

**Herzlich willkommen  
in der Computerlinguistik  
in Heidelberg!**



- 1** Wer wir sind
  - Institut
  - Umgebung
- 2** Computerlinguistik-Studium in Heidelberg
- 3** Computerlinguistik: Inhalte und Methoden
  - CL - Was ist das?
  - Inhalte/Problemstellungen
  - Methoden
- 4** Perspektiven

# Personal Professoren



**Anette Frank**  
Computerlinguistik



**N.N./Yannick Verseley**  
Computerlinguistik

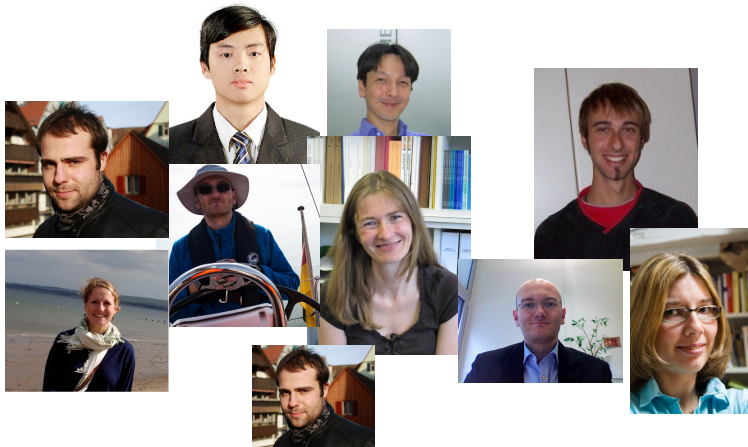


**Stefan Riezler/Alex Fraser**  
Linguistische Informatik



**Michael Strube**  
Honorarprofessor

# Personal Wissenschaftliche Mitarbeiter



**PD Dr. Karin Haenelt**, Fraunhofer Institut, Darmstadt

**PD Dr. Kurt Eberle**, Lingenio, Heidelberg und Universität Leeds

**Dr. Michael Herweg**, IBM Deutschland

**Dr. Andreas Harth**, AIFB, Universität Karlsruhe

**Dr. Andreas Witt**, Institut für Deutsche Sprache, Mannheim

**Dr. Andrea Zielinski**, Fraunhofer Institut, Karlsruhe



**Sandra Suchowitz & Cornelia Conrad**

Email: [sekretariat@cl.uni-heidelberg.de](mailto:sekretariat@cl.uni-heidelberg.de)

Öffnungszeiten: Mo – Fr, 9:00 – 15:00



**Markus Kirschner**  
Systemadministrator

... mit Unterstützung von  
studentischen Hilfskräften:  
Mirko & Lauritz

- Hält unsere Rechner am Laufen und auf dem Laufenden
- Email: [technik@cl.uni-heidelberg.de](mailto:technik@cl.uni-heidelberg.de)



RUPRECHT-KARLS-UNIVERSITÄT HEIDELBERG

INSTITUT FÜR  
COMPUTER-  
LINGUISTIK

Personen

Startseite

In Studien

Kolloquium

Bibliothek

Links

Kontakt



An Institut für Computational Linguistics an der Universität Heidelberg lehrt und forscht ein junges dynamisches Team von Wissenschaftlern zu neuen Themen und Methoden der maschinellen Verarbeitung natürlicher Sprache. Mit drei Professoruren, einem wachsenden Mitarbeiterstab von Doktoranden und Postdoktoranden sowie internationalen Gastwissenschaftlern zählt das Institut zu den wichtigsten Zentren der Computational Linguistics in Deutschland.

Neues am Institut

Five three-year PhD scholarships  
Five three-year PhD scholarships for work in the area of "semantics beyond the sentence: coherence in language processing" from October 1st 2013. ... mehr

Lehrveranstaltungen

HSPDS und LSF

Signet

Fachschaft

Ressourcen

AKTUELLES

Posterpreise im Wintersemester 2011/12

Programmprüfung

Lehrveranstaltungen WGLI - Kurse für Studierende in der Sommersemesterzeit (05. September - 26. Oktober)

Informationen zur Sommersemesterprüfung (07. - 14. Oktober)

Geänderte Immatrikulationsfrist für BA-Studiengänge (01. September - 6. Oktober)

[www.cl.uni-heidelberg.de](http://www.cl.uni-heidelberg.de)

**Neu: Wiki:** <https://wiki.cl.uni-heidelberg.de>

**Enthält alle wichtigen Informationen, z.B.:**

- Vorlesungsverzeichnis und Kursmaterialien
- Studien- und Prüfungsordnung und Hinweise dazu
- Mitarbeiter und ihre Sprechstunden
- Praktika, Auslandsstudium
- Aktuelle Veranstaltungen und Neues am Institut
- Infos zur Technik, Links, und, und, und. . .

RUPRECHT-KARLS-UNIVERSITÄT HEIDELBERG

INSTITUT FÜR  
COMPUTER-  
LINGUISTIK

Suche

Personen

Startseite

In Studium

Kolloquium

Bibliothek

Links

Kontakt



An Institut für Computational Linguistics an der Universität Heidelberg lehrt und forscht ein junges dynamisches Team von Wissenschaftlern zu neuen Themen und Methoden der maschinellen Verarbeitung natürlicher Sprache. Mit drei Professoruren, einem wachsenden Mitarbeiterstab von Doktoranden und Postdoktoranden sowie internationalen Gastwissenschaftlern zählt das Institut zu den wichtigsten Zentren der Computational Linguistics in Deutschland.

