



UNIVERSITÄT  
ZU KÖLN

# Introduction

VL Sprachliche Informationsverarbeitung

Nils Reiter

`nils.reiter@uni-koeln.de`

October 12, 2023

Winter term 2023/24

## Historisch-Kulturwissenschaftliche Informationsverarbeitung

- ▶ Prof. Dr. Øyvind Eide
- ▶ Keywords
  - ▶ Maps
  - ▶ Models and modeling
  - ▶ Cultural heritage
  - ▶ Simulation

## Sprachliche Informationsverarbeitung

- ▶ Prof. Dr. Nils Reiter
- ▶ Keywords
  - ▶ Geschriebene und gesprochene Sprache
  - ▶ Textanalyse
  - ▶ Machine/deep learning
  - ▶ Operationalisierung

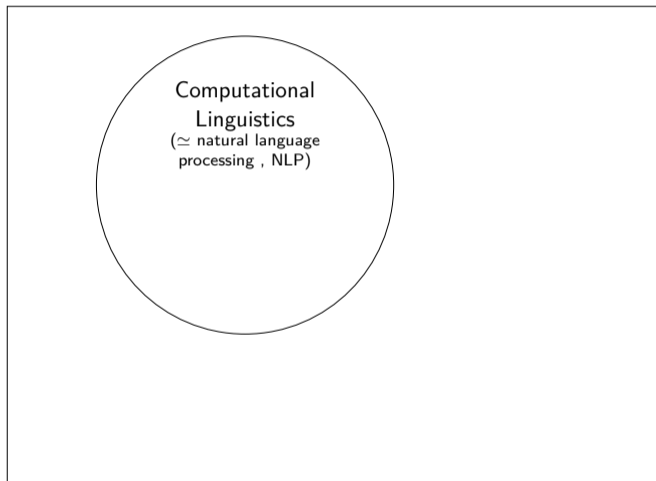
# Orga

- ▶ Donnerstag, 12:00–13:30
- ▶ Module: ...
- ▶ Studienleistung: fünf Hausaufgaben, Abgabe via Ilias
- ▶ Prüfung: Klausur (01.02.2023)

