

Einführung in die Computerlinguistik Überblick

Robert Zangenfeind

Center for Information and Language Processing

2023-10-16

Die Grundfassung dieses Foliensatzes wurde von Dr. Benjamin Roth erstellt. Fehler und Mängel liegen ausschließlich in meiner Verantwortung.

1 Organisatorisches

2 Linguistik

- Vorlesung/Übung
 - Dr. Robert Zangenfeind
 - Dr. Robert Litschko
 - Yihong Liu
- Tutorium/Übungsaufgaben (Moodle)
 - Elisa Rieß (Tutorium 1)
 - Adrian Mülthaler (Tutorium 2)
 - Barbara Pfeffer (Moodle-Betreuung)
- Sie erreichen uns unter der gemeinsamen Adresse:
eic12324 (at) cis.uni-muenchen.de

- Zur Klärung von Fragen zu Übungsaufgaben und zur Vorlesung:
- Übung: Montag, 15–15:45h
- Tutorien: ?

- Bearbeitung in Moodle
- Freischaltung: Jeweils Montag nach der Übung.
- Abgabetermin: Montag (eine Woche später) vor der Vorlesung.
- Übungsblätter müssen von den Teilnehmern **eigenständig** bearbeitet werden.
- Klausurbonus: Abhängig von den erreichten Übungspunkten wird ein Klausurbonus von bis zu 10% der maximal erreichbaren Klausurpunkte gewährt.

- Lektüre: für die meisten Vorlesungen wird es einen ergänzenden kurzen Abschnitt (je ca. 10 Seiten) aus einem Lehrbuch geben. Dieser wird auf der Moodle-Seite bekanntgegeben und soll bis zum angegebenen Datum gelesen werden.
- Inverted classroom: ergänzende Internet-Tutorials, die auf der Moodle-Seite verlinkt und bis zum angegebenen Datum angesehen werden sollen.

Teil 1: Sprachwissenschaft

- ① Phonetik/Phonologie
Merkmale sprachlicher Laute, Lautsystem, Lautstrukturen
- ② Morphologie
Wortbildung, Flexion, Wortarten, Wortstrukturen
- ③ Syntax
Größere sprachliche Einheiten und deren Zusammenhang
Satzstrukturen
- ④ Semantik
Bedeutung sprachlicher Einheiten
Bedeutungsstrukturen
- ⑤ Pragmatik
Sprache im kommunikativen Kontext
Kommunikative Bedeutung

Teil 2: Statistical NLP

Probekl. voraussichtlich 29.1.2024

Klausur voraussichtlich 5.2.2024

Moodle

Fragen?

Herausforderungen für die Computerlinguistik: Variabilität und Ambiguität (Mehrdeutigkeit)

Schwierigkeiten für Sprachverarbeitungssysteme:

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

Typen von Ambiguität (1)

- **Phonetische Ambiguität (Homophone):**
Miene – Mine, Meer – mehr, viel – fiel
⇒ Unterschiedliche Wörter haben dieselbe lautliche Form.
- **Orthographische Ambiguität (Homographen):**
übersetzen – übersetzen, umfahren – umfahren
⇒ Unterschiedliche Wörter werden gleich geschrieben.
- **Morphologische Ambiguität:**
Staub-ecken – Stau-becken
⇒ Eine Wortform kann auf unterschiedliche Arten in seine Bestandteile zerlegt werden.
- **Lexikalische Ambiguität (Homonyme):**
*Maria geht zum **Ball**.*
⇒ Ein Wort hat mehrere verschiedene Bedeutungen.

Typen von Ambiguität (2)

- Strukturelle/syntaktische Ambiguität:

- ① *Visiting relatives can be boring.*
- ② *Peter fuhr seinen Freund sturzbetrunk nach Hause.*
- ③ *Sie traf die Nachbarin mit ihrer Tochter.*

⇒ Die Grammatikregeln lassen verschiedene Analysen zur Kombination der Satzelemente zu.

