

Informationsveranstaltung

BA Computerlinguistik

Korpus- und Computerlinguistik

Prof. Dr. Stephanie Evert

Lehrstuhl für Korpus- und Computerlinguistik

<https://www.linguistik.phil.fau.de/>



FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG

PHILOSOPHISCHE FAKULTÄT
UND FACHBEREICH THEOLOGIE

Was genau ist eigentlich ein Korpus?

Das Korpus
Die Korpora

- Korpus¹ (im weiten Sinn)
= Sammlung von Sprachdaten / Texten in maschinenlesbarer Form
 - sehr große Korpora (≥ 100 M Wörter) sind besonders nützlich
 - Auswertung mit statistischen Methoden und maschinellen Lernverfahren
- Korpus² (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*
- **Korpus- und Computerlinguistik**
= Erstellung, maschinelle Verarbeitung und Auswertung von Korpora¹

Sprachtechnologie: Maschinelle Übersetzung

≡ Google Translate



Text

Documents

DETECT LANGUAGE

ENGLISH

SPANISH

FRENCH



GERMAN

ENGLISH

SPANISH



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. ✕

Zum Beispiel hatte der Mensch auf dem Planeten Erde immer angenommen, er sei intelligenter als Delfine, weil er so viel erreicht hatte - das Rad, New York, Kriege und so weiter -, während alles, was die Delfine jemals getan hatten, im Wasser herumgespielt hatte eine gute Zeit haben. Umgekehrt hatten die Delfine immer geglaubt, sie seien weitaus intelligenter als der Mensch - aus genau den gleichen Gründen. ☆



380 / 5000





Sprachtechnologie: Maschinelle Übersetzung



DeepL

Übersetzer ▾

DeepL Pro

Für Unternehmen ▾

Kostenloses Probeabo starten



Anmelden



Text übersetzen

31 Sprachen



Dateien übersetzen

.pdf, .docx, .pptx



DeepL Write

KI-Schreibassistent

Englisch (erkannt) ▾



Deutsch ▾

automatisch ▾

Glossar

For instance, on 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 much 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.



Auf dem Planeten Erde zum Beispiel hatte der Mensch immer angenommen, dass er intelligenter sei als die Delphine, weil er so viel erreicht hatte - das Rad, New York, Kriege und so weiter - während alles, was die Delphine je getan hatten, darin bestand, im Wasser herumzutollen und sich zu amüsieren. Umgekehrt hatten die Delfine immer geglaubt, sie seien viel intelligenter als der Mensch - aus genau denselben Gründen.



Sprachtec

Google Translate

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.

Künstliche Intelligenz: virtuelle Assistenten

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



Künstliche Intelligenz: virtuelle Assistenten

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



Künstliche Intelligenz: ChatGPT



How can I train a bag-of-words classifier in Python?



Training a bag-of-words (BoW) classifier in Python involves several steps, including text preprocessing, feature extraction, and model training. Here's a step-by-step guide on how to do it using popular libraries like scikit-learn:



1. Import Libraries:

Start by importing the necessary libraries, including scikit-learn and any other relevant modules.

python

Copy code

```
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score, classification_report
```

2. Data Preparation:

Load and preprocess your text data. This may include cleaning the text, tokenizing it, and organizing it into a format suitable for the BoW model. Here's a simple example:

python

Copy code

```
# Sample data
text_data = ["This is a positive review.", "This is a negative review.", "A

# Corresponding labels (binary classification)
labels = [1, 0, 1]
```

3. Split the Data:

Split your data into a training set and a testing set to evaluate the classifier.

Digital Humanities: Digitale Editionen

PARATEXTS
SEARCH | INDEX
RESULTS
DIE FACKEL CONTENTS

DIE FACKEL CONTENTS

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

OVERVIEW

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

2. Jg. 1 Der begabte Czernin
23 ► Glossen
23 Niemand geringerer als
25 Autor und Direktor
26 Von einem Mann namens Ernst Posse
26 Ein Kronzeuge für die österreichische Regierung
26 Ein Ausspruch
27 Nimmermehr
27 Auf Deutsches
28 Österreich 1918
28 Eine neue Naturgewalt
28 Fremdwörterchutz
31 Sauerb
32 Die Entschädigten
33 Rassenunterschiede
33 Unterricht
34 Gut getroffen
35 Die überlegene Wirkung unserer Gase
36 Wir Deutschen begrüßen alle Versuche, dem Völkerrecht und der Menschlichkeit zum Siege zu verhelfen, mit Freude, wir lehnen es aber ab, uns übertölpeln zu lassen
36 Mit G.
37 Ein Wunder
37 Biblisches
38 Mein Ehrenwort, daß das nicht von mir ist
38 Die große Kanone oder Beweis gegen Barbaentum
39 Der Praeceptor Germaniae
40 Lionardo da Vinci
41 Das technoromantische Abenteuer
46 Für Lammasch
50 ▼ Inschriften
52 Der darbenende Bürger
57 ▼ Glossen
69 ▼ Notizen
78 Bange Stunde
81 Halbschlaß
82 Das zweite Sonett der Louise Labé
83 An eine Falte
84 Suchen und Finden
85 Die Flamme der Epimeleia
87 ▼ Programme
95 ▼ Glossen
113 Ein Staatsstreich
119 ▼ Inschriften

