

# Information Service Engineering

# Lecture 2: Natural Language Processing - 1



Leibniz Institute for Information Infrastructure

Prof. Dr. Harald Sack

FIZ Karlsruhe - Leibniz Institute for Information Infrastructure

AIFB - Karlsruhe Institute of Technology

Summer Semester 2021

## Last Lecture: Information, Natural Language, and the Web

- 1.1 How to get Information (from the Web)?
- 1.2 Communication, Language, and Understanding
- 1.3 How to measure Information?
- 1.4 The ever-growing Web of Information
- 1.5 Search Engines on the Web
- 1.6 The Meaning of Information

## Lecture 2: Natural Language Processing (1)

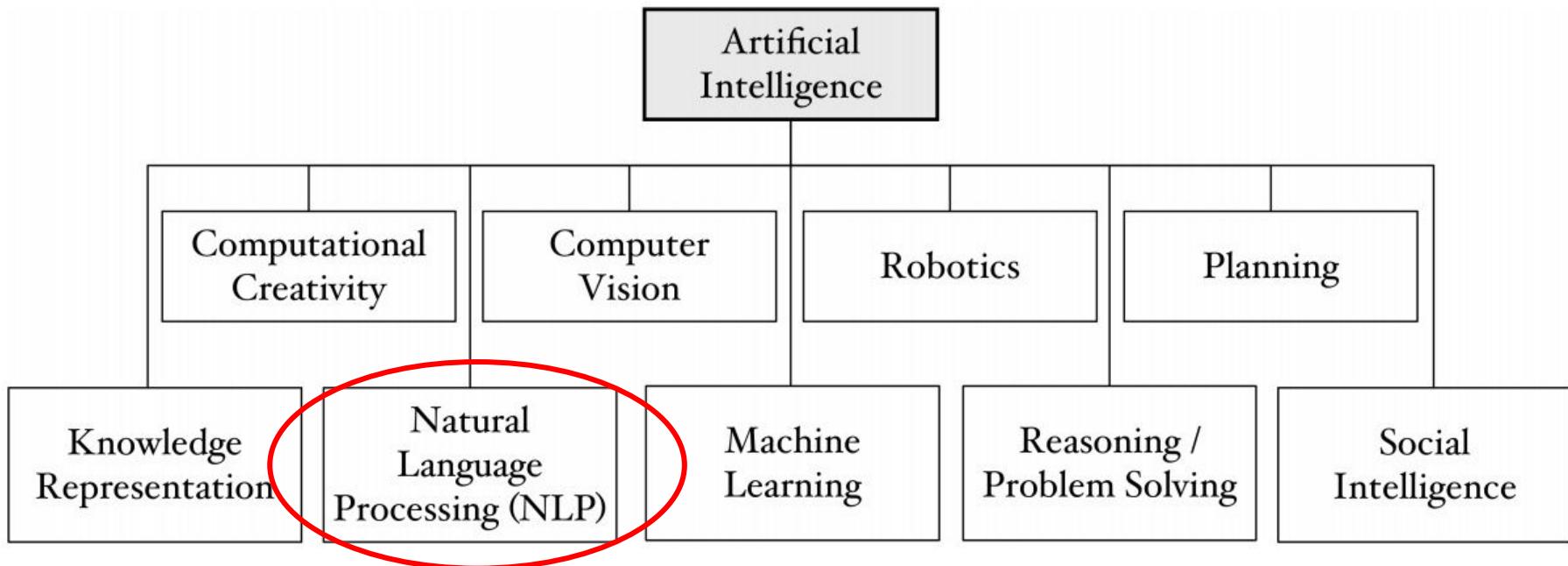
### 2.0 What is Natural Language Processing?

- 2.1 NLP and Basic Linguistic Knowledge
- 2.2 Morphology
- 2.3 NLP Applications
- 2.4 NLP Techniques
- 2.5 NLP Challenges
- 2.6 Evaluation, Precision and Recall
- 2.7 Regular Expressions
- 2.8 Finite State Automata
- 2.9 Tokenization
- 2.10 Language Model and N-Grams
- 2.11 Part-of-Speech Tagging
- 2.12 Word Embeddings

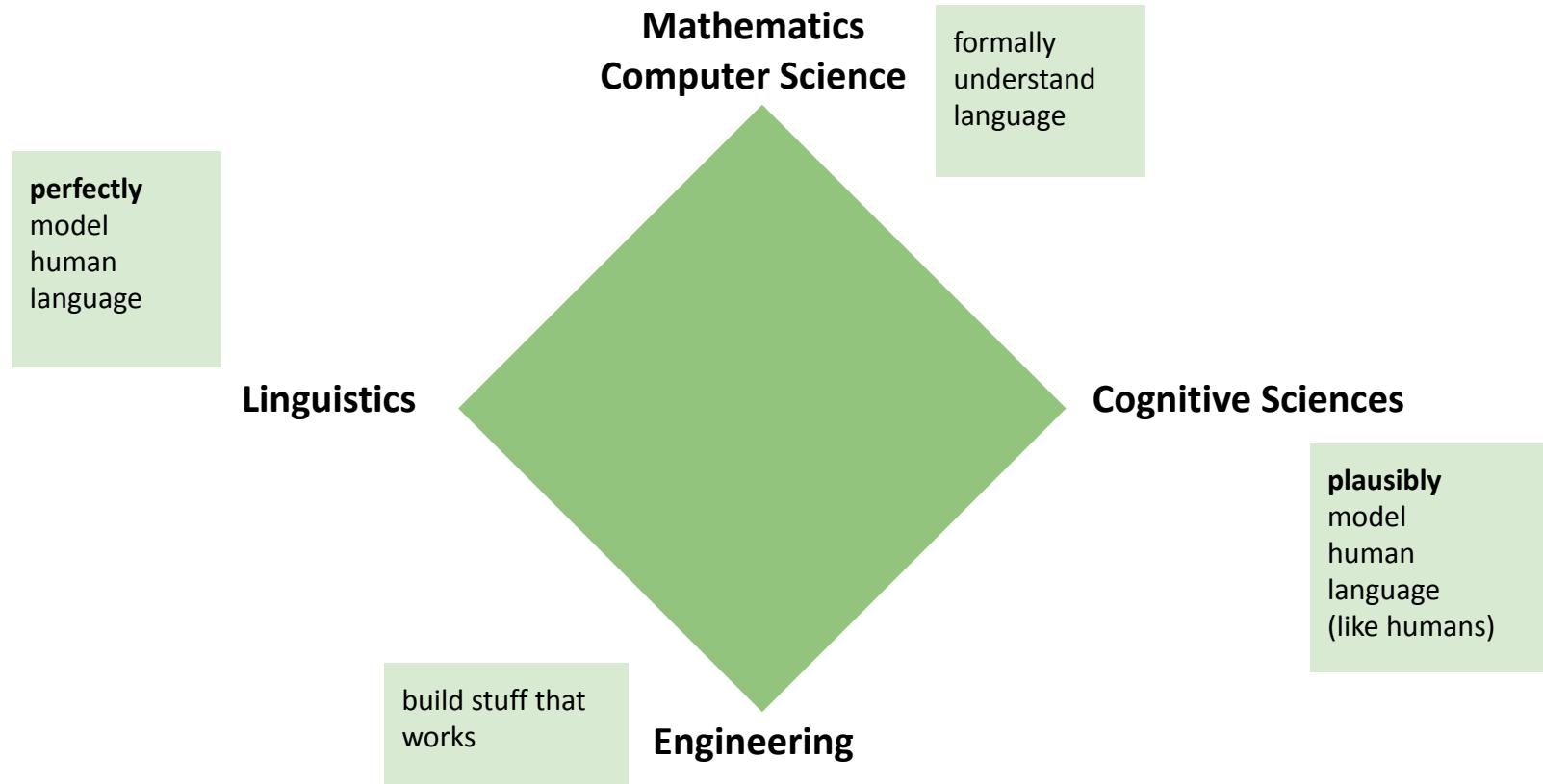
# Natural Language Processing (NLP)

- is a field of computer science, artificial intelligence, and computational linguistics and
- is concerned with the **interactions between computers and human (natural) languages** and, in particular,
- is concerned with **programming computers to fruitfully process large natural language corpora**.
- Specifically the task **to extract meaningful information from natural language input** or **to produce natural language output**.

# Positioning of NLP in Artificial Intelligence



# NLP is an Interdisciplinary Science

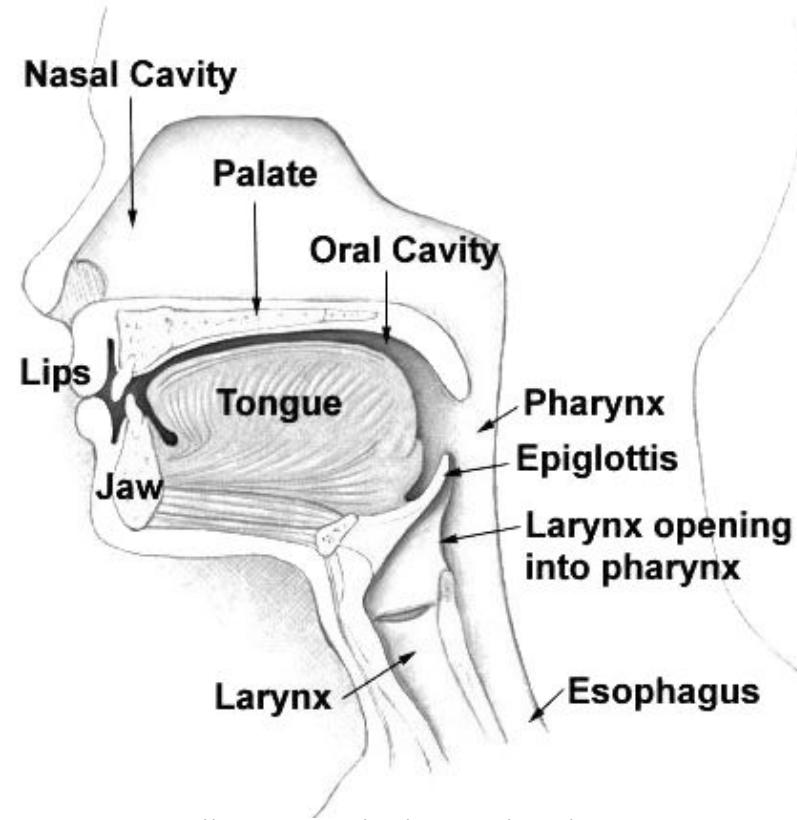


## Natural Languages

- are mainly built on three different knowledge components:
  - **Phonology:** the **sound** of words
  - **Semantics:** the **meaning** of words
  - **Syntax:** the **grammatical rules** according to which words are put together

# Phonetics vs Phonology

- **Phonetics** is a branch of linguistics that comprises
  - the **study of the sounds of (all) human speech**
  - and is concerned with the **physical properties of speech sounds**, i.e.
    - their physiological production,
    - acoustic properties,
    - auditory perception, and
    - neurophysiological status.

