



Computerlinguistische Grundlagen

Jürgen Hermes

Wintersemester 19/20

Sprachliche Informationsverarbeitung

Institut für Digital Humanities

Universität zu Köln



Was ist Computerlinguistik?

- **Definition**
- **Anwendungen**
- **Fragestellung**
- **Nachbardisziplinen**
- **Wissensbereiche**
- **Semesterprogramm**



Computerlinguistik an der Schnittstelle

Computerlinguistik aus *wissenschaftlicher* Sicht:

Schnittstelle zwischen Linguistik und Informatik

Computerlinguistik aus *anwendungsbezogener* Sicht:

Schnittstelle zwischen Sprache und Rechnersystemen





Computerlinguistik in der Anwendung

- **Maschinelle bzw. Computergestützte Übersetzung**
 - Übersetzungsspeicher, Terminologie-Datenbanken,
 - Abgleichwerkzeuge
- **Information Retrieval**
 - Text Mining, Text Classification, Text Summarization,
 - Information Extraktion
- **Mensch-Maschine-Kommunikation**
 - Spracherkennung, Sprachsynthese, Dialogsysteme
- **Aufbau von Ressourcen**
 - Korpora, Lexika, Wortnetze, Baumbanken

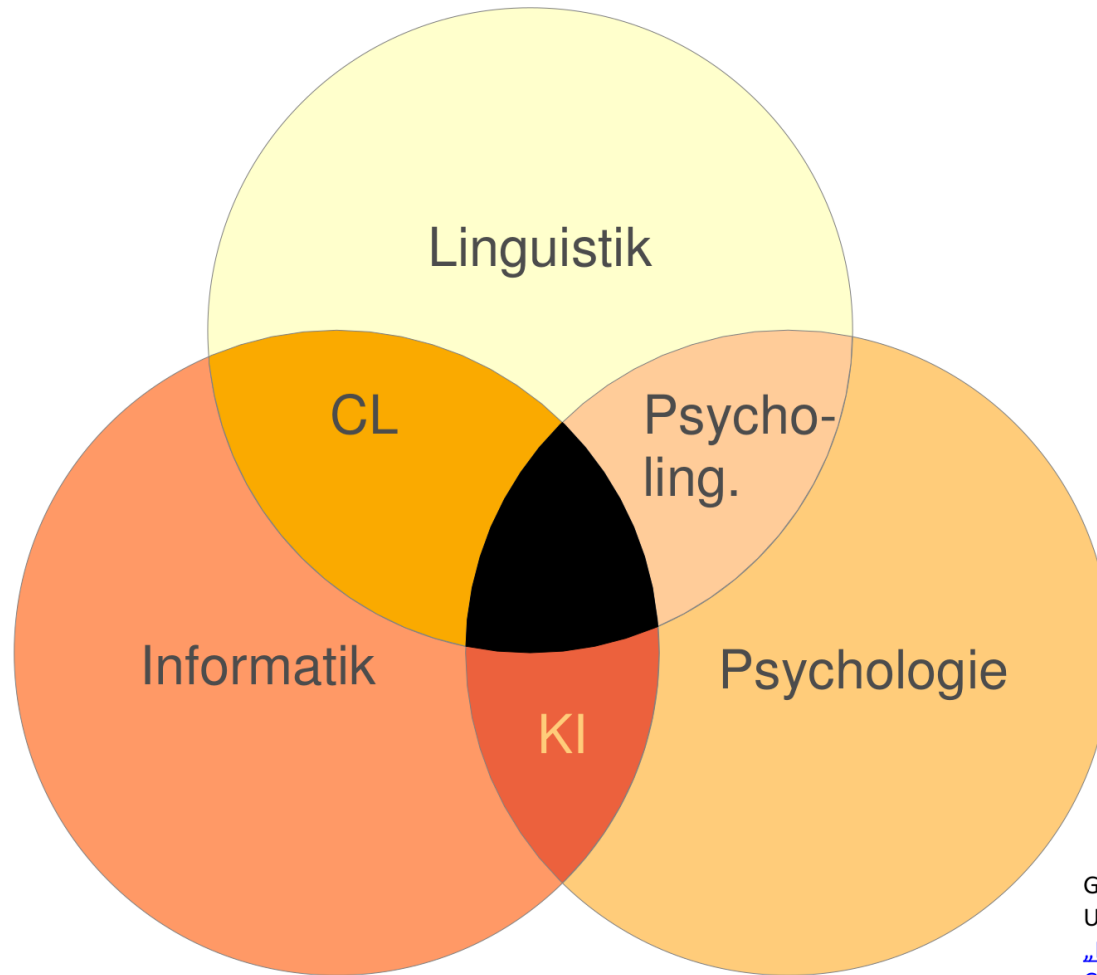


Computerlinguistik als ...