DIE FACKEL TEXT

PAGE TEXT ISSUE VOLUME
Heft 474-483 23.5.1918 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 darbenende Bürger / Glossen / Notizen / Bange Stunde / Halbschlaß / 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.

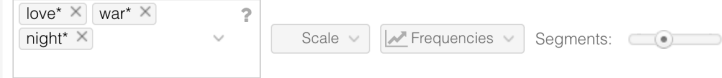
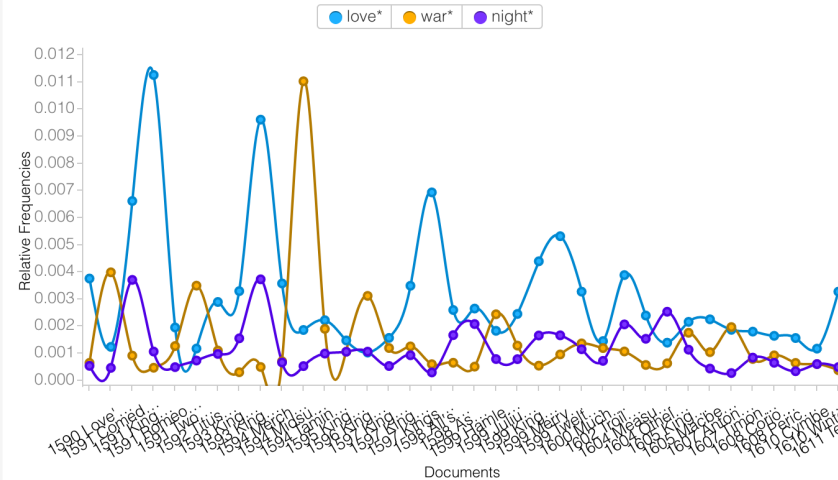
IMAGE

<https://www.oeaw.ac.at/fackellex/aac/aac.html>

Digital Humanities: Analyse & Visualisierung



which once unangled, much misfortune becometh.
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,
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,



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 (0.168); 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)

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) 1592 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* 32 context expand Scale



Team

Department

Profil

Team

Kontakt



Prof. Dr. Stephanie Evert

Zweitmitglied



Prof. Dr. Peter Gentzel

Zweitmitglied



Prof. Dr. Georg Glasze

Zweitmitglied



Prof. Dr. Svenja Hagenhoff

Departmentsprecherin



Prof. Dr. Michael Kohlhasse

Zweitmitglied



Dr. Dominik Kremer

Wissenschaftlicher Mitarbeiter



Dr. Sabine Lang

Wissenschaftliche Mitarbeiterin



Prof. Dr. Ayaka Löschke

Zweitmitglied



Prof. Dr.-Ing. habil. Andreas Maier

Zweitmitglied



Christian Sandig, M.A.

Geschäftsführer

Sprachwissenschaft: Empirische Sprachbeschreibung

BNCweb Query result

http://localhost/bncweb-cgi/main.pl?program=sort&textOrSpeaker=&theData=%5Bword%3D%22linguistics%22%5D

Your query "linguistics" returned 784 hits in 100 different positions. **on position -1 with tag-restriction any adjective (224 hits)**

Sort parameters: Position: 1 Left Tag restriction:

No	Filename	
201	HGR 481	Generative grammar rejects the empirical na
202	CGY 1266	In the semiological phase his work is inspir
203	CGY 1183	The structuralism that Lévi-Strauss applies i
204	H8V 441	But by integrating these theories with their v
205	CGY 1271	But in his second phase the issues of custom
206	G1N 89	CM2: Minds, machines and evolution. Sample containing about 38482 words from a book (domain: belief and thought)
207	CGY 14	Spoken or Written: Written Number of Words (tagged items): 38,616 Average sentence length (w-tags per s-unit): 27.8816 Derived text type: Academic prose Genre: W:ac:humanities_arts Text type: Written books and periodicals Publication date: 1985-1993
208	CMR 1	Age of Author: unknown Domicile of Author: unknown Sex of Author: unknown Type of Author: Multiple Age of Audience: Adult Text Domain: Informative: Belief and thought Perceived level of difficulty: High Medium of Text: Book Place of publication: UK: South (south of Bristol Channel-Wash line) Text Sample: Whole text Estimated circulation size: High Target audience sex: Mixed
209	CGY 1	
210	KAM 6	
211	HH3 13	
212	J7F 18	
213	CM2 563	Even if my assessment of its implications concerning the relative order of emergence of the intentional ingredient and of syntactic structure were held to be incorrect, the mechanism of that evolution might be of independent interest, and be seen as bearing on other problems besides (especially in developmental psychology and theoretical_AJ0 linguistics).

bucket *noun*

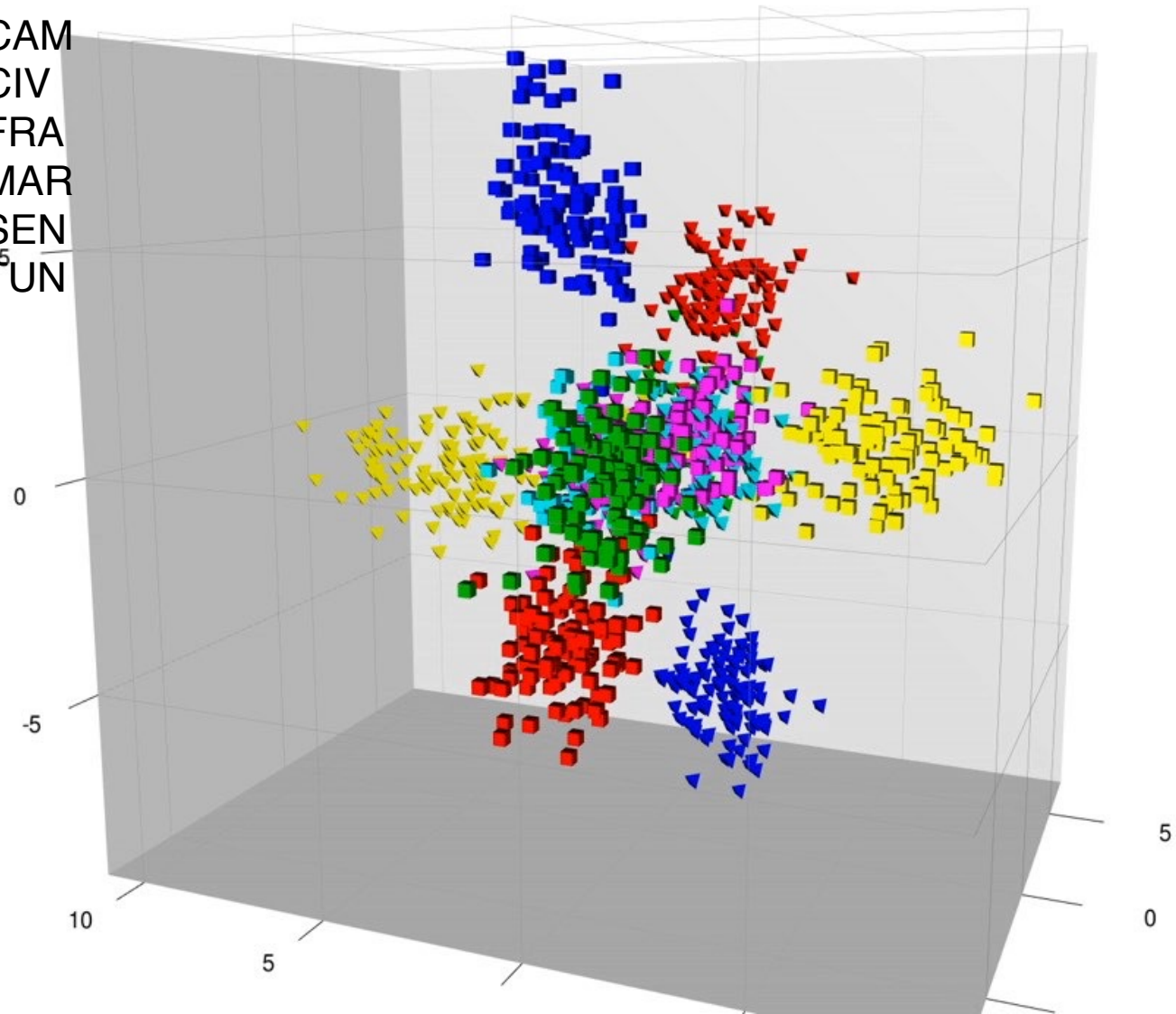
- ADJ. **empty, full | galvanized, metal, plastic | leaky | champagne, ice | water | slop, waste** *a slop bucket full of scraps of food | mop | coal | fire* *The sand had spilt from the fire bucket.*
- VERB + BUCKET **fill** *She filled the bucket with fresh water. | carry | empty, pour, throw* *She poured the bucket of dirty water down the drain.*
- BUCKET + VERB **be filled with/full of sth, contain sth, hold sth | overflow**
- PREP. **in a/the ~ | ~ of a bucket of oats for the horses**
- PHRASES **a bucket and spade** *The children ran down to the beach with their buckets and spades. mop and bucket* *The cleaner put down his mop and bucket and sat down.*

