

Vorläufiges Programm

Einführung in die Computerlinguistik

WS 2011/12

Manfred Pinkal

25.10.11	Einführung
8.11.11	Morphologie und Automaten
15.11.11	Morphologie und Automaten
22.11.11	Syntax
29.11.11	Parsing
6.12.11	Merkmale und Unifikation
13.12.11	Grammatische Verarbeitung
20.12.11	Semantik

Im neuen Jahr: Statistische Verfahren, Anwendungen

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Technisches: Vorlesung und Übung

- **Vorlesungsskript** (auf der Homepage des Kurses)
<http://www.coli.uni-saarland.de/courses/I2CL-11/>
- Ausgewählte **Kurztexte** in englischer und deutscher Sprache
- **Übungsaufgaben:**
 - Ausgabe: Übungsblatt wird zur Vorlesung am Dienstag auf die Homepage gestellt (tendenziell wöchentlich)
 - Einreichen der Lösungen: bis zum Dienstag der folgenden Woche (Vorlesungsbeginn), als PDF oder auf Papier
(Übungsblatt 1 wegen Feiertag bis Montag, 31.10., 18 Uhr!)
 - Besprechung: in der nächsten Übungssitzung am Freitag
- **Übungsgruppen**

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Technisches: Prüfungsvoraussetzungen

Prüfungsvoraussetzung: Schriftliche Bearbeitung der Übungsaufgaben, das heißt genauer:

1. Alle Aufgabenblätter (mit höchstens einer Ausnahme) müssen bearbeitet sein. Aufgabenblatt zählt als bearbeitet, wenn für alle Aufgaben ein ernsthafter Lösungsversuch vorliegt. Dies schließt Teilaufgaben ein, wenn sie unabhängig gelöst werden können.
2. Insgesamt müssen mindestens 50% der Punkte erreicht sein.
3. Aufgaben können in Gruppen mit bis zu drei Studierenden bearbeitet werden (Näheres auf der Homepage des Kurses)

Abschreiben ist nicht erlaubt: Bei allen Beteiligten wird das Blatt als nicht bearbeitet gewertet. Das führt im Wiederholungsfall automatisch zum Verlust der Klausurzulassung.

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Technisches: Klausur

- **Anmeldung zur Prüfung bis zum 27.1.2011**
Wichtig: Ohne fristgerechte Meldung keine Teilnahme möglich!
- **Prüfungsleistung:** Klausur über den Stoff der Vorlesung, der
 - im Vorlesungsskript
 - den Übungen und
 - den Lektüretexten
 vorkommt.
- Klausurtermin: 14.2.2011, 10-12 Uhr (erste Woche der vorlesungsfreien Zeit)

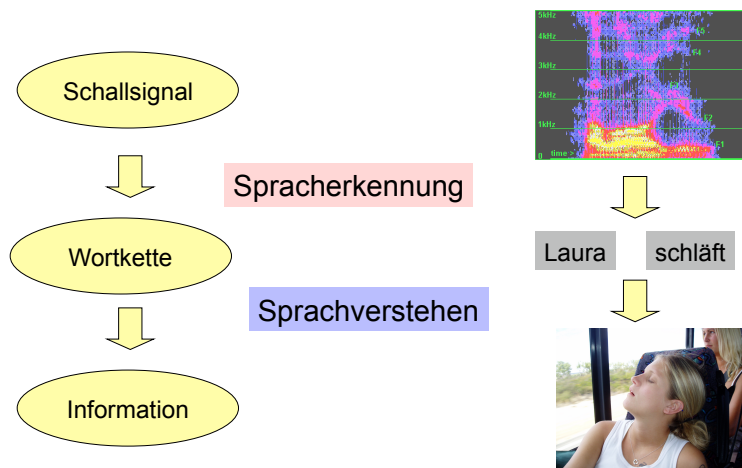
Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Einführungsliteratur und andere Informationsquellen

