

Einführung in die Computerlinguistik

Was ist Computerlinguistik?

Dr. Benjamin Roth

CIS LMU München

Gliederung

- 1 Organisatorisches
- 2 Was ist Computerlinguistik?
- 3 Vorlesungsplan
- 4 Sprachwissenschaft
- 5 Computerlinguistische Methoden
- 6 Computerlinguistische Anwendungen / Sprachtechnologie
- 7 Geschichte, Herausforderungen und Forschungsstand der Computerlinguistik

Outline

- 1 Organisatorisches
- 2 Was ist Computerlinguistik?
- 3 Vorlesungsplan
- 4 Sprachwissenschaft
- 5 Computerlinguistische Methoden
- 6 Computerlinguistische Anwendungen / Sprachtechnologie
- 7 Geschichte, Herausforderungen und Forschungsstand der Computerlinguistik

Dozent und Tutoren

- Dozent:
 - ▶ Dr. Benjamin Roth
Sprechstunde: Dienstags 14:00-15:00, Raum C 105
- Tutoren:
 - ▶ Simon Preißner
(Moodle & VL)
 - ▶ Falk Spellerberg
(Moodle)
 - ▶ Ivana Daskalovska
(Extra Tutorat)
- Sie erreichen den Dozenten und die Tutoren unter:
`eic11718@cis.uni-muenchen.de`

Übungstermine

- Zur Klärung von Fragen zu Übungsblättern und Vorlesung.
- Tafelübung:
 - ▶ Freitags 11:00-11:45
- Extra-Tutorat (Ivana Daskalovska):
 - ▶ Mittwochs 12:15-13:45

Übungsaufgaben

- Bearbeitung in Moodle
- Freischaltung: Jeweils Freitags nach der Vorlesung.
- Bearbeitungsfrist: Freitags (eine Woche später) vor der Vorlesung.
- Übungsblätter müssen von den Teilnehmern **eigenständig** bearbeitet werden.
- Klausurbonus: In Abhängigkeit der erreichten Übungspunkte wird ein Klausurbonus von bis zu 10% der maximal erreichbaren Klausurpunkte gewährt, **wenn die Klausur auch ohne die Bonuspunkte als bestanden gewertet würde.**

Sonstiges

- Für die meisten Vorlesungen wird es einen zu lesenden kurzen (je ca. 10 Seiten) Abschnitt aus einem Lehrbuch geben, den Sie im **Vorlesungsordner in der Institutsbibliothek** (Oettingenstr.) finden können.
- Vor Weihnachten wird es eine Probeklausur geben.

Outline

- 1 Organisatorisches
- 2 Was ist Computerlinguistik?**
- 3 Vorlesungsplan
- 4 Sprachwissenschaft
- 5 Computerlinguistische Methoden
- 6 Computerlinguistische Anwendungen / Sprachtechnologie
- 7 Geschichte, Herausforderungen und Forschungsstand der Computerlinguistik

Was ist Computerlinguistik?

Definition (Hinrich Schütze)

Computational linguistics is the scientific study of models and methods for automatic processing of natural language. Computational linguistics is an interdisciplinary field that shares a large part of its subject matter with computer science and linguistics. However, computational linguists also work on theories, models and methods that are not part of core linguistics or core computer science.

Bereiche der Computerlinguistik

Zwei große Teilbereiche:

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

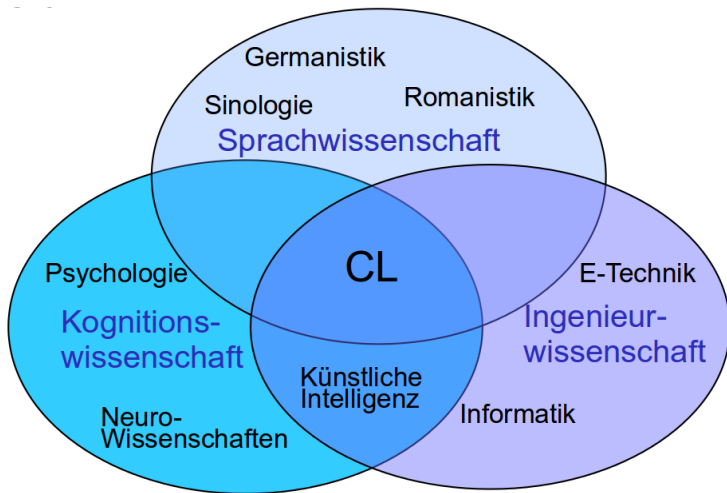
Anwendungen der Computerlinguistik

- Häufigkeitsanalysen von Vorkommen von Wörtern und linguistischen Phänomenen
- Lexikographie (Thesauri, Wörterbücher)
- Internetsuchmaschinen: sehr große Menge an Information, aber hochgradig unstrukturiert → direkter Zugang zu relevanten Daten ist schwierig.
- Dialoganwendungen: Zugang zu komplexen Systemen, z.B. Bestellung eines Bahn- oder Flugtickets, Interaktion mit Bank, auch mit natürlichsprachlichen Anwendungen
- Übersetzungssysteme: fremdsprachige Web-Seiten, Gebrauchsanweisungen, Social Media etc.
- automatische Silbentrennung, Rechtschreibprüfung und -korrektur
- automatische Spracherkennung
- Informationsextraktion, z.B. relevante Qualifikationen aus Bewerbungsschreiben und Lebensläufen maschinell extrahieren

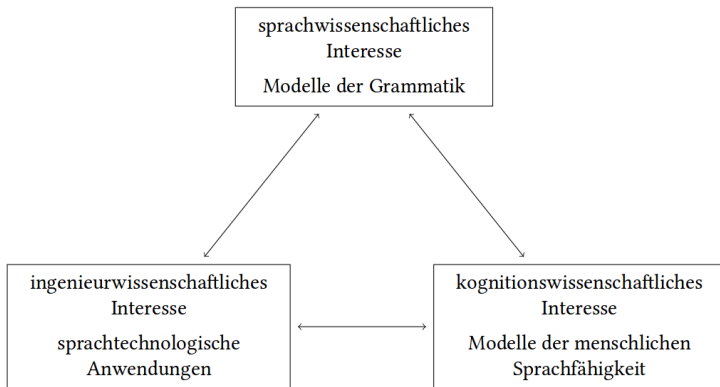
Berufsfelder für Computerlinguisten

- Verarbeitung gesprochener Sprache für die Interaktion mit Computern
- Verarbeitung von Texten (suchen, bearbeiten und verwalten)
- Einsatz sprachtechnologischer Software und Ressourcen (in Verlagen, Übersetzungsbüros, Verwaltungen etc.): Maschinelle Übersetzung, elektronische Wörterbücher, Spracherkennung, Sprachgenerierung, lexikonbasierte Optimierung von Optical-Character-Recognition-Verfahren (OCR)
- akademischer Bereich
- Bedarf an Experten steigt tendenziell

Disziplinen / Untersuchungsgegenstände



Disziplinen / Untersuchungsgegenstände



Typische Forschungsgegenstände

- Entwicklung von Methoden (Theorie)
- Entwicklung realistischer Anwendungen (Praxis)
- Aufbau und Verwaltung großer wiederverwendbarer Korpora (Daten)
- Konzeption effektiver Evaluierungsmechanismen (Experimente)

Nachbardisziplinen (1)

- Linguistik
 - ▶ Die Wissenschaft, die sich mit menschlicher Sprache beschäftigt
 - ▶ Grundinventar linguistischer Termini
 - ▶ Teilgebiete: Phonetik/Phonologie, Morphologie, Syntax, Semantik, Pragmatik; Korpuslinguistik
- Informatik (Algorithmen, Datenstrukturen, Software Engineering)
- Philosophie (Verbindung von Sprache, Denken und Handeln; Relation zu außersprachlichen Gegebenheiten)
- Künstliche Intelligenz (knowledge representation, reasoning, learning)

Nachbardisziplinen (2)

- Kognitionswissenschaft (Sprachbeherrschung ist spezieller Teilbereich der kognitiven Fähigkeiten des Menschen)
- Mathematik
 - ▶ Insbesondere: Logik, Wahrscheinlichkeitstheorie, Statistik, Graphentheorie
- Sprache ist oft nicht logisch:

(1) *Ein großer Berg* vs. *Eine große Ameise*

→ Vagheit des Adjektivs (kein Problem für Menschen) → Logik muss modifiziert werden

(2) *Vögel fliegen.* / *Pinguine sind Vögel.* / *Pinguine fliegen.*

→ scheinbar widersprüchliche Aussagen (Mensch hat wenig Probleme damit)

Outline

- 1 Organisatorisches
- 2 Was ist Computerlinguistik?
- 3 Vorlesungsplan**
- 4 Sprachwissenschaft
- 5 Computerlinguistische Methoden
- 6 Computerlinguistische Anwendungen / Sprachtechnologie
- 7 Geschichte, Herausforderungen und Forschungsstand der Computerlinguistik

Vorlesungsgliederung

Teil 1: Sprachwissenschaft

Klassische Aufteilung von sprachlicher "Form" zu kommunikativer "Funktion"

Teil 2: Computerlinguistische Methoden

Computerlinguistische Techniken, die in verschiedenen Kontexten genutzt werden

Teil 3: Computerlinguistische Anwendungen

Praktische Anwendungen, wie z.B. automatische Übersetzungssysteme.

