



CLARIAH-DE



Während und nach Corona: Digitale Lehre in der Germanistik

CLARIAH-DE in der digitalen Lehre

26. August 2020

Prof. Dr. Andreas Witt und Dr. Antonina Werthmann

Leibniz-Institut für Deutsche Sprache, Mannheim

Sina Bock und Prof. Dr. Fotis Jannidis

Julius-Maximilians-Universität Würzburg

clariah.de

 [@CLARIAHde](https://twitter.com/CLARIAHde)

Gliederung



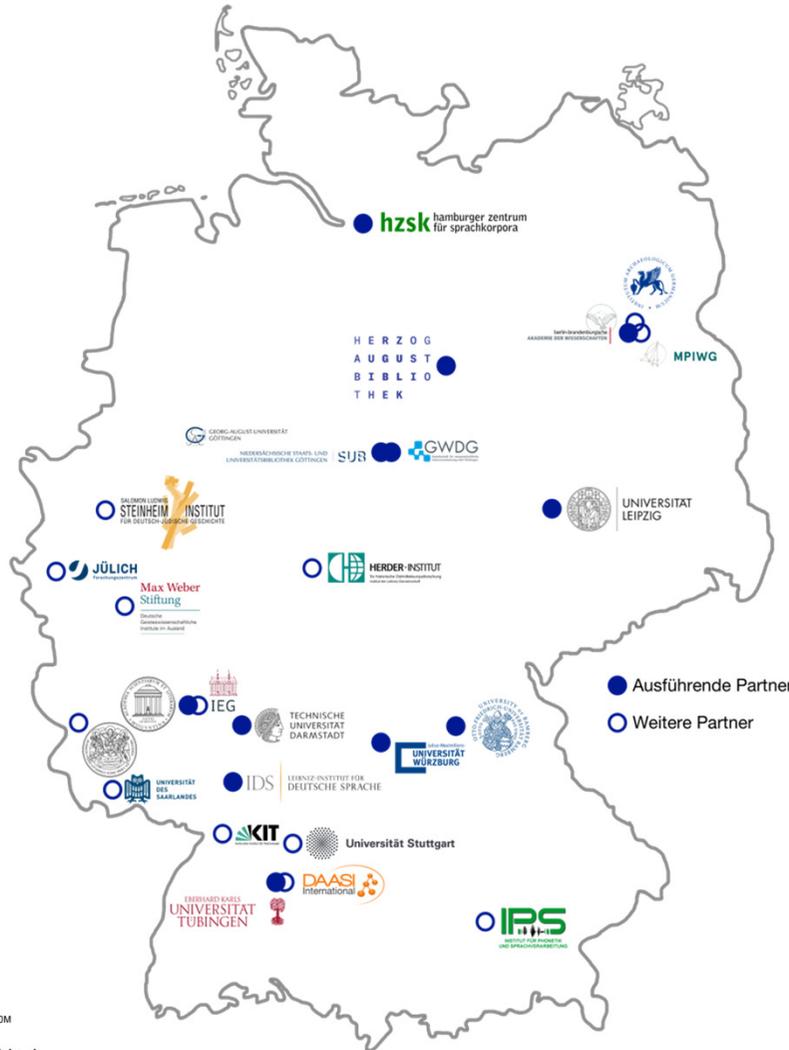
- Forschungsinfrastruktur CLARIAH-DE
- Angebot von Forschungsdaten: Nutzbar in Forschung und Lehre
- Nutzen der Daten
- Perspektiven

CLARIAH-DE



- Digitale Forschungsinfrastruktur für die Geisteswissenschaften
- Förderung durch das Bundesministerium für Bildung und Forschung (BMBF)
- Projektlaufzeit: 1. März 2019 bis 31. März 2021
- Beitrag zu einer Nationalen Forschungsdateninfrastruktur (NFDI) als Partner der NFDI-Initiative Text+

CLARIAH-DE Partner



- Zehn Partner und zwei Unterauftragnehmer
- 13 assoziierte Partner
- Gesamtkoordination:
 - Niedersächsische Staats- und Universitätsbibliothek Göttingen
 - Eberhard Karls Universität Tübingen (Seminar für Sprachwissenschaft, Abt. Computerlinguistik)