- Eine ausgezeichnetes englisch-sprachiges Einführungswerk: Jurafsky, D./ Martin, J.: *Speech and Language Processing*, 2009
- Ein aktuelles deutsches *Handbuch der Computerlinguistik*: Carstensen, Kai-Uwe et al.: *Computerlinguistik und Sprachtechnologie - Eine Einführung*, 2009
- Ein linguistisches Wörterbuch: H. Bussmann: *Lexikon der Sprachwissenschaft*, 2008
- Das Online-Wörterbuch: [LEO](#)
- Und: Die [WikiPedia](#)

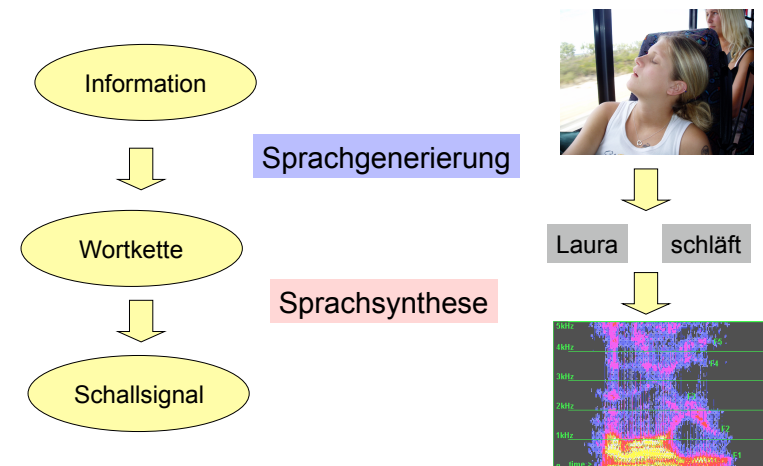
Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Was ist Sprachverarbeitung?



Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Was ist Sprachverarbeitung ?



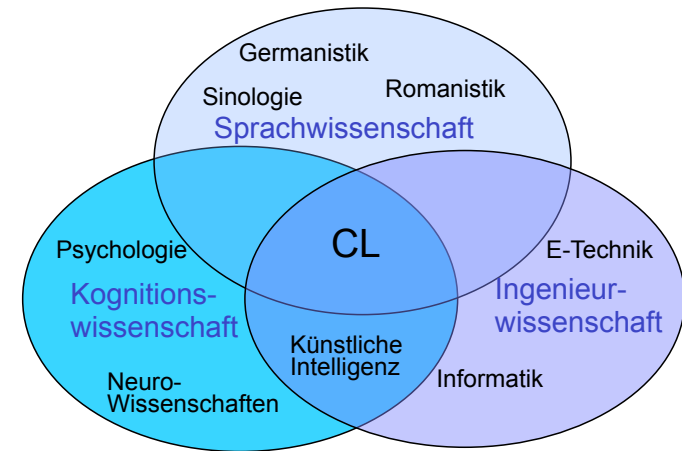
Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Aufgaben der Computerlinguistik

- Die Modellierung und Implementierung der komplexen Zusammenhänge und Abläufe bei:
 - Sprachverstehen
 - Sprachproduktion
 - Spracherwerb
- 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 Diskurs
- Die Entwicklung von **natürlich-sprachlichen Anwendungssystemen**.