Vorlesungsgliederung

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: Computerlinguistische Methoden

- 1 Reguläre Sprachen, Endliche Automaten
Beschreibungsmittel für einfache Zeichenketten.
- 2 Hidden Markov Models, Wortartenzuweisung
Statistisches Verfahren, Wortarten zu bestimmen.
- 3 Kontextfreie Grammatiken, Parsing
Automatische syntaktische Analyse.

Teil 3: Computerlinguistische Anwendungen

- 1 Maschinelle Übersetzung.
- 2 Suchmaschinen.

Outline

- 1 Organisatorisches
- 2 Was ist Computerlinguistik?
- 3 Vorlesungsplan
- 4 Sprachwissenschaft**
- 5 Computerlinguistische Methoden
- 6 Computerlinguistische Anwendungen / Sprachtechnologie
- 7 Geschichte, Herausforderungen und Forschungsstand der Computerlinguistik

Sprachliche Einheiten

geschrieben	gesprochen
Buchstabe	Laut
Silbe	Silbe
Wort	Wort
Satz	Äußerung
Paragraph	Discourse

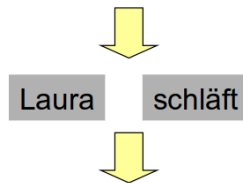
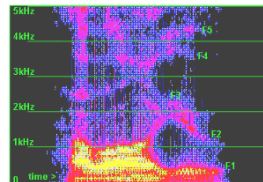
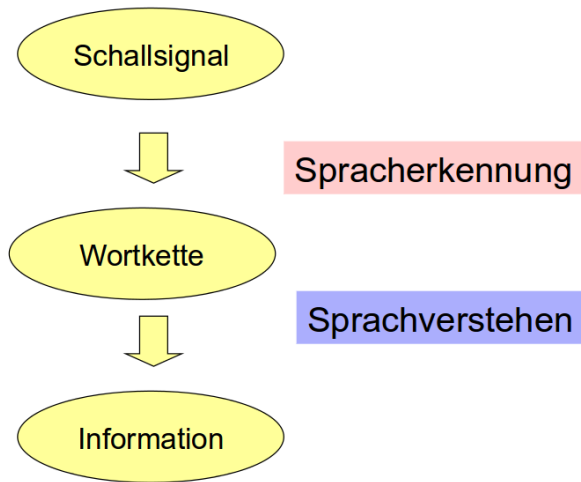
beschreibt Strukturen der Sprache(n) auf den Ebenen

- des Lautsystems
- des Aufbaus von Wörtern
- der Kombination von Wörtern zu Phrasen und Äußerungen
- der Bedeutung dieser Äußerungen, insbesondere im (außersprachlichen) Kontext (\Rightarrow Pragmatik)
- erklärt das Funktionieren dieser Strukturen als Kommunikationsmittel
- zunächst einzelsprachlich, aber durch Vergleich und Generalisierungen auch sprachübergreifend (universell)

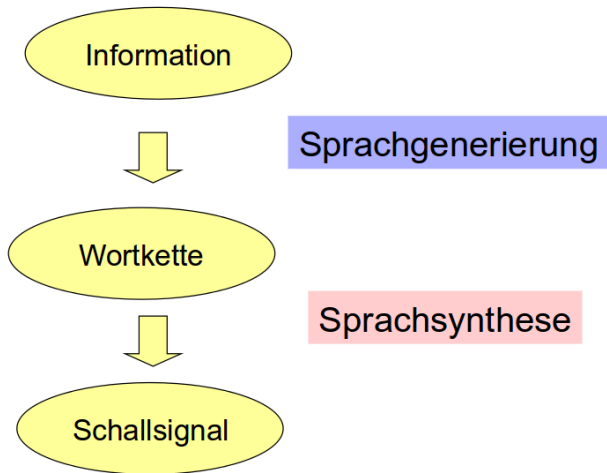
Sprachliche Einheiten

Einheit	Teildisziplin
Laut	Phonetik, Phonologie
Silbe	Phonetik, Phonologie
Wort	Morphologie
Phrase	Syntax
Satz, Äußerung	Syntax, Semantik
Discourse	Pragmatik

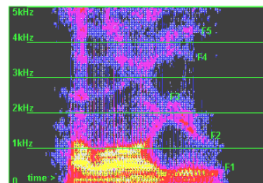
Verstehen von Sprache



Erzeugen von Sprache



Laura schläft



Teilgebiete der Linguistik

- Phonetik und Phonologie
- Morphologie
- Syntax
- Semantik
- Pragmatik
- Jedes dieser Teilgebiete hat auch eine Entsprechung in der Computerlinguistik.

Phonetik und Phonologie

- artikulatorische Merkmale
- Lautstruktur natürlicher Sprachen
- Spracherkennung: Erkennung und Produktion gesprochener Sprache
- modellieren, welche Segmente ein Wort enthält und wie sich deren Struktur auf die Aussprache auswirkt
- z.B. wenn ein im Prinzip stimmhafter Konsonant am Wortende stimmlos wird (“Auslautverhärtung”):

(3) *Dieb* /Di:p/ vs. *Diebe* /Di:be/

Morphologie

- Bildung und Struktur von Wörtern
- Was ist die lexikalische Wurzel einzelner Wörter?
- Welche Prozesse sind verantwortlich für unterschiedliche Erscheinungsformen an der Oberfläche?
- Veränderung der Verwendung und Bedeutung des Wortes durch Oberflächenmodifikationen
- z.B. Suffix -e als Pluralmarkierung:

(4) *Dieb-e* → *Dieb-pl* → “Mehr als ein Dieb”

- Strukturbildung von Sätzen
- traditionell am stärksten vertretene Teildisziplin der Computerlinguistik
- Erkennung von Grammatikalität und darauf folgende Bedeutungserschließung
- z.B.

(5) *Der gewitzte Dieb stahl das Geld.*

vs.

**Der Dieb gewitzte stahl das Geld.*

- Bedeutung sprachlicher Einheiten (Wort, Satz etc.)
- z.B.

(6) *Die Polizei beschlagnahmte das Diebesgut.*

vs.

Das Diebesgut beschlagnahmte die Polizei.

→ gleiche Bedeutung

- Zweck einer Äußerung in der Welt, z.B.
Wissen Sie, wie spät es ist?
- Bestimmung des Bezugs von Wörtern: Antezedens eines Pronomens, z.B.:
Die Katze schnurrt. **Sie** hat Hunger.
- implizite Annahmen (Präsuppositionen), z.B.:
“Welche Drogen hat Peter genommen?”
Präsupponiert: Peter hat Drogen genommen.

- Methode, die auf alle Beschreibungsebenen angewandt werden kann
- seit Anfang 1980er
- Fortschritte bei Erkennung gesprochener Sprache
- Wortartendisambiguierung (Tagging)
- syntaktische Analyse (Parsing)
- semantische Lesartendisambiguierung (z.B. *Bank 1* vs. *Bank 2*)
- maschinelle Übersetzung

Text corpus

Definition

A corpus (plural corpora) or text corpus is a large and structured set of texts, nowadays usually electronically stored and processed.

- Corpora are used to do statistical analysis and hypothesis testing, checking occurrences or validating linguistic rules within a specific language territory.
- A corpus may contain texts in a single language (monolingual corpus) or text data in multiple languages (multilingual corpus).
- (from Wikipedia)

Outline

- 1 Organisatorisches
- 2 Was ist Computerlinguistik?
- 3 Vorlesungsplan
- 4 Sprachwissenschaft
- 5 Computerlinguistische Methoden**
- 6 Computerlinguistische Anwendungen / Sprachtechnologie
- 7 Geschichte, Herausforderungen und Forschungsstand der Computerlinguistik

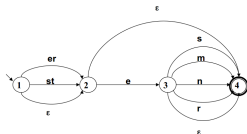
Für jede Sprachliche Beschreibungsebene gibt es passende Computerlinguistische Methoden

In dieser Vorlesung:

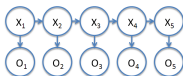
- Phonetik - Signalverarbeitung (Spracherkennung - Sprachsynthese)
- Morphologie - **Wortzerlegung, Wortartenbestimmung**
- Syntax - **computerlesbare Grammatiken, automatische Syntaktische Analyse**
- Semantik - Wissensdatenbanken, automatische semantische Analyse
- Pragmatik - Koreferenzresolution, Kontextmodellierung (Dialogsysteme, Sprachliche Schnittstellen z.B. in der Robotik))

Computerlinguistische Methoden zu den Beschreibungsebenen

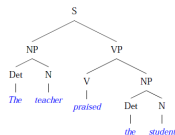
- **Wortzerlegung** - Endliche Automaten



- **Wortartenbestimmung** - HMMs



- **computerlesbare Grammatiken** - CFGs



- **automatische Syntaktische Analyse** - Parsing

Mehr Terminologie

Verarbeitung	Modalität	Mensch	Maschine
Produktion	Lautsprache Schriftsprache	Sprechen Schreiben	Synthese Generierung
Rezeption	Lautsprache Schriftsprache	Hören Lesen	Erkennung Analyse

Outline

- 1 Organisatorisches
- 2 Was ist Computerlinguistik?
- 3 Vorlesungsplan
- 4 Sprachwissenschaft
- 5 Computerlinguistische Methoden
- 6 Computerlinguistische Anwendungen / Sprachtechnologie**
- 7 Geschichte, Herausforderungen und Forschungsstand der Computerlinguistik

Anwendungen kombinieren oft viele verschiedene Methoden, um eine Bestimmte Aufgabe zu lösen

- Spracherkennung (Diktiersysteme, Transkription z.B. Untertitel)
- Sprachsynthese
- Rechtschreibkorrektur
- Maschinelle Übersetzung DeepL:
<https://www.deepl.com/translator>
Google translate: <https://translate.google.com/>
- Sprachlernsysteme
- Automatische Zusammenfassung (Summarisation)
- Suchmaschinen / Information Retrieval

Anwendungen kombinieren oft viele verschiedene Methoden, um eine Bestimmte Aufgabe zu lösen

- Dokumentklassifikation
- Strukturierte Gliederung von Information / Relations-Extraktion
- Frage-Beantwortung (Question Answering)
Start-System (MIT): <http://start.csail.mit.edu/>
- Sentiment-Analyse
Bsp.: Ist dies eine gute oder eine schlechte Bewertung? “Der Film hat mich ja nicht so richtig begeistert, auch wenn manche behaupten, er wäre ganz toll.”
- Dialogsysteme
 - ▶ Telefonie-Systeme: Telefonbanking, Fahrplanauskunft
 - ▶ Gerätebedienung
 - ▶ Interaktion mit virtuellen Agenten und Robotern
- Weitere Anwendungen?

Geschichte der Methoden der Maschinellen Übersetzung (Machine Translation, MT)

- Wörter nachschlagen, aneinanderreihen
- Morphologische Anpassungen
- Syntaktische Umstellungsregeln
- Volle syntaktische Analyse (“parsing”)

⇒

Transfer

- Semantische Analyse (Disambiguierung)

