

# Introduction & Overview

## Einführung in die Informationsverarbeitung

Nils Reiter

October 13, 2022

## Sprachliche Informationsverarbeitung

# Course topics

- ▶ Overview: Language processing
- ▶ Linguistic areas and phenomena
- ▶ Corpus linguistics and statistics
- ▶ Annotation workflow
- ▶ Machine learning

# Sprachliche Informationsverarbeitung

- ▶ Information often encoded in language
  - ▶ E.g., on this slide or in this lecture

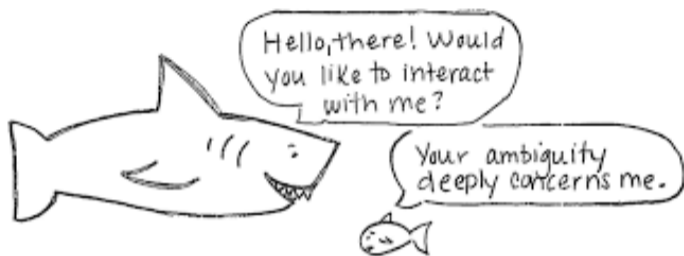
# Sprachliche Informationsverarbeitung

- ▶ Information often encoded in language
  - ▶ E.g., on this slide or in this lecture
- ▶ Harvesting language automatically is difficult
  - ▶ Language is in many ways ambiguous
  - ▶ Meaning of words can change (Mouse as animal vs. as input device)
  - ▶ Language rules are evolving

# Sprachliche Informationsverarbeitung

- ▶ Information often encoded in language
    - ▶ E.g., on this slide or in this lecture
  - ▶ Harvesting language automatically is difficult
    - ▶ Language is in many ways ambiguous
    - ▶ Meaning of words can change (Mouse as animal vs. as input device)
    - ▶ Language rules are evolving
  - ▶ Text production increases
    - ▶ Average student, average day: 15k (spoken) words
    - ▶ Average US-American: 94 text messages per day
- ⇒ There is a gigantic amount of words out there!

Mehl et al. (2007)  
TextRequest



Language Ambiguity

# Language Ambiguity

## Ambiguity / Ambiguität

- ▶ A sentence is ambiguous: There are multiple possible readings
- ▶ »Ein Satz ist ambig: Es gibt mehrere Lesarten, er ist **mehrdeutig**«



# Language Ambiguity

## Ambiguity / Ambiguität

- ▶ A sentence is ambiguous: There are multiple possible readings
- ▶ »Ein Satz ist ambig: Es gibt mehrere Lesarten, er ist **mehrdeutig**«
- ▶ Fundamental property of natural language
- ▶ Often basis for humor
- ▶ Takes place on all language levels
  - ▶ Sentences can be ambiguous
    - ▶ What is their syntactic structure?
  - ▶ Words can be ambiguous
    - ▶ What is their morphological structure?
  - ▶ Words can be ambiguous in their context
    - ▶ To which character does a pronoun refer to?

# Language Ambiguity

## Examples

Der Jäger traf den Mann mit dem Gewehr.

# Language Ambiguity

## Examples

Landesmusikdirektor:in

# Language Ambiguity

## Examples

Landesmusikdirektor:in

Musikdirektor:in des Landes

Direktor:in für Landesmusik

# Language Ambiguity

## Examples

Maria hat Petra beim Einkaufen getroffen. Sie hat ihr Geld geliehen.

# Language Ambiguity

## Examples

Maria ging zur Bank.

# Language Ambiguity

## Examples

Maria ging zur Bank und setzte sich hin.

# Language Ambiguity

## Examples

Maria ging zur Bank und raubte sie aus.



# Language Ambiguity

## Examples

hubert hat dort liebe genossen.

# Language Ambiguity

## Examples

hubert hat dort liebe genossen.

Hubert hat dort Liebe genossen.

Hubert hat dort liebe Genossen.

# Language Ambiguity

## Examples

Time flies like an arrow.

Questions?

(My) next session, December 1, 2022:  
Big picture overview on language technology