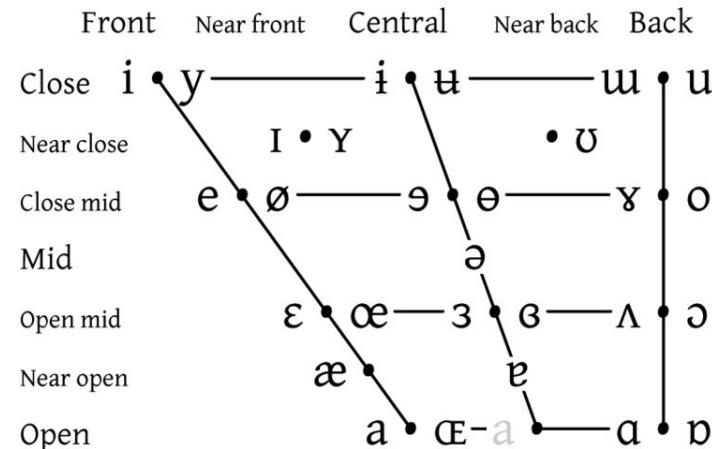


[https://en.wikipedia.org/wiki/Language#/media/File:Illu01\\_head\\_neck.jpg](https://en.wikipedia.org/wiki/Language#/media/File:Illu01_head_neck.jpg)

# Phonetics vs Phonology

- **Phonology** is defined as
  - **the study of speech sounds of a language or languages,**
  - **and the laws governing them,**
  - particularly the laws governing the **composition and combination of speech sounds in language.**

## VOWELS



Vowels at right & left of bullets are rounded & unrounded.

## Lecture 2: Natural Language Processing (1)

2.0 What is Natural Language Processing?

### **2.1 NLP and Basic Linguistic Knowledge**

2.2 Morphology

2.3 NLP Applications

2.4 NLP Techniques

2.5 NLP Challenges

2.6 Evaluation, Precision and Recall

2.7 Regular Expressions

2.8 Finite State Automata

2.9 Tokenization

2.10 Language Model and N-Grams

2.11 Part-of-Speech Tagging

2.12 Word Embeddings

# Phonology Basics

## Phone

- **Any distinct speech sound**, regardless of whether the exact sound is critical to the meanings of words

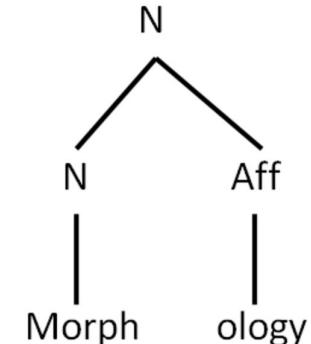
## Phoneme

- **Smallest (abstract cognitive) sound unit** in a language that is able of **conveying a distinct meaning**
- **Example:**
  - “s” and “r” in “sing” and “ring”
  - “ss” and “ll” in “kiss” and “kill”

# Morphology

- Morphology is the study of **internal structures (formation) of words** and how they can be modified.
- Morphology determines how to **parse complex words into their components**.

⇒ **What is a word?**

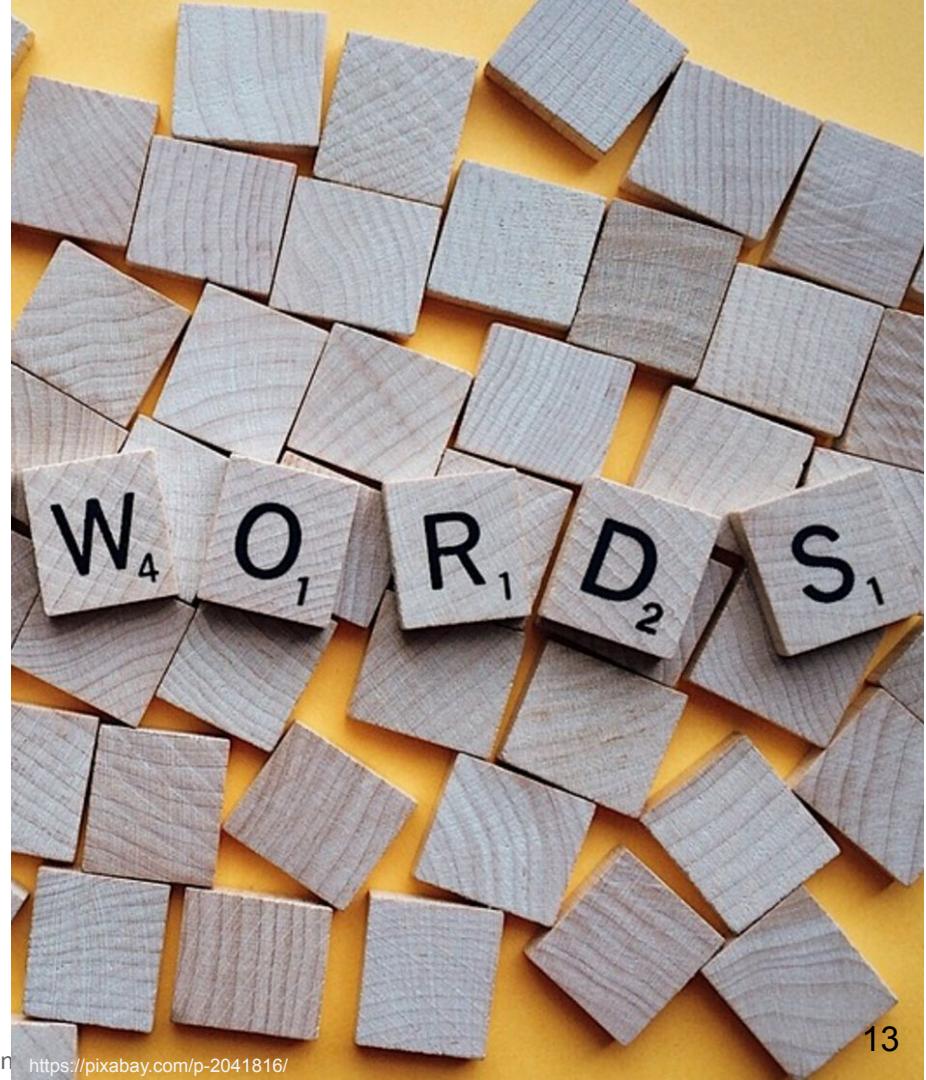


# Linguistic Basics

A **word** ( $w_i$ ) is the smallest **independent** unit of language.

**“Independent”?**

- do not depend on other words
- can be separated from other units
- can change position

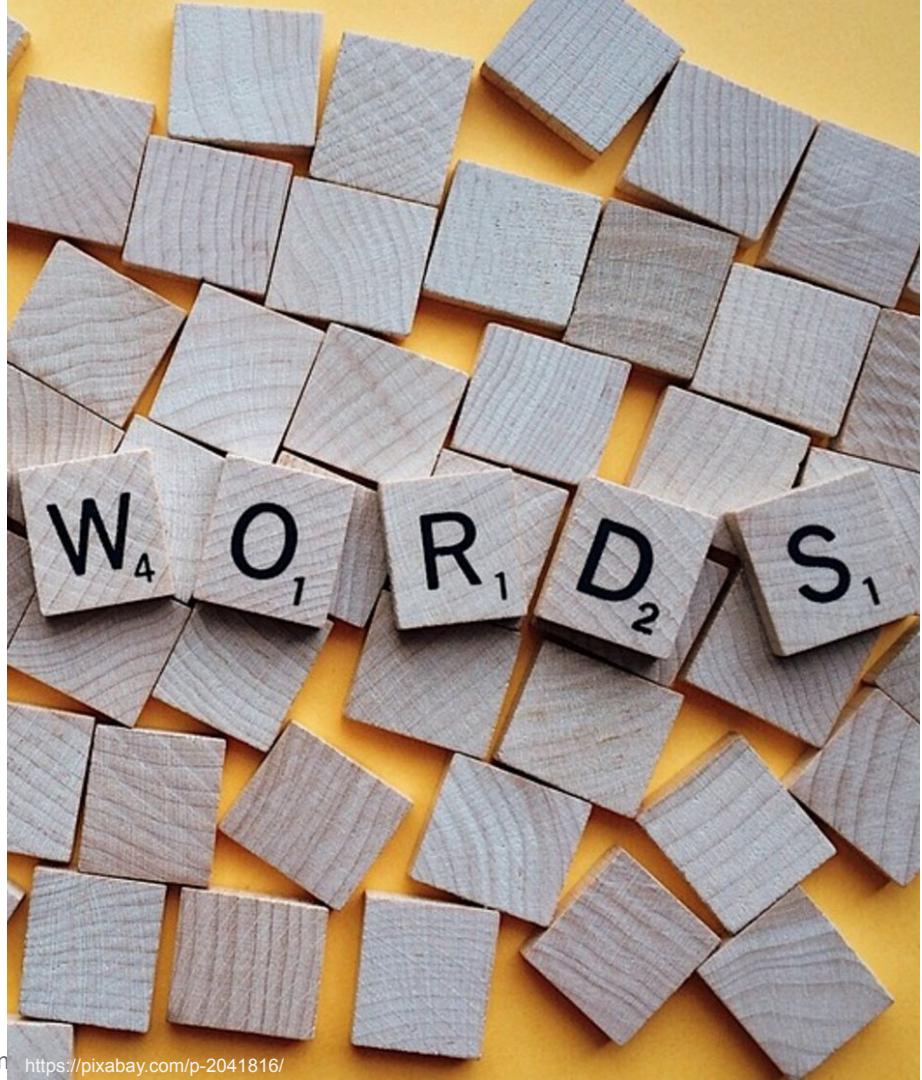


# Linguistic Basics

- Example:

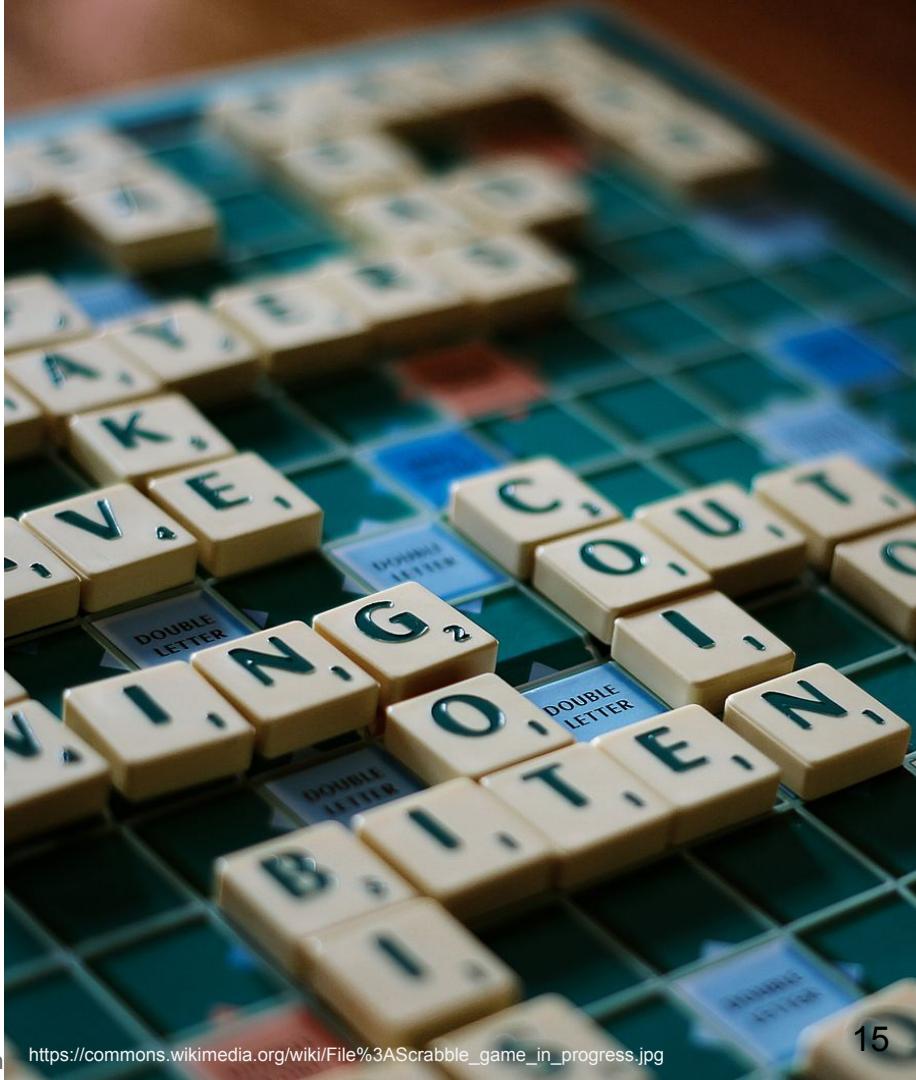
*The man looked at the horses.*

- *s* is the plural marker, dependent on the noun **horse** to receive meaning.
- **Horses** is a word: it can occur in other positions or stand on its own.



# Linguistic Basics

A **vocabulary** consists of a set of **words** ( $w_i$ ).



# Linguistic Basics

A **text** is composed of a sequence of **words** from a **vocabulary**.

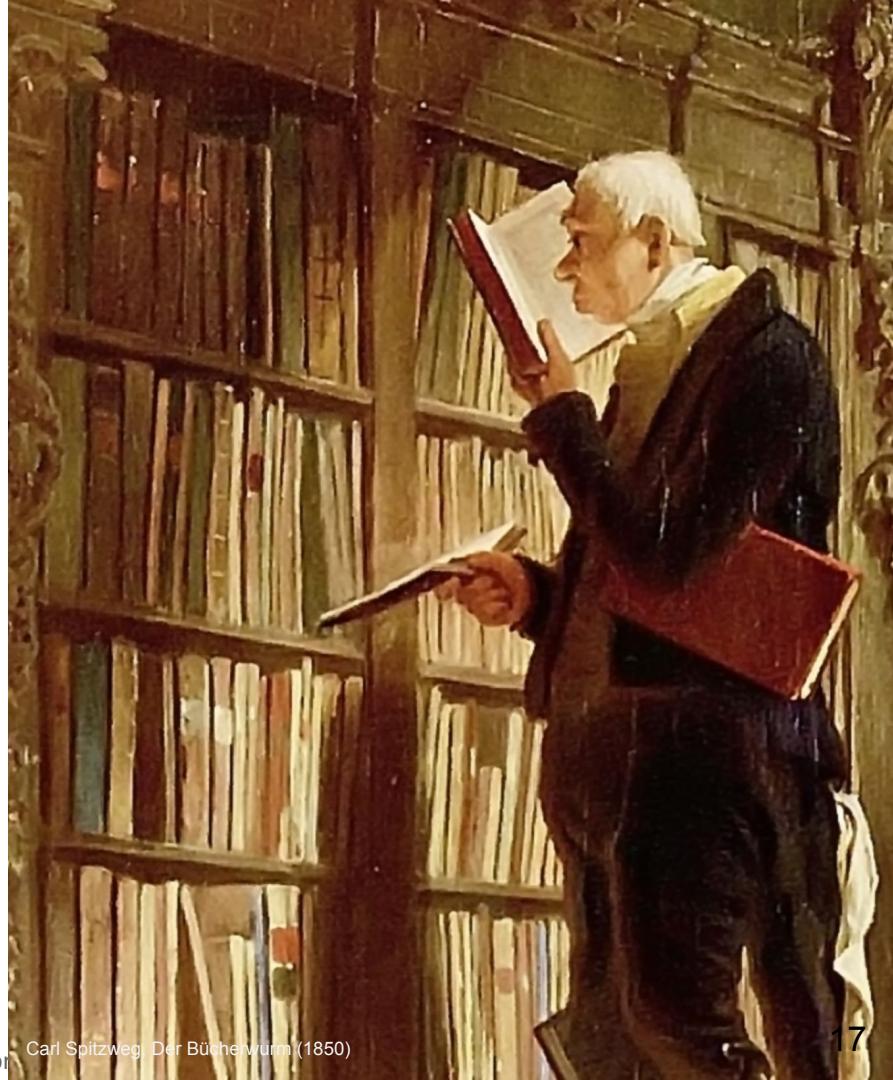
## SHAKESPEARES, SONNETS.

From fairest creatures we desire increase,  
 That thereby beauties Rose might neuer die,  
 But as the riper shoulde by time decease,  
 His tender heire might beare his memory:  
 But thou contracted to thine owne bright eyes,  
 Feed'st thy lights flame with selfe substantiall fewell,  
 Making a famine where abundance lies,  
 Thy selfe thy foe, to thy sweet selfe too cruell:  
 Thou that art now the worlds fresh ornament,  
 And only herald to the gaudy spring,  
 Within thine owne bud buriesth thy content,  
 And tender chorle makst wast in niggarding:  
 Pitty the world, or else this glutton be,  
 To eate the worlds due, by the graue and thee.

VVhen fortie Winters shall besiege thy brow,  
 And digge deep trenches in thy beauties field,  
 Thy youthes proud liuery so gaz'd on now,  
 Wil be a totter'd weed of small worth held:  
 Then being askt, where all thy beautie lies,  
 Where all the treasure of thy lusty daies;  
 To say within thine owne deepe funken eyes,  
 Were an all-eating shame, and thirstlesse praise.  
 How much more praise deseru'd thy beauties vse,  
 If thou couldst answere this faire child of mine  
 Shall sum my count, and make my old excuse  
 Proouing his beautie by succession thine.

# Linguistic Basics

A **language** is  
constructed of a  
**set of all possible texts.**



Carl Spitzweg, Der Bücherwurm (1850)

# Linguistic Basics

A **discourse** is  
 the study of linguistic units  
**larger than a single statement**  
 (i.e. a coherent sequences of  
 sentences).

DISCOURS  
 DE LA METHODE  
 Pour bien conduire sa raison,& chercher  
 la vérité dans les sciences.  
 PLUS  
 LA DIOPTRIQUE.  
 LES METEORES.  
 ET  
 LA GEOMETRIE.  
*Qui sont des essais de cette METHODE.*



A LEYDE  
 De l'Imprimerie de IAN MAIRE.  
 C I O C X X X V I I .  
*Avec Privilège.*

## Lecture 2: Natural Language Processing (1)

2.0 What is Natural Language Processing?

2.1 NLP and Basic Linguistic Knowledge

### **2.2 Morphology**

2.3 NLP Applications

2.4 NLP Techniques

2.5 NLP Challenges

2.6 Evaluation, Precision and Recall

2.7 Regular Expressions

2.8 Finite State Automata

2.9 Tokenization

2.10 Language Model and N-Grams

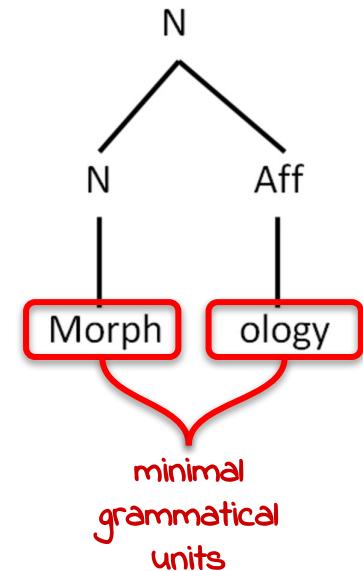
2.11 Part-of-Speech Tagging

2.12 Word Embeddings

# Morphology

## Morphemes

- The **smallest grammatical unit** in a language,  
i.e. the smallest meaningful unit of a language
- We distinguish:
  - **Simple words:** consist of a single morpheme  
E.g. *work, build, run*, etc.
  - **Complex words:** have internal structure,  
i.e. consist of 2 or more morphemes  
E.g. ***morphology***, affix ***-ology*** added to root ***morph***



# Morphology

## Morphemes

- The smallest meaningful units of language
- We distinguish:
  - **Simple words:** consist of a single morpheme  
E.g. *work, build, run*, etc.
  - **Complex words:** have internal structure,  
i.e. consist of 2 or more morphemes  
E.g. *morphology*, **affix**-*ology* added to **root** *morph*

A bound morpheme that is part of a complex word but does not belong to any lexical category (i.e., is not a verb, a noun, an adjective)

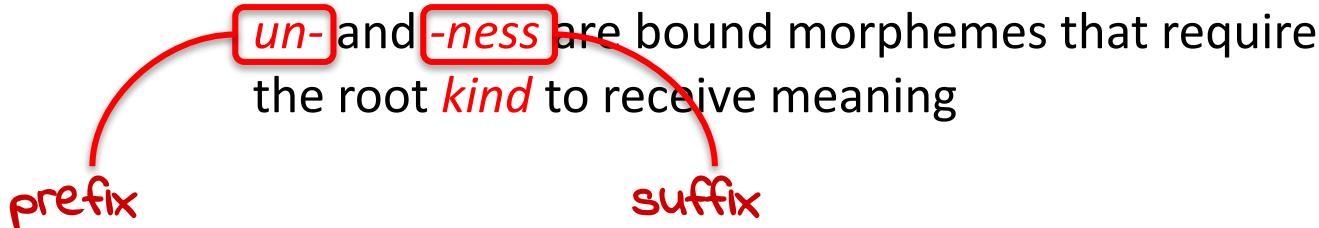
i.e. the smallest

Core part of a complex word, the part that carries the major component of its meaning