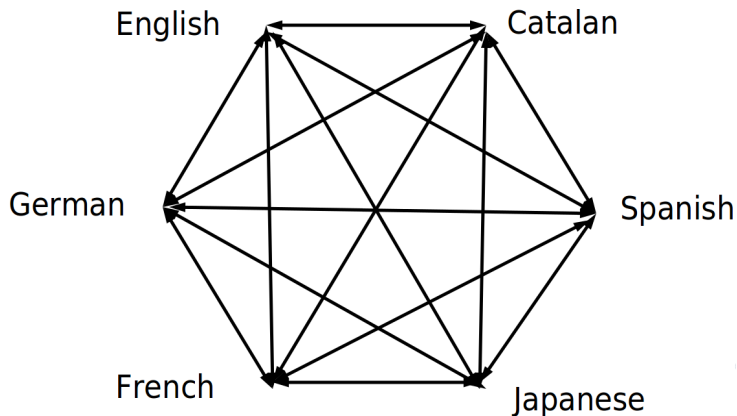
⇒

Interlingua

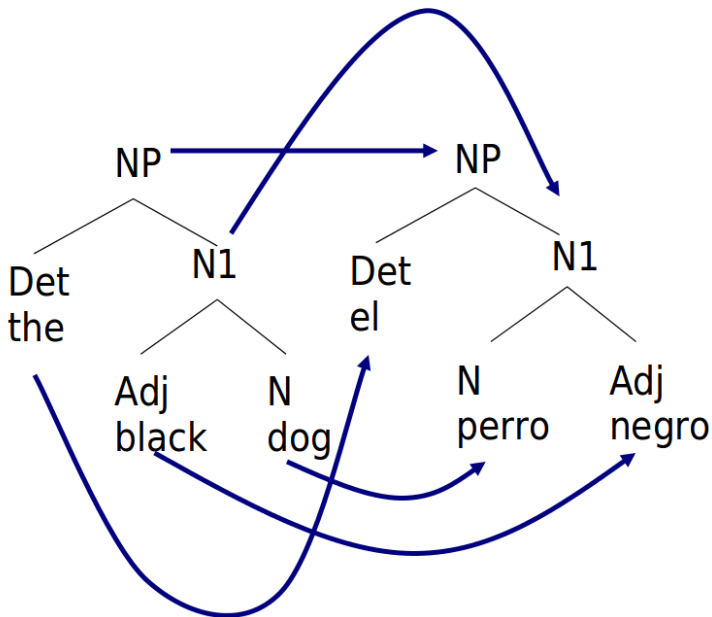
- Unterstützung durch Welt-Wissen
- Übersetzen mit Statistik/Wahrscheinlichkeiten

MT Transfer

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

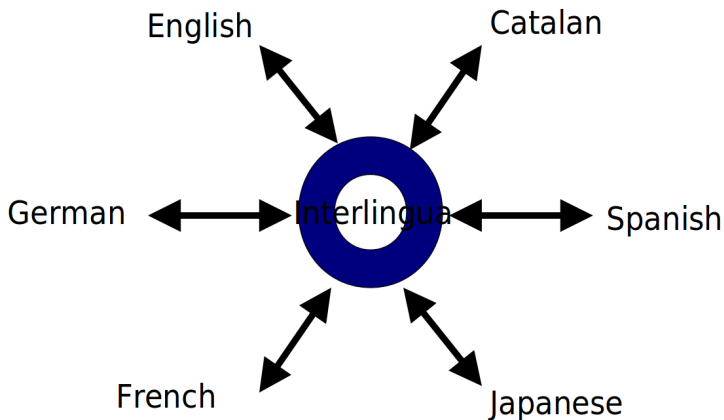


MT Syntaktischer Transfer

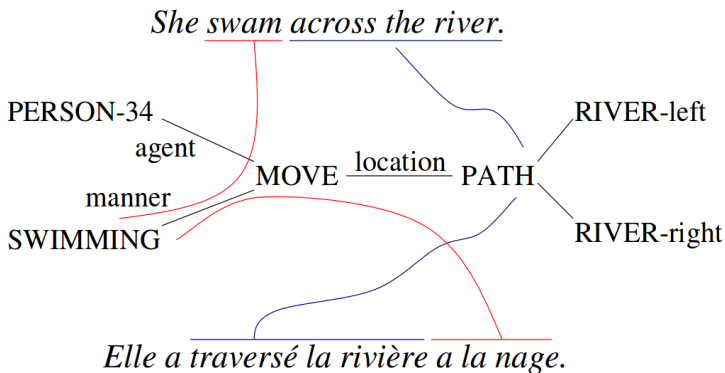


MT Interlingua

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



MT Interlingua (Beispiel)



Outline

- 1 Organisatorisches
- 2 Was ist Computerlinguistik?
- 3 Vorlesungsplan
- 4 Sprachwissenschaft
- 5 Computerlinguistische Methoden
- 6 Computerlinguistische Anwendungen / Sprachtechnologie
- 7 Geschichte, Herausforderungen und Forschungsstand der Computerlinguistik**

Kleine Geschichte der Computerlinguistik (1)

- frühe Entwicklung der Computertechnologie (1930er-, 40er-Jahre): numerische Problemstellungen (“Berechnungen”, z.B. ballistische Kurven), auch symbolische Verarbeitungsaufgaben (Dechiffrierung verschlüsselter Nachrichtentexte → maschinelle Übersetzung (MÜ) als Spezialfall einer Dekodierungsaufgabe)
- frühe Ansätze der MÜ haben gemeinsame Wurzel: stochastische Informationstheorie (Betrachtung des fremdsprachlichen Textes als Ergebnis der Übertragung einer Nachricht über gestörten Kanal → Aufgabe: Rekonstruktion des ursprünglichen Nachrichtentextes)
- Statistische Verfahren wurden dann für Jahrzehnte aufgegeben.