Angebot von Forschungsdaten



Einige Beispiele:

- Das Deutsche Referenzkorpus (DeReKo)
- Korpora der Datenbank für Gesprochenes Deutsch (DGD)
- Das Deutsche Textarchiv (DTA)
- Digitales Wörterbuch der deutschen Sprache (DWDS)
- TextGrid Repository

IDS: DeReKo (Dr. Marc Kupietz)



- Weltweit größte linguistisch motivierte Sammlung elektronischer Korpora von geschriebenen deutschsprachigen Texten aus der Gegenwart und der neueren Vergangenheit

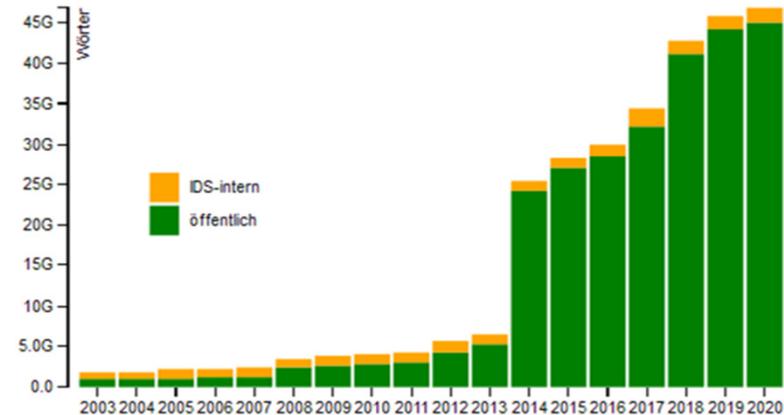


Abb.: DeReKo-Wachstum

- Mit über 46,9 Milliarden Wörtern (Stand 18.01.2020)
- Die Daten sind über COSMAS II und KorAP kostenlos abrufbar
- **Einsatzmöglichkeiten:** Finden von Belegen, Entdecken von Mustern, systematische statistische Auswertungen

IDS: Datenbank für Gesprochenes Deutsch (DGD, Dr. Thomas Schmidt)



- Kontinuierlich aktualisiertes Korpusmanagementsystem
- Sammlung von binnen- und auslandsdeutschen Varietäten und Dialekten
- Tonaufnahmen, Transkripte, Videoaufnahmen und Zusatzmaterialien aus verbalen Interaktionen in verschiedenen sozialen und situativen Kontexten
- **Einsatzmöglichkeiten:** Recherchen und Analysen zum gesprochenen Standarddeutsch sowie zu Umgangssprachen

Korpora in der digitalen Lehre



Herausforderungen

- Den Angeboten wird noch mit Zurückhaltung begegnet
- Nutzung der Portale, deren Suchfunktionen und Abfragesyntax nicht selbstverständlich ist und die eine Einarbeitung benötigen. Es gibt aber bereits viele Online-Tutorials.

Chancen

- Sprache anhand authentischer Belege analysieren
- Vielfältige Korpora sind permanent frei zugänglich
- Komfortable Suchplattformen und Korpusmanagementsysteme (z.B. KorAP und DGD) ermöglichen Suchabfragen und Visualisierungen

BBAW: Deutsches Textarchiv (PD Dr. Alexander Geyken)



- Digitales Volltextkorpus deutschsprachiger Texte ab dem frühen 16. bis zum frühen 20. Jahrhundert
- Diverse Textsorten (Zeitung, Gebrauchsliteratur, diverse wissenschaftliche Disziplinen, Belletristik; gedruckte und handschriftliche Vorlagen)
- Ziel: Abbildung der sprachhistorischen Entwicklungen der deutschen Sprache seit dem Ende der frühneuhochdeutschen Sprachperiode
- Diverse Downloadformate: XML (TEI P5), HTML, Text, TCF (tokenisiert, serialisiert, lemmatisiert, normalisiert)
- **Einsatzmöglichkeiten:** zitierfähige Primärtexte für die Lehre, Korpuszusammenstellung und -auswertung

TextGrid Repository



- Langzeitarchiv für geisteswissenschaftliche Forschungsdaten
- Verfügt über Digitale Bibliothek mit rund 100.000 zitierfähigen Primärtexten aus verschiedenen literarischen Genres und Epochen
- **Einsatzmöglichkeiten:** Volltexte für Unterricht, Korpuszusammenstellung und -auswertung

Repositorien in der digitalen Lehre



Herausforderungen

- Referenzsystem für Texte
- Nutzung ist nicht selbstverständlich und braucht eine Einarbeitung

Chancen

- Ressourcen sind permanent zugänglich sowie zeitlich und örtlich unabhängig

Digitales Wörterbuch der deutschen Sprache



- Wortinformationssystem zur Auskunft über den deutschen Wortschatz in Vergangenheit und Gegenwart
- **Einsatzmöglichkeiten:** Recherchen zu diversen Bedeutungen von Begriffen, deren Etymologie, Anwendungsbeispiele, Angaben über die Worthäufigkeit und Wortverlaufskurve

Wörterbücher

465 000 Einträge in verschiedenen Wörterbüchern.

Konfekt, das

Grammatik Substantiv (Neutrum) - Genitiv Singular: **Konfekts/Konfektes** - Plural: **Konfekte** - wird meist im Singular verwendet

Aussprache **ˈkɔŋfɛkt**

Wortbildung Konfekt (Konfekt)

Zusatz Latein

Wortbildung mit -Konfekt- als Erstglied: [Konfektbrot](#) - mit -Konfekt- als Letztglied: [Eiskonfekt](#)

Bedeutungen Dieser Artikel ist ein Teil des DWDS.

1. **Pralinen o. Ä.**

Beispiele:
Im Hotel **uf** stand **Konfekt** aus Marzipan [...] auf dem **Sak** [...] (Die Welt, 05.07.2009)

Auf dem Tisch stehen zur Begrüßung eine **Fruchtschale**

Textkorpora

25 Milliarden Belege aus historischen und gegenwartssprachlichen Textkorpora.

1-50 von 13157 Treffern

Hamel am Himmel sein wahres Gold	Fell/ oder seine Gold Wolle/ auf En
ein wahres Gold Fell/ oder seine Gold	Wolle/ auf Erden herab.
tund stäts der Aehren schweres Gold	.
Du streust Gold	von deinen Höhen/ unsre Wasen a
Wen solte/ der Rubin in diesem Gold	/ nicht laben?
er andre macht verliebt/ ist von Gold	/ und scharf gespitzt.
ster; lässt/ das mehr als güldne Gold	Ihrer Haare/ ledig fliegen.
rüsten? daß mich solt/ wie nach Gold	/ nach dir gelösten?
Dessen Haupt von feinem Gold	/ die Brust und Arm von Silber/ de

Statistiken

Häufigkeiten, typische Verbindungen und zeitliche Verläufe über 400 Jahre.



<https://www.dwds.de>

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

IDS

LEIBNIZ-INSTITUT FÜR
DEUTSCHE SPRACHE

CLARIAH-DE Tutorial Finder



Die im Tutorial Finder durchsuchbaren Sammlungen mit Lehr- und Schulungsmaterialien

Navigation: Sammlungen | Akteure | Vokabulare | Hilfe | Deutsch | CLARIAH-DE

CLARIAH-DE | COLLECTION REGISTRY

Sammlungen

Zeigen: 10 | Filter:

[SAMMLUNG HINZUFÜGEN](#)

	Bild	Titel	Kollektionstypen	Zugang	Aktuelle Version
Veröffentlicht		DhDKanal Video-Tutorials	Videosammlung	Online file	5. Mai 2020 20:42:57
Veröffentlicht		Teaching and Learning Materials Collection [TeLeMaCo]	Lehrmaterialsammlung	Online file	7. Mai 2020 10:04:33
Veröffentlicht		DARIAH-Campus	Lehrmaterialsammlung	Git Repository	27. Juli 2020 08:48:58
Veröffentlicht		#dariah Teach	Videosammlung	Online file	30. Juli 2020 12:27:23
Veröffentlicht		forTEXT	Videosammlung	Online file	30. Juli 2020 12:29:54
Veröffentlicht		The Programming Historian	Lehrmaterialsammlung	Git Repository	31. Juli 2020 10:58:17

6 Einträge | < 1 >

<https://teaching.clariah.de/colreg-ui/collections/>

Nutzen der Daten



- CLARIAH-DE folgt den FAIR-Prinzipien
 - Auffindbarkeit (Findable), Zugänglichkeit (Accessible), Interoperabilität (Interoperable), Nachnutzbarkeit (Reusable)
- Angebote nach Möglichkeit offen zugänglich – einige Ressourcen jedoch nur unter eingeschränkten Bedingungen
- Zugriff auf die Angebote erfolgt über eine Authentication and Authorization Infrastructure (AAI); derzeit können über 80 Prozent aller deutschen Universitäten das Angebot über AAI nutzen

Perspektiven

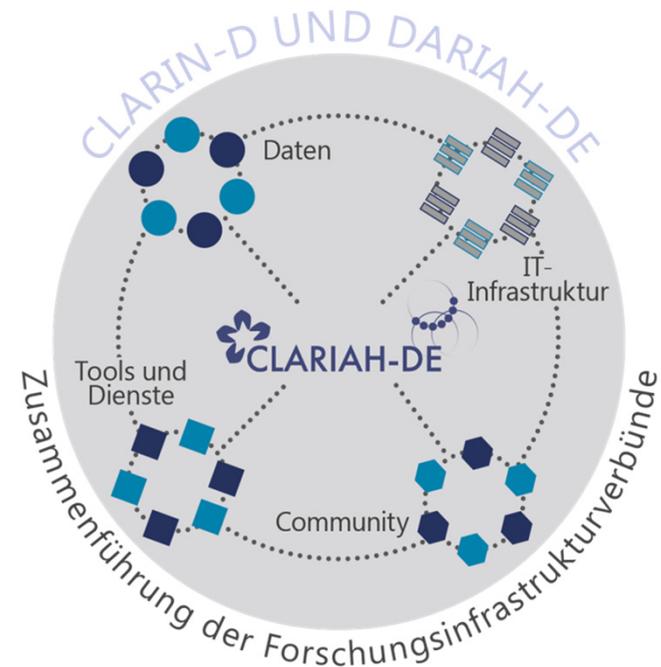


- Forschungsdaten können über den DARIAH-DE Publikator in DARIAH-DE Repository veröffentlicht und einer breiten Community zur Verfügung gestellt werden
- Beratung bei rechtlichen Fragen durch CLARIAH-DE

CLARIAH-DE Angebot



- **Daten:** Forschungsdaten, Standards und Verfahren
- **Tools und Dienste:** Werkzeuge und virtuelle Forschungsumgebungen
- **Community:** Kompetenzvermittlung, Nachwuchsförderung, Community-Engagement
- **IT-Infrastruktur:** Vernetzung und Koordination der technischen Entwicklung



KorAP



- Suchplattform für den Zugriff auf die DeReKo-Daten
- Unterstützung unterschiedlicher Anfragesprachen
- Annotationsassistent, Visualisierung syntaktischer Strukturen (z.B. Konstituenz- und Dependenzrelationen), Metadatenansicht
- Unterstützung der Konstruktion virtueller Korpora anhand von Metadatenvariablen
- **Einsatzmöglichkeiten:** Suche nach speziellen grammatischen Phänomenen, Erforschung grammatischer Variation

beispielsweise auf der Unterseite . Gruß von 15:23, 25. Okt. 2011 (CEST) Hier **ein kleines Beispiel** zu den Referenzen: Anmerkungen "Siehe-auch" Wenn Du Lust hast schau Dir mal meine Seite und WP:SA an. -- 16:23, 27. Okt. 2011 (CEST) Architektur der Renaissance Wow! Beeindruckender Artikel - und das als Erstlingswerk. Ich bin echt beeindruckt! Zwei Kleinigkeiten die

Foundry	Layer	Hier	ein	kleines	Beispiel	zu	den	Referenzen
corenlp	p	ADV	ART	ADJA	NN	APPR	ART	NN
opennlp	p	ADV	ART	ADJA	NN	APPR	ART	NN
tt	l	hier	eine	klein	Beispiel	zu	die	Referenz
tt	p	ADV	ART	ADJA	NN	APPR	ART	NN

corenlp c

malt d

Metadaten Token Relationen Benutzer Diskussion:SovieleRoger von Zacke, u.a. (2017-07-01) [WUD17/S65/38276]

<https://korap.ids-mannheim.de/>

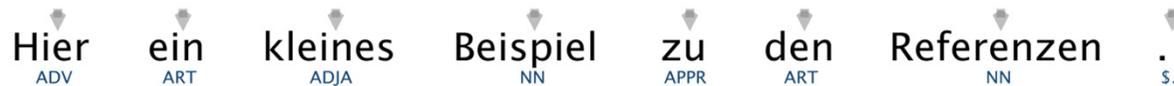
WebLicht



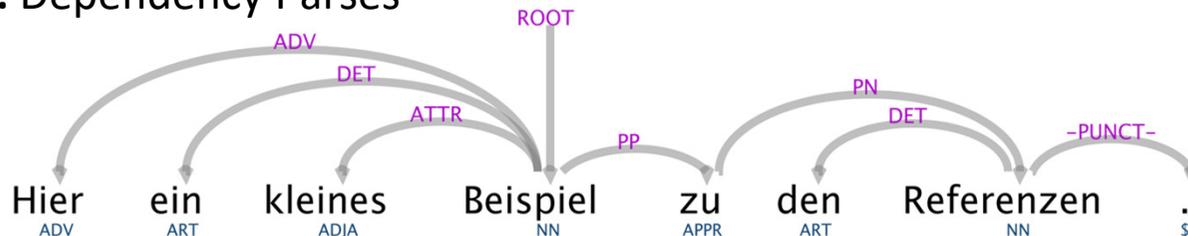
Ein web-basiertes Analysewerkzeug zu Tokenisierung, Lemmatisierung, PoS, Syntaxanalyse (Konstituenz und Dependenz), NER und Geotagging

Bsp. Input

Bsp. Output: POS Tags/Lemmas



Bsp. Output: Dependency Parses



GEFÖRDERT VOM



https://weblight.sfs.uni-tuebingen.de/weblightwiki/index.php/Main_Page

TextGrid Laboratory



- Kostenfreies Software-Paket
- Bereitstellung eines Instruments, mit dem auf die Tools und Services zugegriffen sowie Forschungsdaten erstellt, verwaltet und bearbeitet werden können
- Verfügbar für Windows, Mac OS X und Linux



CLARIAH-DE



Fragen und Diskussion



clariah.de

 [@CLARIAHde](https://twitter.com/CLARIAHde)

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

FÖRDERKENNZEICHEN
01UG1910 A bis I

Linksammlung



- **CLARIAH-DE:** <https://www.clariah.de/>
- **DARIAH-DE Publikator:** <https://de.dariah.eu/web/guest/publikator>
- **DARIAH-DE Repository:** <https://de.dariah.eu/repository>
- **DeReKo:** <https://www1.ids-mannheim.de/kl/projekte/korpora/>
- **DGD:** https://dgd.ids-mannheim.de/dgd/pragdb.dgd_extern.welcome
- **DTA:** <http://www.deutschestextarchiv.de/>
- **DWDS:** <https://www.dwds.de/>
- **KorAP:** <https://korap.ids-mannheim.de/>
- **CLARIAH-DE Tutorial Finder:** <https://teaching.clariah.de>
- **TextGrid Laboratory:** https://textgrid.de/de_DE/download
- **TextGrid Repository:** <https://textgridrep.org/>
- **WebLicht:** https://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Main_Page

Linksammlung



Beispiele für die digitalen Lehr- und Schulungsmaterialiensammlungen:

- **DHdKanal Video-Tutorials:** <https://www.youtube.com/user/dhdkanal/>
- **TeLeMaCo:** <https://fedora.clarin-d.uni-saarland.de/hub/>
- **DARIAH-Campus:** <https://campus.dariah.eu/>
- **#dariah Teach:**
<https://teach.dariah.eu/>
<https://www.youtube.com/channel/UCScSbG7XjiXbZVgilEp0Pkw>
- **forTEXT:** https://www.youtube.com/channel/UC4aOD2eusb_q66LD-Q0km9Q
- **The Programming Historian:** <https://programminghistorian.org/>