

Digitales Forschungsparadigma an der SLF



Ziel: erster (!) Austausch, Information, Vernetzung und Sichtbarkeit zu den Themen:

30.10.23	
10:00-11:00	Digitales Forschungsparadigma
11:00-11:45	Gute wissenschaftliche Praxis
31.10.23	
10:00-10:45	Dienste von UB und CMS
11:00-11:45	IZ D2MCM
01.11.23	
13:00-13:45	KI und FDM für Lehre und Abschlussarbeiten
14:00-14:45	Offene Diskussion/Fragen

Herzlichen Dank an alle Beitragenden!

Digitales Forschungsparadigma an der SLF



Forschung und Forschungswerkzeuge

Lorenzo Filipponio

Hagen Hirschmann

Tomáš Kos

Thomas Krause

Mariia Razguliaeva

Lukas Rösli

Lars Erik Zeige

Labore/Verlag

Georg Lohfink

Katja Maquate

Tine Mooshammer

Daniela Pallesci

Stefan Müller

Gremien

Mingya Liu

Carolin Odebrecht

Nächster Beitrag :-)

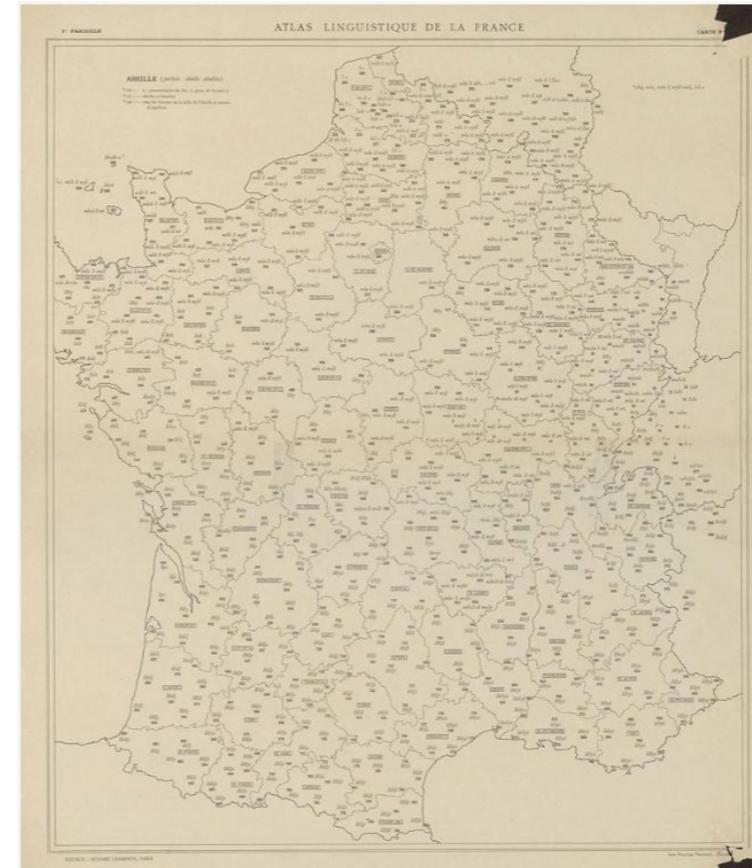
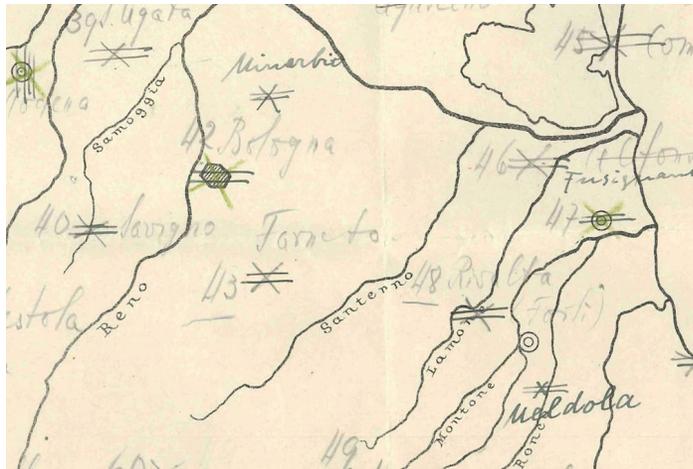


Institut für Romanistik

Digitale Methoden in der romanistischen Forschung



Dialectology: Atlases



Pardun. Formabund Nachmittag.
Bahnhofshôtel 12. Jan. 1918.

Mein lieber Herr Professor!

Seit vorgestern Abend bin ich also wieder an altbekannte
Stätte, so ich vor 2 Jahren als Dolmetscher den Abschied
nahm zu besserer Tätigkeit. Gestern wurden wir vom
Flugplatz Görries und von einem Arbeitskommando
4 wichtige Leute (die sich bei meinem letzten Aufhalt

Dialectology

Banque de Données Langue Corse

LESSICO ITALIEN/CORSE LESSICO CORSE/ITALIEN LOCALITÀ TESTI RICHERCHER TAL

Accoglienza Presentazione Bibliografia Langue

Lexique Corse / Italien

Mon recherché

contexte du mot

- il cane
- un canione
- il cane (del fucile)
- il gattinu

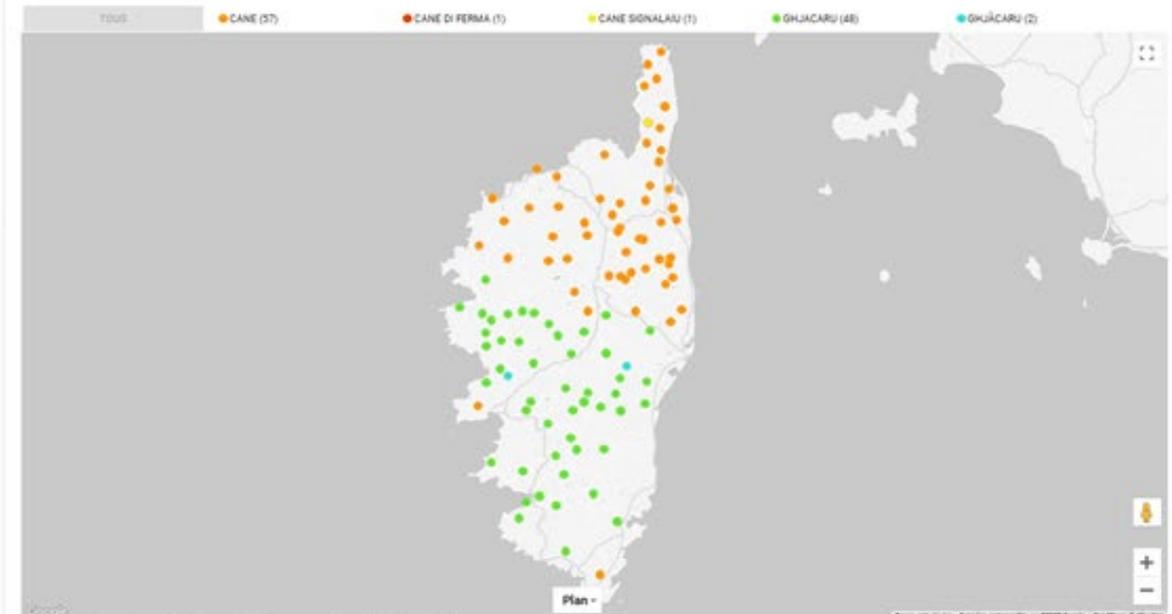
Cartographie

- Carte des lemmes
- Carte phonétique

Traductions et localisations

le chien: 5 lemmes

TOUS CANE (37) CANE DI FERMA (1) CANE SIGNALAU (1) GHJACARU (48) GHJACARU (2)



Plan

1 Cliquez sur un lemme pour plus d'informations. Cliquez sur une classe de couleur pour filtrer.

Dialectology



COSER
Corpus Oral y Sonoro del Español Rural

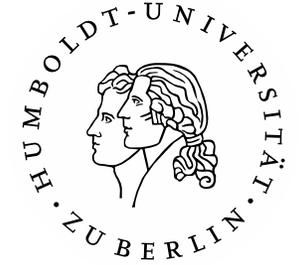
Condiciones de uso Contacto  

El corpus Recogida de datos Transcripción Consulta del corpus Recursos Proyecto COSER

Descripción

El COSER es un corpus dialectal, pero restringido al habla de aquellos informantes que fueron objeto de interés en la dialectología tradicional: hablantes rurales, a ser posible mayores, de escasa escolarización y naturales del lugar en que son entrevistados. El COSER se nutre del mismo tipo de informantes que los atlas lingüísticos. Actualmente (diciembre de 2022), 2.961 informantes están registrados en nuestra base de datos, si bien solo de algo más de la mitad han sido entrevistados en profundidad:

Informantes	Número	Media de edad
Hombres:	1.415 (47,8%)	75 años
Mujeres:	1.546 (52,2%)	73,6 años
Total:	2.961	74,2 años



Dialectology

Inicio [¿Qué es CLICC?](#) [Corpus](#) [Recursos](#) [Guía de uso](#) [Contacto](#) [Iniciar sesión](#)

Inicio > Presentación > Objetivos > **Atlas Lingüístico-Etnográfico de Colombia**

Atlas Lingüístico-Etnográfico de Colombia

ALEC

El corpus ALEC está compuesto por 750 grabaciones recopiladas entre 1954 y 1978 en el marco de la investigación realizada para la construcción del Atlas Lingüístico-Etnográfico de Colombia (ALEC) (1954 - 1983). La recopilación de las muestras fue realizada por 23 encuestadores en 263 localidades distribuidas a lo largo del territorio nacional. La colección contiene encuestas fonéticas, re

[Ver más](#)

[Consultar el corpus](#) [Metodología](#) [Equipo](#) [¿Cómo citar?](#)

[Búsqueda por palabra](#) [Búsqueda por metadatos](#) [Búsqueda por colocaciones](#)

Dialectology



VIVALDI

[Drucken](#)

Willkommen bei VIVALDI

Was ist "Vivaldi"?

