

Computerlinguistik

E01: Einführung in die CL

Was ist Computerlinguistik?

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

Was ist Computerlinguistik?

- Computerlinguistik aus *wissenschaftlicher* Sicht:
Schnittstelle zwischen Linguistik und Informatik
- Computerlinguistik aus *anwendungsbezogener* Sicht:
Schnittstelle zwischen Sprache und Rechnersystemen

Was ist Computerlinguistik?



Was ist Computerlinguistik?



Interpretation GPT4:

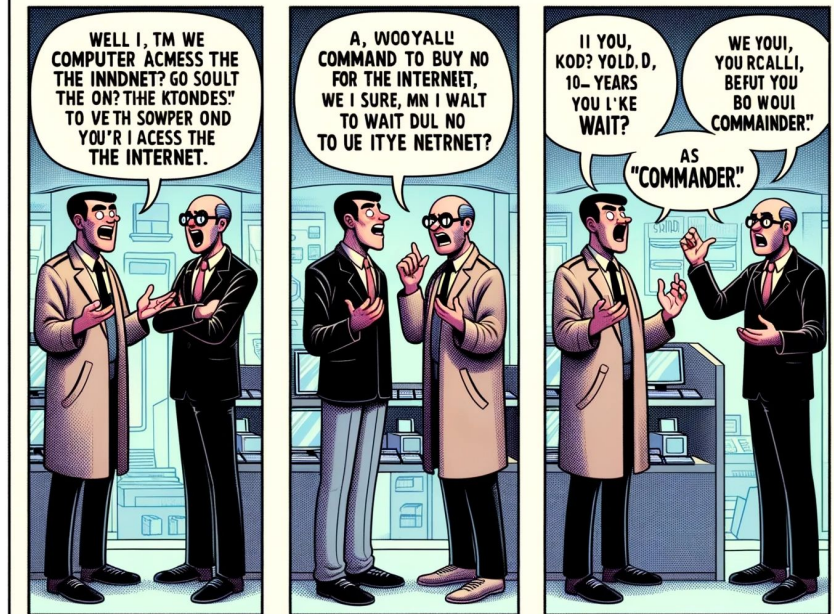
Der Comic beleuchtet die komischen Missverständnisse, die entstehen können, wenn jemand, der wenig bis keine Erfahrung mit moderner Technologie hat, versucht, sie zu nutzen. In der ersten Szene gibt der erste Mann vor, sich mit Computern auszukennen, wahrscheinlich um kompetent oder modern zu wirken. Im zweiten Bild wird sein Mangel an technischem Wissen offensichtlich, als er versucht, mit dem Computer zu sprechen, als wäre es ein menschliches Wesen oder ein Gerät aus einem Science-Fiction-Film. Das dritte Bild liefert die Pointe: Der Verkäufer scherzt, dass es noch Jahrzehnte dauern könnte, bis die Technologie auf dem Niveau ist, dass sie Sprachbefehle in der Art, wie der erste Mann es sich vorstellt, versteht. Der Begriff "Commander" könnte sich auf Star Trek oder ähnliche Sci-Fi-Serien beziehen, in denen Charaktere mit Computern sprechen.

Insgesamt ist der Comic eine humorvolle Darstellung der Kluft zwischen den Erwartungen der Menschen und der tatsächlichen Technologie, insbesondere in den Anfangszeiten des Internets und der persönlichen Computer. Es zeigt auch, wie Menschen manchmal vorgeben, mehr zu wissen, als sie tatsächlich tun, nur um im Trend zu sein oder sich nicht uninformiert zu fühlen.

Was ist Computerlinguistik?



Neuzeichnung durch GPT4:



Computerlinguistik in der Anwendung

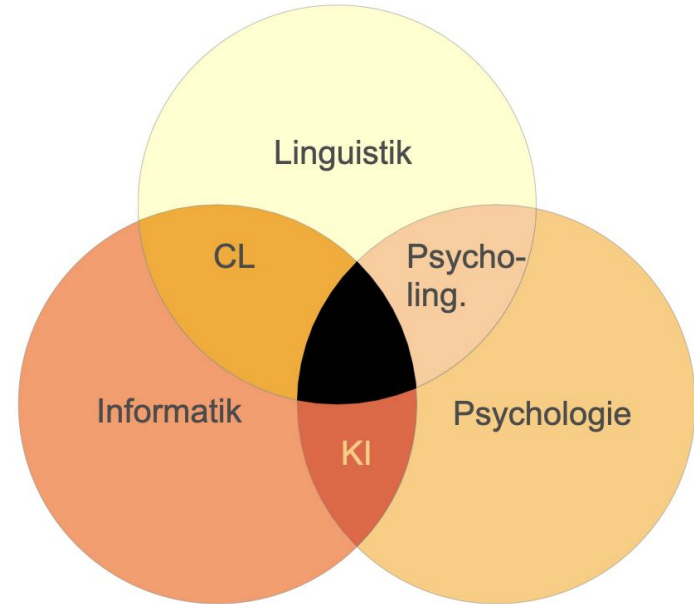
- Mensch-Maschine-Kommunikation
 - Spracherkennung, Sprachsynthese, Dialogsysteme (Assistenten)
- Maschinelle bzw. computergestützte Übersetzung
 - Vollautomatische MÜ, Übersetzungsspeicher, Terminologie-Datenbanken
- Information Retrieval & Text Mining
 - Text Classification, Text Summarization, Information Extraction
- Aufbau und Nutzung von Ressourcen
 - (Annotierte) Korpora, Lexika, Wortnetze, Baumbanken

