

Ivana Daskalovska

Willkommen zur Übung

Einführung in die

Computerlinguistik



Kontakt:
ivana.bt.mk@gmail.com

Betreff: EICL

Wiederholung



- Was ist Computerlinguistik?

- Was ist Computerlinguistik?
 - **Computerlinguistik** ist die wissenschaftliche Lehre, die Modelle und Methoden für automatische Bearbeitung natürlicher Sprache entwickelt.

- Nennen Sie die zwei große Teilbereiche der Computerlinguistik.

- Nennen Sie die zwei große Teilbereiche der Computerlinguistik.
 - **Theoretische Computerlinguistik:** Teildisziplin der Linguistik, die **formale berechenbare Modelle natürlicher Sprache** entwickelt, implementiert und untersucht
 - **Angewandte Computerlinguistik:** interdisziplinäres Forschungsgebiet (Linguistik, Informatik), das **konkrete Algorithmen für die maschinelle Sprachverarbeitung** entwickelt (maschinelle Übersetzung, Spracherkennung ...)

- **Wo findet die Computerlinguistik Anwendung?**

- **Wo findet die Computerlinguistik Anwendung?**
 - Häufigkeitsanalysen von Vorkommen von Wörtern
 - Internetsuchmaschinen
 - Lexikographie
 - Übersetzungssysteme
 - Automatische Spracherkennung
 - Informationsextraktion usw.

Nachbardisziplinen der Computerlinguistik?

Nachbardisziplinen der Computerlinguistik?

- Linguistik
- Informatik
- Philosophie
- Künstliche Intelligenz
- Kognitionswissenschaft
- Mathematik

Was ist Linguistik und womit beschäftigt sie sich?

Was ist Linguistik und womit beschäftigt sie sich?

- **Linguistik** (lateinisch *lingua* ‚Sprache‘, ‚Zunge‘), ist die Wissenschaft, die in verschiedenen Herangehensweisen die **menschliche Sprache untersucht**.
- Sie untersucht die Sprache als System, ihre einzelnen Bestandteile und Einheiten sowie deren Bedeutungen.
- Des Weiteren beschäftigt sich die Sprachwissenschaft mit Entstehung, Herkunft und geschichtlicher Entwicklung von Sprache, mit ihrer vielseitigen Anwendung in der schriftlichen und mündlichen Kommunikation.

- **Nennen Sie die Teilgebiete der Linguistik und beschreiben Sie sie kurz.**

- **Nennen Sie die Teilgebiete der Linguistik und beschreiben Sie sie kurz.**

- **Phonetik und Phonologie:**

- Lautstruktur natürlicher Sprachen
- Artikulatorische Merkmale
- **Computerlinguistische Methoden:** Spracherkennung, Sprachsynthese

- **Morphologie**

- Bildung und Struktur von Wörtern
- Prozesse, die für die unterschiedliche Erscheinungsformen an der Oberfläche verantwortlich sind
- **Computerlinguistische Methoden:** Wortzerlegung(DEA), Wortartenbestimmung (HMM)

➤ Syntax

- Strukturbildung von Sätzen
- **Computerlinguistische Methoden:** Parsing, computerlesbare Grammatiken (GFGs)

➤ Semantik

- Bedeutung sprachlicher Einheiten
- **Computerlinguistische Methoden:** Wissensdatenbanken, automatische Semantische Analyse

➤ Pragmatik

- Zweck einer Äußerung in der Welt
- **Computerlinguistische Methoden:** Koreferenzresolution, Kontextmodellierung

➤ **Korpuslinguistik**

- Methode, die auf alle Beschreibungsebenen angewandt werden kann
- **Wortartendisambiguierung** (Tagging)
- **syntaktische Analyse** (Parsing)
- **semantische Lesartendisambiguierung**
- **maschinelle Übersetzung**

Was versteht man unter Maschinelle Übersetzung?

- **Was versteht man unter Maschinelle Übersetzung?**
 - **Maschinelle Übersetzung** (MÜ oder MT für engl. *machine translation*), auch als **automatische Übersetzung**, bezeichnet die **Übersetzung von Texten durch ein Computerprogramm**.

- **Was ist Transfer?**

- **Was ist Transfer?**

- **Transfer:** Direkte Übersetzung von sprachlichen Elementen, ohne Bedeutungsambiguitäten aufzulösen.

- **Was ist Interlingua?**

- **Was ist Interlingua?**

- **Interlingua:** Sprachunabhängige Repräsentation von Bedeutung, in die Sprache überführt werden kann und umgekehrt..

- **Wodurch haben Sprachverarbeitungssysteme Schwierigkeiten?**

- **Wodurch haben Sprachverarbeitungssysteme Schwierigkeiten?**

- **Variabilität:** Die **selbe Bedeutung** kann durch **viele sprachliche Formen** ausgedrückt werden.
- **Ambiguität:** **Dieselbe sprachliche Form** kann **verschiedene Informationen** ausdrücken (erst durch den Kontext kann erschlossen werden, was gemeint ist).

- **Welche Typen von Ambiguität gibt es?**

- **Welche Typen von Ambiguität gibt es?**
 - **Phonetische Ambiguität (Homophone):**
Miene - Mine, Meer - mehr, viel - fiel ⇒ **Unterschiedliche Wörter** haben **dieselbe lautliche Form**.
 - **Orthographische Ambiguität (Homographen):** übersetzen - übersetzen, umfahren - um-fahren ⇒ **Unterschiedliche Wörter** werden **gleich geschrieben**.
 - **Lexikalische Ambiguität (Homonyme):** Maria geht zur Bank. ⇒ **Ein Wort** hat mehrere **verschiedene Bedeutungen**.
 - **Morphologische Ambiguität:** Staub-ecken - Stau-becken ⇒ **Eine Wortform** kann **auf unterschiedliche Arten analysiert** werden

➤ **Strukturelle/syntaktische Ambiguität:**

Peter fuhr seinen Freund sturzbetrunknen nach Hause.

⇒ Die Grammatikregeln lassen **verschiedene** Analysen zur **Kombination der Satzglieder** zu.

➤ **Kompositionell-semantische Ambiguität bzw. Skopusambiguität**

Alle Politiker sind nicht korrupt.

⇒ **Quantifikatoren (alle, jeder, zwei)** und **Negationen** können sich auf **verschieden große Satzglieder** beziehen.

➤ **Pragmatische Ambiguität:**

Haben Sie eine Uhr?

⇒ Der **Bezug einer Aussage zum außerlinguistischen Kontext** kann auf mehrere Arten hergestellt werden.