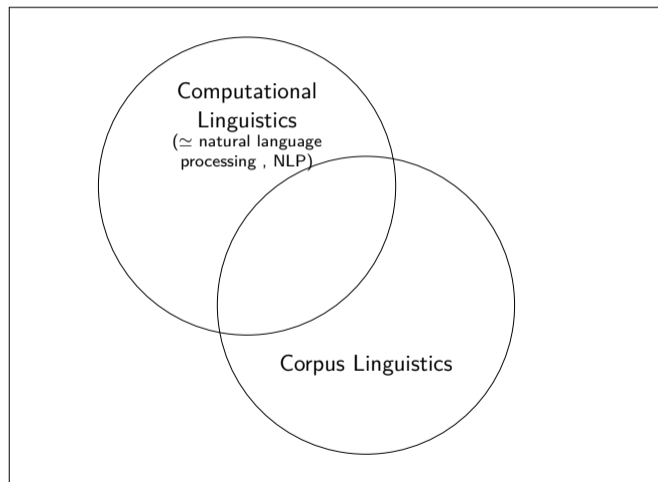
Kurswebseite



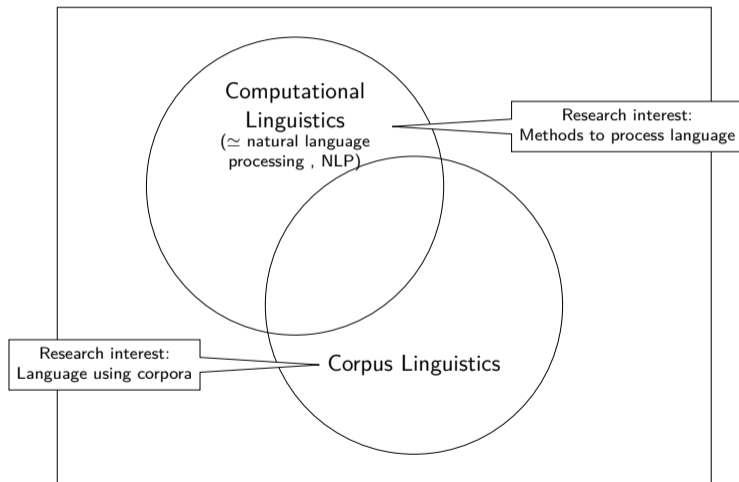
# Disciplinary Placement



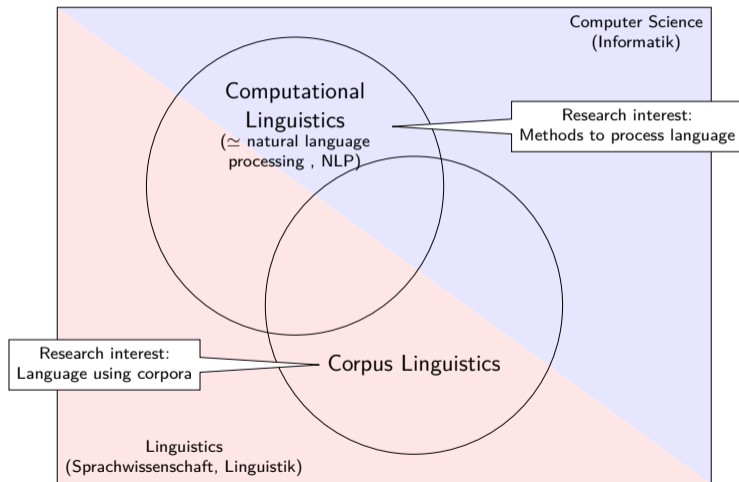
# Disciplinary Placement



# Disciplinary Placement



# Disciplinary Placement



## Brief history of Computational Linguistics I

- ▶ 1950s: DARPA Projects to automatically translate Russian into English
- ▶ 1957/65: Linguistics shifts focus from describing to generating Chomsky (1957, 1965)
- ▶ 1959: Theo Lutz for the first time generates a German poem with a computer Lutz (1959)
- ▶ 1962: Foundation of the „Association for Machine Translation and Computational Linguistics“, 1968 renamed to „Association for Computational Linguistics (ACL)“
- ▶ 1966, ALPAC report: MT more expensive, less accurate and slower than human translation ALPAC (1966)
- ▶ 1968: Foundation of SYSTRAN, first MT company
- ▶ 1975: European commission uses SYSTRAN software (first use of MT on EU level)



## Brief history of Computational Linguistics II

- ▶ 1984: First corpus-based commercial MT system Nagao (1984)
- ▶ 1992: Study programs established in Germany (Saarbrücken/Stuttgart)
- ▶ 2011: IBM Watson beats two humans in Jeopardy [YouTube](#) / Apples Siri launched
- ▶ 2013: Word embeddings (e.g., word2vec) Mikolov u. a. (2013)
- ▶ 2017: Launch of the DeepL Translator
- ▶ 2018: Transformer models: BERT Devlin u. a. (2019)

# Digital Humanities and Computational Linguistics


## Today

- ▶ Digital Humanities, broadly: Working with ‚digital methods‘ on humanities subjects
- ▶ Linguistics: Study of language
- ▶ Computational Linguistics: Pioneer DH area
  - ▶ ... but this is a minority position in CL, often also seen as part of AI

Reiter (2014, 4)

# Digital Humanities and Computational Linguistics

## Today

- ▶ Digital Humanities, broadly: Working with ‚digital methods‘ on humanities subjects
- ▶ Linguistics: Study of language
- ▶ Computational Linguistics: Pioneer DH area Reiter (2014, 4)
  - ▶ ... but this is a minority position in CL, often also seen as part of AI
  - ▶ Historically (and still today) split between engineering (natural language processing, NLP) and science/scholarship (computational linguistics, CL)
  - ▶  Neurolinguistic programming and natural language processing are **not the same** (both use ‚NLP‘ as abbreviation)

# Digital Humanities and Computational Linguistics

## Today

- ▶ Digital Humanities, broadly: Working with ‚digital methods‘ on humanities subjects
- ▶ Linguistics: Study of language
- ▶ Computational Linguistics: Pioneer DH area Reiter (2014, 4)
  - ▶ ... but this is a minority position in CL, often also seen as part of AI
  - ▶ Historically (and still today) split between engineering (natural language processing, NLP) and science/scholarship (computational linguistics, CL)
  - ▶ ⚠ Neurolinguistic programming and natural language processing are **not the same** (both use ‚NLP‘ as abbreviation)

## University of Cologne

For historic reasons, CL and NLP are called „Sprachliche Informationsverarbeitung“

# Experiments

- ▶ Cornerstone of the 'scientific method'
- ▶ Used in many disciplines: Natural sciences, social sciences, medicine, ...