Fragestellungen

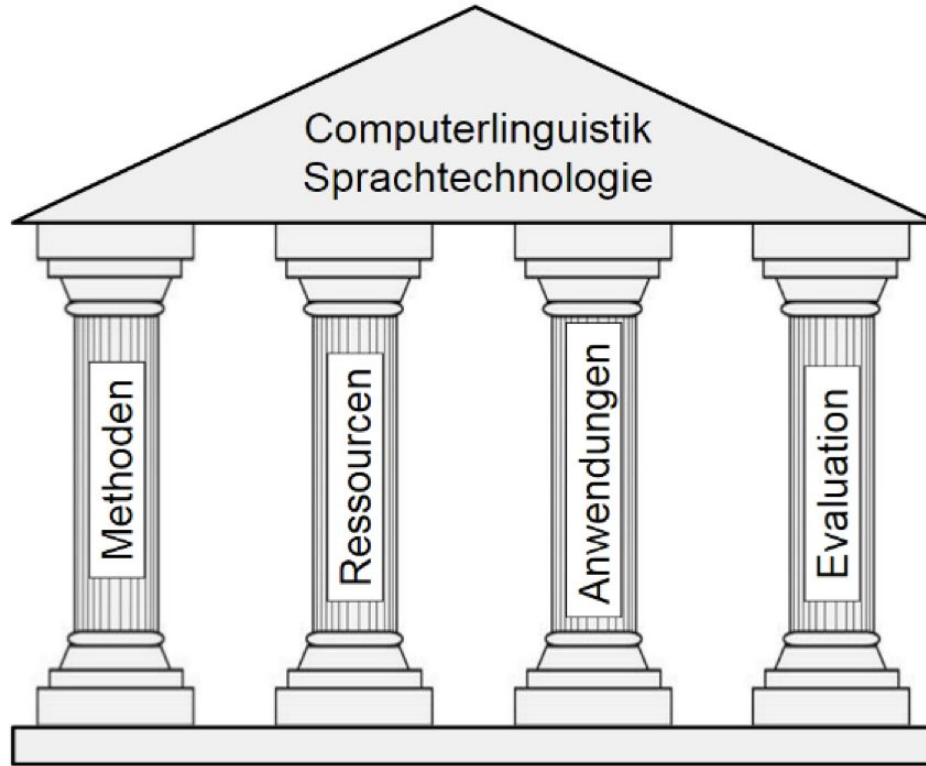
- Computerlinguistik als...
 - Interesse der Sprachwissenschaft: Modelle der Grammatik
 - Interesse der Kognitionswissenschaft: Modelle der menschlichen Sprachfähigkeit
 - Interesse der Ingenieurwissenschaft: Sprachtechnologische Anwendungen

Nachbardisziplinen

- Linguistik (Gegenstand, Terminologie)
- Informatik (Algorithmen, Datenstrukturen)
- Philosophie (Verhältnis Sprache-Denken)
- Künstliche Intelligenz (Wissensrepräsentation und -verarbeitung)
- Kognitionswissenschaft (Sprache-Denken, Architekturen)
- Mathematik (Logik, Graphen, Statistik, Optimierung)