Go to "http://localhost/bncweb-cgi/fileInfo.pl?text=CM2&urlTest=yes"

Sprachwissenschaft: Quantitative Linguistik

- MUTA
- ▲ TRIB
- FRAT
- ▲ VOIE
- LFI
- ▲ LM
- AJD
- ▲ MAT
- SOL
- ▲ WALFA
- LAPRE
- ▲ TEMPS

- CAM
- CIV
- FRA
- MAR
- SEN
- TUN



RTG Dimensions of Constructional Space

Welcome to the DFG-Funded Research Training Group of the Dimensions of Constructional Space at FAU Erlangen-Nürnberg

The Research Training Group will explore a relatively new paradigm in linguistics that has become known under the name of Construction Grammar (CxG). The fundamental premise of this approach is that the totality of speakers' linguistic knowledge is represented in a network of form-meaning pairings called constructions. Constructions differ in size (ranging from morphemes through phrasal and clausal templates to discourse conventions), abstractness (fully elaborated, partially schematic, fully schematic) and entrenchment, as well as in the way they interconnect with other constructions in the network. These properties define a multidimensional space we call 'constructional space'.

FJUEL 2023

📅 6. October 2023 📁 Category: [PhD Researchers](#), [Talks and Events](#)



On September 29th and 30th, 2023, some of the members of our RTG took part in the annual FJUEL conference, hosted at Friedrich-Alexander Universität Erlangen-Nürnberg. The Forum Junge Englische Linguistik in Bayern (FJUEL) is a platform where doctoral and post-doctoral researchers, as well as stu...

[Continue >](#)

Contact



Prof. Dr. Ewa Dabrowska

Project Leader
Department of English and
American Studies
Lehrstuhl für Language und
Cognition (Alexander von
Humboldt-Professur)