Wie funktioniert's?

Vivaldi

- Abruzzen
- Aostatal
- Apulien
- Basilikata
- Emilia-Romagna
- Friaul-Julisch Venetien
- Kalabrien
- Kampanien
- Latium
- Ligurien
- Lombardei
- Molise
- Piemont
- Sardinien
- Sizilien
- Toskana
- Trentino-Südtirol
- Umbrien
- Venetien

Vivaldi Maps

Vivaio Acustico delle Lingue e dei Dialetti d'Italia

Willkommen bei VIVALDI

[Klicken Sie auf eine Region, um auf die Seite der Sprachatlasdarstellungen zu gelangen. Für die Anzeige der Karten sind keine zusätzlichen Plugins notwendig.]

VIVAio
Acustico
delle **L**ingue e
dei **D**ialetti
d'**I**talia

Akustischer Sprachatlas
der Dialekte und
Minderheitensprachen
Italiens





Database

- Corpus Diacrónico del Español (CORDE)
- Corpus del Diccionario histórico de la lengua española (DHLE)
- Tesoro della Lingua Italiana delle Origini (TLIO)
- Frantext (Analyse et Traitement Informatique de la Langue Française)

- Corpus de Referencia del Español Actual (CREA)
- Corpus d'Etude pour le Français Contemporain (CEFC)
- Corpus dell'italiano scritto (CORIS)
- es-ten-ten, it-ten-ten, fr-ten-ten; itWaC, frWaC...

- Corpus Diacrónico y Diatópico del Español de América (CORDIAM)
- Morfologia dell'Italiano in Diacronia (MIDia)



Tools

Morphosyntactic analysis

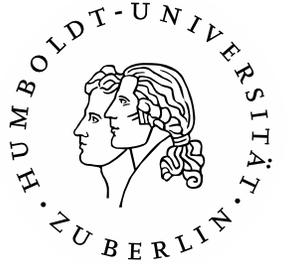
- spaCy (annotation of raw texts)
- Stanza NLP (from raw text to syntactic analysis and entity recognition)
- UDPipe (trainable pipeline for tokenization, tagging, lemmatization and dependency parsing of CoNLL-U files (Computational Natural Language Learning))

Model development and training

- scikit-learn (for predictive data analysis)
- TensorFlow (creation of machine learning models)
- PyTorch (interface for the implementation of machine learning models)
- Hugging Face (platform for the integration of previous models like BERT or GPT)
- LIME (Local Interpretable Model-Agnostic Explanations)

Topic modelling

- Latent Dirichlet Allocation (LDA)
- Mallet-Modell-Wrapper



Treebanks

Corpus

Descàrregues

- ▼ Corpus
 - ▼ AnCora
 - Participants
 - Cerques
 - Documentació
 - Publicacions
 - Competicions
 - IARG-AnCora
 - ▶ CesCa
 - GRERLI
 - CIInt
 - ▶ NIdent
 - ▶ Projecte Araknion
 - ▶ Paràfrasi
 - Hopinion
 - SFU Review-NEG
 - HistoCat
 - DialCat

AnCora

AnCora és un corpus del **català (AnCora-CA)** i de l'**espanyol (AnCora-ES)** amb diferents nivells d'anotació:

- lema i categoria morfològica
- constituents i funcions sintàctiques
- estructura argumental i papers temàtics
- classe semàntica verbal
- tipus denotatiu dels noms deverbals
- sentits de WordNet nominals
- entitats nombrades
- relacions de correferència

El corpus de cada llengua conté **500.000 paraules** i estan constituïts majoritàriament per textos periodístics. Per a més informació vegeu el document [AnCora-corpus](#).

Com a resultat del procés d'anotació es disposa també de **dos** lèxics verbals, **AnCora-Verb**, de 2.647 entrades per a l'espanyol i 2.143 entrades per al català, i d'un lèxic de nominalitzacions deverbals, **AnCora-Nom**, de 1.600 entrades per a l'espanyol. La informació que contenen aquests lèxics és la següent:

Ancora-Verb	Ancora-Nom
Classe semàntica	Tipus denotatiu
Subcategorització	Synset de WordNet
Estructura Argumental i Papers temàtics	Estructura Argumental i Papers temàtics
	Verb del qual deriva

Han participat en l'anotació dels corpus AnCora:

Datamapping

ñ Dialectos del español

Dime cómo hablas y te diré de dónde eres

¿Crees que una app podría adivinar de dónde eres?

En menos de 5 minutos puedes comprobarlo.

Para participar solo hay que contestar **26 preguntas**.

¡Es sencillo y te divertirás un rato!

Si quieres colaborar con nosotros, te pedimos que **reflejes tu modo real de hablar**. Es posible que para algunas preguntas tengas más de una respuesta; en ese caso, selecciona la forma que más usas.

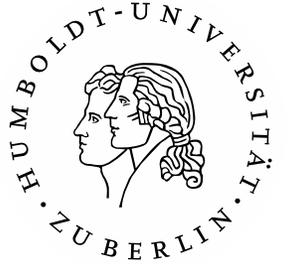


Entrar

Cuando termines, todavía puedes seguir colaborando con nosotros. ¿Cómo?
¡Comparte la app con tus amigos! Cuantas más respuestas obtengamos, mejor funcionará.



Esta app ha sido diseñada para detectar y predecir algunos rasgos generales y característicos de todos los dialectos del mundo hispanohablante. El objetivo del proyecto *Dialectos del español* es obtener información sobre la diversa manera de hablar nuestra lengua.





Development of Games With A Purpose (GWAPs)

Juegos del español Inicio Progreso Juegos Ayuda Proyecto Equipo Contacto Términos de uso Registro Iniciar sesión

¡Juega por los dialectos del español!

Gana carreras, guía a Gummy en su búsqueda de tesoros o salva el mundo de un ataque extraterrestre y nos ayudarás a crear una base de datos que se utilizará para el estudio de las variedades dialectales del español.

No olvides invitar a tus amigos, da click en los siguientes iconos para compartirlos en redes sociales.



Para jugar es necesario que te [registres](#) o si ya tienes una cuenta que [inicies sesión](#).

GHENT UNIVERSITY UHASSELT  fwo

© Games with a purpose 2022

Nächster Beitrag :-)



Korpora und Lernerkorpora am IdSL

Eine Ressourcenübersicht

Hagen Hirschmann

Korpuslinguistik und Morphologie (Professur Anke Lüdeling)

Einführung

- ▶ Korpora am IdSL
 - ▶ Am Institut erstellte Korpora
 - ▶ In Kooperation erstellte Korpora
 - ▶ Gehostete Korpora, vor allem sichtbar uns zugänglich durch das Suchsystem ANNIS (<https://corpus-tools.org/annis/>; allgemeines Interface mit vielen aktuellen Korpora in der Kollektion: <https://korpling.german.hu-berlin.de/annis/>)
- ▶ Vorherrschende Texttypen
 - ▶ Historische Korpora, vor allem als Begleitprodukt von vergangener Forschung der Professur für Sprachgeschichte - <https://www.linguistik.hu-berlin.de/de/institut/professuren/sprachgeschichte/forschung/abgeschlossene-projekte/abgeschlossene-projekte>
 - ▶ Gesprochene Korpora zur phonetisch-phonologischen Forschung, koordiniert an der Professur Phonetik/Phonologie (<https://www.linguistik.hu-berlin.de/de/institut/professuren/phonetik/forschung>)
 - ▶ Lernerkorpora zur Behandlung von Spracherwerbsfragen, koordiniert an der Professur für Korpuslinguistik und Morphologie (<https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/korpora>)

Lernerkorpora am IdSL

- ▶ Lernerkorpora → digital verfügbare, analysierbare Ressourcen mit Sprache von Lernenden einer Sprache (L1-Korpora aus verschiedenen Entwicklungsstufen, L2-Korpora mit Fremdsprachenlernenden, Zweitsprachenlernende mit verschiedenem Background, ...)
- ▶ Seit 2004: Erhebungen und Datenarbeiten zu FALKO („**F**ehlerannotiertes **L**erner**k**orpus“) – handschriftliche DaF-Lernertexte (Textzusammenfassungen, dann Essays); Kooperation zur Georgetown University, FU Berlin, Bangor University
 - ▶ <https://hu-berlin.de/falko>
 - ▶ Durch weitere Kooperationen (Prof. Vyatkina, University of Kansas; Prof. Zinsmeister, Universität Hamburg u.a.) weitere DaF-Korpora nach Falko-Richtlinien (KanDel, Kobalt-DaF, Dulko – siehe <https://hu-berlin.de/falko-familie>)
 - ▶ Bis heute Erweiterungen durch Ausbau von Subkorpora und Aufnahme neuer Daten; Übersicht folgt