Kleine Geschichte der Computerlinguistik (2)

- Abkehr von statistischen Verfahren weil
- Chomsky die Unzulänglichkeit der statistischen Verfahren der 50er und 60er für Sprachmodellierung nachweist.
- die Leistungsfähigkeit der damaligen Hardware nicht ausreichte (Beschränkungen bevorzugen symbolische Ansätze)
- nicht genügend digitalisierte mehrsprachige Textdaten zur Verfügung standen

Herausforderungen der Computerlinguistik: Variabilität und Ambiguität (Mehrdeutigkeit)

Schwierigkeiten für Sprachverarbeitungssysteme:

- 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).

Typen von Ambiguität

- **Phonetische Ambiguität (Homophone):**

Miene - Mine, Meer - mehr, viel - fiel

⇒ Unterschiedliche Wörter haben dieselbe lautliche Form.

- **Orthographische Ambiguität (Homographen):**

übersetzen - über-setzen, 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.

Typen von Ambiguität

- **Strukturelle/syntaktische Ambiguität:**

- ① Visiting relatives can be boring.
- ② Peter fuhr seinen Freund sturzbetrunken nach Hause.
- ③ Ich traf den Sohn des Nachbarn mit dem Gewehr.

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

Typen von Ambiguität

- **Pragmatische Ambiguität:**

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

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

Wie mit Ambiguität umgehen?

- Alle Lesarten berechnen / aufzählen.
Ist in der Regel nicht praktikabel, manchmal aber von theoretischem Interesse.
- Unterspezifizierte Repräsentation verwenden, die alle möglichen Lesarten in einer kompakten Darstellung zusammenfasst.
- Nur die aufgrund des Kontextes präferierte(n) Lesarten berechnen / aufzählen.
Erfordert ein geeignetes gewichtetes oder probabilistisches Modell, oder zusätzliche Information (Kontextwissen oder Weltwissen).
- **Probabilistisches Modell:** Statistisches Modell, das verschiedenen Möglichkeiten Wahrscheinlichkeiten zuweist. Ein System kann sich dann für die wahrscheinlichste Variante entscheiden. Die Wahrscheinlichkeiten können z.B. durch Auswertung von durch Menschen annotierte Trainingsdaten gewonnen werden.

Mehrdeutigkeit, Beispiel MT

- Der Mann ging in den Fluss, weil er schmutzig war.
- Der Mann ging aus dem Fluss, weil er schmutzig war.
- er \Rightarrow he / it

- Wie lösen Menschen Mehrdeutigkeiten auf?
- Welche Möglichkeiten gibt es zur Modellierung mit dem Computer?

Mehrdeutigkeit, Weltwissen, Beispiel MT

- Der Mann ging in den Fluss, weil er schmutzig war.
- Der Mann ging aus dem Fluss, weil er schmutzig war.
- er \Rightarrow he / it

- Wie lösen Menschen Mehrdeutigkeiten auf?
- **Weltwissen:** Wissen darüber, welche Sachverhalte oder Ereignisse möglich oder plausibel sind.

- Welche Möglichkeiten gibt es zur Modellierung mit dem Computer?
- Manuelle Entwicklung von grammatischen, lexikalischen und extralinguistischen Wissensbeständen
- Statistische Modellierung von Wissen durch maschinelle Lernverfahren

Weitere Schwierigkeiten

Quelle: Dan Jurafsky, Stanford

non-standard English

Great job @justinbieber! Were SOO PROUD of what youve accomplished! U taught us 2 #neversaynever & you yourself should never give up either♥

segmentation issues

the New York-New Haven Railroad
the New York-New Haven Railroad

idioms

dark horse
get cold feet
lose face
throw in the towel

neologisms

unfriend
Retweet
bromance

world knowledge

Mary and Sue are sisters.
Mary and Sue are mothers.

tricky entity names

Where is *A Bug's Life* playing ...
Let It Be was recorded ...
... a mutation on the *for* gene ...

Aktueller Forschungsstand

Quelle: Dan Jurafsky, Stanford

mostly solved

Spam detection

Let's go to Agra! ✓

Buy V1AGRA ... ✗

Part-of-speech (POS) tagging

ADJ ADJ NOUN VERB ADV

Colorless green ideas sleep furiously.

Named entity recognition (NER)

PERSON ORG LOC

Einstein met with UN officials in Princeton

making good progress

Sentiment analysis

Best roast chicken in San Francisco! 👍

The waiter ignored us for 20 minutes. 👎

Coreference resolution

Carter told Mubarak he shouldn't run again.