# Morphology

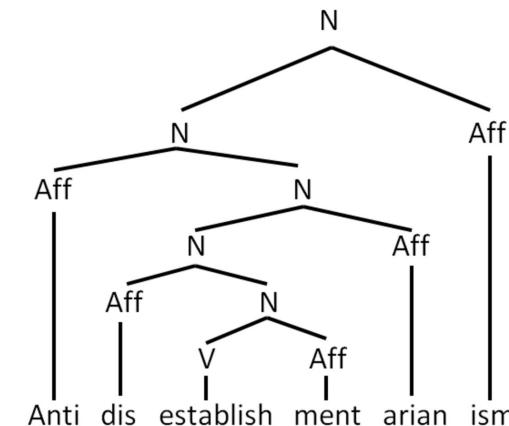
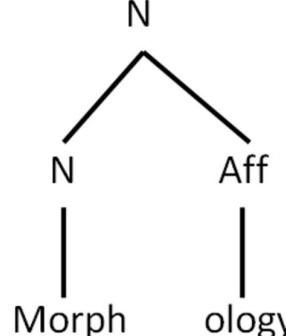
## Free vs. Bound Morphemes

- **Free morpheme:** a simple word, consisting of one morpheme,
  - e.g. *house, work, high, chair, wrap*
- **Bound morpheme:** morphemes that must be attached to another morpheme to receive meaning,
  - e.g. *unkindness*



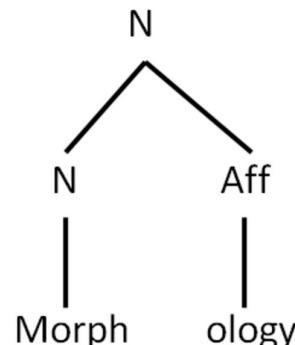
# Morphological Parsing

- = the process of determining the morphemes (and their purpose) from which a given word is constructed.
- Can be visualized in a tree diagram (**morphology tree**).

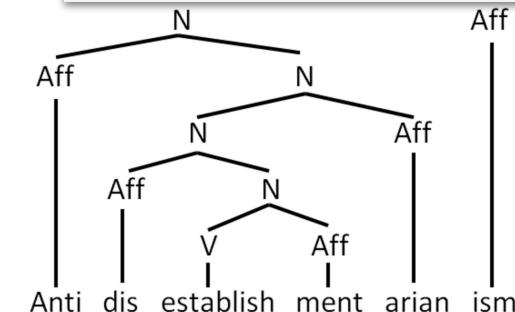


# Morphological Parsing

- In many languages, words can be made up of a main **stem** (carrying the basic dictionary meaning) plus one or more **affixes** carrying grammatical information.
- Surface form:** morphology
- Lexical form:** morph+N+Aff

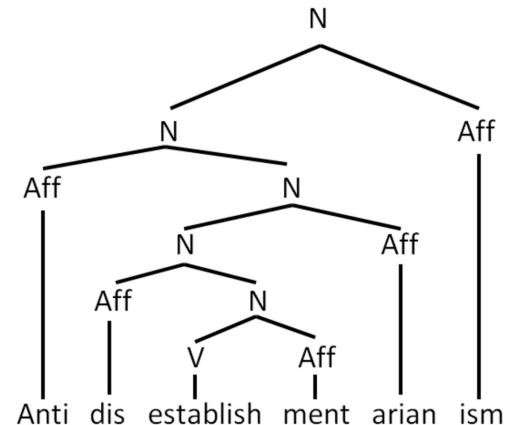


**Morphological Parsing** is the problem of extracting the lexical form from the surface form.



# Morphological Parsing - Applications

- Any NLP tasks involving **grammatical parsing** will typically involve morphology parsing as a prerequisite.
- **Search engines**: e.g. a search for 'fox' should return documents containing 'foxes', and vice versa.
- Even a humble task like **spell checking** can benefit:  
e.g. is "*morphology*" a possible word form?



# Morphological Rules

- Many languages build more complex words out of morphemes via
  - **Derivation** (with derivational morphemes)
  - **Compound**
  - **Inflection** (with inflectional morphemes)

# Derivation

- The process of forming a new word from an existing word by **adding derivational morphemes as affixes**.
- The meaning of the resulting word is different from that of its root.
- Very often there is a change in **word category** involved.

- **Example**

- *teach*

**stem**

**verb**

*-er*

**affix**

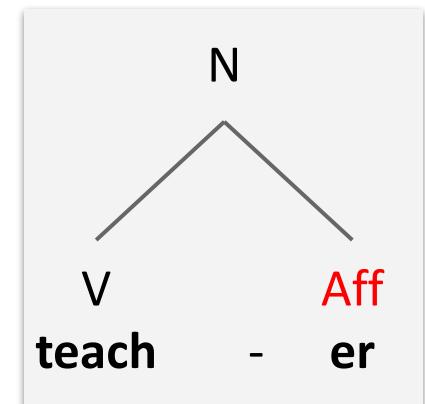
**derivational  
morpheme**



*teacher*

**resulting word**

**noun**

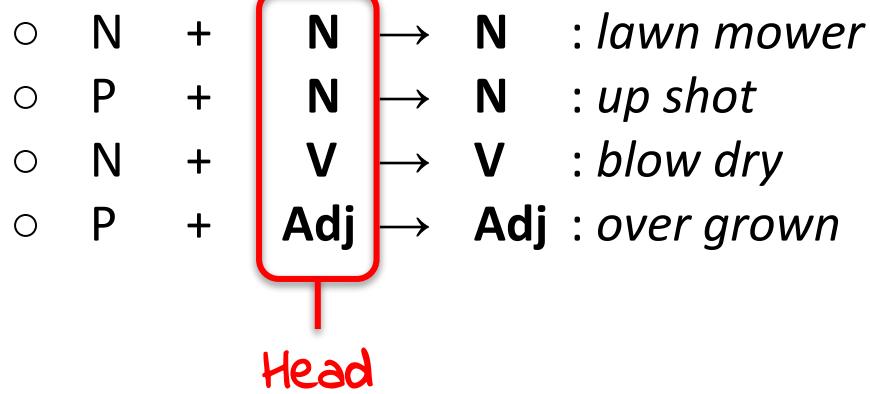


# Compounding

- **Combination of already existing words** into a new one
- There is no affixation but each of the parts can be assigned to a specific **word category**
- Examples:
  - N + N → N : *lawn mower*
  - P + N → N : *up shot*
  - N + V → V : *blow dry*
  - P + Adj → Adj : *over grown*

# Compounding

- Combination of already existing words into a new one
- There is no affixation but each of the parts can be assigned to a specific word category
- Examples:



# Inflection

- Modification of a word to **indicate aspects of the grammatical function of a word** such as **tense, case, voice, aspect, person, number, gender, and mood.**
- In **English, inflection** is predominantly expressed via affixation with **inflectional morphemes**.

# Inflection

## English has only eight inflectional morphemes:

- noun plural {-s} *He has three desserts.*
- noun possessive {-s'} *This is Betty's dessert.*
- verb present tense {-s} *Bill usually eats dessert.*
- verb past tense {-ed} *He baked the dessert yesterday.*
- verb past participle {-en} *He has always eaten dessert.*
- verb present participle {-ing} *He is eating the dessert now.*
- adjective comparative {-er} *His dessert is larger than mine.*
- adjective superlative {-est} *Her dessert is the largest.*

# Morphology

## Inflection vs. Derivation

- **Derivation** often changes the category of the root,  
**inflection** never does that.
- **Derivation** changes the meaning of the root,  
**inflection** does not.
- **Derivation** applies before **inflection**.

# Morphology

## Inflection vs. Derivation

1. The farmer's cows escaped.
2. It was raining.
3. Those socks are inexpensive.
4. Jim needs the newer copy.
5. The strongest rower continued.
6. The pitbull has bitten the cyclist.
7. She quickly closed the book.
8. The alphabet-iz-ation went well.

# Stemming vs. Lemmatization

- **Stemming**

The process of reducing inflected or sometimes derived words to their **word stem**

- Example: **cats** → **cat**

morphological parse of cats **cat + N + PL**

- **Lemmatization**

The process of grouping together the inflected forms of a word so they can be analyzed as a single item, identified by the word's **lemma**, or **dictionary form**

- Example: **better** → **good**

**surface form**

morphological parse of better **good + Adj + Com**

**lexical form**

# Basic Morphology - Summary

- Morphemes - free vs. bound (affixes, suffixes, prefixes)
- Morphological parsing
  - Surface form and lexical form
- Morphological rules
  - Inflection,
  - Derivation, and
  - Compounding
- Stemming vs. Lemmatization