Neues am Institut

Five three-year PhD scholarships  
Five three-year PhD scholarships for work in the area of "semantics beyond the sentence: coherence in language processing" from October 1st 2013. ... mehr

Lehrveranstaltungen

HSPOS und LSF

Signet

Fachschaft

FAQ

Ressourcen

AKTUELLES

Posterpreise im Wintersemester 2011/12

Programmprüfung

Lehrveranstaltungen WGLI - Kurse für Studierende in der Sommersemesterzeit (28. September - 28. Oktober)

Informationen zur Sommersemesterprüfung (27. - 24. Oktober)

Geänderte Immatrikulationsfrist für 444-Studiengänge (2. September - 6. Oktober)

## www.cl.uni-heidelberg.de

**Neu: Wiki:** <https://wiki.cl.uni-heidelberg.de>

**Enthält alle wichtigen Informationen, z.B.:**

- Vorlesungsverzeichnis und Kursmaterialien
- Studien- und Prüfungsordnung und Hinweise dazu
- Mitarbeiter und ihre Sprechstunden
- Praktika, Auslandsstudium
- Aktuelle Veranstaltungen und Neues am Institut
- Infos zur Technik, Links, und, und, und. . .

RUPRECHT-KARLS-UNIVERSITÄT HEIDELBERG

INSTITUT FÜR  
COMPUTER-  
LINGUISTIK

Personen

Startseite

In Studien

Kolloquium

Bibliothek

Links

Kontakt



Lehrveranstaltungen

HSPUS und LSF

Signos

Fachschaft

Ressourcen

AKTUELLES

Posterlesung im Wintersemester 2011/12

Progressprüfung

Lehrerkollegium WGLI - Kurse für Studierende in der Sommersemesterzeit (28. September - 26. Oktober)

Informationen zur Sommer-Exkursionsreise (27. - 14. Oktober)

Geplante Innovationsstudien für die Sommersemesterzeit (2. September - 6. Oktober)

Neues am Institut

Five three-year PhD scholarships

Five three-year PhD scholarships for work in the area of "semantics beyond the sentence: coherence in language processing" from October 1st 2011. ... mehr

[www.cl.uni-heidelberg.de](http://www.cl.uni-heidelberg.de)

**Neu: Wiki:** <https://wiki.cl.uni-heidelberg.de>

**Enthält alle wichtigen Informationen, z.B.:**

- Vorlesungsverzeichnis und Kursmaterialien
- Studien- und Prüfungsordnung und Hinweise dazu
- Mitarbeiter und ihre Sprechstunden
- Praktika, Auslandsstudium
- Aktuelle Veranstaltungen und Neues am Institut
- Infos zur Technik, Links, und, und, und. . .

RUPRECHT-KARLS-UNIVERSITÄT HEIDELBERG

INSTITUT FÜR  
COMPUTER-  
LINGUISTIK

Personen

Startseite

In Studium

Kolloquium

Bibliothek

Links

Kontakt



An Institut für Computational Linguistics an der Universität Heidelberg lehrt und forscht ein junges dynamisches Team von Wissenschaftlern zu neuen Themen und Methoden der maschinellen Verarbeitung natürlicher Sprache. Mit drei Professuren, einem wachsenden Mitarbeiterstab von Doktoranden und Postdoktoranden sowie internationalen Gastwissenschaftlern zählt das Institut zu den wichtigsten Zentren der Computational Linguistics in Deutschland.

Neues am Institut

Five three-year PhD scholarships  
Five three-year PhD scholarships for work in the area of "semantics beyond the sentence: coherence in language processing" from October 1st 2013. ... mehr

Lehrveranstaltungen

HSPDS und LSF

Signos

Fachschaft

Ressourcen

AKTUELLES

Posterpreise im Wintersemester 2011/12

Programmprüfung

Lehrveranstaltungen WGL - Kurse für Studierende in der Sommersemester 2012 (28. September - 26. Oktober)

Informationen zur Sommersemesterprüfung (27. - 14. Oktober)

Geordnete Immatrikulation für 446 Studienanfänger (2. September - 6. Oktober)

# www.cl.uni-heidelberg.de

**Neu: Wiki:** <https://wiki.cl.uni-heidelberg.de>

**Enthält alle wichtigen Informationen, z.B.:**

- Vorlesungsverzeichnis und Kursmaterialien
- Studien- und Prüfungsordnung und Hinweise dazu
- Mitarbeiter und ihre Sprechstunden
- Praktika, Auslandsstudium
- Aktuelle Veranstaltungen und Neues am Institut
- Infos zur Technik, Links, und, und, und. . .

RUPRECHT-KARLS-UNIVERSITÄT HEIDELBERG

INSTITUT FÜR  
COMPUTER-  
LINGUISTIK

Personen

Startseite

In Studien

Kolloquium

Bibliothek

Links

Kontakt



Lehrveranstaltungen:  
HSP10 und LSF 12  
Signat 12 FAQ  
Fachschaft 12 Ressourcen

AKTUELLES

Posterlesung im Wintersemester 2011/12  
Programmprüfung  
Lernerkolleg WGL - Kurse für Studierende in der Sommersemester 2012 (28. September - 26. Oktober)  
Informationen zur Sommer-Examensprüfung (27. - 14. Oktober)  
Geplante Innovationskandidatur für 446-Studiengang (2. September - 6. Oktober)

Neues am Institut

Five three-year PhD scholarships  
Five three-year PhD scholarships for work in the area of "semantics beyond the sentence: coherence in language processing" from October 1st 2013. ... mehr

[www.cl.uni-heidelberg.de](http://www.cl.uni-heidelberg.de)

**Neu: Wiki:** <https://wiki.cl.uni-heidelberg.de>

**Enthält alle wichtigen Informationen, z.B.:**

- Vorlesungsverzeichnis und Kursmaterialien
- Studien- und Prüfungsordnung und Hinweise dazu
- Mitarbeiter und ihre Sprechstunden
- Praktika, Auslandsstudium
- Aktuelle Veranstaltungen und Neues am Institut
- Infos zur Technik, Links, und, und, und. . .