- Teildisziplin der Linguistik
- Linguistische Datenverarbeitung
- Maschinelle Sprachverarbeitung
- Sprachtechnologie



Nachbardisziplinen (I)



Grafik angelehnt an: Hans Uszkoreit: [Vorlesung „Einführung in die Computerlinguistik.“](#)



Nachbardisziplinen (II)

Linguistik: Untersuchungsgegenstand, Termini

Informatik: Algorithmen und Datenstrukturen

Philosophie: Verhältnis Sprache - Denken - Handeln

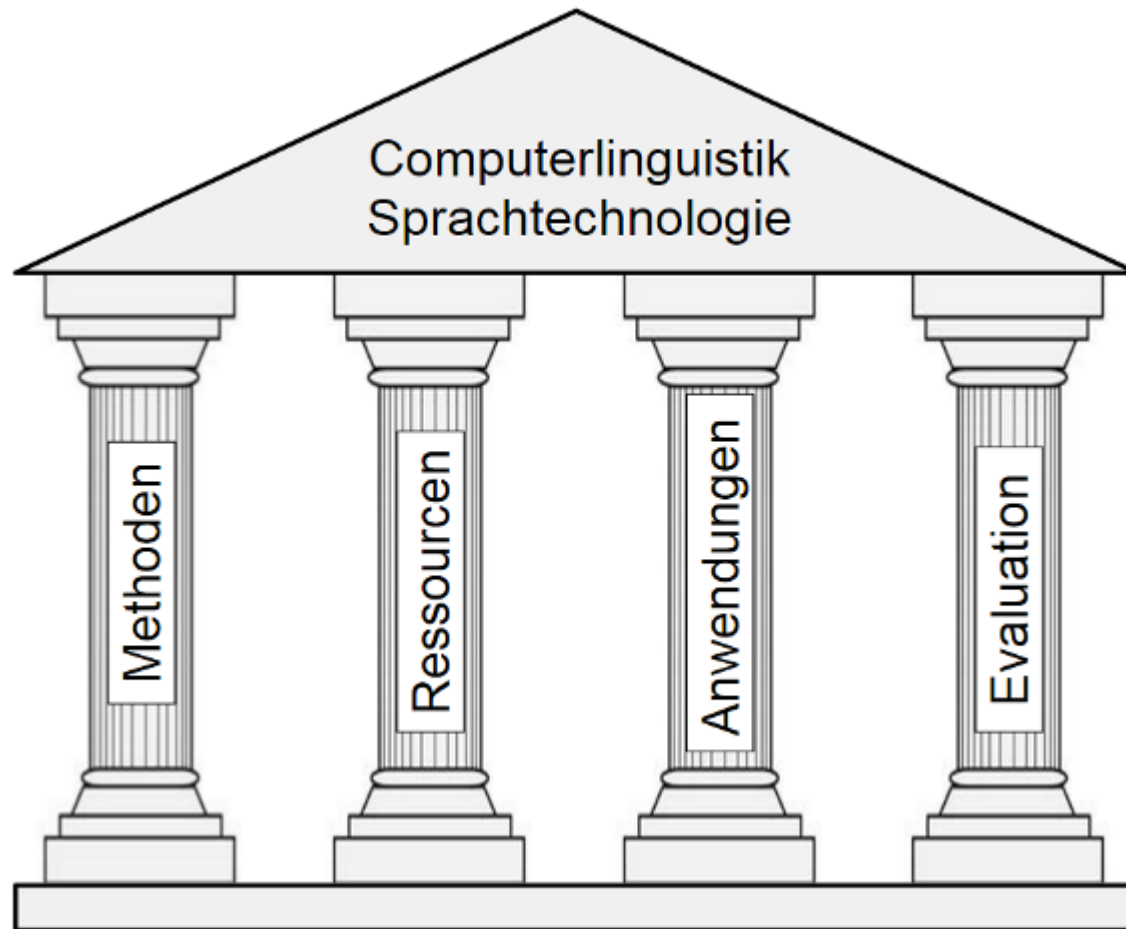
KI: Such- und Planverfahren Wissensrepräsentation,
Wissensverarbeitung

Kognitionswissenschaft: Verhältnis zw. Sprachverarbeitung und allgemeinen Denkprozessen

Mathematik: Logik – Automatentheorie und formale Sprachen – Graphentheorie – Statistik



