

Einführung in die Computerlinguistik

WS 2006/2007

Manfred Pinkal

Vorlesung "Einführung in die CL" 2006/2007 © M. Pinkal UdS Computerlinguistik

Vorlesungsplan

		20.10.05	Prop.: Linux I
24.10.06	Einführung/Überblick	27.10.06	Prop.: Linux II
31.10.06	Wissen in der Sprachverarbeitung	03.11.06	Übung
07.11.06	Verarbeitung gesprochener Sprache	10.11.06	Propädeutikum
14.11.06	Korpora und Annotation (S.Pado)	17.11.06	Übung
21.11.06	Statistische Methoden (S.Pado)	24.11.06	Übung
28.11.06	Morphologie & Endliche Automaten I	01.12.06	Propädeutikum
05.12.06	Morphologie & Endliche Automaten II	08.12.06	Übung
12.12.06	Vorlesung	15.12.06	Übung
19.12.05	Vorlesung	22.12.05	Vorlesungsfrei

Vorlesung "Einführung in die CL" 2006/2007 © M. Pinkal UdS Computerlinguistik

Vorlesungsplan

09.01.06	Vorlesung	12.01.06	Propädeutikum
16.01.06	Vorlesung	19.01.06	Übung
23.01.06	Vorlesung	26.01.06	Übung
30.01.06	Vorlesung	02.02.06	Übung
06.02.06	Vorlesung	09.02.06	Übung
13.02.06	Vorlesung	16.2.06	Klausur

Vorlesung "Einführung in die CL" 2006/2007 © M. Pinkal UdS Computerlinguistik

Technisches

Zur Vorlesung gehören:

- Das **Vorlesungsskript** (auf der Homepage des Kurses)
<http://www.coli.uni-saarland.de/courses/I2CL-06/>
- Ausgewählte **Kurztexte** in englischer und deutscher Sprache
- **Übungsaufgaben**: Sie werden (tendenziell wöchentlich) in der Vorlesung am Dienstag ausgegeben (und auf die Homepage gestellt), sind in der nächsten Vorlesung einzureichen und werden in der darauf folgenden Übungssitzung besprochen.

Vorlesung "Einführung in die CL" 2006/2007 © M. Pinkal UdS Computerlinguistik

Technisches

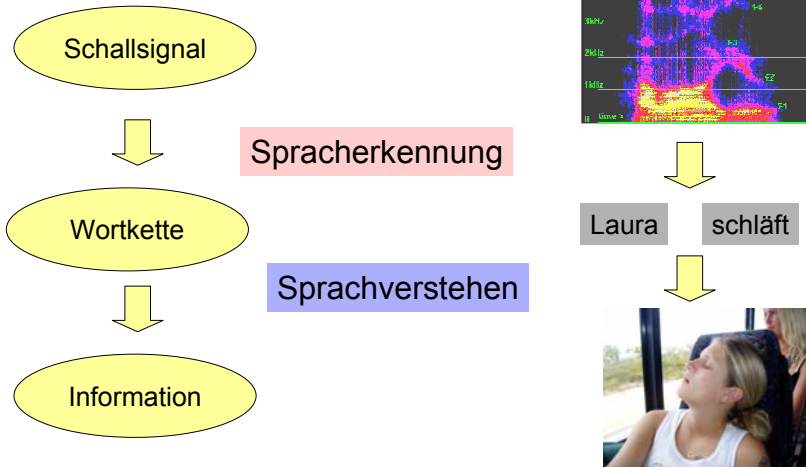
Scheine werden in folgender Weise erworben:

- **Prüfungsvoraussetzung:** Schriftliche Bearbeitung der Übungsaufgaben, d.h.:
 1. Alle Aufgabenblätter (mit höchstens einer Ausnahme) müssen bearbeitet sein. Aufgabenblatt zählt als bearbeitet, wenn für die überwiegende Zahl der Aufgaben ein Lösungsversuch vorliegt.
 2. Insgesamt müssen mindestens 50% der Punkte erreicht sein.
 3. Aufgaben können in Gruppen mit bis zu drei Studierenden bearbeitet werden.
- **Prüfungsleistung:** Klausur über den Stoff der Vorlesung, der im Vorlesungsskript, den Übungen und den Lektüretexten vorkommt. Klausurtermin: voraussichtlich der Vorlesungstermin in der letzten Semesterwoche (16.2.07)

Einführungsliteratur und andere Informationsquellen

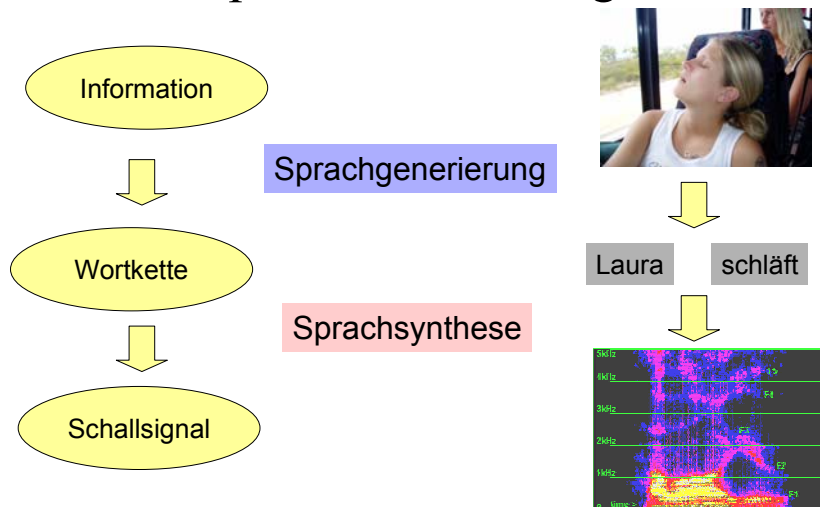
- Eine ausgezeichnetes englisch-sprachiges Einführungswerk: Jurafsky, Daniel/ Martin, James H. 2000. Speech and Language Processing. Prentice-Hall.
- Ein aktuelles deutsches Handbuch zur Computerlinguistik. Carstensen, Kai-Uwe et al. 2001. Computerlinguistik und Sprachtechnologie - Eine Einführung. Heidelberg: Spektrum Akademischer Verlag.
- Ein gutes deutsches Glossar zur Computerlinguistik:
<http://www.ifi.unizh.ch/CL/Glossar/glossary.html>
- Online-Wörterbuch: LEO

Was ist Sprachverarbeitung?



Vorlesung "Einführung in die CL" 2006/2007 © M. Pinkal UdS Computerlinguistik

Was ist Sprachverarbeitung ?



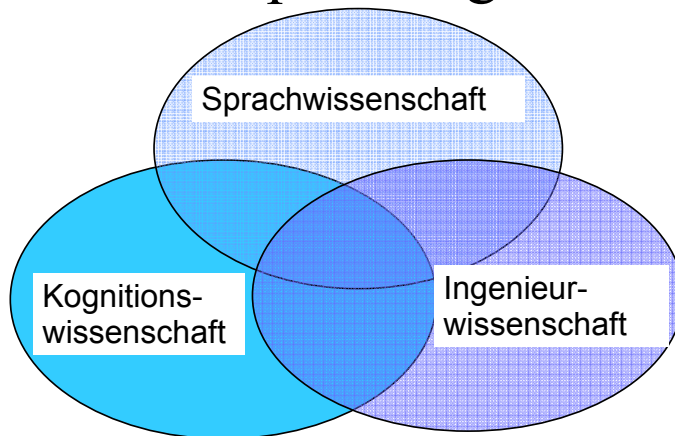
Vorlesung "Einführung in die CL" 2006/2007 © M. Pinkal UdS Computerlinguistik

Gegenstände der Computerlinguistik

- Die Entwicklung von Formalismen und Werkzeugen für die Repräsentation, Verarbeitung und Akquisition von linguistischem Wissen der verschiedenen Ebenen:
 - Phonetik und Phonologie
 - Morphologie und Syntax
 - Semantik
 - Pragmatik und Diskursstruktur
- Die Modellierung und Implementierung der komplexen Zusammenhänge und Abläufe bei:
 - Sprachverstehen
 - Sprachproduktion
 - Spracherwerb
- Die Entwicklung von **natürlich-sprachlichen Anwendungssystemen**.

Vorlesung "Einführung in die CL" 2006/2007 © M. Pinkal UdS Computerlinguistik

Was ist Computerlinguistik?



Vorlesung "Einführung in die CL" 2006/2007 © M. Pinkal UdS Computerlinguistik

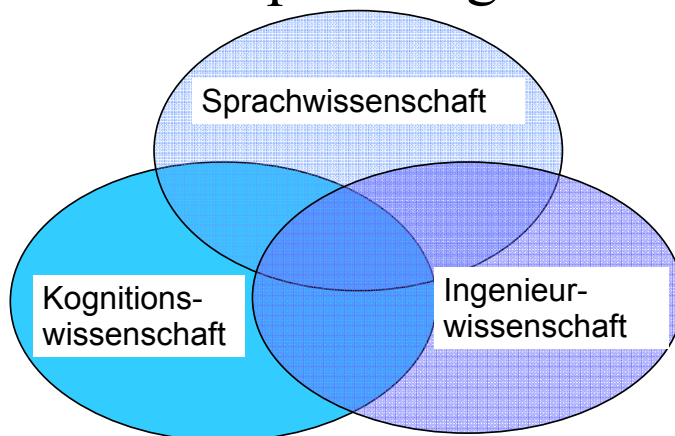
Computerlinguistik als Sprachwissenschaft