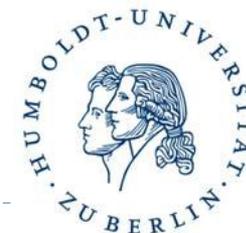
Lernerkorpora am IdSL

- ▶ Seit 2012: Aufbau des gesprochenen BeMaTaC-Korpus mit einer Deutsch-L2-Kohorte - <https://hu-berlin.de/bematac>
- ▶ Seit 2018: Aufbau des RUEG-Korpus (mehrsprachige Sprecher*innen von Heritage- und Majoritätssprachen) - <https://hu-berlin.de/rueg>
- ▶ Seit 2023: L1-Erwerbsdaten aus Grundschulbereich (Prof Binarzer, Hannover; Prof. Langlotz, Kassel) – <https://hu-berlin.de/naleko>

Veröffentlichungskonzept

- ▶ Open access oder zugänglich für Forschende
 - ▶ Zugang zu anonymisierten und ggf. maskierten Primärdaten
 - ▶ Zugang zu Annotationen
 - ▶ Zugang zu einer Forschungsdaten-Infrastruktur, in der niedrighschwelliger Zugang zu eigenen Analysen möglich ist
 - ▶ Dokumentationen zu
 - ▶ Datenerhebungen
 - ▶ Datenaufbereitung
 - ▶ Datennutzung
 - ▶ Bereitstellung von Korpora mit verfügbaren Annotationen im ANNIS-Suchsystem (<https://corpus-tools.org/annis/>)

In ANNIS verfügbare Lernerkorpora am IdSL



Name	Texttyp	Lernertyp	Modalität	Anzahl Token	Anzahl Texte	Link zu Korpus in ANNIS	Link zu Info-Seite
Falko Essay L2 (v2.4)	Essay	DaF	schriftlich	144.619	248		https://hu-berlin.de/falko
Falko Essay LI (v2.3)	Essay	LI	schriftlich	70.615	95		https://hu-berlin.de/falko
Falko Summary L2 (v1.2)	Zusammenfassung	DaF	schriftlich	40.638	106		https://hu-berlin.de/falko
WHiG L2 (v2.1)	Essay	DaF	schriftlich	130.949	196		https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung-whig/WHIG
Kobalt L2 (v1.4)	Essay	DaF	schriftlich	33.368	51		https://hu-berlin.de/kobalt
Kobalt LI (v1.4)	Essay	LI	schriftlich	12.984	20		https://hu-berlin.de/kobalt
KanDel (long.v2)	Bericht	DaF	schriftlich	34.612	185		https://hu-berlin.de/kandel
DulkoEssay (v1.0)	Essay, Übersetzung	DaF	schriftlich	12.319	34		https://www.ids-mannheim.de/gra/projekte/deutung/dulko/
NaLeKo (v1.0)	Erzählungen	LI, L2, 2LI	schriftlich	59.219	288		https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung/naleko
BeMaTaC L2 (v.3.0)	Map-Task-Dialoge	L2	mündlich	12.522	5		https://hu-berlin.de/bematac
RUEG-DE (v0.4)	Unfallbericht	LI, 2LI	mündlich, schriftlich	175.553	1035		https://www.linguistik.hu-berlin.de/en/institut-en/professuren-en/rueg/rueg-corpus

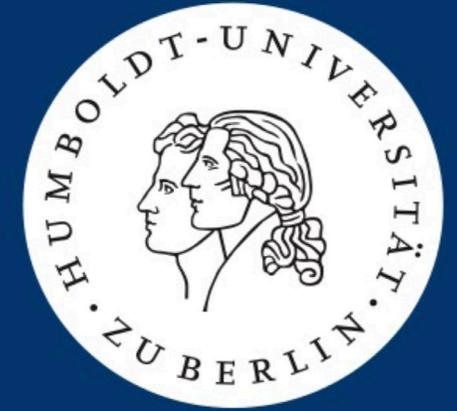
Herzlichen Dank

- Webseite des Fachbereichs Korpuslinguistik und Morphologie:
<https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik>
- Kontakt Hirschmann: hagen.hirschmann@hu-berlin.de

Relevante Referenzen (Lernerkorpora)

- ▶ Granger, Sylviane (2008): Learner corpora. In: Lüdeling, Anke; Kytö, Merja (Hg.): *Corpus Linguistics. An International Handbook*. Vol 1. Berlin; de Gruyter, S. 259-275.
- ▶ Hirschmann, Hagen; Lüdeling, Anke; Shadrova, Anna; Bobeck, Dominique; Klotz, Martin; Akbari, Roodabeh; Schneider, Sarah; Wan, Shujun (2022) FALKO. Eine Familie vielseitig annotierter Lernerkorpora des Deutschen als Fremdsprache. In: *KorDaF (Korpora Deutsch als Fremdsprache)* (2:2), 139-148.
- ▶ Hirschmann, Hagen; Lüdeling, Anke; Rehbein, Ines; Reznicek, Marc; Zeldes, Amir (2013) Underuse of Syntactic Categories in Falko. A Case Study on Modification. In: *20 years of learner corpus research. Looking back, Moving ahead (LCR2011)*. Louvain-la-Neuve, Belgium; Presses universitaires de Louvain.
- ▶ Lüdeling, Anke; Doolittle, Seanna; Hirschmann, Hagen; Schmidt, Karin & Walter, Maik (2008): Das Lernerkorpus Falko. In: *Deutsch als Fremdsprache* 2(2008), S. 67-73.
- ▶ [Falko-Handbuch:]
Reznicek, Marc; Lüdeling, Anke; Krummes, Cedric; Schwantuschke, Franziska; Walter, Maik; Schmidt, Karin; Hirschmann, Hagen; Andreas, Torsten (2012): *Das Falko-Handbuch. Korpusaufbau und Annotationen Version 2.01*

Nächster Beitrag :-)



Junior Prof. Dr. Tomáš Kos

Themenwoche Digitales Forschungsparadigma an der SLF

Institut für Anglistik und Amerikanistik

English Linguistics

Forschung:

- experimentelle Daten, das Z-Labor sowie Crowdsourcing-Webseiten wie Prolific für Online-Studien
- Korpora für Englisch (die unter www.english-corpora.org)
- Sketchengine (www.sketchengine.eu)

Lehre:

- gelegentlich linguistische Korpora; Kahoot
- Studierenden: die englischen Korpora für ihre BA- und MA-Arbeiten

English Language Education

- Audio und Videodaten
 - Interaktion im Unterricht
 - Audio- & Videographie
 - Interviewdaten
- Textdaten
 - Schülerarbeit
 - Tests
- Befragungsdaten: liegen in digitaler Form vor

Green-Screen-Technik

- die Verfügbarkeit von Aufnahmetechnik stellt ein häufiges Nadelöhr dar
- die „Studiozeiten“ der zentralen *Green-Screen* Einrichtungen knapp
- wenig Raum zum Ausprobieren und Experimentieren
- *Green-Screen Technik*
- FDE - Videolernwerkstatt in den Räumen der PSE am Hausvogteiplatz 5-7
- schnell auf- und abzubauen
- Eine kleine Bühne & Weißlichtscheinwerfer, GreenScreen , Tablet Teleprompter, Camcorder, Mikrofon
- steht auch für andere Fremdsprachendidaktiken zur Verfügung!
- im Kontext des Praxissemesters Veranstaltungen, in denen Studierende Erklärvideos kooperativ produzieren
- Die Entwicklung von Erklärvideos – Förderung der Kompetenzen der didaktischen Reduktion & des multimodalen Erklärens



Nächster Beitrag :-)



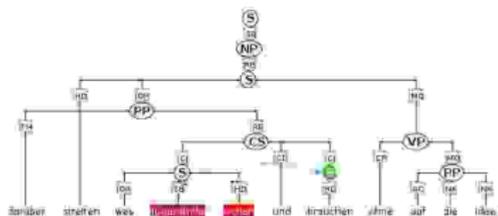
Infrastruktur Korpuslinguistik und Morphologie
Digitales Forschungsparadigma - Themenwoche an der SLF

Thomas Krause

Institut für deutsche Sprache und Linguistik

30. Oktober 2023

- Digitale Methoden sind aus aus den Forschungsprojekten an denen wir beteiligt sind oder waren nicht wegzudenken
- Notwendig zum *Aufbau* und *Auswertung* der Datensätze (meistens Korpora) aus den verschiedensten Bereichen der Linguistik



13 COBERT 81000 0100202011

Feigenblatt: Die Jugendlichen die Zinsen wollen ein Musikerteil. Das forderten sie bei der ersten Zusammen- Runde im Dienstgespräch. Dass die Politiker der Staat dafür Verständnis haben, ist üblich. Mit dem Treffen ist

- Entwicklung von Methoden und Modellierung sind Teil des Forschungsprozess

corpus-tools.org

With corpus-tools.org we provide an infrastructure to annotate, migrate, and analyze linguistic data.