## Lecture 2: Natural Language Processing (1)

2.0 What is Natural Language Processing?

2.1 NLP and Basic Linguistic Knowledge

2.2 Morphology

### 2.3 NLP Applications

2.4 NLP Techniques

2.5 NLP Challenges

2.6 Evaluation, Precision and Recall

2.7 Regular Expressions

2.8 Finite State Automata

2.9 Tokenization

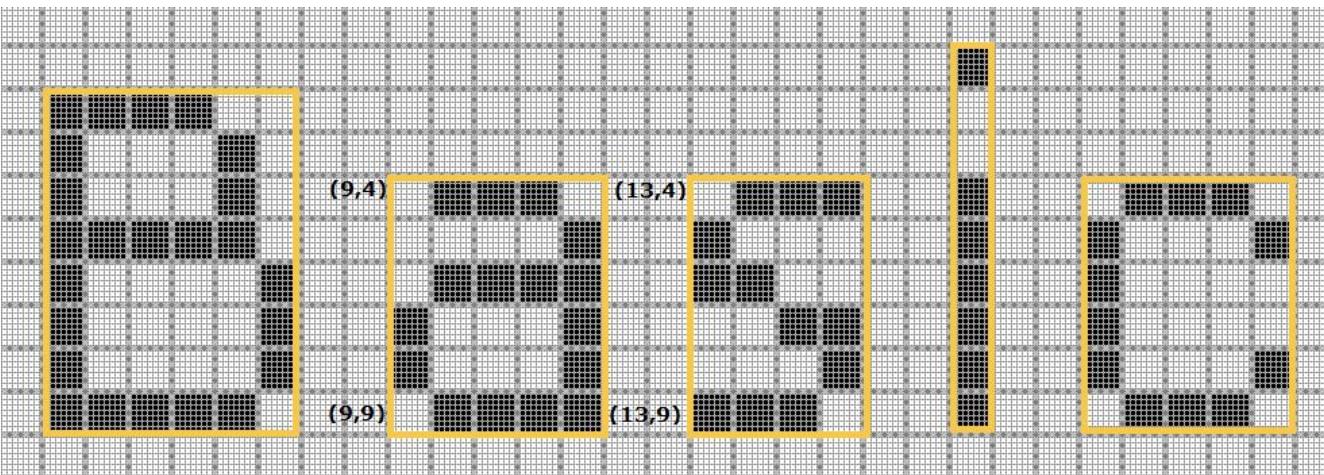
2.10 Language Model and N-Grams

2.11 Part-of-Speech Tagging

2.12 Word Embeddings

# Optical Character Recognition

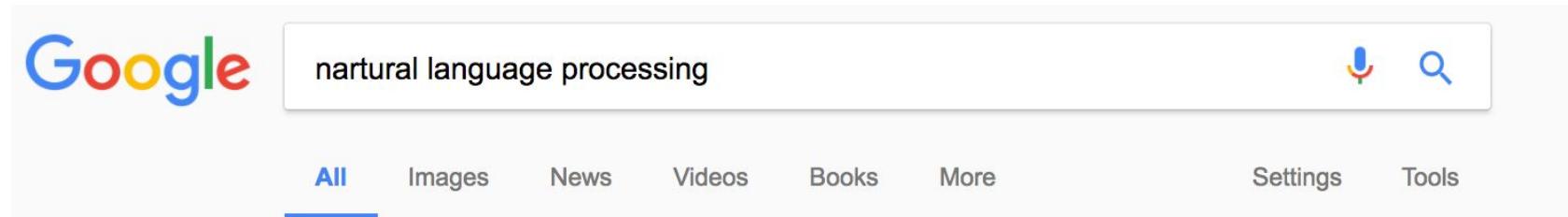
- Conversion of **images** of typed, handwritten or printed text into machine-encoded **text**
- In combination with computer vision, pattern recognition, and artificial intelligence



<https://commons.wikimedia.org/wiki/File:BasicBoundary.png>

# Spelling and Grammar Checking

- Checking spelling and grammar
- Suggesting alternatives for the errors



Showing results for **natural** language processing  
Search instead for nartural language processing

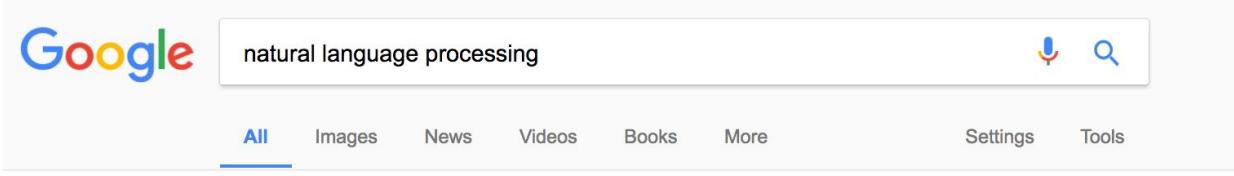
# Word Prediction

- Predicting the next word that is highly probable to be typed by the user



# Information Retrieval

- Finding relevant information according to the user's information need expressed in a query



About 9.090.000 results (0,55 seconds)

**Natural language processing - Wikipedia**  
[https://en.wikipedia.org/wiki/Natural\\_language\\_processing](https://en.wikipedia.org/wiki/Natural_language_processing) ▾  
Natural language processing (NLP) is a field of computer science, artificial intelligence, and computational linguistics concerned with the interactions between computers and human (natural) languages and, in particular, concerned with programming computers to fruitfully process large natural language corpora.  
Natural language understanding · Corpus linguistics · Computational linguistics

**What is natural language processing (NLP)? - Definition from WhatIs ...**  
[searchcontentmanagement.techtarget.com](https://searchcontentmanagement.techtarget.com/definition/natural-language-processing) › Text analytics & NLP › Programming ▾  
Natural language processing (NLP) is the ability of a computer program to understand human speech as it is spoken. NLP is a component of artificial intelligence (AI). ... Most of the research being done on natural language processing revolves around search, especially enterprise search ...

**Natural Language Processing - Research at Google**  
<https://research.google.com/pubs/NaturalLanguageProcessing.html> ▾  
by CAED Parsing - 2015 - Cited by 2 - Related articles  
Natural Language Processing (NLP) research at Google focuses on algorithms ... Proceedings of the ACL Workshop on Statistical NLP and Weighted Automata ...

**Artificial Intelligence Natural Language Processing - TutorialsPoint**  
[https://www.tutorialspoint.com/.../artificial\\_intelligence\\_natural\\_language\\_processing....](https://www.tutorialspoint.com/.../artificial_intelligence_natural_language_processing....) ▾  
Natural Language Processing (NLP) refers to AI method of communicating with an intelligent systems using a natural language such as English.

# Text Categorization

- Assigning one (or more) pre-defined category to a text



[https://upload.wikimedia.org/wikipedia/commons/0/09/Spam\\_can.png](https://upload.wikimedia.org/wikipedia/commons/0/09/Spam_can.png)

# Text Categorization

<https://www.fakenewsai.com/>



<https://www.whitehouse.gov/article> 

FAKE!

This site is probably not a reliable news source.



# Text Summarization

- Generating a short summary from one or more documents, sometimes based on a given query

**RESOOMER**

Service   Extensions   How does it work ?   PREMIUM   Contact   Login   English

Go to the main ideas in your texts, summarize them « relevantly » in 1 Click

[Text example](#)   [Delete text](#)   [Resoomer](#)

Only argumentative texts

http://scih.org/joseph-fourier-greenhouse-effect/

VOLLSTE FLEXIBILITÄT ERLEBEN!  
MEHR ENTDECKEN   

Summary : [Automatic](#) [Manual](#) [Optimized](#) [Analyze](#) [Help](#)

Highlight important sentences

On March 21, 1768, French mathematician and physicist Jean Baptiste Joseph du Fourier was born. He is probably best known for his work in thermodynamics, where he introduced the concept of the Fourier Analysis, named in honor after him. There, he claimed that every mathematical function of a variable can be expanded to a sum of sines of multiples of that variable. What people most likely don't know is that Fourier also was the first to describe the greenhouse effect, which is responsible also for global warming.  
Profound study of nature is the most fertile source of mathematical discoveries.  
Joseph Fourier, The Analytical Theory of Heat , ch.1, p. 7

<https://resoomer.com/en/>

# Question Answering

- Automatically answer questions posed by humans in a natural language



who discovered the green house effect?

ALL IMAGES VIDEOS MAPS NEWS SHOPPING

9,260,000 Results Any time ▾

 Who discovered the greenhouse effect?  
**Joseph Fourier**

Image: wikipedia.org

Green house effect (GHE) The green house effect was discovered by **Joseph Fourier** in 1824 and first investigated quantitative by **Svante Arrhenius** in 1896.

Reference: [ronzusgreenworld.wordpress.com/disasters-and-controls/green-house-effect-ghe/](http://ronzusgreenworld.wordpress.com/disasters-and-controls/green-house-effect-ghe/)

Was this helpful?  

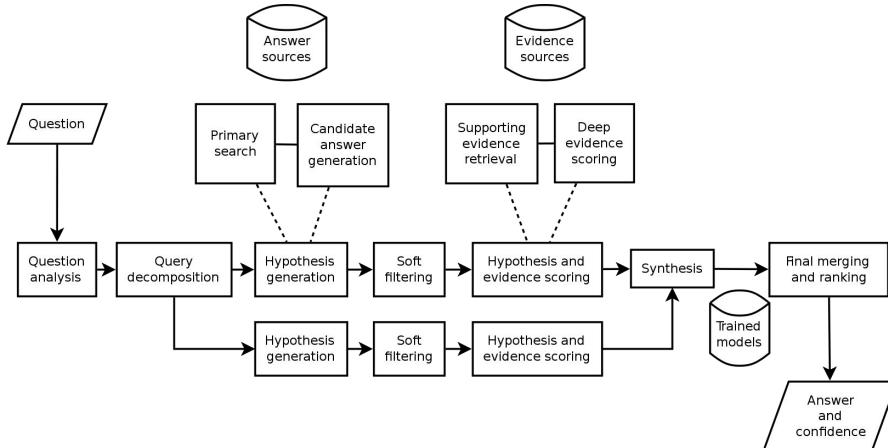
## Greenhouse Effect



The greenhouse effect is the process by which radiation from a planet's atmosphere warms the planet's surface to a temperature above what it would be without this atmosphere. Radiatively active gases (i.e., greenhouse gases) in a planet's atmosphere radiate energy in all directions. Part of this radiation is directed towards the surface, warming it. The intensity of the downward radiation – that is, the strength of the greenhouse effect – will depend on the atmosphere's temperature and on the amount of greenhouse gases that the atmosphere contains.

# Question Answering

- IBM Watson in Jeopardy!



High-level architecture of IBM's DeepQA used in Watson



[https://www.youtube.com/watch?v=WFR3lOm\\_xhE](https://www.youtube.com/watch?v=WFR3lOm_xhE)

Ferrucci, D.; et al. (2010). "Building Watson: An Overview of the DeepQA Project". *AI Magazine*. 31 (3). Retrieved February 19, 2011.

# Information Extraction

- Automatically extracting **structured information** from unstructured and/or semi-structured machine-readable documents



**WIKIPEDIA**  
The Free Encyclopedia

[https://en.wikipedia.org/wiki/Neil\\_Armstrong](https://en.wikipedia.org/wiki/Neil_Armstrong)

Neil Armstrong

From Wikipedia, the free encyclopedia.

For other people named Neil Armstrong, see [Neil Armstrong \(disambiguation\)](#).

**Neil Alden Armstrong** (August 5, 1930 – August 25, 2012) was an American astronaut and the first person to walk on the Moon. He was also an aerospace engineer, naval aviator, test pilot, and university professor. Before becoming an astronaut, Armstrong was an officer in the U.S. Navy and served in the Korean War. After the war, he earned his bachelor's degree at Purdue University and served as a test pilot at the National Advisory Committee for Aeronautics (NACA) High-Speed Flight Station, where he logged over 900 flights. He later completed graduate studies at the University of Southern California.

