

Einführung in die Computerlinguistik

Einführung

Laura Kallmeyer
Heinrich-Heine-Universität Düsseldorf
Sommersemester 2013

Einführung	1	Sommersemester 2013
------------	---	---------------------

Überblick

1. Anwendungen der Computerlinguistik
2. Bereiche der Computerlinguistik
3. Komponenten eines Sprachmodells
4. Ambiguität

[Carstensen et al., 2009, Jurafsky and Martin, 2009]

Einführung	2	Sommersemester 2013
------------	---	---------------------

Anwendungen der Computerlinguistik

[Jurafsky and Martin, 2009]

- Dialogsysteme *conversational agent (CA)*
Bsp.: KomParse <http://komparse.dfki.de/index.html>
- Maschinelle Übersetzung
Bsp.: Systran <http://www.systranet.com>
“This is my first course in Computational Linguistics, and I am quite excited about it.”
“Dieses ist mein erster Kurs in der Computerlinguistik, und ich bin über sie ziemlich aufgeregt.”
- Web-basiertes question answering
Bsp.: START <http://start.csail.mit.edu/>
“Who was the president of the United States in 1940?”
“Franklin Delano Roosevelt: March 4, 1933 to April 12, 1945”

Einführung	3	Sommersemester 2013
------------	---	---------------------

Bereiche der Computerlinguistik (1)

[Carstensen et al., 2009]

- Computerlinguistik als Teilbereich der Linguistik
 - theoriegeleitet
 - Entwicklung formaler Sprachmodelle
 - berechnungsrelevante Aspekte von Sprache und Sprachverarbeitung
 - unabhängig von konkreter Realisierung
 - **theoretische Computerlinguistik**
- Computerlinguistik als Disziplin für die Verarbeitung linguistischer Daten
 - Korpora
 - **linguistische Datenverarbeitung**

Einführung	4	Sommersemester 2013
------------	---	---------------------

Bereiche der Computerlinguistik (2)

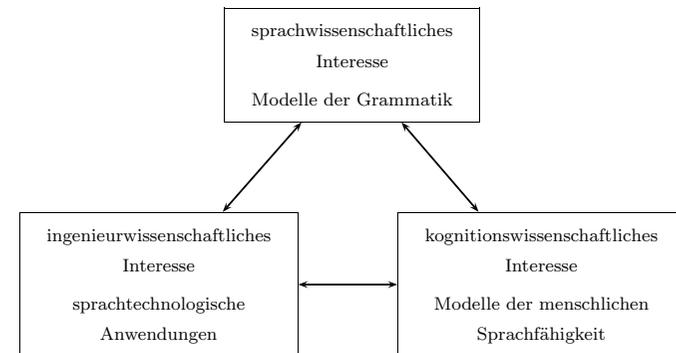
- Computerlinguistik als Realisierung natürlichsprachlicher Phänomene auf dem Computer
 - Nachbardisziplinen: Kognitionswissenschaft, Künstliche Intelligenz
- **maschinelle Sprachverarbeitung**
- Computerlinguistik als praxisorientierte, ingenieurmäßig konzipierte Entwicklung von Sprachsoftware
 - **Sprachtechnologie**

Bereiche der Computerlinguistik (3)

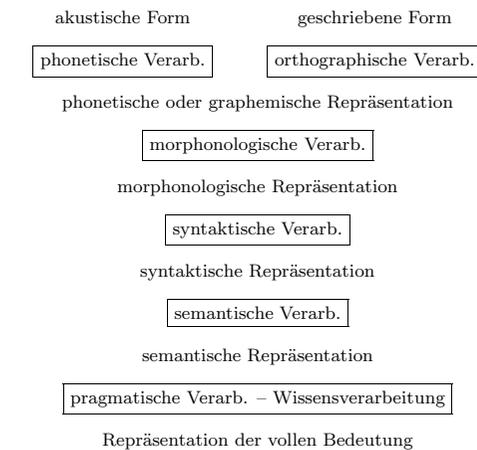
Zwei große Teilbereiche:

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

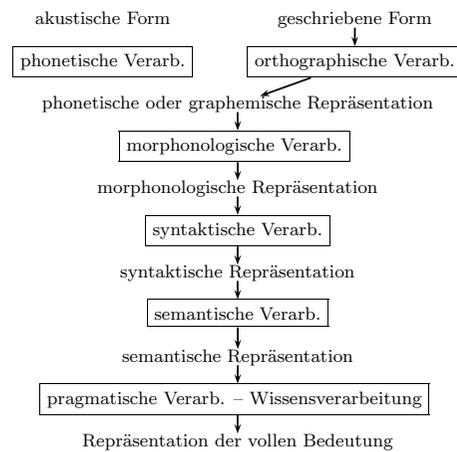
Bereiche der Computerlinguistik (4)



Komponenten eines Sprachmodells



Komponenten eines Sprachmodells: Textverstehen



Ambiguität (1)

Schwierigkeit bei der Sprachverarbeitung: Sprache ist hochgradig ambig.

Wir unterscheiden verschiedene Typen von Ambiguitäten:

- *phonetische* Ambiguität (Homophone)
Miene – Mine
- *orthographische* Ambiguität (Homographen)
übersetzen – übersetzen
- *lexikalische* Ambiguität (Homonyme)
(1) Hans geht zur *Bank*
- *morphologische* Ambiguität
Staubacken, Hauptpostsekretär

Ambiguität (2)

- *strukturelle/syntaktische* Ambiguitäten
(2) he sees the man with the telescope
(3) Peter fuhr seinen Freund sturzbetrunken nach Hause.
(4) Ich traf den Sohn des Nachbarn mit dem Gewehr.
- *kompositionell-semantische* Ambiguität bzw. *Skopusambiguität*
(5) Die zwei Mitarbeiter müssen vier Sprachen beherrschen.
(6) Some student likes every course.
(7) Alle Politiker sind nicht korrupt
- *pragmatische* Ambiguität
(8) Könnten Sie die Aufgabe lösen?

Ambiguität (3)

Wege, mit Ambiguität umzugehen:

- Alle Lesarten berechnen.
→ Ist in der Regel nicht praktikabel, manchmal aber von theoretischem Interesse.
- Unspezifizierte Repräsentationen verwenden, die alle möglichen Lesarten in einer kompakten Darstellung zusammenfassen.
→ Meistens Verwendung von *Constraints*.
- Nur die aufgrund des Kontextes präferierte(n) Lesarten berechnen.
→ Erfordert ein geeignetes gewichtetes oder probabilistisches Modell.

References

- [Carstensen et al., 2009] Carstensen, K.-U., Ebert, C., Ebert, C., Jekat, S., Langer, H., and Klabunde, R., editors (2009). *Computerlinguistik und Sprachtechnologie: Eine Einführung*. Spektrum Akademischer Verlag. 3. überarbeitete und erweiterte Auflage.
- [Jurafsky and Martin, 2009] Jurafsky, D. and Martin, J. H., editors (2009). *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall Series in Artificial Intelligence. Pearson Education International. Second Edition.