ANNIS

ANNIS is an open source, cross-platform (Linux, Mac, Windows), browser-based search and visualization architecture for complex multi-layer linguistic corpora with diverse types of annotation. ANNIS, which stands for ANNotation of Information Structure, was originally designed to provide access to the data of the SFB 632 - "Information Structure: The Linguistic Means for Structuring Utterances, Sentences and Texts". It has since extended to be used by a large number of projects annotating a variety of phenomena. Since complex linguistic phenomena, such as information structure, interact on many levels, ANNIS addresses the need to concurrently annotate, query and visualize data from varied areas such as syntax, semantics, morphology, prosody, referentiality, lexis and more. For projects working with spoken language, support for audio / video annotations is also required.

Pepper

If you need to convert corpora from one linguistic format into another, **Pepper** is your swiss-army knife. When your annotation tool produces a different data format from the one your analysis tool can read, Pepper is there to the rescue.

- Pepper can convert documents in a variety of linguistic formats, such as: [EXMARaLDA](#), [Tiger XML](#), [MMAx2](#), [RST](#), [TCF](#), [TreeTagger format](#), [TEI](#) (subset), [ANNIS](#) format, [PAULA](#) and many many more.
- Pepper comes with a plug-in mechanism which makes it easy to extend it for further formats and data manipulations.

Hexatomic

Hexatomic is an extensible, OS-independent platform for deep multi-layer linguistic annotation of corpora. It is being developed for sustainability, in order to support research software re-use rather than new development of software with each new research project. Using Hexatomic, linguistic research projects can implement what they need on top of an existing platform with high compatibility to other tools and pipelines. Hexatomic is funded by [Deutsche Forschungsgemeinschaft \(DFG\)](#) under grant number 391160252. Development is based at the [Department of English Studies \(Friedrich Schiller University Jena\)](#) and the [Department for German Studies and Linguistics \(Humboldt-Universität zu Berlin\)](#).

Salt

With **Salt** we provide an easily understandable meta model for linguistic data as well as an open source API to store, manipulate and represent data. Salt is an abstract model, poor in linguistic semantics. As a result, it is independent of any linguistic schools or theories. The core model is graph-based, thereby keeping structural restrictions very low and allowing for a wide range of possible linguistic annotations like syntactic, morphological, coreferential annotations, and many more. You can even model your own very personal annotation as long as it fits into a graph structure (and so far we have not seen a linguistic annotation which does not). Furthermore, Salt does not depend on a specific linguistic tagset which allows you to use every tagset you like.

+

<https://github.com/korpling/>

Infrastruktur: Betrieb Plattformen und Verstatigung

Pharmabesitzer

Pharmabesitzer

Pharmabesitzer	Pharmabesitzer	Pharmabesitzer	Pharmabesitzer	Pharmabesitzer	Pharmabesitzer	Pharmabesitzer	Pharmabesitzer	Pharmabesitzer	Pharmabesitzer
Pharmabesitzer									

<https://korpling.german.hu-berlin.de/annis/>



LAUDATIO

Long-term Access and Usage of Deeply Annotated Information

Search metadata (German or English) with text strings. * and ? can be used as wildcards

Anselm Corpus (Version 1.0)
 Dishes, Inflection, Schreibung, Semantik
 435564 Tokens 2018

Take part and publish a corpus within the Laudatio repository
 Please contact: laudatio-support@hu-berlin.de

<https://www.laudatio-repository.org/>

sfb1412

sfb1412

Group ID: 2345

Subgroups and projects

Subgroup	Shared projects	Archived projects
AG3		
AG6		
AG8		
BE1		

<https://scm.cms.hu-berlin.de/>

Herausforderungen der Nachhaltigkeit

- Für bestimmte Bedarfe können bestehende Software und Plattformen nicht nur weiterentwickelt werden, sondern erfordern Neuentwicklungen
 - graphANNIS als Nachfolge von ANNIS und Salt
 - Annatto als Nachfolge von Pepper
 - ...
- Gerade Plattformen müssen aber oft weiter betrieben werden, wenn nicht alle Arten von Bedarfen durch die Neuentwicklungen gedeckt werden
- Ab wann kann die Wartung von Plattformen und Software eingestellt werden?



thomaskrause (maven-release-plugin) prepare for next development iteration ✓ 554787 on Jul 26 🗨 8,742 commits

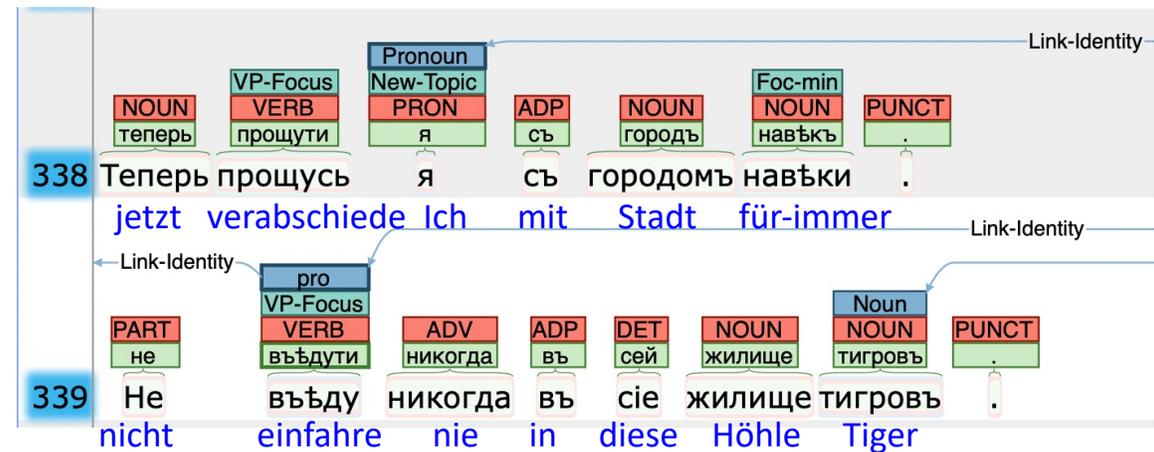
 .github	Use a more recent version of the graphANNIS CLI, because otherwi...	last year
 .vscode	use root package "org.corpus_tools.annis" instead of "annis"	2 years ago
 THIRD-PARTY	Files updated in release process	7 months ago
 docs	Document the configuration of the handling of virtual tokens	last year
 misc	Add parallel test classes	last year
 src	Files updated in release process	1 month ago
 .gitignore	Use midhook 0.4.2, also in deploy	2 years ago
 CHANGELOG.md	Files updated in release process	3 months ago
 CITATION.cff	Files translated in release process	"
 LICENSE	- adding LICENSE, NOTICE and 3rd party licenses.	14 years
 NOTICE	updated changelog	"
 README.md	update README.m...	last year

Nächster Beitrag :-)

Pronoun History: War ICH schon erwähnt und deshalb überflüssig?

University of Oxford und HU-Berlin, (DFG/AHRC, DFG: ME 4125/5-1)

- Wie hat haben sich Pronomina in Subjekt- und Objektposition in den slawischen, keltischen und germanischen Sprachen entwickelt?
- HU: Fokus auf Subjekte im Russischen, Polnischen und Bulgarischen
- Von welchen morphosyntaktischen, semantischen und pragmatischen Faktoren hängt ab, ob ein Subjekt phonetisch leer ist oder als Pronomen auftritt? Wie verändert sich dieses Verhältnis mit der Zeit?
- Erstellen von historischen Korpora, Balancierung nach Jahrhundert und Texttyp, tiefe Annotation (grammatische Eigenschaften, Dependenz, Koreferenz, Informationsstruktur)
- Die annotierten Texte werden öffentlich zugänglich gemacht und können für weitere Forschung wiederverwendet werden



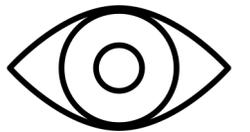
QueSlav: Wie sieht die Wahrnehmung einer Frage aus?

Karls-Universität in Prag, HU-Berlin, DFG/GAČR (DFG: ME 4125/5-1)

- Wie unterscheidet sich die Verarbeitung der Ja-Nein-Fragen im Russischen (und Tschechischen) im Vergleich zu den Aussagen? Welche Rolle spielen dabei Negation, Wortfolge, Partikeln und Intonation?
- Methode: Eye-Tracking-Experiment: Augenbewegungen werden mit einer speziellen Kamera verfolgt und gespeichert. Annahme: Augenbewegungen spiegeln kognitive Prozesse wieder
- Visual World Paradigm (Bilder + Audio):

 **Sprecher A:** Počistila li Katja utrom sapogi? ('Hat Katja am Morgen die Stiefel geputzt?')

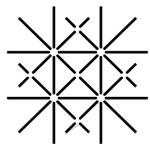
 **Sprecherin B:** Ja dumaju, čto da. (,Ich glaube, dass ja.')