Word sense disambiguation (WSD)

I need new batteries for my *mouse*. 🖱️

Parsing

I can see Alcatraz from the window!

Machine translation (MT)

第13届上海国际电影节开幕...

The 13th Shanghai International Film Festival...

Information extraction (IE)

You're invited to our dinner party, Friday May 27 at 8:30

Party
May 27
add

still really hard

Question answering (QA)

Q. How effective is ibuprofen in reducing fever in patients with acute febrile illness?

Paraphrase

XYZ acquired ABC yesterday

ABC has been taken over by XYZ

Summarization

The Dow Jones is up

The S&P500 jumped

Housing prices rose

Economy is good

Dialog

Where is Citizen Kane playing in SF?

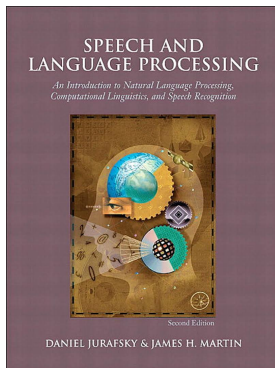
Castro Theatre at 7:30. Do you want a ticket?



TUCHE

by OTOM

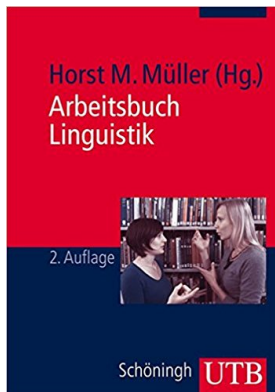




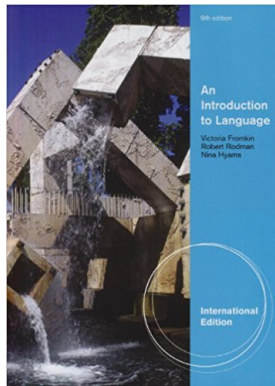
- Jurafsky & Martin: Speech and Language Processing. Pearson Prentice Hall. 2008. (Lehrbuchsammlung)



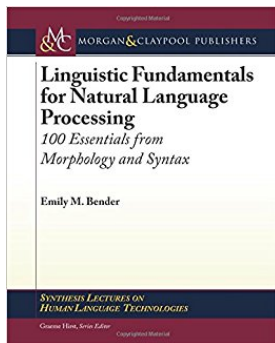
- Carstensen et al.: Computerlinguistik und Sprachtechnologie. Eine Einführung. Heidelberg 2010 (3. Auflage)
(Elektronische Version:
<https://login.emedien.ub.uni-muenchen.de/login>)



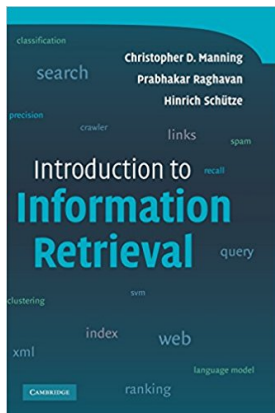
- Müller: Arbeitsbuch Linguistik. Schöningh / UTB. 2009. (Elektronische Version: s.o.)



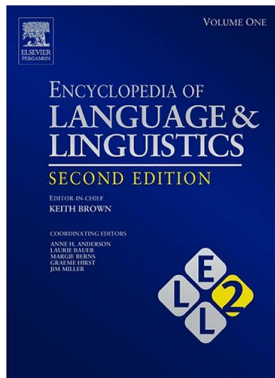
- Fromkin, Rodman, Hyams: An Introduction to Language. 2011.



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



- Manning, Raghavan, Schütze: Foundations of Introduction to Information Retrieval. Cambridge University Press. 2008.
<https://nlp.stanford.edu/IR-book/>



- Keith Brown (ed.): Encyclopedia of Language & linguistics. Elsevier. 2006.
(Elektronische Version: s. Homepage)

Außerdem ...

- 1 Melden Sie sich für die Vorlesung in Moodle an (Anleitung auf der Vorlesungshomepage), und bearbeiten Sie das erste Übungsblatt.
- 2 Melden Sie sich für den Email-Verteiler für Studenten an (optional).
 - ▶ Ankündigung von Vorträgen und Events am CIS (wissenschaftliche Vorträge und Firmenvorträge).
 - ▶ Studentenjobs (als Tutor oder wissenschaftliche Hilfskraft).
 - ▶ http://www.cis.uni-muenchen.de/cis_listen/collect_email_alias.html