Grundpfeiler der Computerlinguistik



Methoden

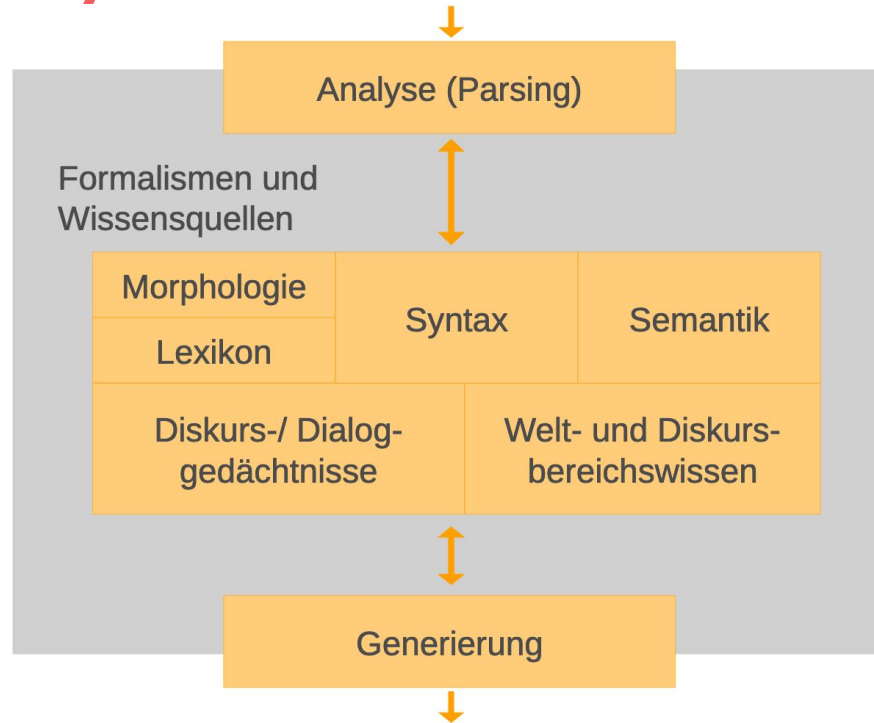
➤ Methoden der Linguistik

- Identifikation sprachlicher Einheiten auf unterschiedlichen Ebenen
 - Zeichenebene (Graphemik / Phonetik / Phonologie)
 - Wortebene (Morphologie)
 - Satzebene (Syntax)
- Untersuchung des Zusammenspiels der Ebenen mit der Ebene der Bedeutung (Semantik)

➤ Methoden der Informatik

- Nutzung etablierter Methoden zur Datenspeicherung und -verarbeitung
- Darunter Automaten- und Graphentheorie, formale Sprachen, Netz- und Vektorrepräsentationen, maschinelle Lernverfahren
- Aktuell vor allem Deep-Learning-Architekturen

Architektur eines natürlichsprachlichen Systems – Symbolischer Ansatz



Architektur eines natürlichsprachlichen Systems – Transformer-based



Architektur eines natürlisprachlichen Systems – Transformer-based

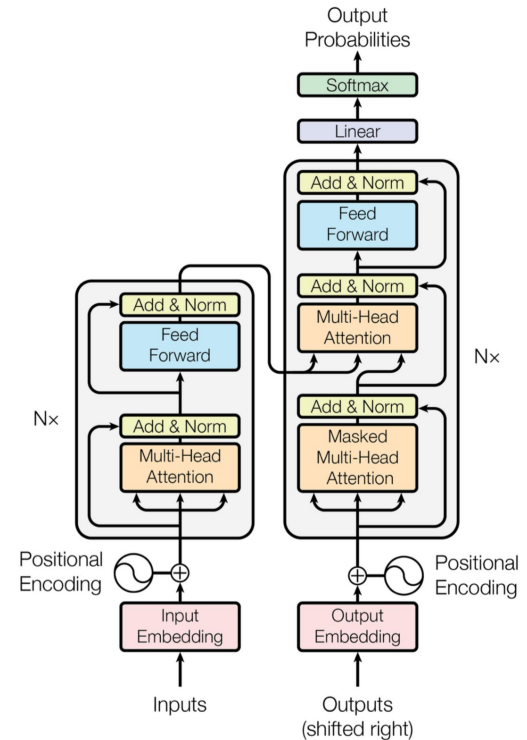


Figure 1: The Transformer - model architecture.

Forschung und Anwendung

➤ Theoretische Computerlinguistik

- Suche nach komplexen Beschreibungsformalismen, die gleichzeitig handhabbar und effizient sind.
- Abwägung über das Verhältnis von linguistischer Beschreibung und Berechnungsoperationen.

➤ Praktische Computerlinguistik

- Entwicklung und Erforschung realitätsnaher Anwendungen.
- Natürlichsprachliche Kommunikation mit oder auf der Maschine

Semesterprogramm

- Allgemeine Einführung in die Grundbegriffe von Computerlinguistik, Sprachwissenschaft und Annotation
- Beschreibungsebenen der (Computer)Linguistik:
Zeichenebene, Wortebene, Satzebene, Textebene
- Annotation zur Erstellung von Ressourcen für 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.