A participant in the U.S. Air Force's Man in Space Soonest and X-20 Dyna-Soar human spaceflight programs, Armstrong joined the NASA Astronaut Corps in 1962. He made his first space flight as command pilot of Gemini 8 in March 1966, becoming NASA's first civilian astronaut to fly in space. He performed the first docking of two spacecraft, with pilot David Scott.<sup>[1]</sup> This mission was aborted after Armstrong used some of his reentry control fuel to prevent a dangerous spin caused by a stuck thruster, in the first in-flight space emergency.

Armstrong's second and last spaceflight was as commander of Apollo 11, the first manned landing mission in July 1969. Armstrong and Lunar Module pilot Buzz Aldrin descended to the lunar surface and spent two and a half hours outside the spacecraft, while Michael Collins remained in lunar orbit in the Command/Service Module. Along with Collins and Aldrin, Armstrong was awarded the Presidential Medal of Freedom by President Richard Nixon; President Jimmy Carter presented Armstrong the Congressional Space Medal of Honor in 1979. Armstrong and his former crewmates received the Congressional Gold Medal in 2009.

Armstrong died in Cincinnati, Ohio on August 25, 2012, at the age of 82, after complications from coronary artery bypass surgery.<sup>[2][3]</sup>

Contents [view]

- 1 Early years
- 2 Navy service
- 3 College years
- 4 Test pilot
- 5 Astronaut career
  - 5.1 Gemini program
    - 5.1.1 Gemini 8
    - 5.1.2 Gemini 11
  - 5.2 Apollo program
    - 5.2.1 Apollo 11
    - 5.2.1.1 Voyage to the Moon
    - 5.2.1.2 First Moon walk
    - 5.2.1.3 Return to Earth
  - 6 Life after Apollo
    - 6.1 Teaching
    - 6.2 NASA accident investigations
    - 6.3 Business activities

Neil Armstrong



Armstrong in July 1969

USAF / NASA astronaut

Neil Armstrong

Born

August 5, 1930 (aged 82)  
Near Wapakoneta, Ohio, U.S.

Died

August 25, 2012 (aged 82)  
Cincinnati, Ohio, U.S.

Previous occupation

Lunar module pilot

Alma mater

Purdue University, B.S. 1955  
University of Southern California, M.S. 1970

Rank

Lieutenant (junior grade),  
United States Navy

Time in space

1 hour, 11 minutes,  
12 seconds, and 30 seconds

Selection

1966 USAF Man in Space



[http://dbpedia.org/page/Neil\\_Armstrong](http://dbpedia.org/page/Neil_Armstrong)

DBpedia

Browse using - Formats -

Faceted Browser Sparql Endpoint

## About: Neil Armstrong

An Entity of Type : [NASA](#), from [Named Graph](#) : <http://dbpedia.org>, within Data Space : [dbpedia.org](#)

Neil Alden Armstrong (August 5, 1930 – August 25, 2012) was an American astronaut and the first person to walk on the Moon. He was also an aerospace engineer, naval aviator, test pilot, and university professor. Before becoming an astronaut, Armstrong was an officer in the U.S. Navy and served in the Korean War. After the war, he earned his bachelor's degree at Purdue University and served as a test pilot at the National Advisory Committee for Aeronautics (NACA) High-Speed Flight Station, where he logged over 900 flights. He later completed graduate studies at the University of Southern California. A participant in the U.S. Air Force's Man in Space Soonest and X-20 Dyna-Soar human spaceflight programs, Armstrong joined the NASA Astronaut Corps in 1962. He made his first space flight as command pilot of Gemini 8 in March 1966, becoming NASA's first civilian astronaut to fly in space. He performed the first docking of two spacecraft, with pilot David Scott. This mission was aborted after Armstrong used some of his reentry control fuel to prevent a dangerous spin caused by a stuck thruster, in the first in-flight space emergency. Armstrong's second and last spaceflight was as commander of Apollo 11, the first manned landing mission in July 1969. Armstrong and Lunar Module pilot Buzz Aldrin descended to the lunar surface and spent two and a half hours outside the spacecraft, while Michael Collins remained in lunar orbit in the Command/Service Module. Along with Collins and Aldrin, Armstrong was awarded the Congressional Space Medal of Honor in 1979. Armstrong and his former crewmates received the Congressional Gold Medal in 2009. Armstrong died in Cincinnati, Ohio on August 25, 2012, at the age of 82, after complications from coronary artery bypass surgery. (en)

### Property

[dbo:Astronaut/timelineSpace](#)

### Value

• 0.5

• 12372.0

### dbo:abstract

- Neil Alden Armstrong (\* 5. August 1930 bei Wapakoneta, Ohio; † 25. August 2012 in Cincinnati, Ohio) war ein US-amerikanischer Testpilot und Astronaut. Er war Kommandant von Apollo 11, die mit Buzz Aldrin und Michael Collins zum Mond flog. Am 21. Juli 1969 wurde er als erster Mensch auf dem Mond ausgestiegen.
- Neil Alden Armstrong (August 5, 1930 – August 25, 2012) was an American astronaut and the first person to walk on the Moon. He was also an aerospace engineer, naval aviator, test pilot, and university professor. Before becoming an astronaut, Armstrong was an officer in the U.S. Navy and served in the Korean War. After the war, he earned his bachelor's degree at Purdue University and served as a test pilot at the National Advisory Committee for Aeronautics (NACA) High-Speed Flight Station, where he logged over 900 flights. He later completed graduate studies at the University of Southern California. A participant in the U.S. Air Force's Man in Space Soonest and X-20 Dyna-Soar human spaceflight programs, Armstrong joined the NASA Astronaut Corps in 1962. He made his first space flight as command pilot of Gemini 8 in March 1966, becoming NASA's first civilian astronaut to fly in space. He performed the first docking of two spacecraft, with pilot David Scott. This mission was aborted after Armstrong used some of his reentry control fuel to prevent a dangerous spin caused by a stuck thruster, in the first in-flight space emergency. Armstrong's second and last spaceflight was as commander of Apollo 11, the first manned landing mission in July 1969. Armstrong and Lunar Module pilot Buzz Aldrin descended to the lunar surface and spent two and a half hours outside the spacecraft, while Michael Collins remained in lunar orbit in the Command/Service Module. Along with Collins and Aldrin, Armstrong was awarded the Congressional Space Medal of Honor in 1979. Armstrong and his former crewmates received the Congressional Gold Medal in 2009. Armstrong died in Cincinnati, Ohio on August 25, 2012, at the age of 82, after complications from coronary artery bypass surgery. (en)

### dbo:occupation

• [dbo:Naval\\_aviation](#)

# Entity Suggestion / Autosuggestion

- Predicting a potential knowledge base entity from (ambiguous) text input according to the intention of the user

es. Fogg tells Passepartout to pack only a few things, while everything else will be bought on the trip. The only will carry about is a carpet bag filled with £2000.

ginning of the fabulous plot and from on we see how Phileas Fogg travels around the world and we witness dventures that he has together with his companion White disengaging in Egypt, he is spotted by a diligent l detective named Fix, who has been dispatched to London to catch of a bank robber. Because Fogg matches n of the bank robber, Fix mistakenly believes him to be the criminal. In order to secure a e, Fix goes on board the steamer crewed by Passepartout and Foggy Nelson. Fix gets h Passepartout, but does not reveal his purpose. Fix's true intentions, who wants to get Fogg to rtout and Fogg have

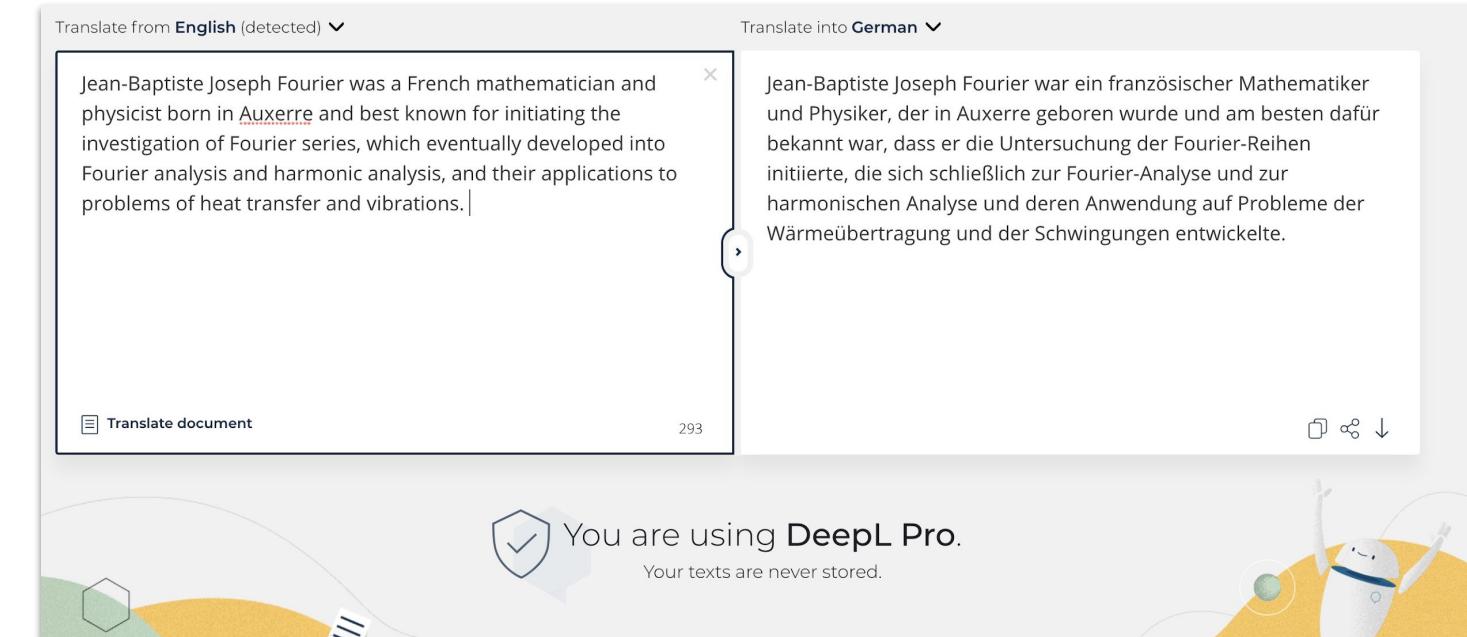


<http://refer.cx>

<http://scifi.org/around-the-world-in-80-days/>

# Machine Translation

- Translating a text from one language to another



The screenshot shows a DeepL Pro interface translating a text from English to German. The English text is:

Jean-Baptiste Joseph Fourier was a French mathematician and physicist born in Auxerre and best known for initiating the investigation of Fourier series, which eventually developed into Fourier analysis and harmonic analysis, and their applications to problems of heat transfer and vibrations.