# Experiments

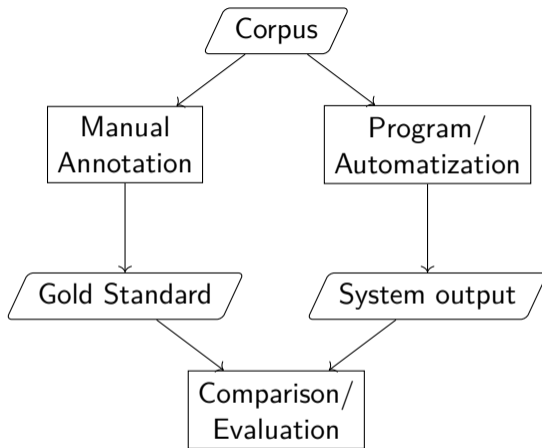
- ▶ Cornerstone of the 'scientific method'
- ▶ Used in many disciplines: Natural sciences, social sciences, medicine, ...
- ▶ Experiments are used to verify or falsify hypotheses
- ▶ Reproducibility: The outcome does not depend on the experimenter

# Experiments

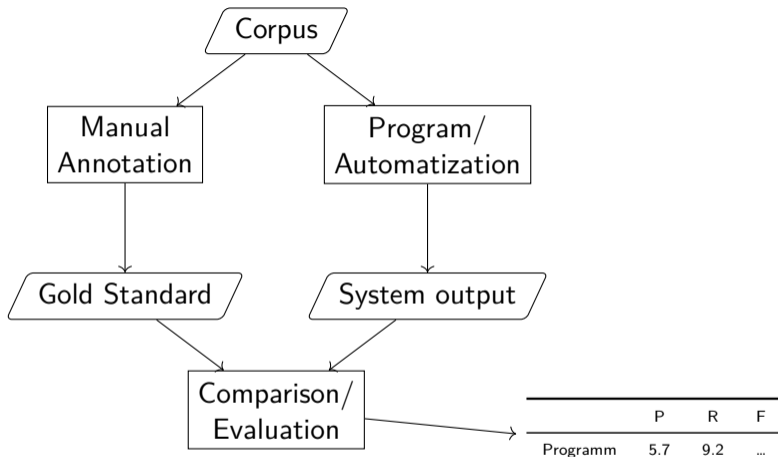
- ▶ Cornerstone of the ‚scientific method‘
- ▶ Used in many disciplines: Natural sciences, social sciences, medicine, ...
- ▶ Experiments are used to verify or falsify hypotheses
- ▶ Reproducibility: The outcome does not depend on the experimenter
- ▶ CL: Hypotheses about the operationalisation of language/text phenomena