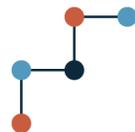
Nächster Beitrag :-)

**NACH
HALLENDE
NETZ
WERKE**

**WORKS
NET
NATUNG
RESSO**



Universität
Basel



Schweizerischer
Nationalfonds

Gefördert durch

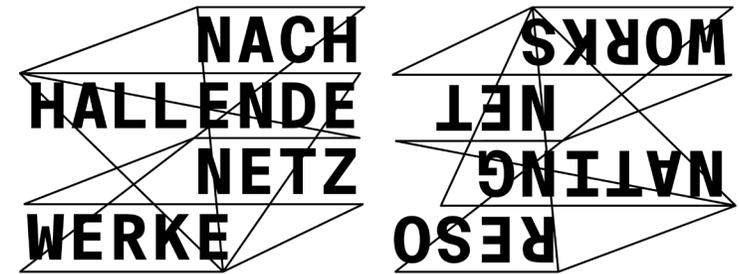


Deutsche
Forschungsgemeinschaft

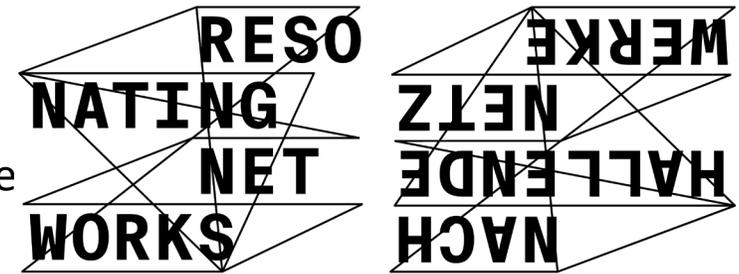
HUMBOLDT-
UNIVERSITÄT
ZU BERLIN



- **Projekttitlel:** *Nachhallende Netzwerke. Diskursive, räumliche und personelle Knotenpunkte altnordistischer Forschungsparadigmen (1650–1950)*
- **Standorte:** Universität Basel und HU Berlin
- **Förderung:** SNF (lead agency) und DFG
- **Projektlaufzeit:** 36 Monate
- **Projektstart:** 1. Dezember 2023
- **PI:** Prof. Dr. Lena Rohrbach (Uni Basel) und Prof. Dr. Lukas Rösli (HU Berlin)
- **Teilprojekte:** 4 PhD, 1 Postdoc, 2 PI
- **Extern:** 2 assoziierte PhD, 4 internationale Projektmitglieder aus Skandinavien

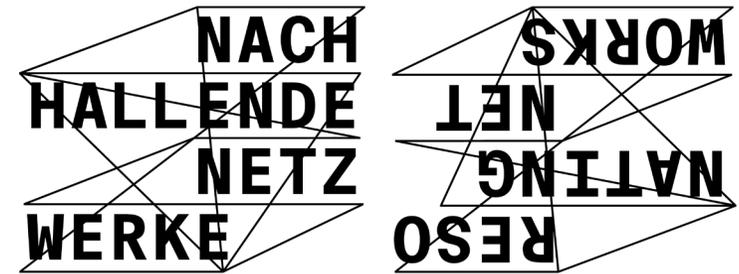


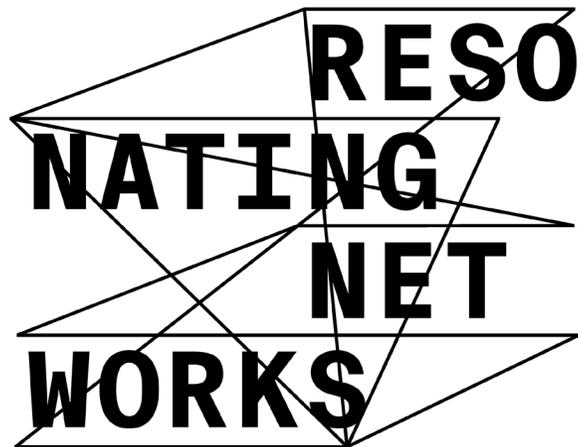
- **Ziel:** Diskursive Setzungen und Implikationen von Forschungsparadigmen zwischen 1650 und 1950 in der Skandinavistischen Mediävistik hinsichtlich der darüber erzeugten Konstruktion eines “altnordischen Mittelalters“ untersuchen
- **Methodologische Grundlagen:** Verschränkung von Diskursanalyse und netzwerktheoretischen Ansätzen der Akteur-Netzwerk-Theorie (ANT)
- **Teilprojekte:** (P1, PhD) Protophilologischer Beginn der Textkritik, (P2, Postdoc) Theoretisierung und nationalistische Diskursivierung des Mythosbegriffs, (P3, PhD) nationale Aneignung und Umdeutung mündliche Erzählungen Islands bei der Konstruktion von Volksmärchen, (P4, PhD) Konrad Maurers Konzept des Freistaats und (P5, PhD) Andreas Heuslers Begriff der germanischen Dichtung; Ergänzend die Metaparadigmen (P6, PI) Autorschaft und (P7, PI) Gattung



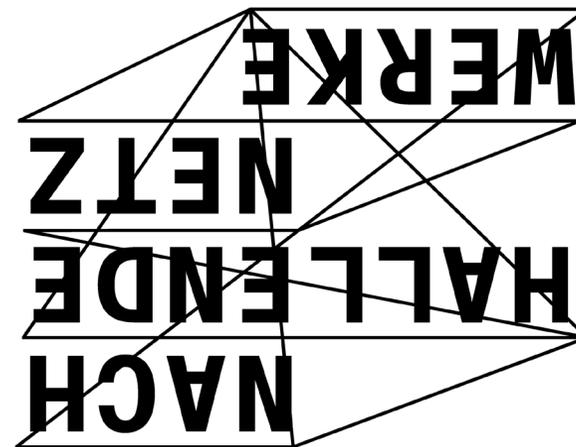
	1650-1700	1700-1750	1750-1800	1800-1850	1850-1900	1900-1950
Textual Criticism (P1)						
Norse Mythology (P2)						
Folk literature (P3)						
Freistaat (P4)						
Germanic Poetry (P5)						
Authorship (P6)						
Genres (P7)						

- **Qualitativer Ansatz:** Das Projekt wird einen substantiellen Beitrag zur Diskurs- und Netzwerkgeschichte der internationalen „Altnordistik“ und Skandinavistischen Mediävistik von der späten Vormoderne bis zur Mitte des 20. Jahrhunderts leisten und Einblicke in die synchrone und diachrone Resonanz zentraler Forschungsparadigmen geben.
- **DH und Public Outreach:** Das Projekt wird eng mit den Bibliotheks- und Archivinstitutionen in Skandinavien, Deutschland und der Schweiz zusammenarbeiten, die die Nachlässe der im Projekt behandelten Forschenden verwalten. Im Rahmen dieser Zusammenarbeit wird das Projekt die Bibliotheken/Archive und die zugehörigen Kataloge (Kalliope, handrit.is u.a.) mit neuen Metadaten versorgen, die im Zuge der Projektarbeit generiert werden. So wird es möglich sein, die wissenschaftlichen Erkenntnisse direkt in bibliothekarische Nachweissysteme (z.B. GND) zu überführen und damit der Allgemeinheit zugänglich zu machen.





**RESO
NATING
NET
WORKS**



**MERKE
NETZ
HALLENDE
NACH**

Mehr Informationen in Kürze unter:

www.resonatingnetworks.com

Besten Dank für Ihre Aufmerksamkeit!

Nächster Beitrag :-)

Das Referenzkorpus Altdeutsch (ReA)

- ein tiefenannotiertes Mehrebenenkorpus der gesamten altdeutschen Textüberlieferung und ihrer lateinischen Vorlagen

edition	lk	gihorta	ōat	seggen	ōat	sih	urhettun
ling	lk	gihorta	ōat	seggen	ōat	sih	urhettun
lemma	ih	gihören	dēr	sagen,sagēn	daʒ	sih	urheiʒo
posLemma	PPER	VV	DD	VV	KO	PRF	NA
pos	PPER	VVFIN	DDA	VVINF	KOUS	PRF	NA
inflectionClassLemma		WK1A		WK1B,WK3			N_MASC
inflectionClass		WK1A		WK1B,WK3			N_MASC
inflection	SG_NOM_1	IND_PAST_SG_1	NEUT_SG_ACC			PL_ACC_3	PL_NOM
lang	osx,goh	goh	osx,goh	osx,goh	osx,goh	goh	osx,goh
clause	CF_U_M				CF_I_O		
document	Hildebrandslied						
verse	1				2		
rhyme							S1
translation	ich	hören	der, die, das	sagen	daß	sich	Herausforderer

Abb. Hildebrandslied mit Annotationen im Referenzkorpus Altdeutsch

 *Deutsch Diachron Digital*
Referenzkorpora zur
deutschen Sprachgeschichte

ReA

ReM

ReF

ReN

ReDI

ddd

Projektverbund www.deutschdiachrondigital.de

Altdeutsch 750–1050, Mittelhochdeutsch 1050–1350, Frühneuhochdeutsch 1350–1650, Niederdeutsch 1200–1650, Deutsche Inschriften

Nächster Beitrag :-)

Zentral-Labor der Sprach- und literaturwissenschaftlichen Fakultät

August-Boekh-Haus, Dorotheenstraße 65, Raum 3.60 & 3.62

Zweck

- Forschung: - Verarbeitung und Gebrauch von Sprache
- Spracherwerb und Sprachentwicklung
- Lehrveranstaltungen: - Studierende sammeln praktische Erfahrungen im Labor
- Möglichkeit der Kooperation zwischen Mitgliedern innerhalb der Fakultät sowie universitätsübergreifende Kooperationen
- Ausleihe von Geräten für die Feldforschung

Laborräume

- Raum 3.60
 - drei separate schallisolierte Kuben, zur parallelen Nutzung
 - Möglichkeiten zur Durchführung von Reaktionszeitexperimenten, Interaktionsstudien und Eye-tracking u.a.
- Aufnahmestudio (3.62)



Eye-tracking

- Eyelink Portable Duo
- mobiles Setup
- Software:
Experiment Builder, Data Viewer



RT

- Jeweils zwei Experimentier-PCs in zwei Kuben
- Software: Presentation, E-Prime, PsychoPy



Raum für Interaktionsstudien

- Wohnlich eingerichtete Kube zur flexiblen Nutzung



Studio (Raum 3.62)

- Möglichkeit der Anfertigung von Audioaufnahmen
z.B. Einzel- & Dialogaufnahmen,
Stimuli, Lehrmaterialien
- Weitere zwei Plätze für RT



Dienstleistungen

- Hilfe beim Skripten von Experimenten, Datenvisualisierung und anderweitige Programmieraufgaben (Python, R)
- Audio- und Videobearbeitung (Praat, Audacity, Adobe Premier Pro)

Beispiel: EEG Trainingsvideo (BUA: sharing resources)

<https://youtu.be/btgwgOjKV E?feature=shared>

- Ausleihe von Audioequipment
- Beratung und Training: Aufnahmen im Studio oder im Feld, Nachbearbeitung, Manipulation von Audiostimuli, Automatisierung

Buchung des Labors

- Anfragen bitte an zentrallabor-admin-slf@hu-berlin.de
- Access Form: Kurzbeschreibung des Vorhabens, Zeitplan, benötigte Ressourcen, Ethik und Datenschutz
- Demnächst Buchung über OpenIris (BUA: sharing resources)

Nächster Beitrag :-)

Labor Psycholinguistik

Leitung: Prof. Pia Knoeferle

Lab Managerin: Katja Maquate

Techniker: Carsten Schlieve

Forschungsfeld Psycholinguistik

- Untersucht mentale Repräsentationen & Prozesse
 - der Sprachverarbeitung & Sprachproduktion
 - des Spracherwerbs & Lernens
- Fokus auf Echtzeitmessungen der Sprachverarbeitung
 - Reaktionszeitmessungen
 - Augenbewegungsmessungen während des Lesens und Hörens
 - Gehirnstrommessungen

Unsere Labore

Reaktionszeitlabor (3.330)

- 4 Teststationen
- Genutzte Software: Opensesame, Presentation & Eprime



Unsere Labore

Eye-tracking Labor (3.403)

- EyeLink 1000 plus
 - desktop mounted mit Kinnstütze
 - Genutzte Software: Experiment Builder, Data Viewer

Training Videos / Tutorien:

<https://www.jove.com/de/t/57694/eye-tracking-during-visually-situated-language-comprehension>

(Ronderos et al., 2018)

Analyse: <https://doi.org/10.3758/s13428-022-01969-3> (Ito & Knoeferle, 2022)

- Für Studierende: regelmäßig angebotene Eye-tracking Projekt- und Analyseseminare



Unsere Labore

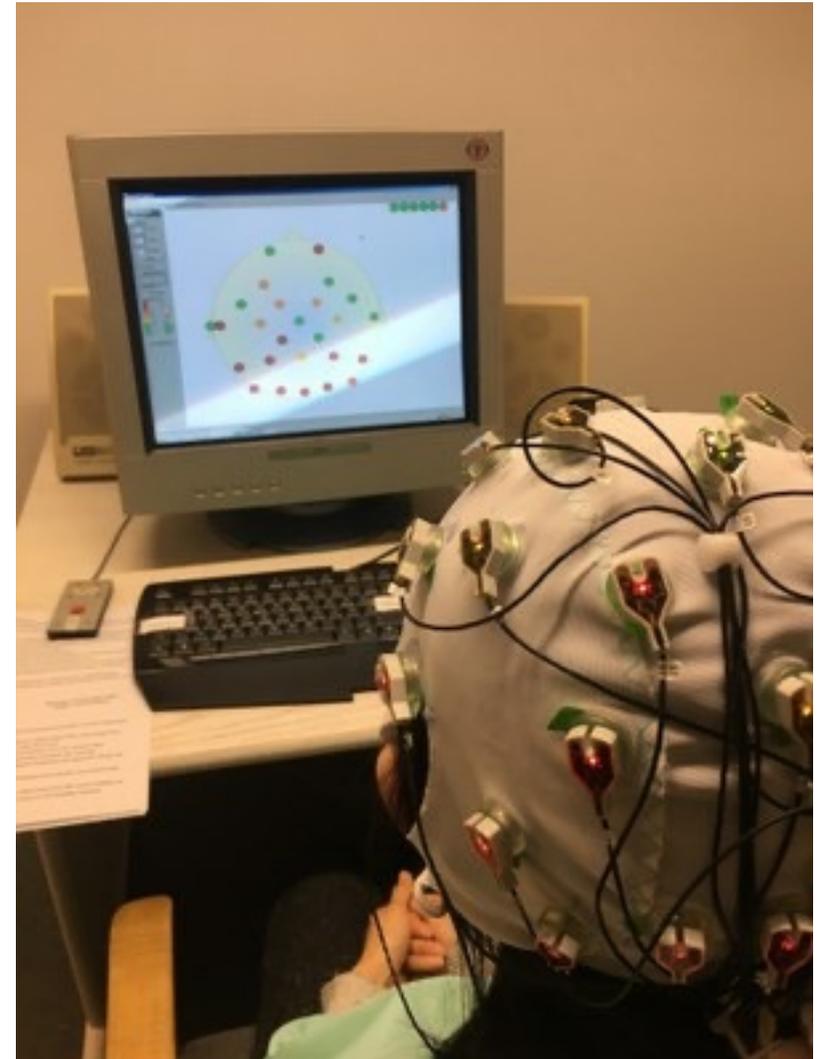
EEG Labor (3.402)

- BrainVision ActiCap system
- 32 Kanäle
- Genutzte Software: Presentation & BrainVision Recorder, Matlab

Training Videos / Tutorien:

https://youtu.be/btgwgOjKV_E?feature=shared (Wyrobnik et al., 2023)

- Für Studierende: regelmäßig angebotene EEG Seminare



Nutzung

Nutzungsanfrage per Email an Prof. Pia Knoeferle

Admin-psycholinguistik@hu-berlin.de

- Labor und Materialbuchungen bald auch über OpenIris möglich!
 - BUA Objective 5: Katalog für Labormaterialien und Laborbuchungen
https://www.ibi.hu-berlin.de/de/forschung/information_behavior/projekteundaktivitaeten/Lab-Know-How
- Weitere laborrelevante Ressourcen
 - Statistikberatung des Instituts (Daniela Palleschi, daniela.palleschi@hu-berlin.de, <https://github.com/daniela-palleschi?tab=repositories>)
 - U.a. frei zugängliche R Kurse, Analyse Scripts für Eye-tracking Leseexperimente, reproducible workflow
 - Anfänger Statistikkurs (BUA) Materialien: <https://osf.io/fe4rq/>

Nächster Beitrag :-)

Phonetiklabor am IdSL

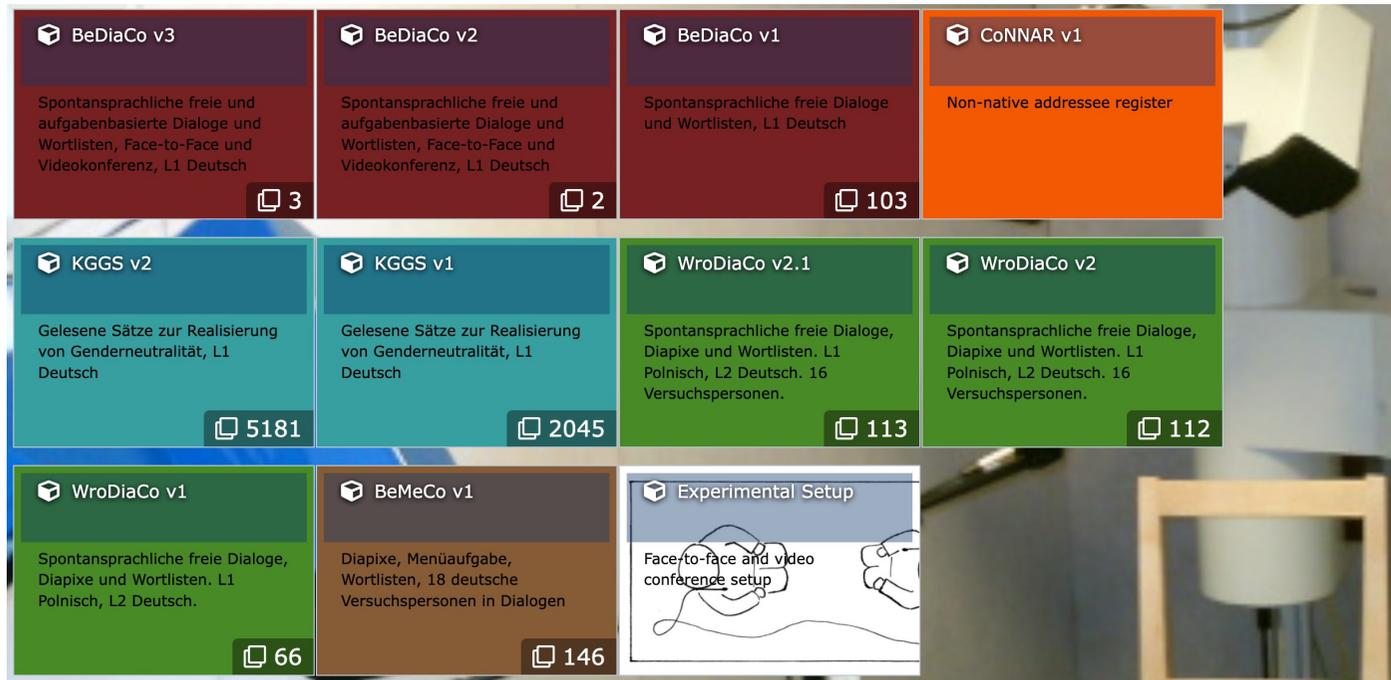
Tine Mooshammer

Malte Belz

(mit Valéria Krepsz, Georg Lohfink, Tom
Offrede, Megumi Terada, Qiang Xia u.a.)

Akustische Aufnahmen

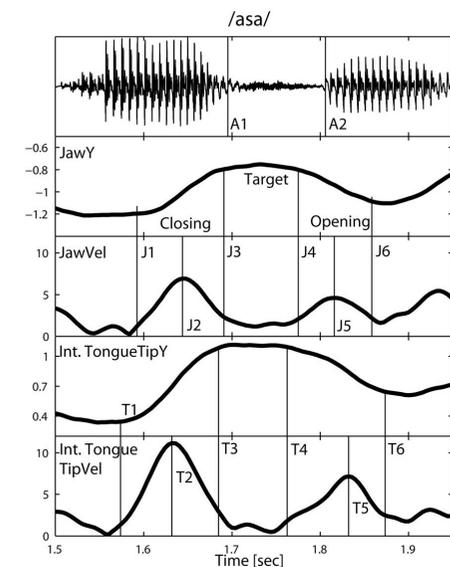
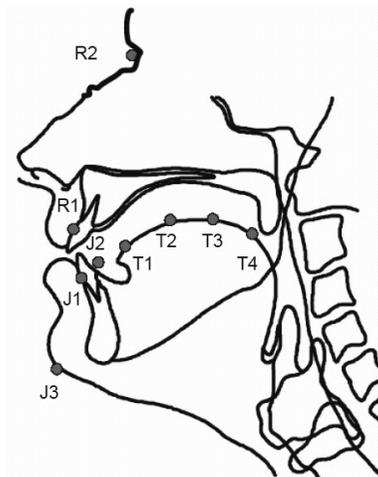
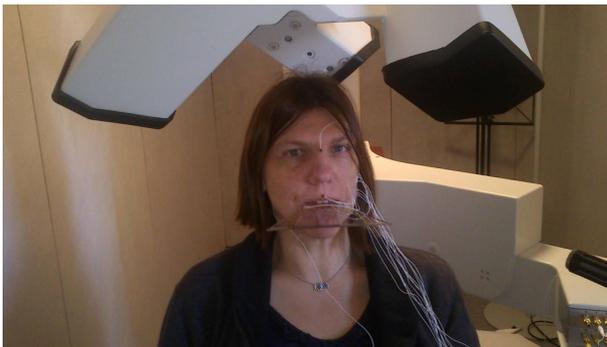
- Schallschutzkabine
- Einzelaufnahmen und Dialoge
- Prompting mit Matlab Scripten
- Korpora: <https://rs.cms.hu-berlin.de/phon/pages/home.php?login=true>



Sprechbewegungen:

Elektromagnetische Artikulographie

- Kleine Spulen bewegen sich in einem elektromagnetischen Feld
- Induzierte Spannung hängt von der Entfernung zwischen den Sensoren und den Sendern ab
- Sensoren werden auf Zunge, Kiefer, Lippen etc. geklebt

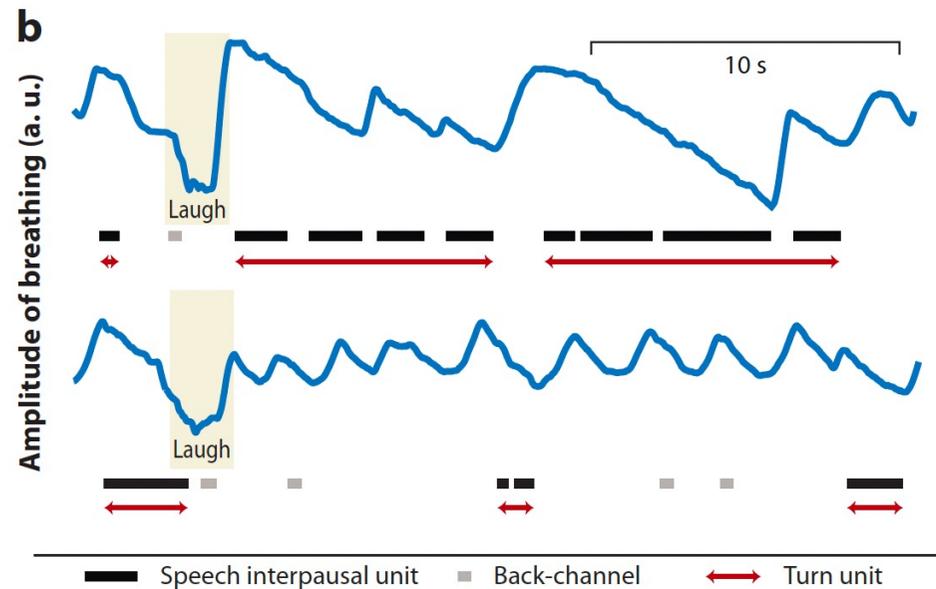
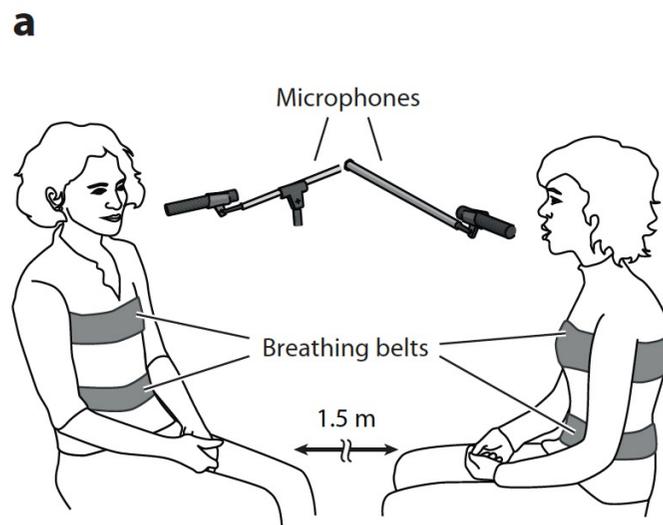


Sprechbewegungen: Elektromagnetische Artikulographie

- Studien:
 - Prosodische Grenzen im Deutschen (mit Jelena Krivokapić, UMich)
 - Koordination von Gesten (mit Peter Birkholz, Dresden)
 - Auditive Perturbation: Wie verändern Sprecher ihre Artikulation, wenn sie sich selbst anders hören? (Jana Brunner, Eugen Klein)

Respiration

- RespTrack System von Columbi
- Brustkorb- und Bauchumfang während der Atmung
- Zwei Systeme für Dialoge



aus Fuchs & Rochet-Capellan (2020)

Respiration

- Studien:
 - Valeria Krepesz: Sprechplanung und Prosodie im Ungarischen
 - Tom Offrede: Atmungsaktivität beim Sprechen und bei körperlicher Anstrengung
 - Oksana Rasskazova (in Kombination mit EMA): Initiierung von Sprechbewegungen
 - in Planung: Projektantrag zum Turntaking mit Dual-EMA und Dual-RespTrack



- Datenschutzplan und Ethikvotum für das Labor (also Kombis von EMA, Atmung, Akustik, Perzeption)
 - auch für Experimente mit neuen Kooperationspartner:innen gültig
- Ansprechpartner: Malte Belz

Nächster Beitrag :-)

Statistik-Beratung im IdSL

Daniela Palleschi

- ▶ Was wird angeboten?
 - ▶ Hilfe bei der **Planung, Durchführung** und **Interpretation** von **Analysen** und **Visualisierungen** in **R**
 - ▶ in der Regel **45-minütige Sitzungen**, um einen **Aktionsplan** zu entwickeln und **Ressourcen** vorzuschlagen, mit Folgeterminen nach Bedarf
 - ▶ Sprachen: Englisch, Deutsch
- ▶ Sprechstunden: **Dienstags 14-16 Uhr / nach Vereinbarung**
- ▶ Sitz: **Raum 3.321**, Dorotheenstr. 24, 10117 Berlin
- ▶ Mail: **daniela.palleschi@hu-berlin.de**
- ▶ Website: daniela-palleschi.github.io

Nächster Beitrag :-)



Language Science Press

Ansatz und Tools

Stefan Müller

22. Oktober 2023

HU Berlin

) Advisory Board entscheidet über Einrichtung von Reihen.