The German translation is:

Jean-Baptiste Joseph Fourier war ein französischer Mathematiker und Physiker, der in Auxerre geboren wurde und am besten dafür bekannt war, dass er die Untersuchung der Fourier-Reihen initiierte, die sich schließlich zur Fourier-Analyse und zur harmonischen Analyse und deren Anwendung auf Probleme der Wärmeübertragung und der Schwingungen entwickelte.

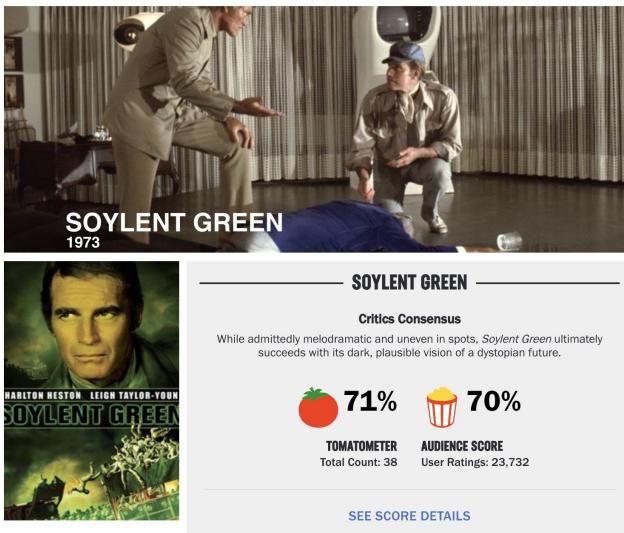
At the bottom, there is a message: "You are using DeepL Pro. Your texts are never stored." and the DeepL logo.



<https://www.deepl.com/>

# Sentiment Analysis

- Identifying sentiments and opinions stated in a text



[http://rottentomatoes.com/m/soylent\\_green](http://rottentomatoes.com/m/soylent_green)



**Test with your own text**

The secret of Soylent Green shouldn't be revealed; suffice it to say that it isn't quite as chilling as it should be, given the energy put into making it mysterious.

However, Soylent Green is a

**Classify Text**

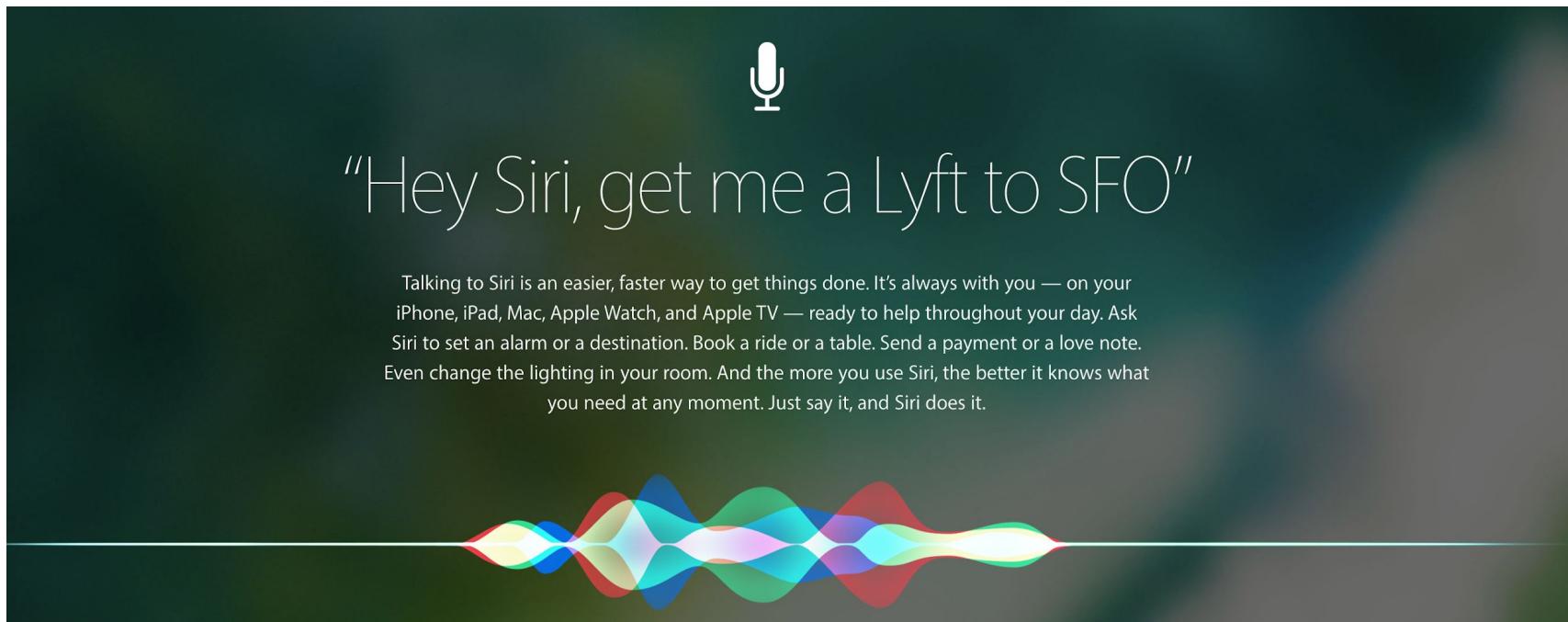
**Results**

TAG	CONFIDENCE
Positive	98.7%

<https://monkeylearn.com/sentiment-analysis-online/>

# Speech Recognition (Speech to Text, STT)

- Recognizing a spoken language and transforming it into a text



# Speech Synthesis

 (Text to Speech, TTS)

- Producing a spoken language from a text

This is a Speech Synthesis Demo

Voice       ▼

Volume

Rate

Pitch

**Speak**

<https://codepen.io/matt-west/pen/wGzuU>

# Dialog Systems

- Running a dialog between the user and the system

Hello, I am the bank's Virtual Agent.

I can help you with a number of banking tasks:

- Making a credit card payment
- Booking an appointment with a banker
- Choosing a credit card

**Can you help me?**

I can help you with questions about our products, specifically credit cards. For demonstration purposes click on one of the tiles to the left.

**what is a credit card?**

We have a great range of Credit Cards available.

Would you like me to help you choose a credit card best suited to your needs?

**yes**

**Great!**

I can help you find a credit card to suit your needs. We have credit cards to build credit, provide rewards, and help you save money. What are you looking for most in a credit card?

Type something

\* This system is for demonstration purposes only and is not intended to process Personal Data. No Personal Data is to be entered into this system

Watson understands

```

32 }

1 {
2   "output": {
3     "generic": [
4       {
5         "response_type": "text",
6         "text": "Great!"
7       },
8       {
9         "response_type": "text",
10        "text": "I can help you find a credit card to suit your n"
11      }
12    ],
13    "intents": [],
14    "entities": [
15      {
16        "entity": "ResponseTypes",
17        "location": [
18          0,
19          3
20        ],
21        "value": "positive",
22        "confidence": 1
23      }
24    ],
25  },
26  "context": {
27    "global": {
28      "system": {
29        "turn_count": 4
30      }
31    },
32    "skills": {

```

<https://assistant-simple.ng.bluemix.net/>

## Lecture 2: Natural Language Processing (1)

2.0 What is Natural Language Processing?

2.1 NLP and Basic Linguistic Knowledge

2.2 Morphology

2.3 NLP Applications

### **2.4 NLP Techniques**

2.5 NLP Challenges

2.6 Evaluation, Precision and Recall

2.7 Regular Expressions

2.8 Finite State Automata

2.9 Tokenization

2.10 Language Model and N-Grams

2.11 Part-of-Speech Tagging

2.12 Word Embeddings

# Sentence Splitting

- Splitting a **text** into **sentences**

## Original Text

On March 21, 1768, French mathematician and physicist Jean Baptiste Joseph du Fourier was born. He is probably best known for his work in thermodynamics, where he introduced the concept of the Fourier Analysis, named in honor after him. There, he claimed that every mathematical function of a variable can be expanded to a sum of sines of multiples of that variable. What people most likely don't know is that Fourier also was the first to describe the greenhouse effect, which is responsible also for global warming.

## Analysis Result

On March 21, 1768, French mathematician and physicist Jean Baptiste Joseph du Fourier was born.

He is probably best known for his work in thermodynamics, where he introduced the concept of the Fourier Analysis, named in honor after him.

There, he claimed that every mathematical function of a variable can be expanded to a sum of sines of multiples of that variable.

What people most likely don't know is that Fourier also was the first to describe the greenhouse effect, which is responsible also for global warming.

<http://textanalysisonline.com/nltk-sentence-segmentation>

# Tokenization

## Tokenizer Text

### Enter text

On March 21, 1768, French mathematician and physicist Jean Baptiste Joseph du Fourier was born. He is probably best known for his work in thermodynamics, where he introduced the concept of the Fourier Analysis, named in honor after him.

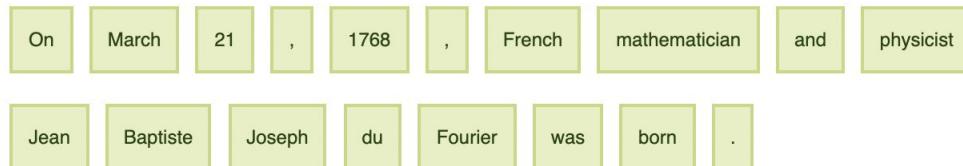
Enter up to 50000 characters

Tokenize

- Splitting a **stream of text** up into words, phrases, symbols, or other meaningful elements called **tokens**

## TreebankWordTokenizer

1.



<http://text-processing.com/demo/tokenize/>

# Part-of-Speech Tagging (POS Tagging)

- Marking up a word in a text as corresponding to a particular **part of speech**, based on both its definition and its context

— Text to annotate —

French mathematician and physicist Jean Baptiste Joseph du Fourier is probably best known for his work in thermodynamics, where he introduced the concept of the Fourier Analysis, named in honor after him. What people most likely don't know is that Fourier also was the first to describe the greenhouse effect, which is responsible also for global

— Annotations —

parts-of-speech

— Language —

English

**Part-of-Speech:**

1 French mathematician and physicist Jean Baptiste Joseph du Fourier is probably best known for his work in thermodynamics , where he introduced the concept of the Fourier Analysis , named in honor after him .

2 What people most likely do n't know is that Fourier also was the first to describe the greenhouse effect , which is responsible also for global warming .

<http://corenlp.run/>

# Constituency Parsing

- Analyzing a string of symbols conforming to the rules of a formal grammar and
- building a **syntactic tree of a sentence**

## Selected Syntactic Categories:

- Noun Phrase (NP)
- Verb Phrase (VP)
- Prepositional Phrase (PP)
- Adverbial Phrases (AdvP)
- Determiner (DT)
- Auxiliary (MD): *have, may*
- Conjunction (CC): *and*
- Noun (NN), Proper Noun (NNP), Verb (VB), Preposition (IN)
- Adjective (JJ), Adverb (RB)
- ...

<http://corenlp.run/>

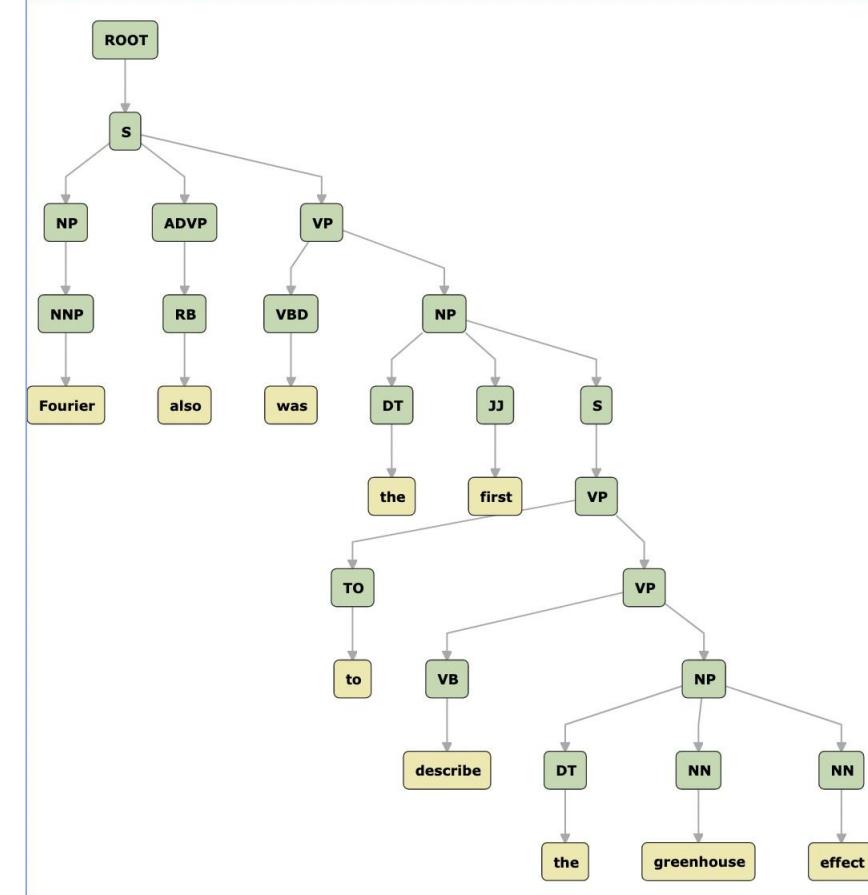
[Jurafsky, chap. 13]

Fourier also was the first to describe the greenhouse effect

— Annotations —

constituency parse ×

## Constituency Parse:



# Dependency Parsing

- Approximation of semantic relations between arguments
- Relies on direct binary grammatical relations among words
- Draws relations from fixed inventory of grammatical relations
- Can handle morphologically rich languages

[Jurafsky, chap. 15]

## Selected Dependency Relations:

- Clausal Argument Relations
  - NSUBJ Nominal subject
  - DOBJ Direct object
  - IOBJ Indirect object
- Nominal Modifier Relations
  - NMOD Nominal modifier
  - AMOD Adjectival modifier
  - DET Determiner
  - CASE Preposition, postpositions, other markers
- Other Relations
  - CONJ Conjunct
  - CC Coordinating conjunction

### — Text to annotate —

What people most likely don't know is that Fourier also was the first to describe the greenhouse effect,

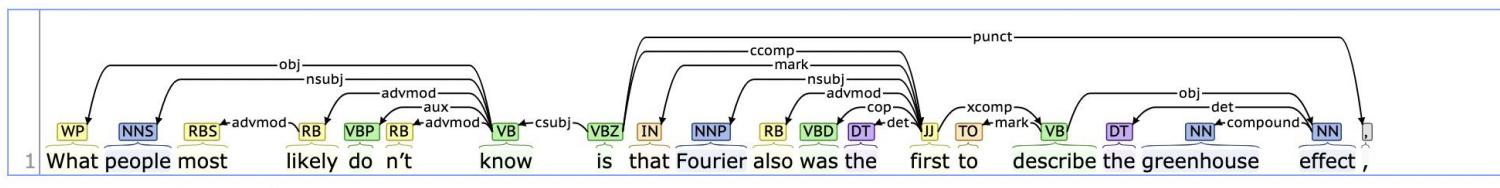
### — Annotations —

dependency parse

### — Language —

English

## Basic Dependencies:



<http://corenlp.run/>

# Named Entity Recognition

- Locating and classifying atomic elements into predefined categories such as **names, persons, organizations, locations, expressions of time, quantities, monetary values, etc.** C.J.Rijsbergen, Information Retrieval (1979)

— Text to annotate —

On March 21, 1768, French mathematician and physicist Jean Baptiste Joseph du Fourier was born.

— Annotations —

named entities

**Named Entity Recognition:**

1 | On March 21 , 1768 , French mathematician and physicist Jean Baptiste Joseph du Fourier was born .

DATE 1768-03-21 NATIONALITY TITLE TITLE PERSON

<http://corenlp.run/>

# Named Entity Resolution / Word Sense Disambiguation

- Finding out the exact meaning of an entity, i.e. **linking a text segment to its corresponding entity** in a knowledge base



On March 21, 1768, French mathematician and physicist **Jean Baptiste Joseph du Fourier** was born. He is probably best known for his work in thermodynamics, the Fourier Analysis, named in honor after him. The Fourier transform of a variable can be expanded to a sum of sines of multiple frequencies. What we don't know is that Fourier also was the first to describe the greenhouse effect.

*"Profound study of nature is the most fertile soil."*  
Joseph Fourier, *The Analytical Theory of Heat*

Joseph Fourier – Early Years

Jean Baptiste Joseph Fourier was born on March 21, 1768, in a small town in France. His father was a tailor. Orphaned already at age nine, Fourier was taken in by a local priest. Through this introduction, he was educated by the Benvenistes of the Convent of St. Mark. While he

<b>Joseph Fourier</b>	
Jean Baptiste Joseph Fourier	(21 March 1768 – 16 May 1830) was a French mathematician and physicist born in Auxerre and best known for his work on the Fourier series, Fourier transform and Fourier's law of conduction of heat.
<b>birth year</b>	1768
<b>death year</b>	1830
<b>is influenced by</b>	Adolphe Quetelet
<b>birth place</b>	Auxerre
<b>death place</b>	Bourbon Restoration
<b>birth place</b>	Burgundy

<http://scihi.org/joseph-fourier-greenhouse-effect/>

FIZ Institute for Information Infrastructure & AIFB - Karlsruhe Institute of Technology



# Semantic Role Labeling

- Extracting subject-predicate-object triples from a sentence

— Text to annotate —

On March 21, 1768, French mathematician and physicist Jean Baptiste Joseph du Fourier was born.

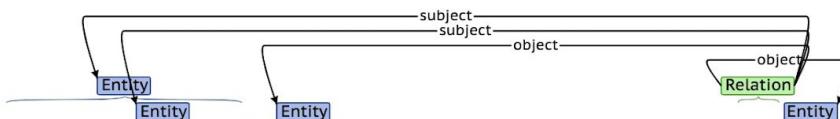
— Annotations —

openie X

— Language —

English ▼

**Open IE:**



1 On March 21 , 1768 , French mathematician and physicist Jean Baptiste Joseph du Fourier was born .

<http://corenlp.run/>

# Co-Reference Resolution

- Also known as **anaphora resolution**
- Determine which phrases in a document or discourse **refer to the same underlying entity**

<http://corenlp.run/>
**— Text to annotate —**

On March 21, 1768, French mathematician and physicist Jean Baptiste Joseph du Fourier was born. He is probably best known for his work in thermodynamics, where he introduced the concept of the Fourier Analysis, named in honor after him.

**— Annotations —**

**— Language —**


**Speakers:**
S(PERO)

1 On March 21 , 1768 , French mathematician and physicist Jean Baptiste Joseph du Fourier was born .

S(PERO)

2 He is probably best known for his work in thermodynamics , where he introduced the concept of the Fourier Analysis , named in honor after him .

**Coreference:**
CorefEntity7

1 On March 21 , 1768 , French mathematician and physicist Jean Baptiste Joseph du Fourier was born .

CorefEntity7
CorefEntity7
CorefEntity7
CorefEntity7

2 He is probably best known for his work in thermodynamics , where he introduced the concept of the Fourier Analysis , named in honor after him .

## Lecture 2: Natural Language Processing (1)

- 2.0 What is Natural Language Processing?
- 2.1 NLP and Basic Linguistic Knowledge
- 2.2 Morphology
- 2.3 NLP Applications
- 2.4 NLP Techniques
- 2.5 NLP Challenges
- 2.6 Evaluation, Precision and Recall
- 2.7 Regular Expressions
- 2.8 Finite State Automata
- 2.9 Tokenization
- 2.10 Language Model and N-Grams
- 2.11 Part-of-Speech Tagging
- 2.12 Word Embeddings

## 2. Natural Language Processing - 1

### Bibliography

- K.-U. Carstensen et al. (eds.), *Computerlinguistik und Sprachtechnologie, eine Einführung*, 3. Aufl., Spektrum Akademischer Verlag, 2010.  
(Teile von Kap.3, Morphologie)
- D. Jurafsky, J. H. Martin, *Speech and Language Processing*, 3nd ed. (draft), 2019,
  - [13 Constituency Parsing](#), p. 232ff
  - [15 Dependency Parsing](#), p. 273ff
- Ferrucci, D.; et al. (2010). "[Building Watson: An Overview of the DeepQA Project](#)". AI Magazine. 31 (3). Retrieved February 19, 2011.

## 2. Natural Language Processing - 1

### Syllabus Questions

- What is natural language processing?
- Why is language a central concept of intelligence?
- What are the main components natural languages are built on?
- What are words, vocabulary, text, language, and discourse?
- What is the difference between phonetics and phonology?
- What is the difference between a phone and a phoneme in phonology?
- What is morphology?
- What are morphemes, free morphemes, and bound morphemes?
- How can new words in a language be created from existing words?
- What is the difference between compounding and derivation?
- What is the difference between derivation and inflection?
- Would you rather use lemmatization or stemming for frequency-based analysis of text? Why?
- What are the most important techniques in NLP?
- What is the difference between Named Entity Recognition and Named Entity Resolution?