Pfeiler der Computerlinguistik



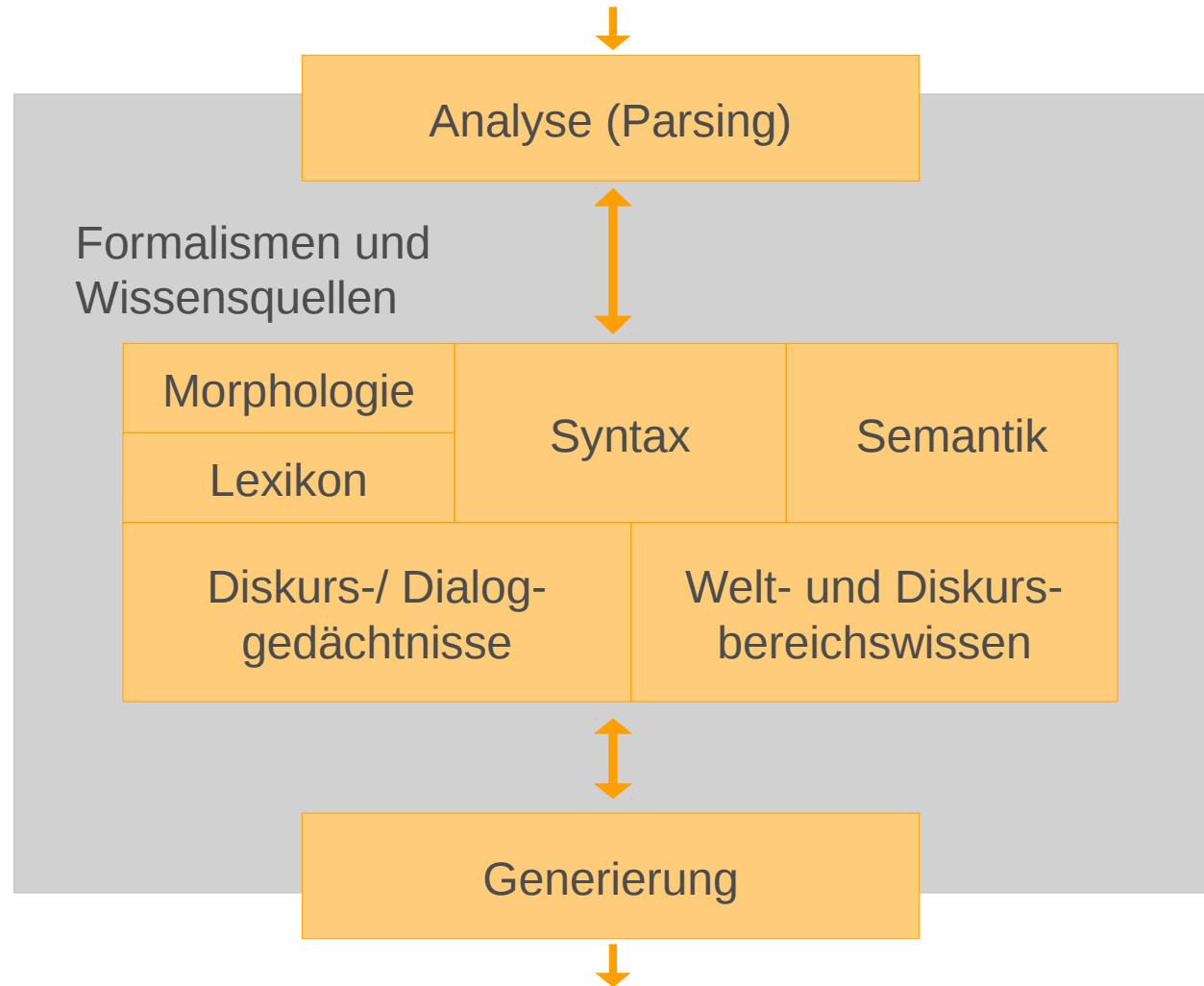


Methoden

- **Linguistische Methoden:** Identifikation von Sprachlichen Einheiten auf unterschiedlichen Ebenen, Untersuchung deren Zusammenspiels (Phonologie, Morphologie, Syntax) und Bedeutungen (Semantik)
- **Informatische Methoden:** Nutzung etablierter Methoden zur Datenspeicher- und -verarbeitung (Datenstrukturen und Algorithmen), z.B. Automaten- und Graphentheorie, Netz- und Vektorrepräsentationen, statistische und konnektionistische Methoden.



Architektur eines natürlichsprachlichen Systems





Forschung und Anwendungen

Theoretische Computerlinguistik:

Suche nach komplexen Beschreibungsformalismen, die gleichzeitig handhabbar und effizient sind.

Praktische Computerlinguistik:

Entwicklung und Erforschung realitätsnaher Anwendungen
-> natürlichsprachliche Kommunikation auf der Maschine



Semesterprogramm (Winter- und Sommersemester)

1. Allgemeine Einführung in Computerlinguistik und Sprachwissenschaft
2. Methoden der (Computer)Linguistik:
Phonologie, Morphologie, Syntax, Semantik
3. Ressourcen für computerlinguistisches Arbeiten und ihre Anwendung
4. Algorithmen und Formalismen für die Verarbeitung natürlichsprachlicher Daten
5. (Begleitend) Computerlinguistische Anwendungen



Literatur / Hausaufgabe

Zur Nachbereitung:

Lesen Sie: Carstensen et al. (2009): Kapitel 1 (S. 1-23)

Zur Vorbereitung:

Lesen Sie: Vater (2002): Kapitel 1 (S. 2-24)

Die Texte finden Sie im Ilias-Seminarordner.