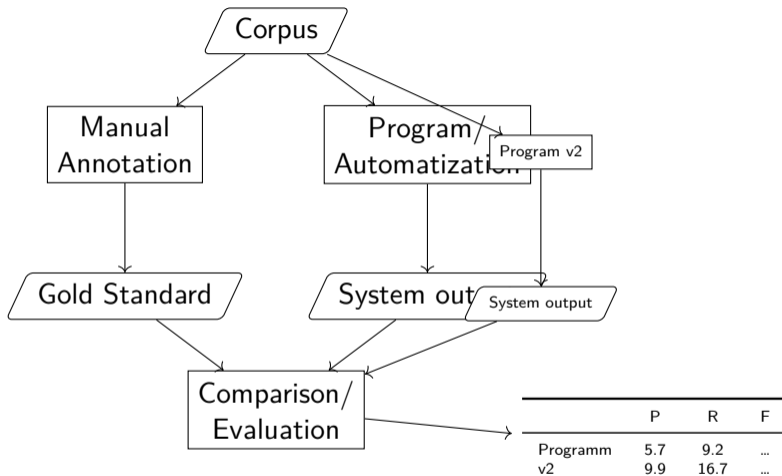
## Example

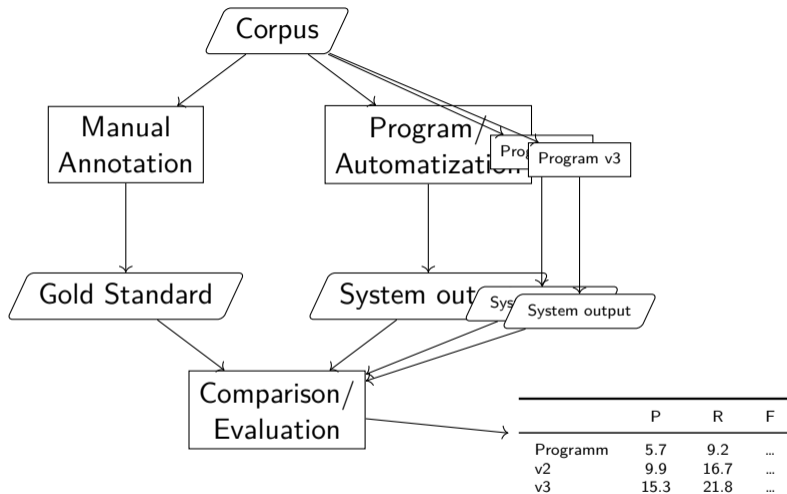
Position within a sentence is indicative for the part of speech

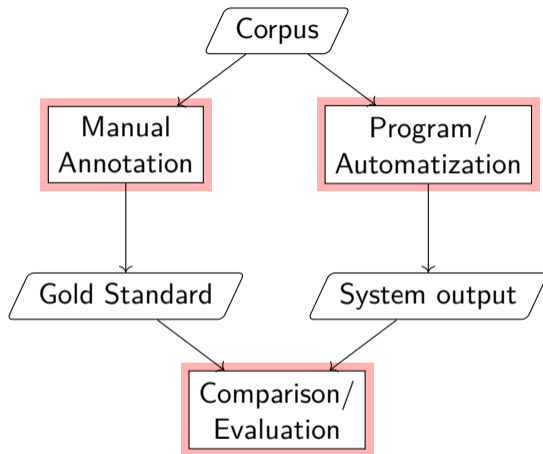












## Section 2

# Language and Linguistics

### Introduction

### Language and Linguistics

Phonology and Phonetics

Morphology

Syntax

# What is Linguistics?

*Linguistics is the scientific study of language.*

Wikipedia, 925699120

# What is Linguistics?

*Linguistics is the scientific study of language.*

Wikipedia, 925699120

- ▶ ‚Scientific study‘
  - ▶ ‚the‘ scientific method
  - ▶ Testable explanations
- ▶ Language
  - ▶ ?

# What is Linguistics?

*Linguistics is the scientific study of language.*

Wikipedia, 925699120

- ▶ ‚Scientific study‘
  - ▶ ‚the‘ scientific method
  - ▶ Testable explanations
- ▶ Language
  - ▶ ?

## Prescriptive vs. descriptive

- ▶ Prescriptive: Telling people how to use language
- ▶ Descriptive: Observing and analysing how people do use language



# What is Linguistics?

*Linguistics is the scientific study of language.*

Wikipedia, 925699120

- ▶ ‚Scientific study‘
  - ▶ ‚the‘ scientific method
  - ▶ Testable explanations
- ▶ Language
  - ▶ ?

## Prescriptive vs. descriptive

- ▶ Prescriptive: Telling people how to use language
- ▶ Descriptive: Observing and analysing how people do use language
- ▶ Academic linguistics: Nowadays mostly descriptive

# What is Language?

- ▶ Communication system
- ▶ Conventionalised: We agree (mostly)
  - ▶ Only partially authoritative

# What is Language?

- ▶ Communication system
- ▶ Conventionalised: We agree (mostly)
  - ▶ Only partially authoritative
- ▶ What do we agree on?
  - ▶ Relation between *signs* and its *meaning* (which is not the same!)
    - ▶ Saussure: Semiotics
  - ▶ E.g.: ‚the students in this class‘ *means* all of you

# Linguistic sign

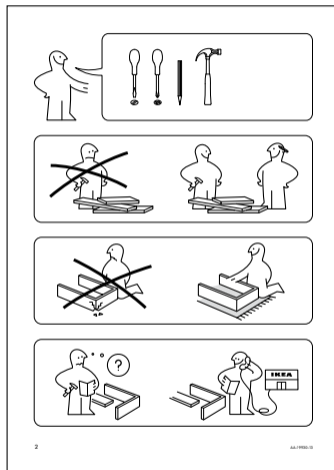
- ▶ Abstract notion
- ▶ Different levels
  - ▶ Texts
  - ▶ Sentences
  - ▶ Words
  - ▶ Syllables, morphemes
  - ▶ Spoken utterances

# Linguistic sign

- ▶ Abstract notion
- ▶ Different levels
  - ▶ Texts
  - ▶ Sentences
  - ▶ Words
  - ▶ Syllables, morphemes
  - ▶ Spoken utterances
  - ▶ Non-textual signs
    - ▶ Emojis 🥰

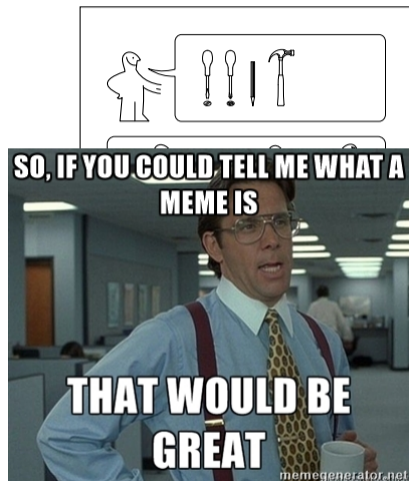
# Linguistic sign

- ▶ Abstract notion
- ▶ Different levels
  - ▶ Texts
  - ▶ Sentences
  - ▶ Words
  - ▶ Syllables, morphemes
  - ▶ Spoken utterances
  - ▶ Non-textual signs
    - ▶ Emojis 🥰
    - ▶ Assembly instructions



# Linguistic sign

- ▶ Abstract notion
- ▶ Different levels
  - ▶ Texts
  - ▶ Sentences
  - ▶ Words
  - ▶ Syllables, morphemes
  - ▶ Spoken utterances
  - ▶ Non-textual signs
    - ▶ Emojis 🥰
    - ▶ Assembly instructions
    - ▶ Memes
    - ▶ ...



# Linguistic sign

Meaning is arbitrary

- ▶ ‚dog‘ refers to a four-legged, wolf-like mammal





# Linguistic sign

Meaning is arbitrary

- ▶ ‚dog‘ refers to a four-legged, wolf-like mammal
- ▶ This is an arbitrary decision
- ▶ The sign could be different, e.g., ‚cat‘
- ▶ No inherent meaning in signs
  - ▶ ...but strongly conventionalised



# Linguistic sign

Meaning is arbitrary

- ▶ ‚dog‘ refers to a four-legged, wolf-like mammal
- ▶ This is an arbitrary decision
- ▶ The sign could be different, e.g., ‚cat‘
- ▶ No inherent meaning in signs
  - ▶ ...but strongly conventionalised
- ▶ Interpreting signs (and language) is something we learn
- ▶ Language is a social construct
  - ▶ Studying language is different from studying gravity



# Ambiguities

- ▶ Der Jäger traf den Mann mit dem Gewehr.

# Ambiguities

- ▶ Der Jäger traf den Mann mit dem Gewehr.
- ▶ Landesmusikdirektor:in

# Ambiguities

- ▶ Der Jäger traf den Mann mit dem Gewehr.
- ▶ Landesmusikdirektor:in
- ▶ Maria hat Petra beim Einkaufen getroffen. Sie hat ihr Geld geliehen.

# Ambiguities

- ▶ Der Jäger traf den Mann mit dem Gewehr.
- ▶ Landesmusikdirektor:in
- ▶ Maria hat Petra beim Einkaufen getroffen. Sie hat ihr Geld geliehen.
- ▶ hubert hat dort liebe genossen.

# Ambiguities

- ▶ Der Jäger traf den Mann mit dem Gewehr.
- ▶ Landesmusikdirektor:in
- ▶ Maria hat Petra beim Einkaufen getroffen. Sie hat ihr Geld geliehen.
- ▶ hubert hat dort liebe genossen.

Linguistics: Let's explain / represent / reproduce these ambiguities