RUPRECHT-KARLS-UNIVERSITÄT HEIDELBERG

INSTITUT FÜR  
COMPUTER-  
LINGUISTIK

Personen

Startseite

In Studium

Kolloquium

Bibliothek

Links

Kontakt



Lehrveranstaltungen

HSPOS und LSF

Signos

Fachschaft

Ressourcen

AKTUELLES

Posterlesung im Wintersemester 2011/12

Programmprüfung

Lehrveranstaltungen WGLI - Kurse für Studierende in der Sommersemester 2012 (28. September - 26. Oktober)

Informationen zur Sommersemesterprüfung (27. - 24. Oktober)

Geplante Innovationskonferenz für die Sommersemester (2. September - 6. Oktober)

Neues am Institut

Five three-year PhD scholarships

Five three-year PhD scholarships for work in the area of "semantics beyond the sentence: coherence in language processing" from October 1st 2011. ... mehr

[www.cl.uni-heidelberg.de](http://www.cl.uni-heidelberg.de)

**Neu: Wiki:** <https://wiki.cl.uni-heidelberg.de>

**Enthält alle wichtigen Informationen, z.B.:**

- Vorlesungsverzeichnis und Kursmaterialien
- Studien- und Prüfungsordnung und Hinweise dazu
- Mitarbeiter und ihre Sprechstunden
- Praktika, Auslandsstudium
- Aktuelle Veranstaltungen und Neues am Institut
- Infos zur Technik, Links, und, und, und. . .

RUPRECHT-KARLS-UNIVERSITÄT HEIDELBERG

INSTITUT FÜR  
COMPUTER-  
LINGUISTIK

Personen

Startseite

In Studien

Kolloquium

Bibliothek

Links

Kontakt



Lehrveranstaltungen

HSPDS und LSF

Signos

Fachschaft

Ressourcen

AKTUELLES

Posterlesung im Wintersemester 2011/12

Progressreporting

Lehrerkollegium WGL - Kurse für Studierende in der Sommersemester 2012 (28. September - 26. Oktober)

Informationen zur Sommer-Orthographie (27. - 24. Oktober)

Geplante Innovationsstudien für 446 Studierende (2. September - 6. Oktober)

Neues am Institut

Five three-year PhD scholarships

Five three-year PhD scholarships for work in the area of "semantics beyond the sentence: coherence in language processing" from October 1st 2013. ... mehr

[www.cl.uni-heidelberg.de](http://www.cl.uni-heidelberg.de)

**Neu: Wiki:** <https://wiki.cl.uni-heidelberg.de>

**Enthält alle wichtigen Informationen, z.B.:**

- Vorlesungsverzeichnis und Kursmaterialien
- Studien- und Prüfungsordnung und Hinweise dazu
- Mitarbeiter und ihre Sprechstunden
- Praktika, Auslandsstudium
- Aktuelle Veranstaltungen und Neues am Institut
- Infos zur Technik, Links, und, und, und. . .

## Bereichsbibliothek Mathematik und Informatik (BMI)

INF 294

## Universitätsbibliothek

- Hauptbibliothek Altstadt: Plöck 107-109
- Zweigstelle Neuenheim: INF 368





# Kein Studium ohne Bücher – aber wo findet man die?

## Heidi – Online-Katalog der UB

**HEIDI** Katalog für die Bibliotheken  
der Universität Heidelberg

The screenshot shows the HEIDI online catalog interface. At the top, there are navigation tabs: 'Einfache Suche', 'Erweiterte Suche', 'Konto', 'Merkliste', 'Fernleihe', 'Leihstelle: UB Altstadt', 'Hilfe', and 'Beenden'. Below these is a search bar containing 'Jurafsky' and a 'Suchen' button. To the right of the search bar, there are links for 'Suchhistorie (1 Recherche)' and 'Treffer aus externen Quellen: 0 Treffer Digital (1836-1972), 41 wissenschaftliche Web-Dokumente (BASE)'. Below the search bar, there are options for 'Optionen', 'Neue Suche', and 'Druckansicht'. The search results are sorted by 'Relevanz' and show 5 hits. The first result is for 'Jurafsky, Daniel: Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition / Daniel Jurafsky ; James H. Martin. -2. ed., [Pearson International Edition] Upper Saddle River, NJ [u.a.] : Prentice Hall, Pearson Education International, 2009. - 1024 S. : Ill., graph. Darst. Themen: Automatische Spracherkennung | Computerlinguistik'. To the right of the search results, there is a 'Treffer einschränken:' section with a 'Jahr' dropdown menu showing years from 2000 to 2009 and a search box.

Zahlreiche weitere Online-Ressourcen

Den CIP-Pool kennen Sie schon

Sie bekommen . . .

- einen Account, damit Sie programmieren können
- eine Mail-Adresse, damit wir Sie benachrichtigen können

**Bitte setzen Sie ein `.forward`, falls Sie Ihre Institutsadresse nicht verwenden wollen!**

## **organisiert**

- Ersti-Frühstück (incl. Kneipentour)
- Coli-Wochenende
- Sommer- und andere Feste
- Studenten-CL-Tagungen (TaCoS)

## **informiert über**

- Tagungen
- Termine
- Tipps

## **protestiert auch mal**

... wenn es sein muss

Das ICL gehört zur **Neuphilologischen Fakultät**

Weitere Fächer der Fakultät:

- Romanistik, Germanistik, Slavistik, Anglistik
- Deutsch als Fremdsprachenphilologie
- Übersetzen und Dolmetschen

## **INF 348 Institut für Informatik**

- gemeinsame Lehrveranstaltungen
- Forschungs Kooperationen

## **INF 368 Interdisziplinäres Zentrum für Wissenschaftliches Rechnen (IWR)**

## **INF 293 Rechenzentrum (URZ)**

- Druckkontingent für Studierende
- Beherbergt unseren Hadoop-Cluster (Vernetzte Rechner, zur Verarbeitung großer Datenmengen)

# Die Verwaltung

- Gemeinsames Prüfungsamt
- Zentrale  
Universitätsverwaltung
- Zentrale Studienberatung
  - Beantwortet Fragen zum Studium allgemein (Studiengebühren, Urlaubssemester etc.)
  - Workshops zu Themen rund ums Studium (Zeitmanagement, Bewerbungstraining etc.)



Zentrale Univerwaltung (oben),  
Gemeinsames Prüfungsamt (unten)



## Heidelberger Institut für Theoretische Studien HITS gGmbH

- Lehre im BA- und MA-Studiengang
- Forschungsk Kooperationen
- **Gemeinsames Doktorandenkolloquium:** wöchentliche Vorträge und geladene Gäste

<http://www.h-its.org/>

# Übersicht Studiengänge

		Bachelor of Arts (B.A.)				
		3 Jahre, 180 Leistungspunkte				
CL		100%	75%	50%	50%	25%
2.Fach	(25% Inf.)	25%	50%	50%	75%	

		Master of Arts (M.A.)		
		2 Jahre, 120 Leistungspunkte		
CL		100%	80%	20%
2. Fach	–	20%	80%	



# Übersicht Studiengänge

Bachelor of Arts (B.A.)					
3 Jahre, 180 Leistungspunkte					
CL	100%	75%	50%	50%	25%
2.Fach	(25% Inf.)	25%	50%	50%	75%

Master of Arts (M.A.)		
2 Jahre, 120 Leistungspunkte		
CL	100%	80%
2. Fach	–	20%

# Übersicht Studiengänge

		Bachelor of Arts (B.A.) 3 Jahre, 180 Leistungspunkte				
CL		100%	75%	50%	50%	25%
2.Fach	(25% Inf.)	25%	50%	50%	75%	

		Master of Arts (M.A.) 2 Jahre, 120 Leistungspunkte		
CL		100%	80%	20%
2. Fach	–	20%	80%	

# Gegenstand und Ziel des BA-Studiums

”**Gegenstand des Bachelor-Studiengangs** Computerlinguistik ist die **Vermittlung der theoretischen und anwendungsorientierten Grundlagen** des Faches Computerlinguistik.”

”**Ziel des Bachelor-Studiengangs** ist es, die Studierenden dazu zu befähigen, **Erkenntnisse der computerlinguistischen Forschung eigenständig auf Probleme und Fragestellungen der Computerlinguistik anzuwenden ...**”

*Aus der Prüfungsordnung für den Bachelor-Studiengang  
Computerlinguistik*

# Gegenstand und Ziel des BA-Studiums

”**‘Gegenstand des Bachelor-Studiengangs** Computerlinguistik ist die **Vermittlung der theoretischen und anwendungsorientierten Grundlagen** des Faches Computerlinguistik.”

”**‘Ziel des Bachelor-Studiengangs** ist es, die Studierenden dazu zu befähigen, **Erkenntnisse der computerlinguistischen Forschung eigenständig auf Probleme und Fragestellungen der Computerlinguistik anzuwenden ...**”

*Aus der Prüfungsordnung für den Bachelor-Studiengang  
Computerlinguistik*

# Gegenstand und Ziel des BA-Studiums

”**Gegenstand des Bachelor-Studiengangs** Computerlinguistik ist die **Vermittlung der theoretischen und anwendungsorientierten Grundlagen** des Faches Computerlinguistik.”

”**Ziel des Bachelor-Studiengangs** ist es, die Studierenden dazu zu befähigen, **Erkenntnisse der computerlinguistischen Forschung eigenständig auf Probleme und Fragestellungen der Computerlinguistik anzuwenden ...**”

Aus der *Prüfungsordnung für den Bachelor-Studiengang  
Computerlinguistik*

# Aufbau des Studiums

- 1.-3. Semester verpflichtende Grundlagen- und Programmierkurse
- 3.-5. Semester wählbare Vertiefungskurse
- 4. oder 5. Semester Softwareprojekt (programmierpraktische Arbeit in Kleingruppen)
- 6. Semester Bachelor-Arbeit: Eigene kleine Forschungsarbeit

Jeder Kurs ist einem oder mehreren **Modulen** zugeordnet (Modul = inhaltliche Einheit). Je nach Studiengang müssen Sie jeweils eine bestimmte Anzahl von **Leistungspunkten** aus verschiedenen Modulen erwerben.

- 1.-3. Semester** verpflichtende Grundlagen- und Programmierkurse
- 3.-5. Semester** wählbare Vertiefungskurse
- 4. oder 5. Semester** Softwareprojekt (programmierpraktische Arbeit in Kleingruppen)
- 6. Semester** Bachelor-Arbeit: Eigene kleine Forschungsarbeit

Jeder Kurs ist einem oder mehreren **Modulen** zugeordnet (Modul = inhaltliche Einheit). Je nach Studiengang müssen Sie jeweils eine bestimmte Anzahl von **Leistungspunkten** aus verschiedenen Modulen erwerben.

- 1.-3. Semester** verpflichtende Grundlagen- und Programmierkurse
- 3.-5. Semester** wählbare Vertiefungskurse
- 4. oder 5. Semester** Softwareprojekt (programmierpraktische Arbeit in Kleingruppen)
- 6. Semester** Bachelor-Arbeit: Eigene kleine Forschungsarbeit

Jeder Kurs ist einem oder mehreren **Modulen** zugeordnet (Modul = inhaltliche Einheit). Je nach Studiengang müssen Sie jeweils eine bestimmte Anzahl von **Leistungspunkten** aus verschiedenen Modulen erwerben.



- 1.-3. **Semester** verpflichtende Grundlagen- und Programmierkurse
- 3.-5. **Semester** wählbare Vertiefungskurse
- 4. oder 5. **Semester** Softwareprojekt (programmierpraktische Arbeit in Kleingruppen)
- 6. **Semester** Bachelor-Arbeit: Eigene kleine Forschungsarbeit

Jeder Kurs ist einem oder mehreren **Modulen** zugeordnet (Modul = inhaltliche Einheit). Je nach Studiengang müssen Sie jeweils eine bestimmte Anzahl von **Leistungspunkten** aus verschiedenen Modulen erwerben.

- 1.-3. **Semester** verpflichtende Grundlagen- und Programmierkurse
- 3.-5. **Semester** wählbare Vertiefungskurse
- 4. oder 5. **Semester** Softwareprojekt (programmierpraktische Arbeit in Kleingruppen)
- 6. **Semester** Bachelor-Arbeit: Eigene kleine Forschungsarbeit

Jeder Kurs ist einem oder mehreren **Modulen** zugeordnet (Modul = inhaltliche Einheit). Je nach Studiengang müssen Sie jeweils eine bestimmte Anzahl von **Leistungspunkten** aus verschiedenen Modulen erwerben.

- 1.-3. **Semester** verpflichtende Grundlagen- und Programmierkurse
- 3.-5. **Semester** wählbare Vertiefungskurse
- 4. oder 5. **Semester** Softwareprojekt (programmierpraktische Arbeit in Kleingruppen)
- 6. **Semester** Bachelor-Arbeit: Eigene kleine Forschungsarbeit

Jeder Kurs ist einem oder mehreren **Modulen** zugeordnet (Modul = inhaltliche Einheit). Je nach Studiengang müssen Sie jeweils eine bestimmte Anzahl von **Leistungspunkten** aus verschiedenen Modulen erwerben.

# Übersicht der Module

Modulübersicht Hauptfach (Fachanteil 75%) → 113 LP (plus 12 LP BA-Arbeit im Hauptfach plus 20 LP ÜK plus 35 LP Begleitfach)

Semester	Computerlinguistische Module	Linguistische Module	Informatische Module		Übergreifende Kompetenzen
6	Computational Linguistics Colloquium (2LP, PM)	BA-Thesis (12 LP, PM) Oral Exam (7 LP, PM)			Erwerb von 20 LP aus dem Bereich der Übergreifenden Kompetenzen
5		Advanced Studies (CL) (8 LP, WPM) or Advanced Studies (FL) (8 LP, WPM)	Core Studies in CS (Theoretical CS) (8 LP, WPM) or Core Studies in CS (Applied CS) (4 + 4 LP, WPM)	Software Project (6 LP CL + 4 LP ÜK, PM)	
4	Core Studies in Computational Linguistics (30 LP, PM) (5 x 6 LP)				
3	Statistical Methods for CL (6 LP, PM)	Formal Semantics (6 LP, PM)			
2	Formal Foundations: Mathematical and Logical Foundations (12 LP, PM)	Formal Syntax (6 LP, PM)	Advanced Programming for CL (6 LP, PM)		
1		Foundations of Linguistic Analysis (4 LP, PM)	Introduction to Programming (6 LP, PM)		

Genauere Erläuterungen zu den Modulen finden Sie auf der Institutshomepage > Im Studium > Studiengänge > Bachelor

# Welchen Kurs für welches Modul?

RUPRECHT-KARLS-UNIVERSITÄT HEIDELBERG

## INSTITUT FÜR COMPUTER- LINGUISTIK

Google Custom Search

Erweiterte Suche | Seiten im Überblick

Startseite

Personen

Forschung

Interesse am Studium

### Im Studium

Lehrveranstaltungen

Studiengänge

Studienberatung

Plagiats

Ressourcen und Technik

Studentische Projekte

Auslandsstudium

Abschlussarbeiten

Auszeichnungen und Preise

Praktika, Jobs und Alumni

Kolloquium

Bibliothek



Lehrveranstaltungen

HSPOS und LSF ☺

Signup ☺

FAQ

Fachschaft ☺

Ressourcen

## Statistical Machine Translation

### Kursbeschreibung

Dozent\*innen Stefan Riezler, Laura Jehl  
Veranstaltungsart Vorlesung/Übung  
Erster Termin 20.10.2011  
Zeit und Ort Do, 11:15–12:45, INF 346 / SR 10  
voraussichtlich Do, 16:15–17:45, INF 346 / SR 10  
ACHTUNG: endgültiger Termin für Übung wird in der 1. Sitzung festgelegt!!!

### Teilnahmevoraussetzungen

Grundlagen der Wahrscheinlichkeitstheorie und Linearen Algebra, z.B. aus Mathematischem Grundkurs oder Statistical Methods.

Studiengang	Modulkürzel	Leistungs- bewertung
BA-2010	AS-CL	8 LP
NBA	AS-CL	8 LP
Master	SS-CL, SS-TAC	8 LP
Magister	-	-

- **Leistungsnachweis** in einem Kurs in der Regel durch Klausur(en); Hausarbeiten oder praktische Arbeit (Softwareprojekt) sind auch möglich
- **Alle** benoteten Veranstaltungen/Module erscheinen im Transkript **mit Note**
- Selbstverständlich müssen alle vorgeschriebenen Veranstaltungen **erfolgreich absolviert (bestanden)** werden

- **Leistungsnachweis** in einem Kurs in der Regel durch Klausur(en); Hausarbeiten oder praktische Arbeit (Softwareprojekt) sind auch möglich
- **Alle** benoteten Veranstaltungen/Module erscheinen im Transkript **mit Note**
- Selbstverständlich müssen alle vorgeschriebenen Veranstaltungen **erfolgreich absolviert (bestanden)** werden

- **Leistungsnachweis** in einem Kurs in der Regel durch Klausur(en); Hausarbeiten oder praktische Arbeit (Softwareprojekt) sind auch möglich
- **Alle** benoteten Veranstaltungen/Module erscheinen im Transkript **mit Note**
- Selbstverständlich müssen alle vorgeschriebenen Veranstaltungen **erfolgreich absolviert (bestanden)** werden



- **Leistungsnachweis** in einem Kurs in der Regel durch Klausur(en); Hausarbeiten oder praktische Arbeit (Softwareprojekt) sind auch möglich
- **Alle** benoteten Veranstaltungen/Module erscheinen im Transkript **mit Note**
- Selbstverständlich müssen alle vorgeschriebenen Veranstaltungen **erfolgreich absolviert (bestanden)** werden

Um für Ihre Lehrveranstaltungen die notwendigen Leistungspunkte (LPs, Credit Points) zu bekommen, müssen Sie

- Im Laufe des Semesters zu einer bestimmten Frist ein **Commitment** abgeben.
- Das heisst: sie legen sich definitiv fest, dass Sie die Klausur/Prüfung ablegen werden.
- Sie müssen dann die Klausur/Prüfung zum festgelegten Termin absolvieren.

**Wichtig: HISPOS Einführung** (diese Woche!), siehe  
Instituts-Homepage > Im Studium: Aktuelles > Informationen zur  
Semester-Einführungswoche

Um für Ihre Lehrveranstaltungen die notwendigen Leistungspunkte (LPs, Credit Points) zu bekommen, müssen Sie

- Im Laufe des Semesters zu einer bestimmten Frist ein **Commitment** abgeben.
- Das heisst: sie legen sich definitiv fest, dass Sie die Klausur/Prüfung ablegen werden.
- Sie müssen dann die Klausur/Prüfung zum festgelegten Termin absolvieren.

**Wichtig: HISPOS Einführung** (diese Woche!), siehe  
Instituts-Homepage > Im Studium: Aktuelles > Informationen zur  
Semester-Einführungswoche

Um für Ihre Lehrveranstaltungen die notwendigen Leistungspunkte (LPs, Credit Points) zu bekommen, müssen Sie

- Im Laufe des Semesters zu einer bestimmten Frist ein **Commitment** abgeben.
- Das heisst: sie legen sich definitiv fest, dass Sie die Klausur/Prüfung ablegen werden.
- Sie müssen dann die Klausur/Prüfung zum festgelegten Termin absolvieren.

**Wichtig:** **HISPOS Einführung** (diese Woche!), siehe  
Instituts-Homepage > Im Studium: Aktuelles > Informationen zur  
Semester-Einführungswoche

Um für Ihre Lehrveranstaltungen die notwendigen Leistungspunkte (LPs, Credit Points) zu bekommen, müssen Sie

- Im Laufe des Semesters zu einer bestimmten Frist ein **Commitment** abgeben.
- Das heisst: sie legen sich definitiv fest, dass Sie die Klausur/Prüfung ablegen werden.
- Sie müssen dann die Klausur/Prüfung zum festgelegten Termin absolvieren.

**Wichtig:** HISPOS Einführung (diese Woche!), siehe  
Instituts-Homepage > Im Studium: Aktuelles > Informationen zur  
Semester-Einführungswoche

Um für Ihre Lehrveranstaltungen die notwendigen Leistungspunkte (LPs, Credit Points) zu bekommen, müssen Sie

- Im Laufe des Semesters zu einer bestimmten Frist ein **Commitment** abgeben.
- Das heisst: sie legen sich definitiv fest, dass Sie die Klausur/Prüfung ablegen werden.
- Sie müssen dann die Klausur/Prüfung zum festgelegten Termin absolvieren.

**Wichtig: HISPOS Einführung** (diese Woche!), siehe Instituts-Homepage > Im Studium: Aktuelles > Informationen zur Semester-Einführungswoche

## ganz wichtig:

- Wer ein Commitment abgibt, aber nicht zur Prüfung erscheint, hat die Prüfung automatisch **nicht bestanden**
- Wenn Sie ein Commitment **nicht fristgerecht** abgegeben haben, dürfen Sie im laufenden Semester **nicht an der Klausur teilnehmen**
- Beachten Sie, dass viele Veranstaltungen nur einmal pro Jahr angeboten werden

Informationen über Commitment-Fristen werden über die **Bachelor-Mailingliste** verschickt

Bitte tragen sie sich ein:

Institutshomepage > Im Studium > Studiengänge > Bachelor > Modulübersicht > (BA-)Mailinglist

oder direkt:

<http://lists.cl.uni-heidelberg.de:8080/mailman/listinfo/clba>



# Die Orientierungsprüfung

- Indikator für Sie und uns, ob Computerlinguistik für Sie das richtige Fach ist
- keine eigene Prüfung, sondern studienbegleitend
- Spätestens im 3. Fachsemester müssen Sie folgende Module bestanden haben:
  - Einführung in die Computerlinguistik
  - Programmieren I
- Fortsetzung des Computerlinguistik-Studiums ist nur mit bestandener Orientierungsprüfung möglich
- Speziell für Module der Orientierungsprüfung:  
Wiederholung bei Nicht-Bestehen **nur einmal** möglich

# Nicht-Bestehen einer Prüfung

Was, wenn Sie eine Prüfung nicht bestehen?

- Die Prüfung darf **einmal** wiederholt werden, spätestens im folgenden Semester
- Falls Sie (aus welchem Grund auch immer) die Prüfung **ein zweites Mal** nicht bestehen, können Sie einen dritten Versuch beim Prüfungsamt beantragen
- Ein Antrag auf einen dritten Versuch ist aber **höchstens zweimal** im Studium möglich
- Ausnahme: Orientierungsprüfung

## Pflichtmodule

- Softwareprojekt (4 LP)
- Praktikum (bis zu 8 LP)

## Weitere Möglichkeiten

- Arbeitstechniken, mathematische Grundlagen
- Konferenzen / Sommerschulen, auch studentische Tagungen
  - TaCoS Tagung der Computerlinguistik Studierenden (jährlich)
  - StuTS <http://www.stuts.de/>
  - DGfS Computerlinguistik Herbstschule(n)
- Sprachkurse
- Eigeninitiative
- Fakultätsübergreifendes Angebot von Veranstaltungen (im Vorlesungsverzeichnis)

# Wer hilft bei Fragen/Problemen weiter?

- 1 Institutshomepage!**
- 2 Kommilitonen!**
- 3 Fachschaft!** Einstieg: Institutshomepage
- 4 Fachstudienberater**

## Fachstudienberatung:

**Eva Mujdricza-Maydt, Madeline Remse**, Fachstudienberatung

Bachelor studienberatung-bachelor@cl.uni-heidelberg.de

**Hiko Schamoni**, Fachstudienberatung Master

studienberatung-master@cl.uni-heidelberg.de

**FAQ** <http://www.cl.uni-heidelberg.de/studies/faq/faqBa.mhtml>

*Die Fachstudienberatung unterstützt vor allem bei Unklarheiten bezüglich der Prüfungsordnung, bei der Anrechnung von Leistungen aus vorangegangenen Studiengängen, bei Fragen zum Ablauf der Abschlussphase, und berät beim Fachwechsel ("Umschreibung")*

# Wer hilft bei Fragen/Problemen weiter?

- 1 **Institutshomepage!**
- 2 **Kommilitonen!**
- 3 **Fachschaft!** Einstieg: Institutshomepage
- 4 **Fachstudienberater**

## **Fachstudienberatung:**

**Eva Mujdricza-Maydt, Madeline Remse**, Fachstudienberatung

Bachelor `studienberatung-bachelor@cl.uni-heidelberg.de`

**Hiko Schamoni**, Fachstudienberatung Master

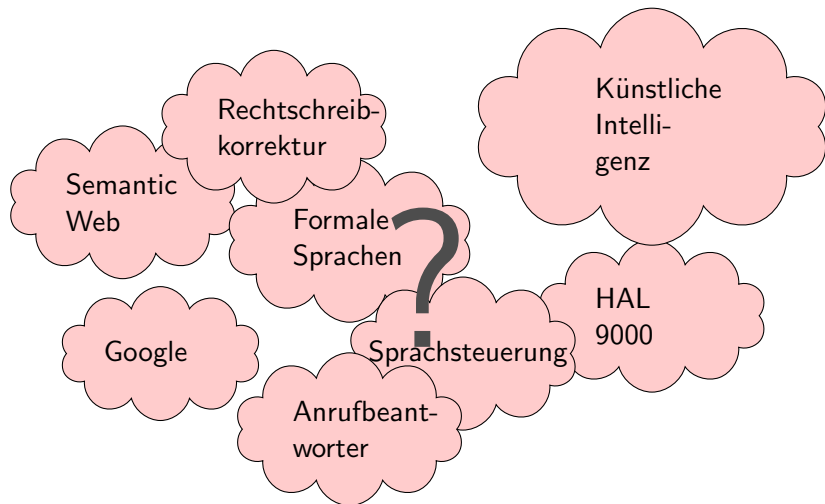
`studienberatung-master@cl.uni-heidelberg.de`

**FAQ** <http://www.cl.uni-heidelberg.de/studies/faq/faqBa.mhtml>

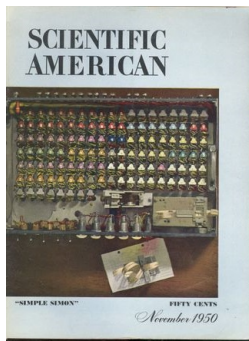
*Die Fachstudienberatung unterstützt vor allem bei Unklarheiten bezüglich der Prüfungsordnung, bei der Anrechnung von Leistungen aus vorangegangenen Studiengängen, bei Fragen zum Ablauf der Abschlussphase, und berät beim Fachwechsel ("Umschreibung")*

# Computerlinguistik

## – Was ist das?



# Computerlinguistik ist ein junges Fach, das 1947 seinen Anfang nahm...



*When I look at an article in Russian, I say to myself: This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.*

Warren Weaver, Mathematiker,  
*Translation* (1949)

Statistische Verfahren der Informationstheorie  
und Kryptographie als Modell für die  
automatische Übersetzung natürlicher  
Sprachen

## ... und das **heute** die moderne Gesellschaft mitgestaltet und prägt

- Sprachbasierte Internetsuche: Google, Yahoo, Bing ...
- Maschinelle Übersetzung: [translate.google.com](http://translate.google.com)
- Business Intelligence: Meinungen im WWW suchen
- eScience: Fakten für Wissenschaft & Firmen
- eLearning: Sprachlernprogramme und Wörterbuch auf Knopfdruck
- automatisierte Kommunikation und Informationshilfen: Warteschleifen, Siri



Computerlinguistik hat zum Ziel:

- die Natur, Funktion und Diversität von Sprache(n) zu verstehen und formal zu modellieren
- und diese Erkenntnisse für neue Aufgaben in der Wissenschaft und der IT-Gesellschaft nutzbar zu machen

## **Computerlinguistik/Computational Linguistics**

→ die akademische Disziplin

## **Maschinelle Sprachverarbeitung/Natural Language Processing (NLP)**

→ Betont den Anwendungsaspekt

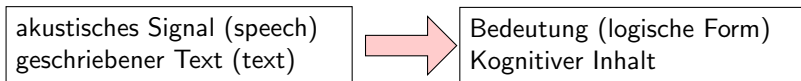
## **Sprachtechnologie/Language Technology, Speech Technology**

→ Anwendungen

- Struktur von Sprache verstehen, modellieren und automatisch ermitteln
- Wissen aus Texten gewinnen
- Sprachtechnologische Anwendungen verbessern

# Struktur von Sprache verstehen und modellieren

Modellierung aller Sprachebenen: von Laut zu Bedeutung



*Johanna(x)  
Buch(y)  
lesen(x,y)*

Empirische Validierung linguistischer Theorien anhand von Korpora

# Struktur von Sprache automatisch ermitteln

## Ambiguität auf allen Ebenen

### Phonologie

→ *Wreck a nice beach - Recognize speech*

### Morphologie

→ *Time flies (V/N) like (V/P) an arrow*



### Syntax

→ *Der Mann sah die Frau mit dem Fernglas* – Wer hat das Fernglas?

### Semantik

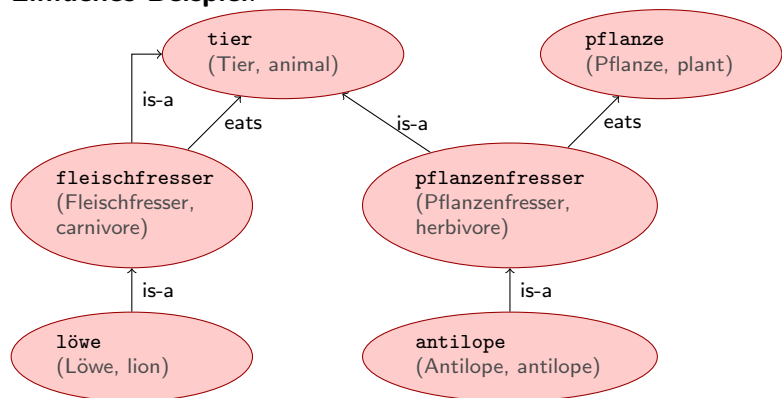
→ *Jeder Holländer liebt eine Frau* – wie viele Frauen werden geliebt? → *Wir sitzen im Zug* – Luftzug oder Eisenbahn?

### Diskurs

→ *Fritz und Maria fahren nach Paris. Dort trafen/trennten sie sich.*

- Modellierung von Wissen in **Ontologien** ("Begriffsnetzwerken")
- Ontologien automatisch aus Wikipedia oder anderen Textsammlungen extrahieren
- Übersetzungs- und anderes multilinguales Wissen

## Einfaches Beispiel:



**Echtes Beispiel:** Wikinet → Multilinguale Ressource aus Wikipedia, am HITS erstellt.

Erfolgreiche computerlinguistische Anwendungen, z.B.:

- Websuche
- Maschinelle Übersetzung
- Spracherkennung



Seit den 90er Jahren: **Regelbasierte Verfahren** werden zunehmend ersetzt durch **maschinelle Lernverfahren**

## Maschinelles Lernen

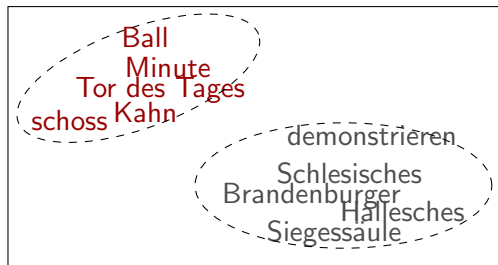
Mittels mathematischer Modelle Muster in unbekanntem Daten finden

- Der Computer **lernt von** annotierten oder unannotierten **Daten**
- Maschinelles Lernen kommt in vielen Bereichen, auch außerhalb der Computerlinguistik zum Einsatz:
  - Text- und Bilderkennung
  - Robotik
  - 'Big Data' in allen Lebensbereichen

# Maschinelles Lernen: Beispiel

## Problemstellungen: z.B.

- Daten klassifizieren (Disambiguierung, Annotierung) → *Clustering, Classification*
- Daten in eine Rangfolge bringen (z.B. Ergebnisse einer Suchmaschine, Übersetzungskandidaten) → *Ranking*



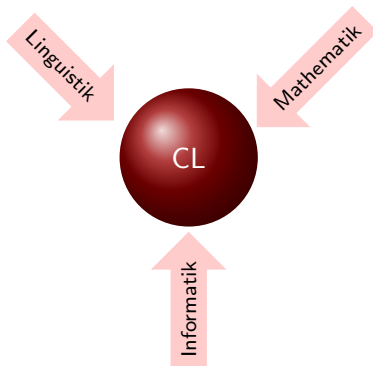
(ausgedachtes) Clustering-Beispiel:  
*Zwei Bedeutungen von "Tor"*

- Meistens gilt: Je besser die Datengrundlage, desto besser die Ergebnisse
- Das Trainieren von Modellen auf großen Datenmengen erfordert effiziente Algorithmen
- Oft kann ein normaler Rechner diese Aufgaben nicht mehr bewältigen



Hadoop-Cluster der Computerlinguistik

**Computerlinguistik ist interdisziplinär!**



Forschung und Lehre

Forschung und Entwicklung

Entwicklung und Anwendung

## Forschung und Lehre

- Universitäten
  - Computerlinguistik
  - Digitale Geisteswissenschaften (Digital Humanities)
- Forschungsinstitute (*DFKI, Fraunhofer Institut, Institut für deutsche Sprache*)

## Forschung und Entwicklung

## Entwicklung und Anwendung

## Forschung und Lehre

## Forschung und Entwicklung

- Forschungsabteilungen der (IT-)Industrie (*Google, Yahoo, Microsoft, Apple, SAP, IBM, Telekom, Siemens, Daimler*)
- Große Sprachtechnologiefirmen (*Nuance Communications, SDL*)

## Entwicklung und Anwendung

Forschung und Lehre

Forschung und Entwicklung

Entwicklung und Anwendung

- Kleinere Sprachtechnologiefirmen
- Verlagshäuser, Museen
- Anbieter standardisierter Sprachtests



## Knowledge Engineer

- Terminologie und „Lokalisierung“ für internationale Produkte
- Aufbau multilingualer Wissensbasen
- Multilinguale Generierung von Produktdokumentationen



## Data Analyst

- Begriffshierarchien & Text Mining in den Lebenswissenschaften
- Text Mining für Risikoforschung
- Marktforschung und Meinungsanalyse



## eLearning im Verlagswesen

- Entwickler von Sprachlernprogrammen
- Erstellung von Lexika



## Speech Specialist

- Spracherkennungs- und -synthesysteme
- Übersetzungstechnologie
- Benutzeradaptive Diktiersysteme



# Unternehmen in der Region

NETBASE

HMi

IBM®

lingenio

SAP®



temis

VICO  
RESEARCH CONSULTING

**Wir wünschen Ihnen viel Erfolg im Studium!**