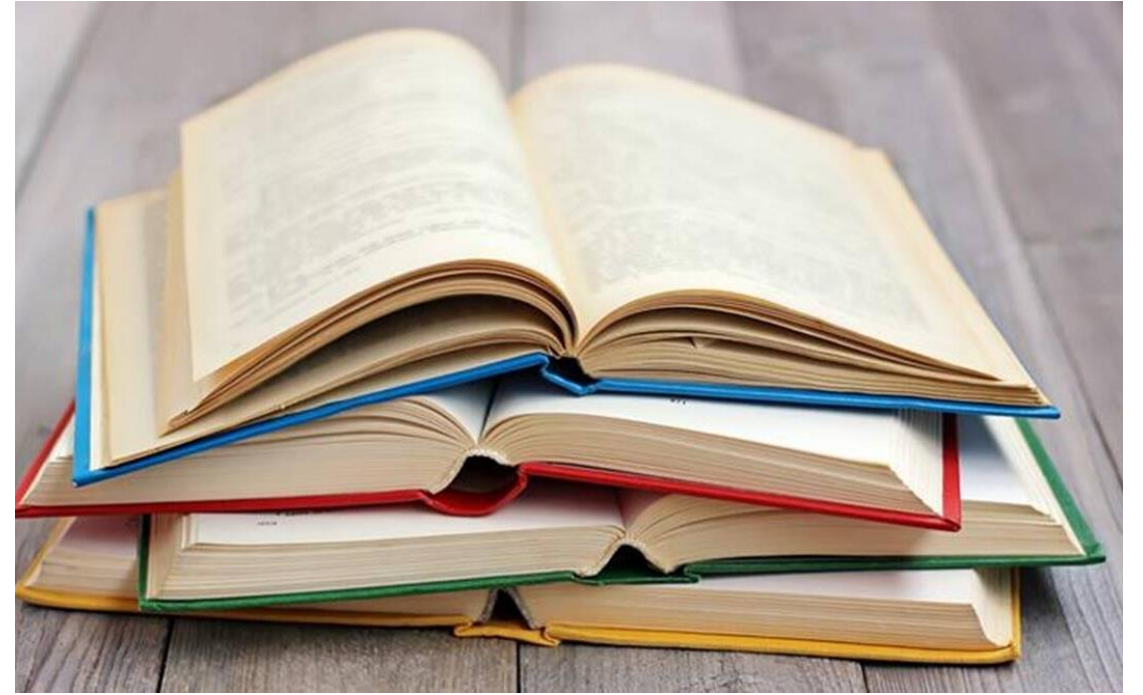
Martin is **not** a Data Scientist

Martin is **not** a Mad Scientist



Ambiguity

- Across all levels of Description
- **Phonology:** Ship // Sheep



Ambiguity

- Across all levels of Description
 - **Morphology:** books, parks
- Verb Phrase // Noun Phrase



Ambiguity

- Across all levels of Description
 - **Morphology (Lexical):** Bank
- Noun // Noun