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Computerlinguistik als Schnittstellenfach



Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Angewandte Computerlinguistik: Sprachtechnologie

Wichtige Teilbereiche der Sprachtechnologie:

- Informationszugriff und -management
- Gesprochene Sprache
- Multilinguale Anwendungen

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Informationszugriff und -management

- Information Retrieval
- Informations-Extraktion
- Question Answering (Frage-Antwort-Systeme)
- Summarisation (Automatische Zusammenfassung)
- Dokumentklassifikation

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Multilinguale Anwendungen

- Maschinelle Übersetzung
- Multilinguale Informationssuche
- Sprachlernsysteme

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Anwendungen für gesprochene Sprache

- Diktiersysteme, Spracheingabe für medizinische Diagnose, technische Wartungssysteme
- Telefonie-Dialogsysteme: Call-Center, Telefon-Banking, Fahrplanauskunft, ...
- Gerätebedienung: Sprachbedienung für Auto, mobile Geräte, Haushalt
- Interaktion mit virtuellen Agenten und Robotern

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Frühgeschichte der Computerlinguistik

- In den späten 50er Jahren begann in den USA aus politischer Motivation die Förderung der [Maschinellen Übersetzung](#) (Russisch-Englisch) als erster computerlinguistischer Anwendungsdisziplin. Die Resultate waren extrem schlecht, die Förderung wurde eingestellt.
- In den späten 60er und den frühen 70er Jahren entstand die zweite Generation von computerlinguistischen Systemen: [Dialogsysteme](#)
Die beiden vermutlich berühmtesten Systeme:
 - ELIZA, 1966, von Joseph Weizenbaum, und
 - SHRDLU, 1970, von T. Winograd

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

ELIZA: Beispieldialog [1]

- | | |
|--|--|
| (1) U: I have been very depressed lately
S: <i>You say you have been very depressed lately.</i> | (4) U: My father calls me nasty names
S: <i>Who else in your family calls you nasty names</i> |
| (2) U: Yes, my mother beats me and my father calls me names
S: <i>You`re being somewhat short with me</i> | (5) U: I think the dog would if he could
S: <i>Do you think its likely that he could</i> |
| (3) U: What about my mother beating me?
S: <i>Tell me more about your family</i> | (6) U: No, you do
S: <i>You don`t seem very talkative today</i> |

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

ELIZA: Beispieldialog [2]

- (7) U: I guess what is really bothering me is that I don't have any money
S: *Can you elaborate on that*
- (8) U: The problem is that I know psychiatrists are expensive and I may not be able to pay you
S: *Please continue, this may be interesting*
- (9) U: I thought you might think so
S: *Perhaps in your fantasy we thought each other*

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

ELIZA: Ein sprachverstehendes System?

- ELIZA ist ein Dialogsystem, das beliebig komplexe Eingaben mit beliebigem Wortschatz zu beliebigen Themen akzeptiert.
- ELIZA arbeitet mit einfachen Mustervergleichs-Techniken (**Pattern Matching**), ohne Einsatz von Wissen:
 - **Templates**: Muster mit variablen Teilen, die mit der Benutzereingabe abgeglichen werden, und
 - Template-basierten System-Äußerungen (Prompts)
- ELIZA hat in gewisser Hinsicht den **Turing-Test** absolviert (s. Lektüre), dies aber unter besonderen Rahmenbedingungen.
- ELIZA funktioniert besonders gut mit englischem Dialog und dem Psychotherapie-Szenario. Wieso?

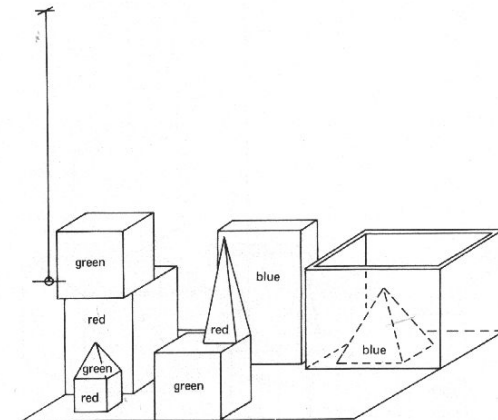
Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Was steckt hinter ELIZA?

- TEMPLATE: I _X1_
RESPONSE: You say you _X1_
- TEMPLATE: _X1_ my _X2_(category family) _X3_
RESPONSE: Who else in your family _X3_ ?
- TEMPLATE: _X1_ if _X2_
RESPONSE: Do you think its likely that _X2_ ?
- TEMPLATE: _X1_
RESPONSE: You're being somewhat short with me.