Eine wesentliche Voraussetzung für die Computerlinguistik ist die systematische und einheitliche Beschreibung von sprachlichem Wissen und sprachlichen Strukturen. Umgekehrt stellt die Computerlinguistik für die Erhebung und Erfassung komplexer sprachlicher Struktur Theorien und Werkzeuge zur Verfügung. Insofern gehört Computerlinguistik zu den sprachwissenschaftlichen Disziplinen, zusammen mit

- Theoretischer Linguistik / allgemeiner Sprachwissenschaft
- Historischer und vergleichender Sprachwissenschaft
- Phonetik
- Germanistischer, romanistischer, japanischer ... Sprachwissenschaft

Vorlesung "Einführung in die CL" 2006/2007 © M. Pinkal UdS Computerlinguistik

Was ist Computerlinguistik?



Vorlesung "Einführung in die CL" 2006/2007 © M. Pinkal UdS Computerlinguistik

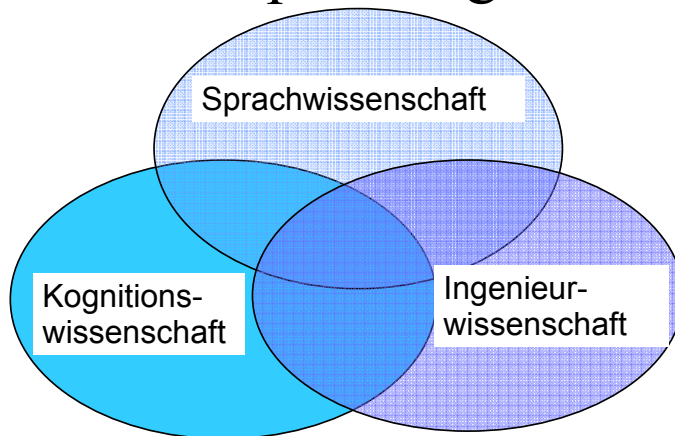
Computerlinguistik als Kognitionswissenschaft

Das übergeordnete Erkenntnisziel der Computerlinguistik ist die **Erforschung der menschlichen Sprachfähigkeit**: Wie ist sprachliches Wissen beim Menschen organisiert, und wie wird Sprache produziert, verstanden, und gelernt? Insofern gehört die Computerlinguistik zu den Kognitionswissenschaften, die die "kognitiven" Fähigkeiten des Menschen erforschen, zusammen mit den Fächern und Forschungsbereichen:

- Psycholinguistik
- kognitive Psychologie
- Neuropsychologie
- Künstliche Intelligenz (in der Informatik)
- Sprachphilosophie

Vorlesung "Einführung in die CL" 2006/2007 © M. Pinkal UdS Computerlinguistik

Was ist Computerlinguistik?



Vorlesung "Einführung in die CL" 2006/2007 © M. Pinkal UdS Computerlinguistik

Computerlinguistik als Ingenieurwissenschaft

Die praktische Zielsetzung der Computerlinguistik ist die Realisierung von Computersystemen, die sprachliches Wissen und sprachliche Fertigkeiten einsetzen, um den Menschen in der Kommunikation, beim Verwenden von Sprache und beim Umgang mit sprachlichen Dokumenten zu unterstützen. Computerlinguistik als Sprachtechnologie gehört in den Bereich der Informationstechnologie, zusammen mit den Fächern und Forschungsbereichen

- Angewandte Informatik
- Informationswissenschaft
- Signalverarbeitung (in der Elektrotechnik)

Vorlesung "Einführung in die CL" 2006/2007 © M. Pinkal UdS Computerlinguistik

Sprachtechnologie

- Wichtige Teilbereiche der Sprachtechnologie
 - Informationszugriff und -management
 - Multilinguale Anwendungen
 - Gesprochene Sprache und Dialog

Vorlesung "Einführung in die CL" 2006/2007 © M. Pinkal UdS Computerlinguistik

Informationszugriff und -management

- Information Retrieval
- Informationsextraktion
- Frage-Antwort-Systeme (Question Answering)
- Summarisation
- Dokumentklassifikation

Vorlesung "Einführung in die CL" 2006/2007 © M. Pinkal UdS Computerlinguistik

Multilinguale Anwendungen

- Maschinelle Übersetzung
- Multilinguale Informationssuche
- Sprachlernsysteme

Vorlesung "Einführung in die CL" 2006/2007 © M. Pinkal UdS Computerlinguistik

Anwendungen für gesprochen Sprache

- Sprachsynthese: Text-to-Speech Systeme (TTS)
- Spracherkennung: Diktiersysteme
- Sprachdialog:
 - Telefonie-Anwendungen: Fahrplan-Information, Call-Center, Telebanking
 - Dialog für Gerätebedienung: Mensch-Computer-Interaktion, Sprache im Fahrzeug (In-car), Intelligentes Haus, Robotersteuerung