Bismarckstr. 6
91054 Erlangen

☎ +49 9131 85-29263

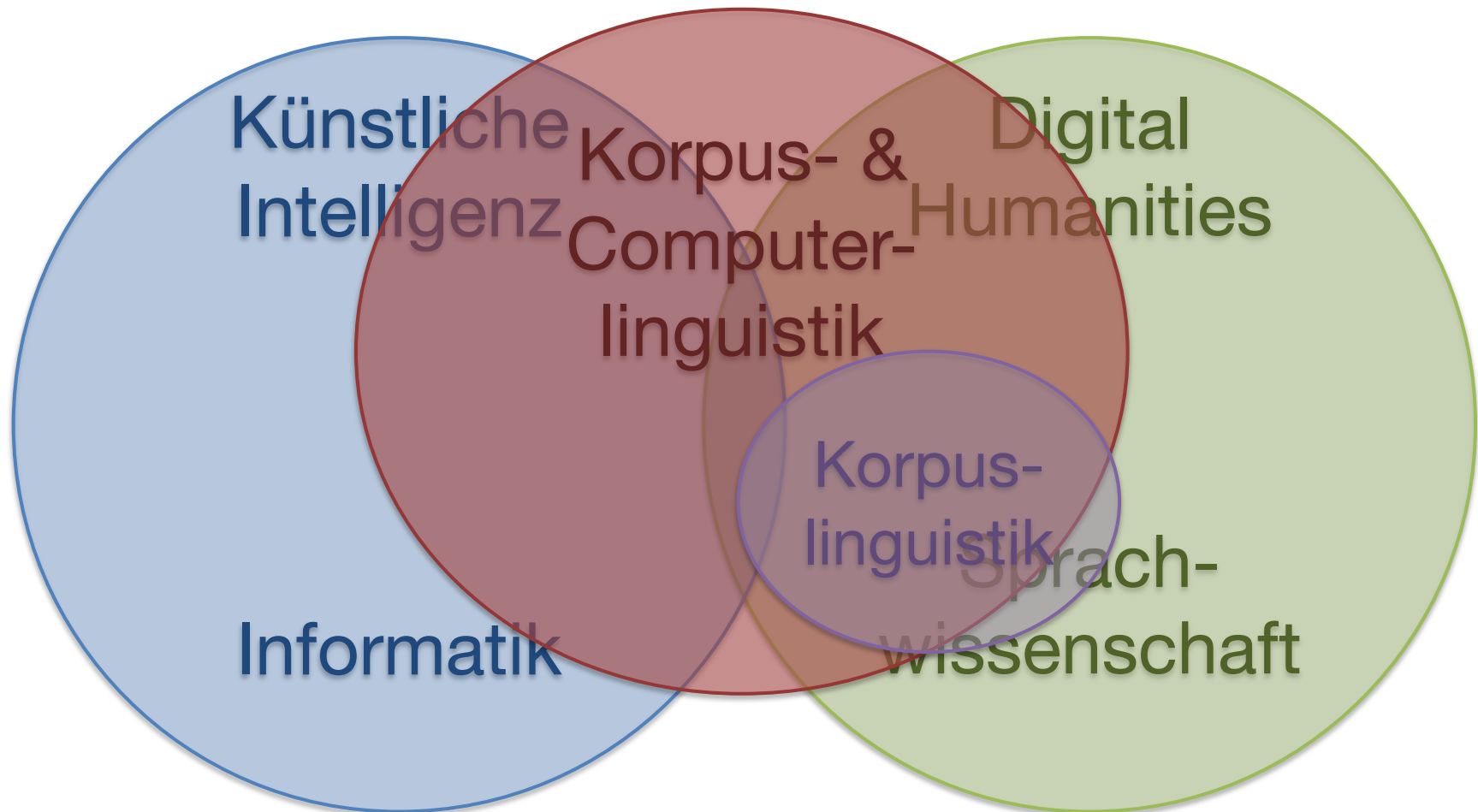
✉
ewa.dabrowska@fau.de

Dr. Susanne Dyka

Project Coordinator
Lehrstuhl für Language and Cognition
Institut für Anglistik und Amerikanistik

Bismarckstraße 6
91054 Erlangen

Wie war das jetzt mit Korpus- und Computerlinguistik?



BA Computerlinguistik

- Fokus auf maschinelle Lernverfahren (insb. **Deep Learning**)
 - folgt dem Trend der modernen Computerlinguistik (und KI)
 - Schwerpunkt des Studiums: Operationalisierung von computerlinguistischen Aufgaben als Lernprobleme
 - „Baukasten“ geeignet für zahlreiche Anwendungen in Sprachtechnologie, DH, Korpuslinguistik
- Programmierpraxis: **Python**
 - konsequenter Einsatz von Python und entsprechenden Frameworks im ganzen Studienverlauf
 - entsprechende Kenntnisse werden z.B. in praktischen Hauptseminaren vorausgesetzt
- Aber keine Reduktion auf *prompt engineering* für ChatGPT
 - formale und theoretische Grundlagen der Computerlinguistik
 - interdisziplinäre Anwendungen v.a. im Rahmen der Hauptseminare (politische Diskurse, Digital Humanities, Konstruktionsgrammatik, ...)

Grundstudium

- **Grundlagen der Computerlinguistik I – III**
 - GdCL I = klassische symbolische Ansätze, Mengenlehre, Logik
 - GdCL II = statistische Ansätze, maschinelle Lernverfahren, Wahrscheinlichkeitstheorie, lineare Algebra
 - GdCL III = Deep Learning = neuronale Netze, Tensoranalysis, praktische Implementierung
- **Programmierung & Infrastrukturen I + II**
 - Unix-Betriebssystem, Kommandozeile, Editor, reguläre Ausdrücke
 - Einführung in Python
 - Nutzung von Python-Bibliotheken und NLP-Standardwerkzeugen
- **Linguistische Grundkompetenzen**
 - Grundlagen der germ. Linguistik 1 oder Basismodul II: Linguistics (A)
 - bei Germanistik, Anglistik, Romanistik: DH-Modul 1 „Sprache & Text“
- **Proseminar** Computerlinguistik
 - Rezeption von Originalarbeiten, Präsentation, konstruktive Diskussion
 - Schreiben von Hausarbeiten, Literaturrecherche & Zitierung, LaTeX