- Kompositionell-semantische Ambiguität bzw. Skopusambiguität:

- ① *Die zwei Mitarbeiter müssen vier Sprachen beherrschen.*
- ② *Some student likes every course.*
- ③ *Alle Politiker sind nicht korrupt.*

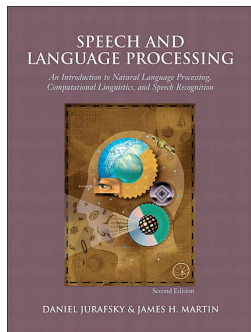
⇒ Quantifikatoren (z.B. *alle, jeder, zwei*) und Negationen können sich auf verschieden große Satzteile beziehen.

Typen von Ambiguität (3)

- **Pragmatische Ambiguität:**

- ① *Könnten Sie die Aufgabe lösen?*
- ② *Haben Sie eine Uhr?*

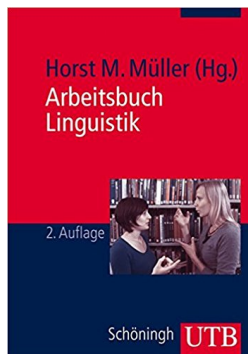
⇒ Der Bezug einer Aussage zum außersprachlichen Kontext kann auf mehrere Arten hergestellt werden.



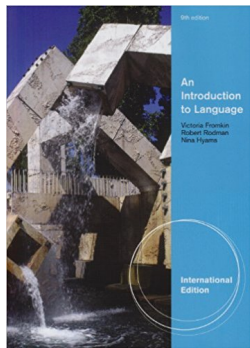
- D. Jurafsky & J. Martin: Speech and Language Processing. Pearson. Prentice Hall. 2009.
(Elektronische Version, 3. Auflage 2020: <https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf>)
(Lehrbuchsammlung)



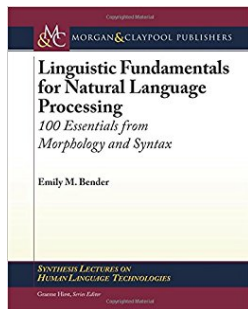
- K.U. Carstensen et al.: Computerlinguistik und Sprachtechnologie. Eine Einführung. Heidelberg 2010 (3. Auflage)
(Elektronische Version:
<https://login.emedien.ub.uni-muenchen.de/login>)



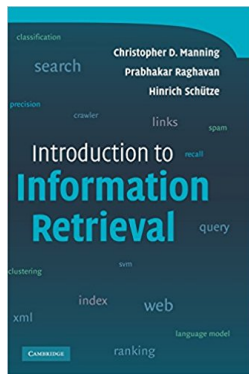
- H. Müller: Arbeitsbuch Linguistik. Schöningh / UTB. 2009.
(Elektronische Version: s.o.)



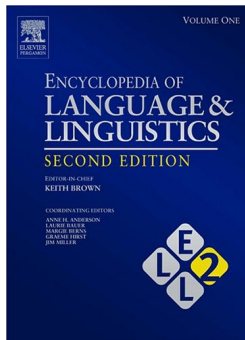
- V. Fromkin, R. Rodman, N. Hyams: An Introduction to Language. Boston et al. 2011.



- E. Bender: Linguistic Fundamentals for Natural Language Processing. Morgan & Claypool. 2013.



- Chr. Manning, P. Raghavan, H. Schütze: Introduction to Information Retrieval. Cambridge University Press. 2009.
<https://nlp.stanford.edu/IR-book/>



- K. Brown (ed.): Encyclopedia of Language & Linguistics. Elsevier. 2006.

- ① Anmeldung im Moodle
- ② Aufgabenbearbeitung im Moodle
- ③ Anmeldung für den Email-Verteiler für Studierende (optional)
 - Ankündigung von Vorträgen und Events am CIS (wissenschaftliche Vorträge und Firmenvorträge)
 - Jobs (als Tutor oder wissenschaftliche Hilfskraft)
 - <http://www.cis.uni-muenchen.de/ba/erstsemester-infos/index.html#verteiler>