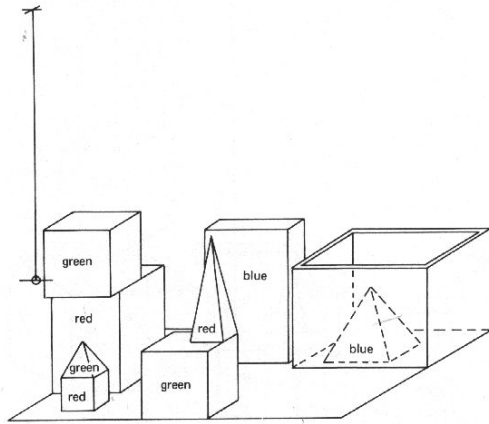
Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

SHRDLU: Ein wissensbasiertes Dialogsystem



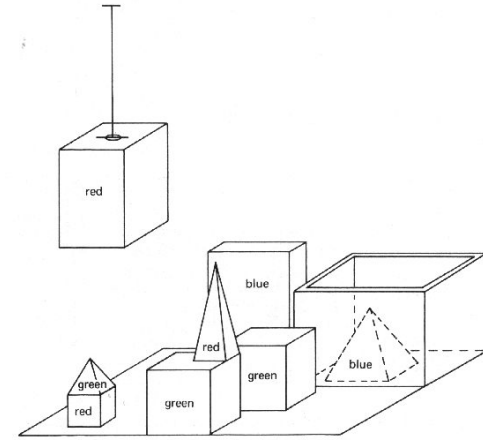
Winograds "Blocks World"

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik



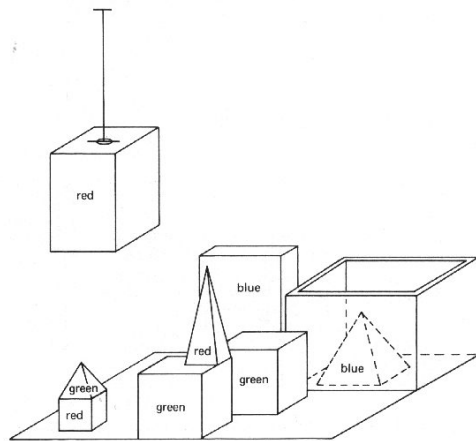
U: Pick up a big red block
 S: OK.

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik



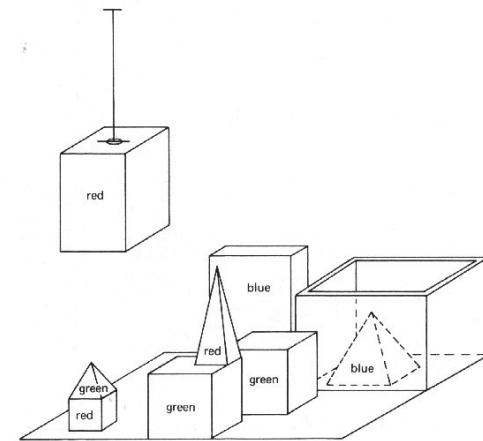
(Pick up a big red block)

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik



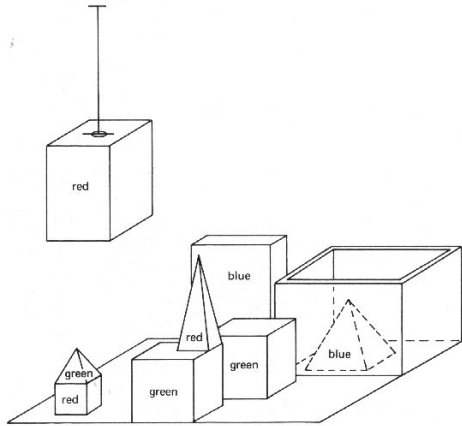
U: Grasp the pyramid.