- › Advisory Board entscheidet über Einrichtung von Reihen.
- › 31 Reihen (+2 neue in diesem Jahr)

Bücher und Reihen

- › Advisory Board entscheidet über Einrichtung von Reihen.
- › 31 Reihen (+2 neue in diesem Jahr)
- › 386 Editorial Board Members aus 52 Ländern und 5 Kontinenten

Bücher und Reihen

- › Advisory Board entscheidet über Einrichtung von Reihen.
- › 31 Reihen (+2 neue in diesem Jahr)
- › 386 Editorial Board Members aus 52 Ländern und 5 Kontinenten
- › 1412 Autor*innen

Bücher und Reihen

- › Advisory Board entscheidet über Einrichtung von Reihen.
- › 31 Reihen (+2 neue in diesem Jahr)
- › 386 Editorial Board Members aus 52 Ländern und 5 Kontinenten
- › 1412 Autor*innen
- › 850 Interessensbekundungen (07.09.2023)

Bücher und Reihen

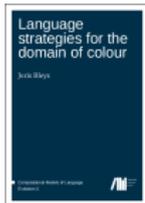
- › Advisory Board entscheidet über Einrichtung von Reihen.
- › 31 Reihen (+2 neue in diesem Jahr)
- › 386 Editorial Board Members aus 52 Ländern und 5 Kontinenten
- › 1412 Autor*innen
- › 850 Interessensbekundungen (07.09.2023)
- › 158 desk rejections

Bücher und Reihen

- › Advisory Board entscheidet über Einrichtung von Reihen.
- › 31 Reihen (+2 neue in diesem Jahr)
- › 386 Editorial Board Members aus 52 Ländern und 5 Kontinenten
- › 1412 Autor*innen
- › 850 Interessensbekundungen (07.09.2023)
- › 158 desk rejections
- › 292 Einreichungen
 - › 37 in Arbeit/im Druck
 - › 26 nach Begutachtung abgelehnt
 - › 12 in Begutachtung

Bücher und Reihen

- › Advisory Board entscheidet über Einrichtung von Reihen.
- › 31 Reihen (+2 neue in diesem Jahr)
- › 386 Editorial Board Members aus 52 Ländern und 5 Kontinenten
- › 1412 Autor*innen
- › 850 Interessensbekundungen (07.09.2023)
- › 158 desk rejections
- › 292 Einreichungen
 - › 37 in Arbeit/im Druck
 - › 26 nach Begutachtung abgelehnt
 - › 12 in Begutachtung
- › 217 Bücher veröffentlicht



) Templates für \LaTeX , LibreOffice und MS Word

- › Templates für \LaTeX , LibreOffice und MS Word
- › Webservice zur Konversion von Officedokumenten nach \LaTeX

- › Templates für \LaTeX , LibreOffice und MS Word
- › Webservice zur Konversion von Officedokumenten nach \LaTeX
- › Integration von Overleaf und Github

- › Templates für \LaTeX , LibreOffice und MS Word
- › Webservice zur Konversion von Officedokumenten nach \LaTeX
- › Integration von Overleaf und Github
- › Open Review und Gamification

- › Autor*innen arbeiten mit \LaTeX , eventuell OverLeaf
- › Word-Manuskripte werden nach \LaTeX konvertiert und können dann in OverLeaf nachbearbeitet werden (Webinterface für \LaTeX mit vorinstallierten LangSci-Paketen).
- › Overleaf ist mit GitHub (Versionsverwaltungssystem) gekoppelt.
- › Alle Bücher haben spätestens für den finalen Satz ein GitHub-Repositoryum.
- › PDFs kommen auf PaperHive (kollaboratives Kommentieren, Open Review, Community Proofreading)
- › PaperHive ist über DocLoop mit GitHub gekoppelt, so dass Kommentare automatisch zu GitHub-Issues werden können.
- › Große Datenbanken mit verifizierten \BibTeX -Einträgen.

Nächster Beitrag :-)



Ethikkommission der Sprach- und literaturwissenschaftlichen Fakultät

Die EK beurteilt die ethische Vertretbarkeit von Forschungsvorhaben und Promotionsprojekten gemäß der Satzung der HU zur Sicherung guter wissenschaftlicher Praxis (GWP) und zum Umgang mit Vorwürfen wissenschaftlichen Fehlverhaltens in der gültigen Fassung. Seit Gründung im April 2021 hat die EK 14 Ethikanträge begutachtet (2021: 3, 2022: 6, 2023: 5).

Website: <https://fakultaeten.hu-berlin.de/de/sprachlit/gremien/ethikkomm>

■ **Mitglieder:** derzeit aus Instituten für Anglistik und Amerikanistik / deutsche Sprache und Linguistik / Romanistik / Slawistik und Hungarologie, Vorsitzende/ Ansprechperson: Prof. Mingya Liu

■ **Kooperationen:**

■ **SLF:** Beauftragte für Forschungsdatenmanagement, allgemeiner Austausch mit anderen

■ **HU:** (1) behördliche Datenschutzbeauftragte – Datenschutzkonzept und datenschutzrechtliche Bestätigung, (2) SZF – Beratung zur Antragstellung der Drittmittelprojekte, (3) GWP – Satzung usw.

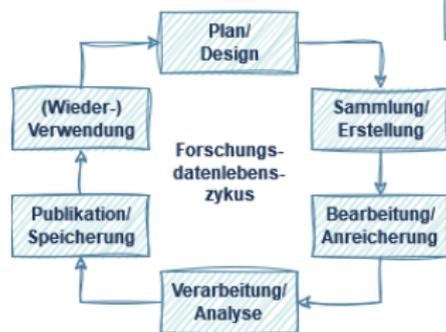
■ **Andere Institute:** Ethikkommission des Instituts für Psychologie – Kurzfragebogen/Beratung, weitere Zusammenarbeit mit den anderen Ethikkommissionen zur Sicherung GWP geplant

■ **Updates:** ab WiSe23/24 vereinfachte Antragstellung mit Kurzfragebogen oder Vollantrag

Nächster Beitrag :-)

Beauftragte für Forschungsdatenmanagement der SLF

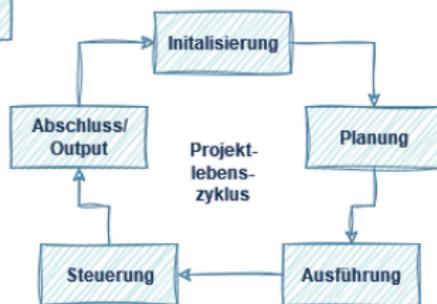
Sprach- und
literaturwissenschaftliche
Fakultät



Forschung
Forschungsfrage
Forschungsleitbild
Lehrleitbild



HUMBOLDT-
UNIVERSITÄT
ZU BERLIN



Beratung, Workshops, Lehre, Forschung
Beauftragte Forschungsdatenmanagement der SLF

- ▶ Kontakt: carolin.odebrecht@hu-berlin.de
- ▶ <https://hu-berlin.de/fdmslf>

Kommission Forschungsdatenmanagement

Die KFDM wird in den Fachbereichen des Forschungsdatenmanagements dem Fakultätsrat und dem Dekanat zuarbeiten und ggf. Empfehlungen aussprechen. [...] Weiterhin versteht sich die FDMK auch als Informationsnetzwerk zwischen den Instituten und auf Ebene der Fakultät.

- ▶ Mitglieder: pro Institut eine Vertretung, tagt hochschulöffentlich
- ▶ Kooperationen: Ethikkommission, CMS
- ▶ Geschäftsführung: Carolin Odebrecht
- ▶ Kontakt: Geschäftsführung und Ihre Institutsvertretungen
- ▶ <https://fakultaeten.hu-berlin.de/de/sprachlit/gremien/KFDM>
- ▶ Aktuelles Thema z. B.: Forschungsdatenmanagement für Abschlussarbeiten

Austausch, Information, Vernetzung und Sichtbarkeit

- ▶ Erster Anfang mit dieser Themenwoche, (noch) Kein vollständiges Bild möglich
- ▶ Hinweise auf weitere Forschung, Forschungswerkzeuge, Labore etc. an mich
- ▶ Plan: Aufsetzen und Pflegen einer Übersichtsseite der SLF zu diesen Themen
- ▶ Bitte Feedback zum Format sehr gern an mich!

Ihnen und Euch eine (weiterhin)
spannende Themenwoche!