Ambiguity

- Across all levels of Description
- **Syntactical:**

I ate pizza **with olives**

I ate pizza **with friends**



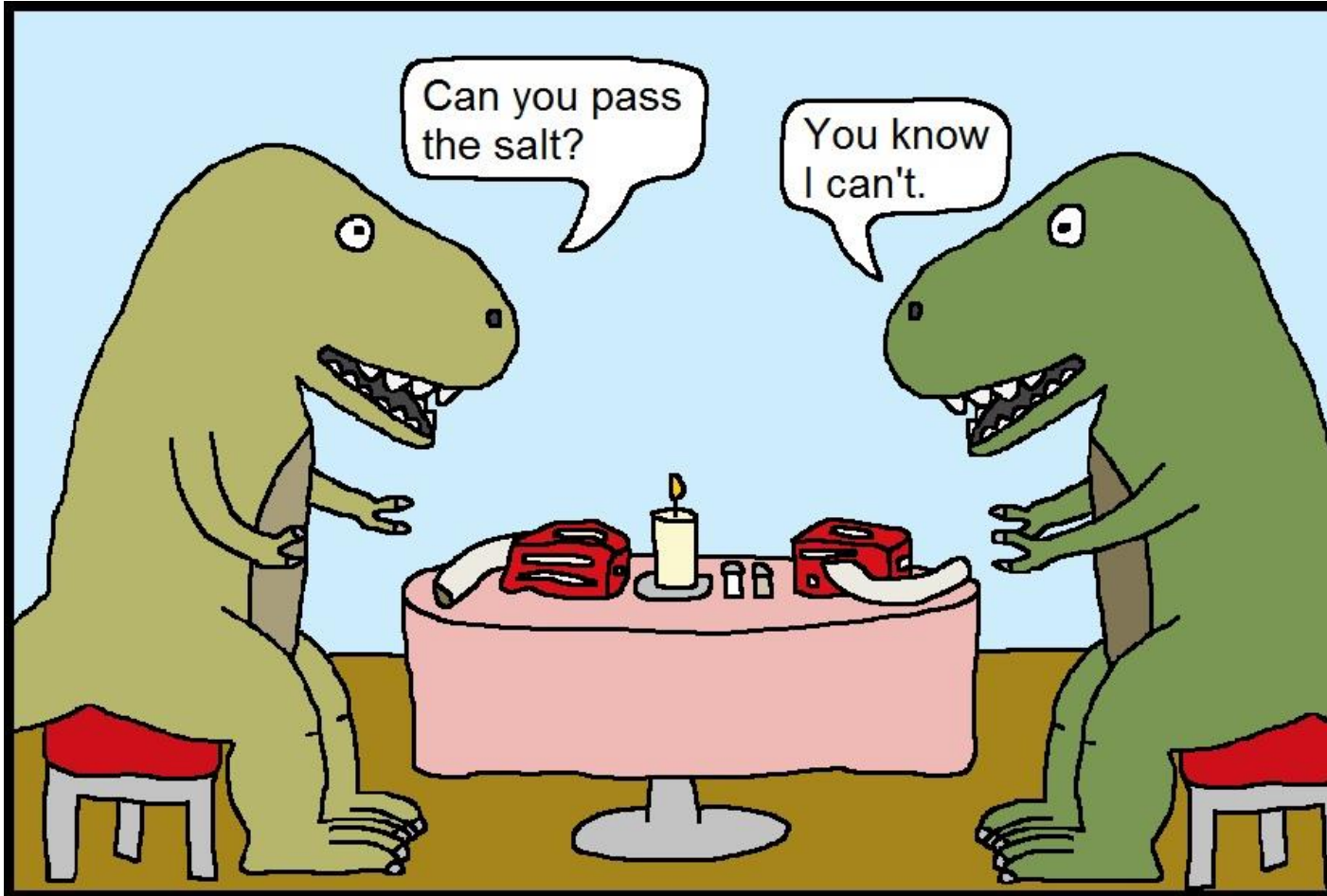
Ambiguity

- Across all levels of Description
- **Semantical:**

The chicken is ready to eat



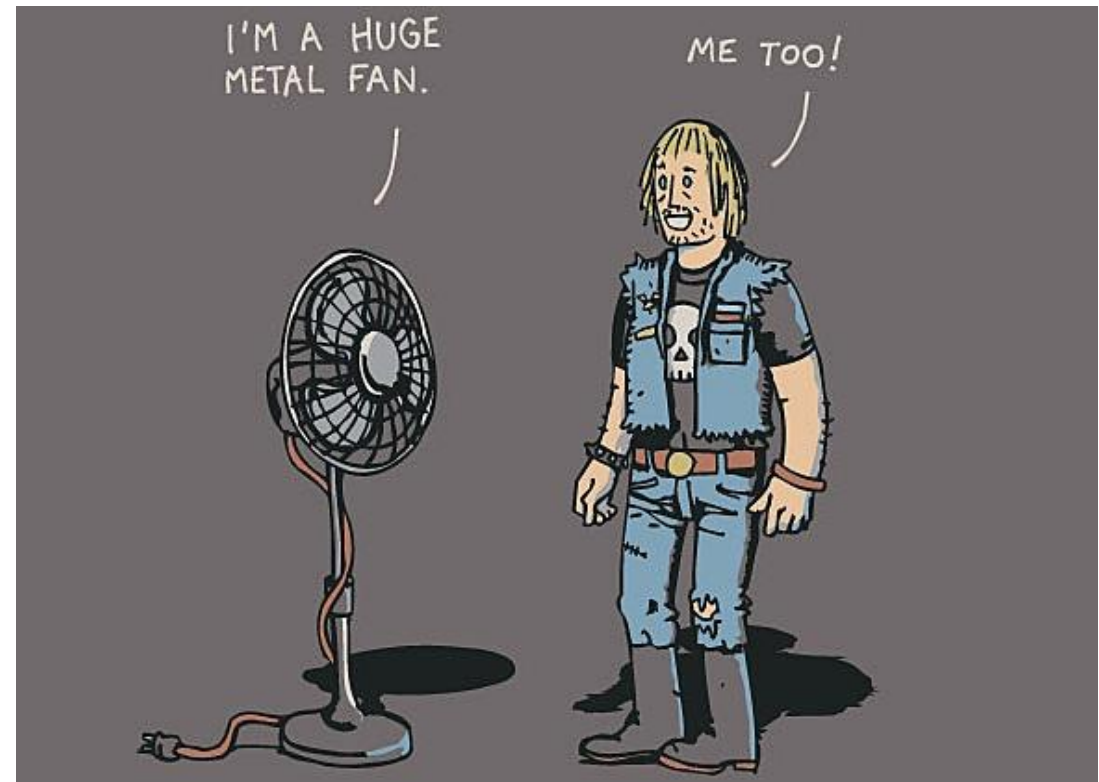
ambiguity



- Across all levels of Description
- **Pragmatics**

ambiguity

- Across all levels of Description
- **Pragmatics**





Extreme Ambiguity (real headlines!)

