

# What is Computational Linguistics?

An Overview of the Field and its Applications

Eberhard Karls University of Tübingen  
Faculty of Philosophy  
Department of Linguistics  
BA Computational Linguistics/General Linguistics  
Wintersemester 2023



# Definition

“Computational linguistics is the scientific study of language from a computational perspective. Computational linguists are interested in providing computational models of various kinds of linguistic phenomena. These models may be “knowledge-based” (“hand-crafted”) or “data-driven” (“statistical” or “empirical”). Work in computational linguistics is in some cases motivated from a scientific perspective in that one is trying to provide a computational explanation for a particular linguistic or psycholinguistic phenomenon; and in other cases the motivation may be more purely technological in that one wants to provide a working component of a speech or natural language system.”

- *Association for Computational Linguistics*

<https://www.aclweb.org/portal/>

# Origins

- History of computers and computational linguistics are closely aligned, the field also overlapped with artificial intelligence since the first efforts to use computers to automatically translate texts from a foreign language.
- The rise of the internet significantly boosted the relevance of the field and its applications; the first “usable” tools appear (web applications, dialogue systems, translation services, intelligence gathering).
- During the last decade there was a stronger integration of (and focus on) machine learning.

# Goals

- The improvement of the applications depends on the integration of linguistic data structures.
- Modelling and simulating human language to make it understandable for computers.
- Human language is prone to ambiguities and variations. On the other hand, computers are restricted to more formal systems and are vulnerable to any ambiguities or spontaneous changes.
- → An adequate representation of the properties of human language into a formal system is needed.


# Natural vs. Formal Language

→ **Natural Languages** are the languages that we speak (English, German, French, Chinese); they were not designed by people, they evolved naturally as they pass from generation to generation.

→ **Formal Languages** are artificial languages that were designed by people for a specific purpose; for example, for denoting relationships between numbers, symbols, or molecules. Programming languages are formal languages that have been designed to express computations.

The process of evaluating the structure of a sentence (in a natural or formal language) and resolving it into its component parts is called **parsing**.

# Natural or Formal?

- $a^2 + b^2 = c^2$
- A bissle isch emmer no bessr wia gar nix.
- ```
public static void main (String[ ] args){  
    System.out.print("Hello, World!") }
```
- $\neg \exists x \neg \forall y ((\exists x P xy \rightarrow Qx) \wedge Ry)$
- Преступление и наказание
- ```
print("Hello, World!")
```
- 
- 0100100 01101001
- Formal  $\rightarrow$  Mathematics
- Natural  $\rightarrow$  German (Swabian)
- Formal  $\rightarrow$  Programming (Java)
- Formal  $\rightarrow$  Predicate Logic
- Natural  $\rightarrow$  Russian
- Formal  $\rightarrow$  Programming (Python)
- Formal/Natural  $\rightarrow$  Wingdings font
- Formal  $\rightarrow$  Binary Code



# Computational Linguistics includes...

- Linguistics
- Computer Science
- Artificial Intelligence
- Mathematics
- Logic
- Philosophy
- Cognitive Science
- Cognitive Psychology
- Psycholinguistics
- Anthropology
- Neuroscience
- Natural Language Processing



# Natural Language Processing

→ **Natural Language Understanding (NLU)**: Mapping the given input in natural language into useful representations, which involves analyzing different aspects of the input (POS tagging, lemmatization, dependency parsing etc).

→ **Natural Language Generation (NLG)**: Producing meaningful natural language text from some more abstract representation (less developed than NLU, very domain-specific).

# Main Fields of Application

- Machine Translations (*Google Translate, DeepL, Linguee*)
- Text Editors/Spellcheckers (*NotePad, Grammarly, Wordtune*)
- Chatbots (*ChatGPT, Customer Support Systems*)
- Speech Recognition Systems/Text-to-Speech Synthesizers (*Speech Services, Google Translate, 'Read Aloud' options in Browsers*)
- Web Search Engines (*basically everywhere*)

# Machine Translation

- Rule-based approach
- Interlingual approach
- Dictionary-based approach
- Statistical approach
- Deep learning-based approach (neural machine translation)

# Chatbots

- Implementation is heavily related to artificial intelligence, machine learning and natural language processing.
- Require a large amount of conversational data to train.
- The input/output database is usually fixed.
- More simplistic chatbots have been around for decades; the field gained more attention in the last years due to OpenAI's ChatGPT.

# Web Search Engines

- Search the World Wide Web for a particular information that is specified in a textual query.
- Web crawling → indexing → searching
- Types of engines: search by keyword/author/title/date (using natural language)
- Query volume: Google 79%, Bing 12%, Baidu 5%, Yahoo! 2%

Me : \*searches cool gaming mouse\*

**Google:**



**Bing:**



# Suggested Literature and Sources

## Introduction to CL/General:

Daniel Jurafsky and James H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall, Upper Saddle River, NJ, 2nd edition edition, 2009.

Ralph Grishman. *Computational linguistics: an introduction*. Cambridge University Press, 1986.

Turing, Alan (1950), "Computing Machinery and Intelligence", *Mind*, LIX (236).

John R. Searle. Minds, brains, and programs. *Behavioral and Brain Sciences* 3, 1980.

Dickinson, Markus, et al. *Language and Computers*. Wiley, 2012.

## Logic:

Magnus, P. D. *FORALLX: An introduction to formal logic.*, 2017.

L.T.F. Gamut. *Logic, Language, and Meaning, Volume 1: Introduction to Logic*.

## Java/DSA I. :

Savitch, Walter. *Java: An Introduction to Problem Solving and Programming*. Pearson, 2010.