Hauptstudium

- **3 Hauptseminare** zu wechselnden Themen
 - 1x mit klassischer Hausarbeit
 - 1x mit mündlicher Prüfung
 - 1x mit praktischem Projekt + Projektbericht
- **1 Praxisseminar**
 - angewandtes Gruppenprojekt (z.B. Teilnahme an Shared Task)
 - gemeinsamer Projektbericht in Form eines Konferenzbeitrags
 - Schwerpunkt auf Teamwork, Selbstorganisation, Arbeitsteilung, ...
- Angebot: mind. **2 Seminare / Semester**
- **Oberseminar** (verpflichtend)
 - Besuch der Vorträge über 2 Semester (→ Anmeldung in StudOn)
 - Essay über zwei ausgewählte Vorträge
- **Praktikum** (150h = 1 Monat Vollzeit)
 - bei externer Firma oder internes Forschungspraktikum

Erstfach ≠ Zweitfach

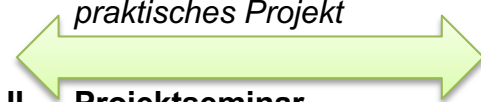
- **Zweitfach**: anwendungsorientierte Ergänzung zu geistes- oder sozialwissenschaftlichem Fach
 - komplett ohne Informatik-Importe
 - Vermittlung der mathematischen und informatischen Grundlagen ganz auf CL zugeschnitten
- **Erstfach**: Interesse an Informatik und NLP-Methoden
 - + 20 ECTS statt Schlüsselqualifikationen
 - 7,5 ECTS **Grundlagen der Informatik** (verpflichtend)
 - 7,5 ECTS + 5,0 ECTS **Wahlpflichtbereich**
 - zahlreiche Module aus dem Bachelor Informatik stehen zur Wahl
 - *Mathematik für Naturwissenschaftler (7,5) und Modellbildung & Statistik (5)*

Studienplan

Computerlinguistik (Zweifach)



Semester 1 (WiSe)	Semester 2 (SoSe)	Semester 3 (WiSe)	Semester 4 (SoSe)	Semester 5 (WiSe)	Semester 6 (SoSe)
VL Grundlagen der CL 1 2 SWS; 2 ECTS	VL Grundlagen der CL 2 2 SWS; 2 ECTS	VL Grundlagen der CL 3 2 SWS; 3 ECTS	Hauptseminar I 2 SWS; 5 ECTS; <i>Hausarbeit</i>	Hauptseminar III 2 SWS; 5 ECTS; <i>praktisches Projekt</i>	
Ü Grundlagen der CL 1 2 SWS; 3 ECTS	Ü Grundlagen der CL 2 2 SWS; 3 ECTS	Ü Grundlagen der CL 3 2 SWS; 7 ECTS	Hauptseminar II 2 SWS; 5 ECTS; <i>mündl. Prüfung</i>	Projektseminar 2 SWS; 5 ECTS; <i>Teamprojekt</i>	
S Grundkurs Programmierung 2 SWS; 5 ECTS	S Aufbaukurs Programmierung 2 SWS; 5 ECTS	Proseminar Computerling. 2 SWS; 5 ECTS	Oberseminar Computerlinguistik 2×1 SWS; 5 ECTS		
Einführung Linguistik 4 SWS; 5 ECTS; <i>Import Angl./Germ.</i>	DH-Modul 1: Sprache & Text 4 SWS; 5 ECTS; <i>Import DGuS</i>			Praktikum (extern / intern) 150 h = 1 Monat Vollzeit; 5 SWS; unbenotet	
10 SWS 15 ECTS	6 SWS 10 ECTS	6 SWS 15 ECTS	5 SWS 12,5 ECTS	5 SWS 12,5 ECTS	0 SWS 5 ECTS



Studienplan Computerlinguistik (Erstfach)



Semester 1 (WiSe)	Semester 2 (SoSe)	Semester 3 (WiSe)	Semester 4 (SoSe)	Semester 5 (WiSe)	Semester 6 (SoSe)
VL Grundlagen der CL 1 2 SWS; 2 ECTS	VL Grundlagen der CL 2 2 SWS; 2 ECTS	VL Grundlagen der CL 3 2 SWS; 3 ECTS	Hauptseminar I 2 SWS; 5 ECTS; <i>Hausarbeit</i>	Hauptseminar III 2 SWS; 5 ECTS; <i>praktisches Projekt</i>	
Ü Grundlagen der CL 1 2 SWS; 3 ECTS	Ü Grundlagen der CL 2 2 SWS; 3 ECTS	Ü Grundlagen der CL 3 2 SWS; 7 ECTS	Hauptseminar II 2 SWS; 5 ECTS; <i>mündl. Prüfung</i>	Projektseminar 2 SWS; 5 ECTS; <i>Teamprojekt</i>	
S Grundkurs Programmierung 2 SWS; 5 ECTS	S Aufbaukurs Programmierung 2 SWS; 5 ECTS	Proseminar Computerling. 2 SWS; 5 ECTS	Oberseminar Computerlinguistik 2×1 SWS; 5 ECTS		
Einführung Linguistik 4 SWS; 5 ECTS; <i>Import Angl./Germ.</i>	DH-Modul 1: Sprache & Text 4 SWS; 5 ECTS; <i>Import DGuS</i>			Praktikum (extern / intern) 150 h = 1 Monat Vollzeit; 5 SWS; unbenotet	
	Grundlagen der Informatik (Gdl) 6 SWS; 7,5 ECTS; <i>Import Informatik</i>	Wahlpflichtbereich Informatik <i>ca. 6 SWS; 12,5 ECTS; Import Informatik (+ andere)</i>			
10 SWS 15 ECTS	12 SWS 17,5 ECTS	8 SWS 20 ECTS	9 SWS 20 ECTS	5 SWS 12,5 ECTS	0 SWS 5 ECTS

Nach dem Studium

- Masterstudium / Promotion
 - Computerlinguistik
 - Sprachwissenschaft / Linguistik
 - Digital Humanities
 - Angewandte Informatik
- Tätigkeitsbereiche in der Wirtschaft (→ Sprachtechnologie)
 - Google, Microsoft, Facebook, Amazon, 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 und der KI suchen Computerlinguist:innen!

Über uns

Unser Kernteam

- Prof. Dr. Stephanie Evert
stephanie.evert@fau.de
- Dr. Besim Kabashi
besim.kabashi@fau.de
- Philipp Heinrich, M.Sc.
philipp.heinrich@fau.de
- Nathan Dykes, M.A.
nathan.dykes@fau.de
- Anke Lutz
ccl-sekretariat@fau.de

www.linguistik.phil.fau.de/team/research-assistants/

Bismarckstr. 6
Raum 4.000
(Lehrstuhlbüro)

Philologien

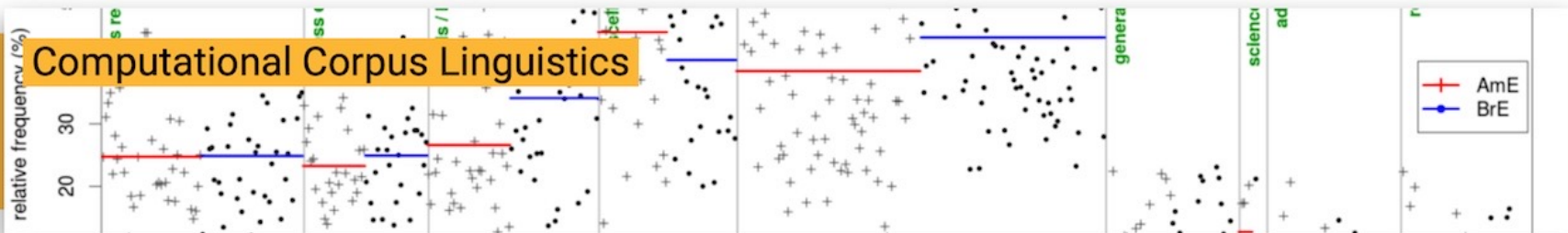


(CIP-Pool)

Bismarckstr. 12

Raum 0.320

Ganz wichtig: <https://www.linguistik.fau.de/>



Welcome!

The computational corpus linguistics group carries out foundational methodological research on the quantitative analysis of large text corpora. Our research lies at the intersection of corpus linguistics, computational linguistics, and the digital humanities.

We offer a Bachelor's degree program in computational linguistics ("Linguistische Informatik"). Teaching usually takes place in German. We currently offer courses in the teaching portal.

Computational Corpus Linguistics

Lehrstuhl für Korpus- und Computerlinguistik
Department Germanistik und Komparatistik

Bismarckstr. 6
91054 Erlangen

+49913185-22426

info@linguistik.uni-erlangen.de

http://linguistik.fau.de

- Sehr kleiner Studiengang (~ 30 Studierende) ... noch

- Aktive Fachschaftsinitiative:

<https://www.linguistik.phil.fau.de/teaching/fsi-computerlinguistik/>

- FSI-Stammtisch

- Grundlagen- und Orientierungsstudium:

<https://www.ziwis.fau.de/lehreundangebote/grundlagen-und-orientierungsstudium/>

OCT 17 4:15 PM - 6:00 PM
Einführungsveranstaltung BA
Computerlinguistik

Stundenplan

siehe <https://www.linguistik.fau.de/teaching/lehrveranstaltungen/>

	Montag	Dienstag	Mittwoch	Donnerstag	Freitag
08–10				09:30–10:15 Fragestunde zu GdCL 1 (Bismarckstr. 12, 0.320)	
10–12			Sprechstunde Prof. Evert (Anmeldung über StudOn)	Übung GdCL 1: Traditionelle Verfahren (Bismarckstr. 12, 0.320)	Vorlesung GdCL 3: Deep Learning (Bismarckstr. 12, 0.320)
12–14					Übung GdCL 3: Deep Learning (Bismarckstr. 12, 0.320)
14–16		Proseminar Computerlinguistik (Bismarckstr. 12, 0.320)	HS Mehrwortausdrücke und Kollokationen aus computerlinguistischer Sicht (Bismarckstr. 12, 0.320)		HS Die algorithmische Öffentlichkeit: Gemeinwille, Populismus und ChatGPT (Bismarckstr. 12, 0320)
16–18	Programmierung & Infrastrukturen I (Bismarckstr. 12, 0.320)		Oberseminar Computerlinguistik (Bismarckstr. 12, 0.320)		

Lehrangebot im WS 2023/24

- VL + Ü Grundlagen der Computerlinguistik 1
 - PS Programmierung & Infrastrukturen 1
 - Linguistische Grundkompetenzen (teilweise)
 - Modul Grundlagen der Informatik (Erstfach, optional)
1. Sem.
- VL + Ü Grundlagen der Computerlinguistik 3
 - PS Computerlinguistik
 - PS Aufbauseminar Python (für Übergang)
3. Sem.
- HS: Die algorithmische Öffentlichkeit (mit Japanologie)
 - HS: Mehrwortausdrücke & Kollokationen
 - weitere HS als Importe möglich (→ selbst anfragen)
5. Sem.
- Oberseminar Computerlinguistik
 - Wahlpflichtbereich Informatik
 - Schlüsselqualifikationen
- allgemein

Informationen für Erstsemester

Grundlagen der Computerlinguistik I

- **Vorlesung:** Mo 14:00 – Do 9:00 (Zeitfenster)
 - Bereitstellung als Screencasts von 45–60 Minuten Dauer
 - Dozent: Prof. Dr. Stefan Evert 😊
- **Fragestunde:** Do 9:30–10:15 (CIP-Pool)
 - Fragen und Diskussion zur Vorlesung
- **Übung:** Do 10:15–11:45 (CIP-Pool)
 - Veranschaulichung mit Beispielen und interaktiven Übungen
 - wöchentliche Übungsaufgaben zur schriftlichen Bearbeitung
 - Dozentin: Prof. Dr. Stephanie Evert
- **Tutorium:** n.V. (Tutorin: Anta Dzene)

Informationen für Erstsemester

Programmierung & Infrastrukturen I

- **Grundkurs Programmierung:** Mo 16:15–17:45 (CIP-Pool)
 - erste Sitzung am Montag 23.10. (→ weitere Organisation)
 - Dozent: Dr. Besim Kabashi

- **Tutorium:** n.V. (Tutorin: Amalia Pannu)

Informationen für Erstsemester

Importmodule

- **Grundlagen der germanistischen Linguistik 1 oder Basismodul II: Linguistics (A)**
 - falls zweites Fach nicht Germanistik, Anglistik oder Romanistik
 - jeweils mehrere Gruppen in verschiedenen Zeitfenstern
- **Grundlagen der Informatik (nur Erstfach)**
 - Vorlesung: Do 8:15–11:45 / „asynchron“ als Videoaufzeichnung
 - Fragestunde zu Gdl: Di 10:15–11:45 (online)
 - Sprechstunden zu Gdl: verschiedene Termine (Chat / virtuell)
 - alternativ: im 2. Semester belegen

Nächste Schritte

- Alle LVen werden durch **StudOn-Kurse** begleitet
 - separater Kurs für jede Lehrveranstaltung in Campo
 - dort aktuelle Informationen, Materialien, Zugang Videos + Zoom
 - gilt unseres Wissens als offizielle Anmeldung
- Lehrveranstaltungen für 1. Semester in dieser Woche
 - Vorlesung GdCL1: ersten Screencast bis Donnerstag schauen
 - Übung GdCL 1: Organisatorisches & Semesterüberblick
 - Linguistische Grundkompetenzen / GdI nicht vergessen!
- Prüfungsanmeldung über **Campo** nicht vergessen!
 - Anmeldezeitraum: **20.11.–10.12.2023**
- **Accounts** für CIP-Pool = IdM-Kennung
- **Sprechstunde** Prof. Dr. Stephanie Evert: **hybrid** (→ StudOn)
 - Mi 10:00–11:30 (Voranmeldung) + 11:30–12:00 (offen)