- Scientists count whales from space
- The pope's baby step on gays
- Girl hit by car in hospital
- Iraqi head seeks arms
- Lung cancer in women mushrooms
- Student get first hand job experience
- Squad helps dog bite victim
- Miners refuse to work after death
- Stolen painting found by tree
- Actor sent to jail for not finishing sentence
- Stadium air conditioning fails — Fans protest

Your turn!

Morphology

- Find at least 2 examples of an **ambiguity** for each of the 5 linguistic levels:
 - Phonology
 - Morphology
 - Syntax
 - Semantics
 - Pragmatics

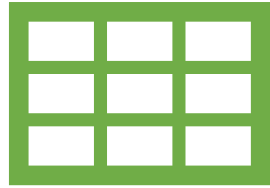


Ambiguity

- On which linguistic level is this example?



NLP – How is it done?



Rule-based modeling

Many rules (with many exceptions)



Statistical Modeling (Data Driven)

Like children who learn by listening and trying
Input sets of 'correct' samples (X, Y)



Learning Goals



Learning Goals

Write

- Write efficient Regular Expressions, to solve (almost) any kind of text-extraction task.

Apply

- Apply Topic Modeling & Text-Distance algorithms to clustering and text-sequence problems.

Train

- Train a spelling / grammar corrector.

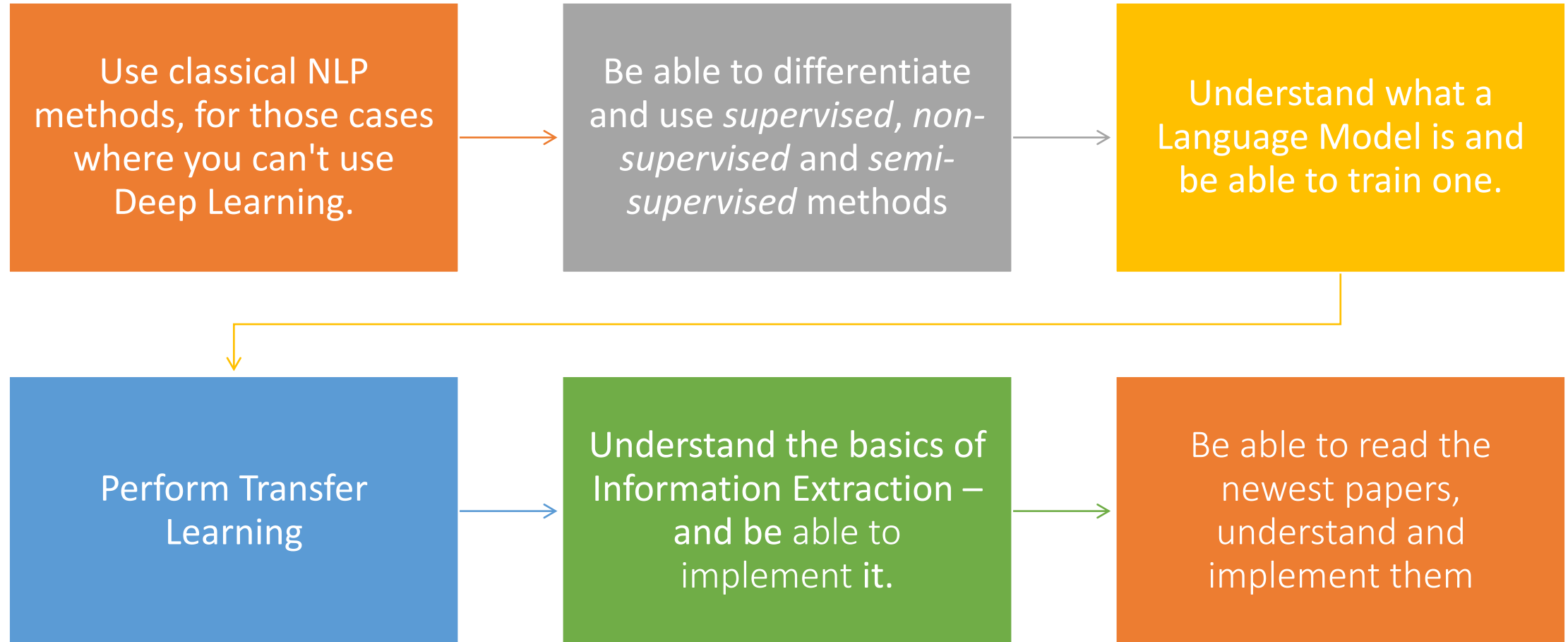
Classify

- Build a supervised classifier to solve problems such as Sentiment Analysis or a Spam-detector.

Build

- Build a recommendation engine.

Learning (hidden) Goals



Learning (hidden) main goal



SET OF TOOLS



ANALYZE TEXT



PREDICTIVE MODELS



Are You Ready?

Let's begin!