

# Informationsveranstaltung

## Linguistische Informatik & Korpuslinguistik

**Prof. Dr. Stefan Evert**

Professur für Korpuslinguistik

<http://www.linguistik.uni-erlangen.de/>



FRIEDRICH-ALEXANDER  
UNIVERSITÄT  
ERLANGEN-NÜRNBERG

PHILOSOPHISCHE FAKULTÄT  
UND FACHBEREICH THEOLOGIE

---

# Was ist „Linguistische Informatik“?

## Korpuslinguistik $\neq$ Computerlinguistik?

bzw. Sprachtechnologie (Natural Language Processing, NLP)

# Sprachtechnologie: Maschinelle Übersetzung



Testen Sie einen neuen Browser mit automatischer Übersetzungsfunktion. [Chrome herunterladen](#) [Schließen](#)

Anmelden

Übersetzer

Von: Englisch ▾



Nach: Deutsch ▾

Übersetzen

Englisch **Spanisch** Deutsch Sprache erkennen

For instance, on the planet Earth, man had always assumed that he was more intelligent than dolphins because he had achieved so much — the wheel, New York, wars and so on — whilst all the dolphins had ever done was muck about in the water having a good time. But conversely, the dolphins had always believed that they were far more intelligent than man — for precisely the same reasons.



Deutsch **Englisch** Französisch

Zum Beispiel auf dem Planeten Erde, hatte man immer angenommen, dass er intelligenter als Delfine war, weil er so viel erreicht hatte - das Rad, New York, Kriege und so weiter - während alle Delfine jemals getan hatte war Dreck im Wasser eine gute Zeit. Aber umgekehrt waren die Delfine immer geglaubt, dass sie viel intelligenter als der Mensch - für genau den gleichen Gründen.



**Neu!** Klicken Sie auf die Wörter oben, um Alternativübersetzungen zu bearbeiten und anzusehen. [Schließen](#)

# Sprachtec

## Übersetzer

| Sprache erkennen | Chinesisch  | Gujarati    | Japanisch   | Litauisch      | Polnisch      | Suaheli       | Walisisch    |
|------------------|-------------|-------------|-------------|----------------|---------------|---------------|--------------|
| Afrikaans        | Dänisch     | Haitianisch | Javanesisch | Malaysisch     | Portugiesisch | Tagalog       | Weißrussisch |
| Albanisch        | Deutsch     | Hausa       | Jiddisch    | Maltesisch     | Punjabi       | Tamil         | Yoruba       |
| Arabisch         | Englisch    | Hebräisch   | Kannada     | Maori          | Rumänisch     | Telugu        | Zulu         |
| Armenisch        | Esperanto   | Hindi       | Katalanisch | Marathisch     | Russisch      | Thailändisch  |              |
| Aserbaidshanisch | Estrnisch   | Hmong       | Khmer       | Mazedonisch    | Schwedisch    | Tschechisch   |              |
| Baskisch         | Finnisch    | Igbo        | Koreanisch  | Mongolisch     | Serbisch      | Türkisch      |              |
| Bengalisch       | Französisch | Indonesisch | Kroatisch   | Nepalesisch    | Slowakisch    | Ukrainisch    |              |
| Bosnisch         | Galizisch   | Irish       | Lao         | Niederländisch | Slowenisch    | Ungarisch     |              |
| Bulgarisch       | Georgisch   | Isländisch  | Lateinisch  | Norwegisch     | Somali        | Urdu          |              |
| Cebuano          | Griechisch  | Italienisch | Lettisch    | Persisch       | Spanisch      | Vietnamesisch |              |



Deutsch Französisch Englisch **Tamil - erkannt**



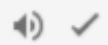
Deutsch Englisch **Französisch**

**Übersetzen**

திருவனந்தபுரம்: கேரளாவின் பூரண மதுவிலக்கை நோக்கிய பயணத்தின் ஒரு கட்டமாக மதுக்கடைகளுக்கு ஞாயிறு தோறும் விடுமுறை விடப் படுவது இன்று முதல் அமல் படுத்தப் பட்டுள்ளது. கேரளாவில் வரும் 10 ஆண்டுகளுக்குள் பூரண மதுவிலக்கை அமல் படுத்த அம்மாநில அரசு திட்டமிட்டுள்ளது. அதன் படி, ஆண்டு தோறும் பத்து பத்து சதவீதமாக மது விற்பனையைக் குறைக்க முடிவு செய்யப்பட்டுள்ளது. இதனால், கேரளாவில் புதிய மது பார்களுக்கு லைசென்சு வழங்கப்பட வில்லை. மேலும் 418 மது பார்களுக்கான லைசென்சு ரத்து செய்யப்பட்டு மது பார்களும் மூடப்பட்டன. காந்தி ஜெயந்தி தினமான கடந்த 2-ந் தேதி கேரளாவில் 10 சதவீத அரசு மதுக்கடைகள் மற்றும் மது பார்கள் மூடப்பட்டதாக அம்மாநில அரசு தெரிவித்துள்ளது. இந்நிலையில், இத்திட்டத்தின் அடுத்த கட்டமாக ஒவ்வொரு ஞாயிற்றுக் கிழமையும் மது கடைகள், பார்களுக்கு விடுமுறை விட வேண்டும் என்றும் கேரள அரசு அறிவித்திருந்தது.

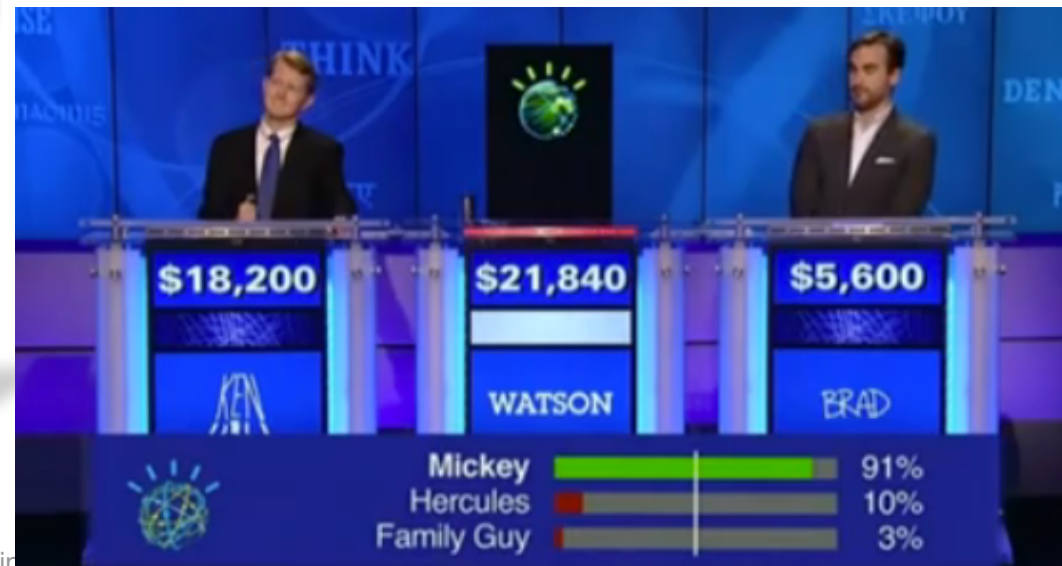


Thiruvananthapuram: Kerala ist ein Schritt auf dem Weg hin zu einer vollständigen matuvilakkai Bars über die Feiertage, als es in der Sonntags hat mit Wirkung ab heute eingeführt. Kerala, ist die Landesregierung plant, in den nächsten 10 Jahren zu implementieren, eine umfassende matuvilakkai. Dementsprechend wurde beschlossen, den Verkauf von Alkohol jährlich auf zehn Prozent zu reduzieren. So hat Kerala Lizenz für den neuen Wein-Bars gewährt. Siehe auch 418 Weinweinstuben geschlossen für die Lizenz wird abgebrochen. Gandhi Jayanti Tag, 10 Prozent der Bundesstaat Kerala auf den letzten 2 Bars und Weinstuben, hat die Landesregierung stillgelegt. In diesem Fall wird die nächste Phase des Projektes jeden Sonntag Spirituosen-Läden, Bars, und die Regierung hat erklärt, Kerala ein Feiertag sein sollte.



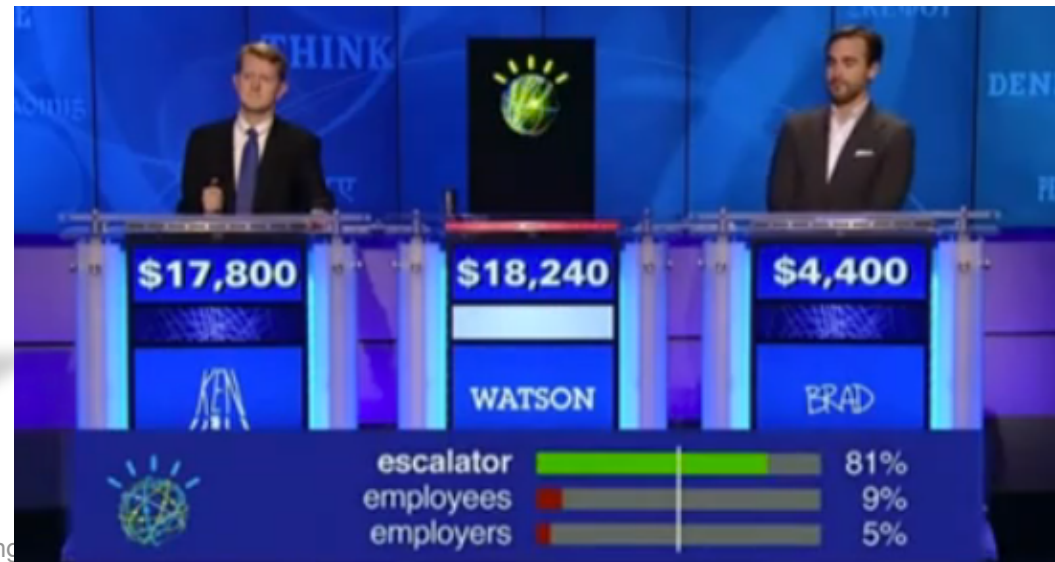
# Artificial Intelligence: Intelligente Assistenten

IN 1939's CARTOON "THE POINTER", THIS GUY GOT A NEW, MORE PEAR-SHAPED BODY & PUPILS WERE ADDED TO HIS EYES



# Artificial Intelligence: Intelligente Assistenten

THIS CLAUSE IN A UNION CONTRACT SAYS THAT WAGES WILL RISE OR FALL DEPENDING ON A STANDARD SUCH AS COST OF LIVING



# Digital Humanities: Digitale Editionen

PARATEXTS  
SEARCH | INDEX  
RESULTS  
DIE FACKEL CONTENTS

**DIE FACKEL CONTENTS**

DATE: 5.1918 ISSUE: XX. JAHR, Heft 474 PAGES: 1-160 29 GO SYNC

**OVERVIEW**

1. Jg. 474-483 (Front cover: contents) (Inside front cover: adverts: contents)

2. Jg.

3. Jg.

4. Jg.

5. Jg. 484-488

6. Jg.

7. Jg.

8. Jg.

9. Jg.

10. Jg.

11. Jg.

12. Jg.

13. Jg.

14. Jg.

15. Jg. 493-500

16. Jg.

17. Jg. 501-507

18. Jg.

19. Jg.

**20. Jg.**

21. Jg.

22. Jg.

23. Jg.

24. Jg.

25. Jg.

26. Jg.

27. Jg.

28. Jg.

29. Jg.

30. Jg.

31. Jg.

32. Jg.

33. Jg.

34. Jg.

35. Jg.

36. Jg.

37. Jg.

**DIE FACKEL TEXT**

PAGE: 1 | TEXT: | ISSUE: Heft 474-483 23.5.1918 | VOLUME: XX. JAHR

**NR. 474—483 MAI 1918 XX. JAHR**

**DIE FACKEL**

**HERAUSGEBER  
KARL KRAUS**

**INHALT:**

Der begabte Czernin / Glossen / Das technoromantische Abenteuer / Für Lammasch / Inschriften / Der darbende Bürger / Glossen / Notizen / Bange Stunde / Halbschlaf / Das zweite Sonett / An eine Falte / Suchen und Finden / Die Flamme der Epimeleia / Programme / Glossen / Ein Staatsstreich / Inschriften / Am Sarg Alexander Girardis / Der Weltspiegel / Glossen / Zum ewigen Frieden.

**NACHDRUCK VERBOTEN**

Preis dieses Heftes:

**3 Kronen = 2 Mark 50 Pf.**

VERLAG ‚DIE FACKEL‘, WIEN  
III/2, HINTERE ZOLLAMTSSTRASSE 3 TELEPHON NR. 187

ERSCHEINT MINDESTENS VIERMAL IM JAHRE.

**DIE FACKEL TEXT**

IMAGE

# Digital Humanities: Analyse & Visualisierung

Voyant Tools

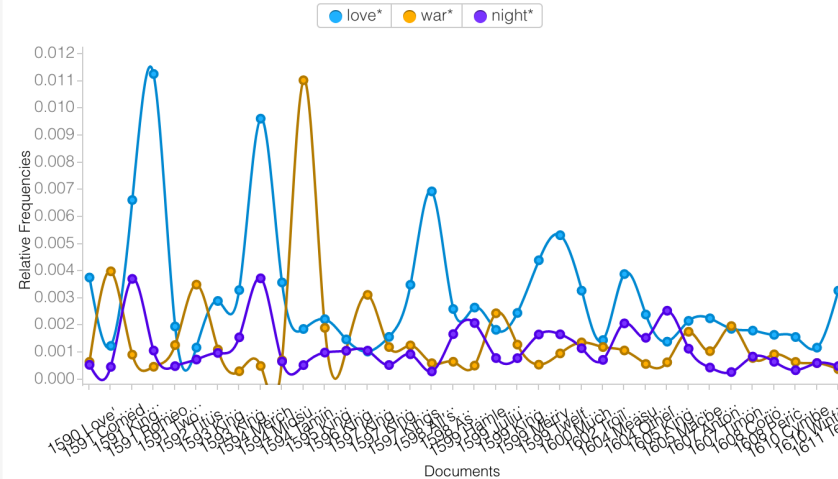
Cirrus Terms Links

Reader TermsBerry

Trends Document Terms



which once ungarbled, much misfortune begets.  
This is the hag, when maids lie on their backs,  
That presses them and learns them first to bear,  
Making them women of good carriage:  
This is she--  
ROMEO Peace, peace, Mercutio, peace!  
Thou talk'st of nothing.  
MERCUTIO True, I talk of dreams,  
Which are the children of an idle brain,  
And, being anger'd, puffs away from thence,  
Which is as thin of substance as the air  
And more inconstant than the wind, who woos  
Even now the frozen bosom of the north,  
And, being anger'd, puffs away from thence,  
Turning his face to the dew-dropping south.  
BENVOLIO This wind, you talk of, blows us from ourselves;  
Supper is done, and we shall come too late.  
ROMEO I fear, too early: for my mind misgives  
Some consequence yet hanging in the stars  
Shall bitterly begin his fearful date  
With this night's revels and expire the term  
Of a despised life closed in my breast  
By some vile forfeit of untimely death.  
But He, that hath the steerage of my course,



love\* x war\* x night\* x ?  
Scale Frequencies Segments: [slider]

Summary Documents Phrases

This corpus has 37 documents with 895,737 total words and 25,114 unique word forms. Created about 7 hours ago.

### Document Length:

- Longest: 1599 Hamlet (32396); 1593 King Richard III (31846); 1608 Coriolanus (29485); 1610 Cymbeline (29204); 1597 King Henry IV (Part... (28225)
- Shortest: 1591 Comedy of Errors (16339); 1594 Midsummer Night's... (17332); 1611 Tempest (17582); 1605 Macbeth (18414); 1591 Two Gentlemen of... (18443)

### Vocabulary Density:

- Highest: 1611 Tempest (0.182); 1605 Macbeth (0.182); 1594 Midsummer Night's... (0.173); 1607 Timon of Athens (0.169); 1591 King Henry VI (Part... (0.167)
- Lowest: 1593 King Richard III (0.129); 1600 Much Ado about Nothi... (0.131); 1604 Othello (0.135); 1599 Merry Wives of Winds... (0.136); 1595 King Henry VI (Part... (0.137)

Contexts Correlations

| Document         | Left                              | Term     | Right                        |
|------------------|-----------------------------------|----------|------------------------------|
| 3) 1591 King...  | Richard, Edward's son, The first- | begotten | and the lawful heir, Of      |
| 3) 1591 King...  | thus removed, Leaving no heir     | begotten | of his body-- I was          |
| 3) 1591 King...  | you have condemn'd: Not me        | begotten | of a shepherd swain, But     |
| 4) 1591 Romeo... | children of an idle brain,        | begot    | of nothing but vain fantasy  |
| 5) 1591 Two...   | thee. Tell me this: who           | begot    | thee? SPEED Marry, the son   |
| 6) 1592 Titus... | AARON First know thou, I          | begot    | him on the empress. LUCIUS   |
| 8) 1593 King...  | the issue was not his             | begot    | ; Which well appeared in his |
| 9) 1594 Merch... | heavens, this is my true-         | begotten | father! who, being more than |

begot\* x ? 32 context expand Scale



## Das Erlanger Modell

„Linguistische Informatik“ =

sprachwissenschaftlich fundierte **Computerlinguistik**  
mit Schwerpunkt auf **korpuslinguistischen** Ansätzen

- Sprachwissenschaft mit dem Computer
- Informatik für Anwendungen, die mit Sprache zu tun haben

... beschäftigt sich mit Methoden, Werkzeugen und Anwendungen der automatischen Verarbeitung und Analyse natürlicher Sprache

- computerlinguistische Methoden als Forschungsgebiet *sui generis*
- bessere Sprachtechnologie durch linguistisches Verständnis, keine „blindes“ *language engineering* oder *deep learning*
- Sprachwissenschaft: objektivierbare Aussagen über Sprache, empirische Überprüfung linguistischer Theorien
- auch Anwendungen in Kognitionswissenschaft, Digital Humanities, ...

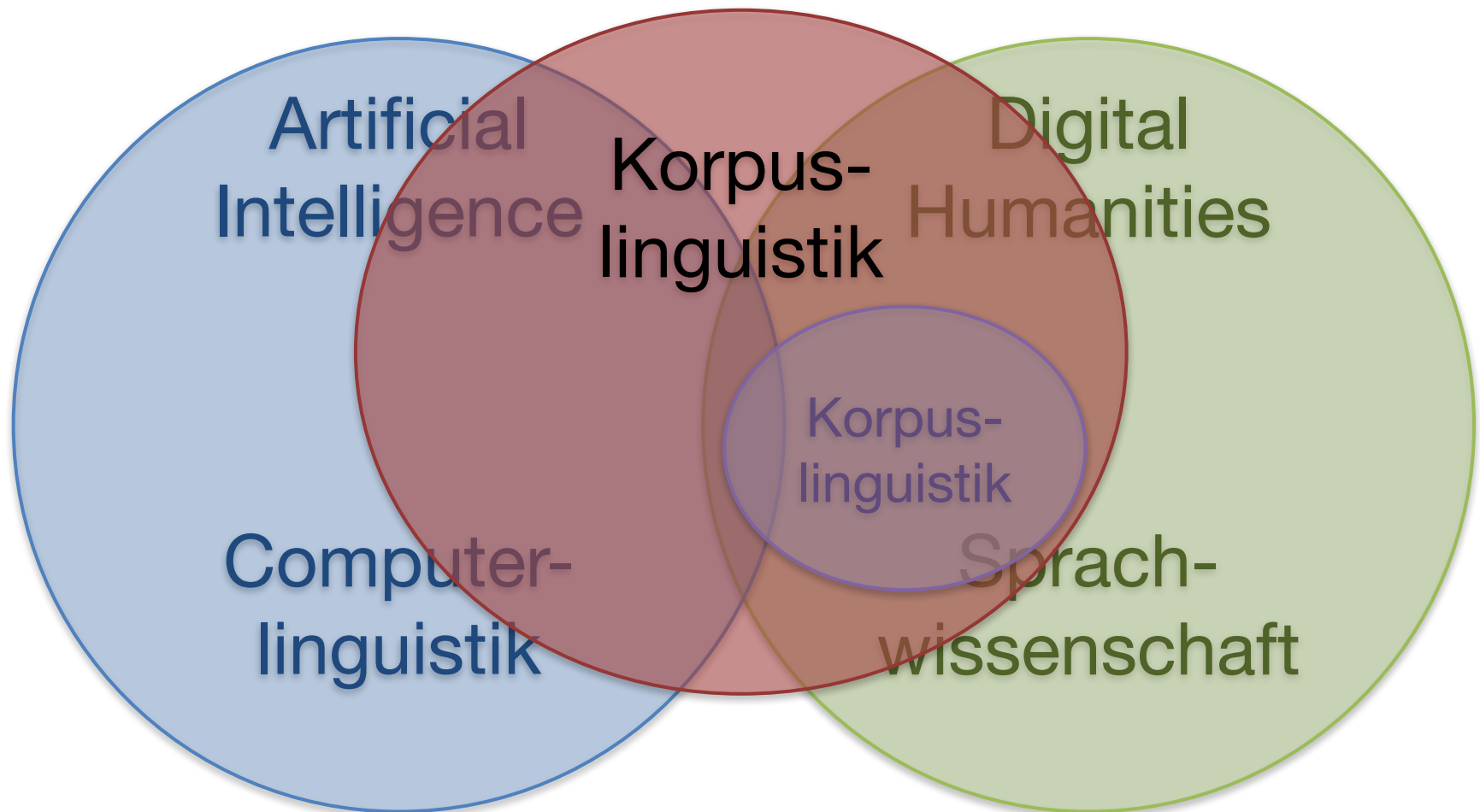
## Korpuslinguistik ≠ Computerlinguistik?

- Korpus (im weiten Sinn)  
= Sammlung von Sprachdaten / Texten in maschinenlesbarer Form
  - sehr große Korpora ( $\geq 100$  M Wörter) sind besonders nützlich
  - Auswertung mit statistischen Methoden und maschinellen Lernverfahren
- Korpus (im engen Sinn)  
= Stichprobe authentischer Sprachdaten / Texte, die für eine bestimmte Sprache oder Sprachvarietät repräsentativ ist
  - z.B. literarische Korpora, Dialekte, gesprochene Sprache, IBK, ...
  - Basis für empirisch fundierte sprachwissenschaftliche Studien
  - zentral für Korpuslinguistik (im engen Sinn) und *Digital Humanities*
- **Korpuslinguistik** (im weiten Sinn)  
= Erstellung, maschinelle Verarbeitung und Auswertung von Korpora

## Korpuslinguistische Fragestellungen (im engen Sinn)

- Kernziel: Erforschung von authentischem Sprachgebrauch
- Empirische Überprüfung linguistischer Theorien
- Lexikographie (neue Einträge, Kollokationen, Präferenzen, ...)
  - IZ Lexikografie, Valenz und Kollokation
- Korpusbasierte (Lerner-)Grammatiken
- Erforschung von Sprachvariation
  - IZ für Dialekte und Sprachvariation
- Erst- und Zweitspracherwerb (→ Lernerkorpora)
- Psycholinguistik (→ Häufigkeitsnormen, Assoziation, ...)
- Historische Sprachwissenschaft
- Digital Humanities (→ Auswertung digitalisierter Bestände)
  - IZ Digitale Geistes- und Sozialwissenschaften (IZdigital)
- Soziologie, Politikwissenschaft und Kulturwissenschaft

## Was ist Korpuslinguistik?



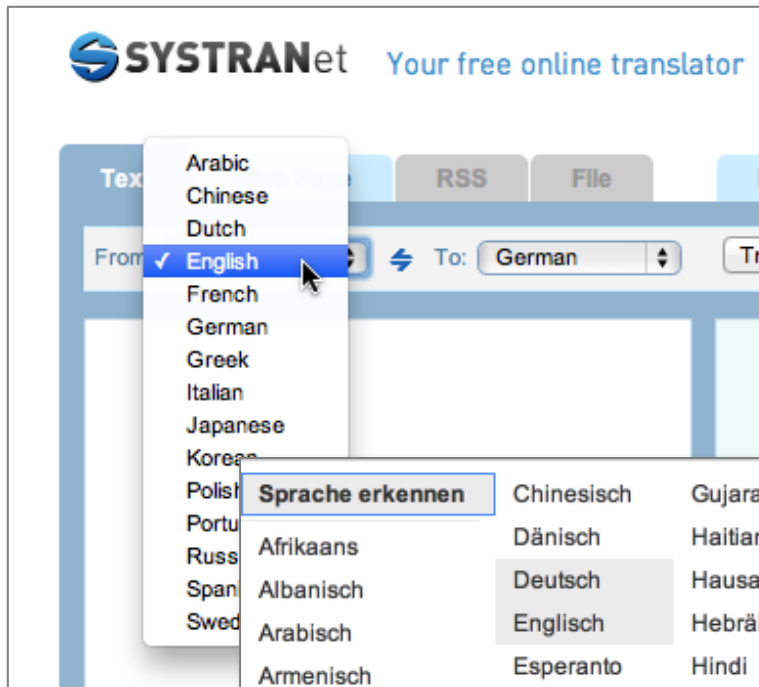
## Was kann man mit Computerlinguistik anfangen?

- Maschinelle Übersetzung (Google, Babel Fish, ...)
- Rechtschreibkorrektur und Grammatikprüfung
- Diktieren & Sprachsteuerung
- Sprachausgabe (z.B. Navi)
- Sprachdialogsystem (z.B. im Auto, Fahrplanauskunft, ...)
- Wörterbücher (zweisprachige, Lerner-WB, Kollokations-WB)
- Semantische Websuche
- Künstliche Intelligenz: Siri, Alexa, Cortana, Watson, ...
- Meinungs- & Marktforschung
- Information Retrieval & Question Answering
- Text / Knowledge Mining (z.B. Biomedizin, Patentrecherche)
- Social Bots & Fake News
- Kontrollierte Terminologie
- Plagiaterkennung
- Automatische Bewertung von Klausuren und Hausarbeiten
- Spam schreiben & erkennen
- Forensik & Aufklärung (z.B. linguistische Steganographie)

## Wo wird Korpustechnologie praktisch eingesetzt?

- Maschinelle Übersetzung (Google, Babel Fish, ...) ✓
- Rechtschreibkorrektur und Grammatikprüfung ✓
- Diktieren & Sprachsteuerung ✓
- Sprachausgabe (z.B. Navi) ✓
- Sprachdialogsystem (z.B. im Auto, Fahrplanauskunft, ...) ✓
- Wörterbücher (zweisprachige, Lerner-WB, Kollokations-WB) ✓
- Semantische Websuche ✓
- Künstliche Intelligenz: Siri, Alexa, Cortana, Watson, ... ✓
- Meinungs- & Marktforschung ✓
- Information Retrieval & Question Answering ✓
- Text / Knowledge Mining (z.B. Biomedizin, Patentrecherche) ✓
- Social Bots & Fake News ✓
- Kontrollierte Terminologie ✓
- Plagiaterkennung ✓
- Automatische Bewertung von Klausuren und Hausarbeiten ✓
- Spam schreiben & erkennen ✓
- Forensik & Aufklärung (z.B. linguistische Steganographie) ✓

# Beispiel: Vorteile korpusbasierter Sprachtechnologie



| Sprache erkennen  |             | Chinesisch  | Gujarati    | Japanisch      | Litauisch     | Polnisch      | Suaheli      | Walisisch |
|-------------------|-------------|-------------|-------------|----------------|---------------|---------------|--------------|-----------|
| Afrikaans         | Dänisch     | Haitianisch | Javanesisch | Malaysisch     | Portugiesisch | Tagalog       | Weißrussisch |           |
| Albanisch         | Deutsch     | Hausa       | Jiddisch    | Maltesisch     | Punjabi       | Tamil         | Yoruba       |           |
| Arabisch          | Englisch    | Hebräisch   | Kannada     | Maori          | Rumänisch     | Telugu        | Zulu         |           |
| Armenisch         | Esperanto   | Hindi       | Katalanisch | Marathisch     | Russisch      | Thailändisch  |              |           |
| Aserbaidshjanisch | Estnisch    | Hmong       | Khmer       | Mazedonisch    | Schwedisch    | Tschechisch   |              |           |
| Baskisch          | Finnisch    | Igbo        | Koreanisch  | Mongolisch     | Serbisch      | Türkisch      |              |           |
| Bengalisch        | Französisch | Indonesisch | Kroatisch   | Nepalesisch    | Slowakisch    | Ukrainisch    |              |           |
| Bosnisch          | Galizisch   | Irish       | Lao         | Niederländisch | Slowenisch    | Ungarisch     |              |           |
| Bulgarisch        | Georgisch   | Isländisch  | Lateinisch  | Norwegisch     | Somali        | Urdu          |              |           |
| Cebuano           | Griechisch  | Italienisch | Lettisch    | Persisch       | Spanisch      | Vietnamesisch |              |           |

# Lernen von Übersetzungsmustern aus Parallelkorpora

Linguee

<http://www.linguee.de/>

Deutsch

↔

Englisch

ä ö ü ß

kick the bucket

Suchen

**Übersetzungsbeispiele aus fremden Quellen für 'kick the bucket':**

| Englisch   | Deutsch  |
|--|--|
| <p>⚠ That's what two cancer-ward roommates, and irascible billionaire (Jack Nicholson) and a scholarly mechanic (Morgan Freeman), decide when they get the bad news. They compose a bucket list - things to do before <b>you kick the bucket</b> - and head off for the around-the-world adventure of their lives. <span style="float: right;">↳ lobigo.com</span></p> | <p>⚠ [...] auf der Krebsstation, als sie die unerbittliche Diagnose hören: Ein jähzorniger Milliardär (Jack Nicholson) und ein gebildeter Automechaniker (Morgan Freeman) stellen eine Löffel-Liste zusammen mit all den Dingen, die sie noch tun wollen, bevor sie <b>den Löffel</b> abgeben. <span style="float: right;">↳ lobigo.com</span></p> |
| <p>⚠ When some of the characters <b>'kick the bucket'</b> it will hurt also the audience that has become accustomed to the [...]. <span style="float: right;">↳ tonycaulfield.com</span></p>   | <p>⚠ Wenn einige der Figuren <b>'in's Gras beißen'</b> wird es weh tun, auch dem Zuschauer, der sich an die Figuren gewöhnt hat und mit ihnen mitfiebert. <span style="float: right;">↳ tonycaulfield.com</span></p>   |
| <p>⚠ If I was alone here, I would probably have to <b>kick the bucket</b> and nobody would care. <span style="float: right;">↳ denis-katzer.de</span></p>  | <p>⚠ [...] wäre hier alleine müsste ich <b>wahrscheinlich verrecken und keiner würde sich darum kümmern</b>. <span style="float: right;">↳ denis-katzer.de</span></p>  |
| <p>⚠ [...] day they just come down with something, or out of the blue they have problems, and there's even a chance one day they'll just develop psychosis or maybe <b>even kick the bucket</b>. <span style="float: right;">↳ falundafa.org</span></p>  | <p>⚠ Ein gewöhnlicher Mensch wird ja doch eines Tages krank sein oder eines Tages auf irgendwelche Probleme <b>stoßen</b>, eines Tages wird er vielleicht geisteskrank oder dahingehen. <span style="float: right;">↳ falundafa.de</span></p>  |
| <p>⚠ Were the next to <b>kick the bucket</b> if we dont get a lift by some guardian angel who happens to fly by -</p>  | <p>⚠ Wir beißen als nächste <b>ins Gras</b>, wenn uns nicht ein vorbeifliegender Schutzengel aufammelt - wir</p>   |



## Der Studiengang Linguistische Informatik

- Bachelorstudiengang (6 Semester)
  - 10 Module mit insgesamt 70 ECTS
  - Praktikum in Forschungsprojekt oder bei Industrieunternehmen
- Kombination mit einem sprachwissenschaftlichen Fach
  - z.B. Germanistik, Anglistik, Romanistik, Japanologie, ...
  - inzwischen auch andere Kombinationsfächer erlaubt
- **Doppelqualifikation** durch
  - linguistische Ausbildung im sprachwissenschaftlichen Fach
  - praktische Anwendung der linguistischen Theorie und computerlinguistischen Methoden im Fach Linguistische Informatik
  - zusätzliche Vorlesungen und Übungen aus der Informatik

# Studienziele

- Theoretische Kompetenzen
  - Logik, Mengenlehre, formale Sprachen, Grammatiken, Wk-Theorie
  - computerlinguistische Aufgabenstellungen und Lösungsansätze
  - Korpuslinguistik, deskriptive und analytische Statistik
  - Verfahren der maschinellen Sprachverarbeitung, maschinelles Lernen
  - sowie Sprachwissenschaft, Informatik, Humanities (Import / 2. Fach)
- Praktische Kompetenzen
  - Linux-Shell, Emacs/LaTeX, reguläre Ausdrücke, Datenformate (z.B. XML)
  - Skriptprogrammierung mit Python
  - Verwendung computerlinguistischer Werkzeuge und Ressourcen
  - Implementierung sprachtechnologischer Anwendungen
  - Durchführung quantitativer korpuslinguistischer Studien

## Studienziele

- Soft skills
  - Teamwork
  - konstruktive Diskussion
  - Vermittlung von Wissen und Ergebnissen in wissenschaftlichen Arbeiten und Vorträgen
  - Berufserfahrung (→ Praktikum)
  - sowie weitere Soft skills aus zweitem Fach
- Allgemeine Ziele
  - praxis- und forschungsorientiertes Studium
  - Verzahnung mit Infrastruktur und Forschungsthemen der Professur

# Studienplan



| Semester 1<br>(WiSe)  | Semester 2<br>(SoSe)                                | Semester 3<br>(WiSe)   | Semester 4<br>(SoSe)   | Semester 5<br>(WiSe)                           | Semester 6<br>(SoSe) |
|---|---|--|--|--|----------------------|
| <b>VL Grundlagen der CL 1</b><br>2 SWS; 2 ECTS                                    | <b>VL Grundlagen der CL 2</b><br>2 SWS; 2 ECTS      | <b>Proseminar Computerling.</b><br>2 SWS; 5 ECTS                               | <b>HS Korpuslinguistik</b><br>2 SWS; 5 ECTS                            |  |                      |
| <b>Ü Grundlagen der CL 1</b><br>2 SWS; 3 ECTS                                     | <b>Ü Grundlagen der CL 2</b><br>2 SWS; 3 ECTS       | <b>Ü Computerling. Werkzeuge und Infrastrukturen</b><br>2 SWS; 5 ECTS          | <b>Ü Statistik</b><br>2 SWS; 5 ECTS                                    |  |                      |
| <b>Ü Arbeitstechniken der CL</b><br>2 SWS; 2,5 ECTS                               | <b>Grundseminar Programmierung</b><br>2 SWS; 5 ECTS | <b>Aufbauseminar Programmierung</b><br>2 SWS; 5 ECTS                           | <b>Hauptseminar theoretisch</b><br>2 SWS; 5 ECTS;<br>Semester 4 oder 6 | <b>Hauptseminar praktisch</b><br>2 SWS; 5 ECTS |                      |
| <b>Vorlesung Grundlagen der Informatik (Gdl)</b><br>3 SWS; insg. 7,5 ECTS; Import |   | <b>V Konzeptionelle Modellierung</b><br>2 SWS; 2,5 ECTS;<br>Import; Sem. 2 – 4 |  |  |                      |
| <b>Tafelübung Gdl</b><br>2 SWS; s.o. ECTS; Import                                 |   | <b>U Konz. Mod.</b><br>2 SWS; 2,5 ECTS;<br>Import; Sem. 2 – 4                  |  |  |                      |
| <b>Rechnerübung Gdl</b><br>1 SWS; s.o. ECTS; Import                               |   |  |  |  |                      |
| 12 SWS<br>15 ECTS   | 6 SWS<br>10 ECTS                                    | 10 SWS<br>20 ECTS  | 6 SWS<br>15 ECTS   | 2 SWS<br>5 ECTS                                | 0 SWS<br>0 ECTS      |

**Praktikum**  
150h; 5 ECTS

# Der Studiengang Linguistische Informatik

- (Noch) kleiner Studiengang: ca. 50–60 Studierende
- Aktive Fachschaftsinitiative:  
<https://blogs.fau.de/klue/>
- KLUE-Stammtisch
  - erstes Treffen: heute, 16.10., 18:00 im *Murphy's Law*
- Weitere Informationen zum Studium unter  
<http://www.linguistik.uni-erlangen.de/studium-lehre/>

## Nach dem Studium

- Masterstudium / Promotion
  - Computerlinguistik
  - Sprachwissenschaft
  - Digital Humanities
  - Angewandte Informatik
- Tätigkeitsbereiche in der Wirtschaft (→ Sprachtechnologie)
  - Google, Microsoft, Yahoo, Twitter, Siemens, ...
  - Text Mining, Information Retrieval, Search Engines, ...
  - Lexikographie und Terminologie
  - Spracherkennung und Sprachsynthese, Dialogsysteme
  - Computergestützter Sprachunterricht (CALL)
  - Viele Start-Up-Unternehmen im IT-Bereich suchen Computerlinguisten!

## Praktische Informationen

### Unser Team

- Prof. Dr. Stefan Evert  
[stefan.evert@fau.de](mailto:stefan.evert@fau.de)
- Dr. Besim Kabashi  
[besim.kabashi@fau.de](mailto:besim.kabashi@fau.de)
- Thomas Proisl, M.A.  
[thomas.proisl@fau.de](mailto:thomas.proisl@fau.de)
- Paul Greiner, M.A.  
[paul.greiner@fau.de](mailto:paul.greiner@fau.de)
- Philipp Heinrich, M.Sc.  
[philipp.heinrich@fau.de](mailto:philipp.heinrich@fau.de)

Bismarckstr. 6  
Raum 4.000  
(Büro Professur)

Philologien



(CIP-Pool)  
Bismarckstr. 12  
Raum 0.320

Ganz wichtig: <http://www.linguistik.uni-erlangen.de/>



## Professur für Korpuslinguistik Studiengang Linguistische Informatik

[Kontakt](#) | [Impressum](#) | [Inhaltsübersicht](#)

Sie befinden sich hier: [▶ Startseite](#) [▶ Aktuelles](#)

### Aktuelles

#### Homepage wird z.Zt. überarbeitet

Die Webseiten der Professur für Korpuslinguistik befinden sich derzeit in Überarbeitung und können daher nicht in allen Punkten aktualisiert werden. Ein neuer Webauftritt wird in Kürze online gehen.

### Einführungsveranstaltung des Faches *Linguistische Informatik* im WS 2017/18

findet am Montag, den 16.10., von 16.15 bis 17.45 Uhr im Raum 0.320, Bismarckstraße 12, 91054 Erlangen statt.

Herzlich willkommen!

Alte Folien (WS 2015/16), PDF, 4.4 MB

#### Aktuelles

- **Einführungsveranstaltung** für Studienanfänger am **Mo, 16.10., 16:15–17:45** (Bismarckstr. 12, Raum 0.320)
- **Stundenplan** im WS 2017/18
- Empfehlungen für **Schlüsselqualifikationen**: *Austerität im Vereinigten Königreich, Automatic Question Answering Using IBM Watson, Web4Science (Webdesign), Deep Learning, Informationsvisualisierung, Logik-basierte Sprachverarbeitung*
- **Sprechstunden** von Prof. Evert jeweils **Mo 12–14** (Bismarckstr. 6, Raum 4.000)



# Lehrangebot im WS 2017/18

- VL + Ü Grundlagen der Computerlinguistik 1
  - Ü Arbeitstechniken der Computerlinguistik
  - VL + TÜ + RÜ Grundlagen der Informatik
- } **1. Sem.**
- PS Computerlinguistik
  - Ü Computerling. Werkzeuge und Infrastrukturen
  - PS Aufbauseminar Python
  - VL + Ü Konzeptionelle Modellierung
- } **3. Sem.**
- Praktisches HS: Social Bots & Fake News
  - Oberseminar Computerlinguistik
- } **5. Sem.**
- S Automatic Question Answering Using IBM Watson
  - Web4Science – Visualisierung mit modernem Webdesign
  - VL + Ü Informationsvisualisierung
  - VL + Ü Deep Learning
  - VL + Ü Logikbasierte Sprachverarbeitung
- } **SQ**

# Stundenplan

|              | Mo   | Di  | Mi   | Do  | Fr  |
|--------------|--|---|--|---|---|
| <b>08:00</b> |  |   |  | 08:15 - 10:45<br><b>Grundlagen der Informatik</b><br>(Bauer) H7, H8, H10<br><br>10:15 - 11:45<br><b>Konzeptionelle Modellierung</b><br>(Lenz) H11 |   |
| <b>09:00</b> |  |   |  |   |   |
| <b>10:00</b> |  |   |  |   | 10:15 - 11:45<br><b>Computerlinguistische<br/>Werkzeuge und<br/>Infrastrukturen</b><br>(Proisl) 0.320 Bismarckstr. 12 |
| <b>11:00</b> |  |   |  |   |   |
| <b>12:00</b> |  |   |  | 12:15 - 13:45<br><b>Übung Grundlagen der<br/>Computerlinguistik 1</b><br>(Kabashi) 0.320 Bismarckstr.<br>12                                       |   |
| <b>13:00</b> |  |   |  |   |   |
| <b>14:00</b> | 14:15 - 15:45<br><b>Arbeitstechniken der<br/>Computerlinguistik</b><br>(Kabashi) 0.320 Bismarckstr.<br>12  | 14:15 - 15:45<br><b>Aufbauseminar<br/>Programmierung Python</b><br>(Proisl) 0.320 Bismarckstr. 12               | 14:15 - 15:45<br><b>Social Bots &amp; Fake News</b><br>(Evert) 0.320 Bismarckstr. 12       |   |   |
| <b>15:00</b> |  |   |  |   |   |
| <b>16:00</b> | 16:15 - 17:45<br><b>Vorlesung Grundlagen der<br/>Computerlinguistik 1</b><br>(Evert) 0.320 Bismarckstr. 12 | 16:15 - 17:45<br><b>Proseminar<br/>Computerlinguistik</b><br>(ab 24.10.)<br>(Heinrich) 0.320 Bismarckstr.<br>12 | 16:15 - 17:45<br><b>Oberseminar<br/>Computerlinguistik</b><br>(Evert) 4.000 Bismarckstr. 6 |   |   |
| <b>17:00</b> |  |   |  |   |   |

# Grundlagen- und Orientierungsstudium (GOS)

## Unsere Themen

- Orientierungsangebote für Studienanfänger/-innen
  - Orientierungswoche
- Selbstorganisation und selbstreguliertes Lernen
  - Verwertung von Lehrveranstaltungen
  - Prüfungsvorbereitung
  - Zeitmanagement
- Prinzipien des wissenschaftlichen Arbeitens und Schreibens
  - Wissenschaftliche Texte lesen
  - Hausarbeiten planen und konzipieren
  - Schreibprojekte aktiv gestalten und managen

## Veranstaltungen im WiSe 2017/18

- Das Studium erfolgreich managen mit Lerntechniken für die Prüfungsvorbereitung
- Selbstkompetenzen für Studium und Beruf stärken
- Souveränes Konfliktmanagement
- Studieren mit guter Work-Life-Balance
- In den Schreibfluss kommen
- Schreibtutor\*innenausbildung

Weitere Informationen:

<http://ziwis.fau.de/gos>

## Nächste Schritte

- LVen werden in der Regel durch **StudOn-Kurse** begleitet
  - bitte treten Sie diesen Kursen jetzt bei!
  - keine formale Anmeldung erforderlich
- Prüfungsanmeldung über meinCampus nicht vergessen!
  - Anmeldezeitraum: 3 Wochen ca. Ende Nov. – Anfang Dez.
- **Accounts** für CIP-Pool (über IdM-Kennung)
  - Anmeldeformulare jetzt oder in Ü Arbeitstechniken der CL
- Sprechstunde Prof. Dr. Stefan Evert
  - Montag, 12–14 Uhr, Raum 4.000 (Bismarckstr. 6)
  - kurzfristige Änderungen auf Homepage („Aktuelles“)