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

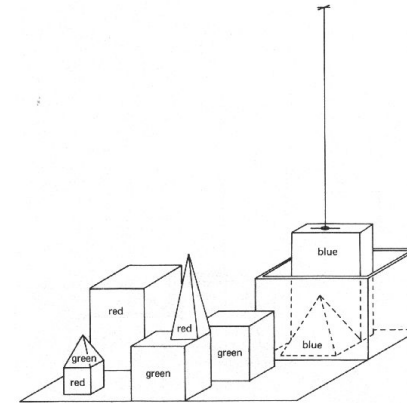


U: Grasp the pyramid.
 S: I don't understand which pyramid you mean.

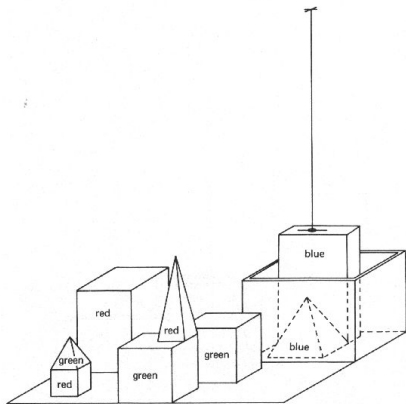
Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik



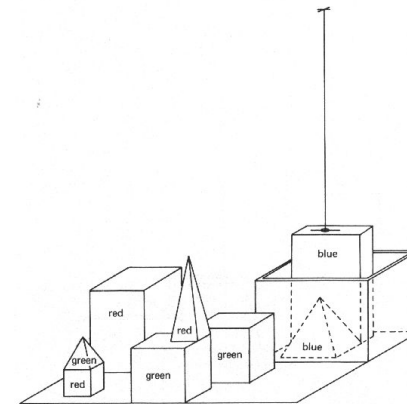
U: Find a block which is taller than the one you are holding and put it into the box.
 S: By „it“, I assume you mean the block which is taller than the one I am holding. OK



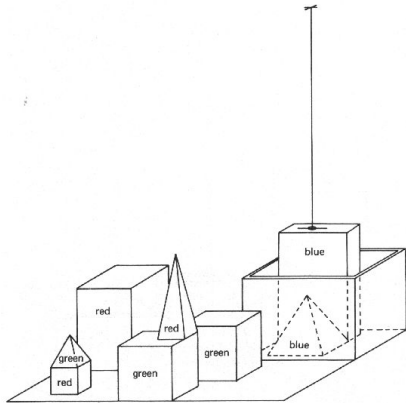
(U: Find a block which is taller than the one you are holding and put it into the box.)



U: What does the box contain?
 S: The blue pyramid and the blue block



U: What is the pyramid supported by?
 S: The box



U: How many blocks are not in the box?
S: *Four of them*

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

SHRDLU

SHRDLU beantwortet Fragen, führt Anweisungen aus und lernt Begriffe.

Wichtige Programmkomponenten von SHRDLU sind:

- (Linguistische) Analyse
- Generierung
- (Handlungs-)Planung
- (grafische) Visualisierung

Winograds SHRDLU-System arbeitet in einer kleinen, eingeschränkten Mini-Welt oder -Domäne ("Blocks World").

Interessant ist die Interaktion von Analyse und Planung; die Generierungskomponente ist sehr einfach (patternbasiert); die Grafik ist computerlinguistisch nicht sehr interessant, hat aber zum durchschlagenden Erfolg des Systems beigetragen.

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

SHRDLU

SHRDLU ist ein **wissensbasiertes Dialogsystem**.

Im Unterschied z.B. zu ELIZA verwendet es bei der Dialogführung unterschiedliche Arten von Wissen, und zwar

- Sprachliches/linguistisches Wissen
- Kontextwissen
- Weltwissen

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik