

CLARIAH-DE in der digitalen Lehre

Antonina Werthmann, Andreas Witt (Mannheim)

Sina Bock, Fotis Jannidis (Würzburg)

Abstract

Die durch die Covid-19-Pandemie bedingte Umstellung der Präsenzlehre auf digitale Lehr- und Lernformate stellte Lehrende und Studierende gleichermaßen vor eine Herausforderung. Innerhalb kürzester Zeit musste die Nutzung von Plattformen und digitalen Tools erlernt und getestet werden. Der Beitrag stellt exemplarisch Dienste und Werkzeuge von CLARIAH-DE vor und erläutert, wie die digitale Forschungsinfrastruktur Lehrende und Studierende auch im Rahmen der digitalen Lehre unterstützen kann.

Schlagwörter: CLARIAH-DE, Forschungsinfrastruktur, digitale Forschungsdaten, digitale Werkzeuge, digitale Lehre, Germanistik, Literaturwissenschaft, DaF-/DaZ-Unterricht

CLARIAH-DE¹ ist ein vom Bundesministerium für Bildung und Forschung (BMBF) gefördertes, zweijähriges (2019–2021) Verbundprojekt, in dessen Rahmen die Ressourcen der Forschungsinfrastrukturen CLARIN-D² und DARIAH-DE³ zusammengeführt werden, um die Forschung der Geisteswissenschaften und ihre benachbarten Disziplinen bestmöglich zu unterstützen. Daher beteiligt sich CLARIAH-DE auch am Aufbau der Forschungsdateninfrastruktur Text+⁴, die sich auf digitale Sammlungen textbasierter Forschungsdaten (einschließlich Korpora), lexikalische Ressourcen und Editionen konzentriert. Am Projekt CLARIAH-DE beteiligen sich über 25 universitäre und außeruniversitäre Partnerinstitutionen. Die Gesamtkoordination liegt dabei gemeinschaftlich bei der Niedersächsischen Staats- und Universitätsbibliothek Göttingen und der Eberhard Karls Universität Tübingen (Seminar für Sprachwissenschaft, Abt. Computerlinguistik).

¹ CLARIAH-DE, URL: <https://www.clariah.de/> (Zugriffsdatum: 27.10.2020).

² CLARIN-D ist der deutsche Beitrag zur *Common Language Resources and Technology Infrastructure* (CLARIN), die als *European Research Infrastructure Consortium* (ERIC) ein Netzwerk aus 20 Ländern neben weiteren Partnern bildet, URL: <https://www.clarin-d.net/de> (Zugriffsdatum: 27.10.2020).

³ DARIAH-DE ist der deutsche Partner der *Digital Research Infrastructure for the Arts and Humanities* (DARIAH-EU), die mit 19 Ländern und weiteren Partnern ebenfalls ein ERIC auf europäischer Ebene ist, URL: <https://de.dariah.eu> (Zugriffsdatum: 27.10.2020).

⁴ Text+, URL: <https://www.text-plus.org/> (Zugriffsdatum: 27.10.2020).

Ein Aufgabenbereich von CLARIAH-DE ist die Entwicklung von digitalen Werkzeugen zum Editieren, Annotieren und Visualisieren von textbasierten Forschungsdaten sowie das Erstellen von Lehr- und Schulungsmaterialien zur Anwendung von quantitativen Methoden. Die Auseinandersetzung mit der Erstellung digitaler Forschungsdaten, der Entwicklung von Werkzeugen sowie der nachhaltigen Bereitstellung von Informationsangeboten für textbasierte Forschungsdaten allein ermöglicht es uns jedoch nicht, die Frage eindeutig zu beantworten, wie digitale Lehre der Germanistik erfolgreich gestaltet werden kann. Das Ziel des vorliegenden Beitrags ist es, anhand von ausgewählten Beispielen das Potenzial von CLARIAH-DE-Angeboten in der digitalen Lehre aufzuzeigen und dadurch einen Einblick zu geben, welchen Beitrag die CLARIAH-DE-Forschungsdateninfrastruktur dabei leisten kann. Anschließend werden Herausforderungen und Chancen skizziert, die die Anwendung der digitalen Ressourcen in der Lehre mit sich bringt.

Die hier dargestellten Fallbeispiele lassen sich vier Anwendungsbereichen zuordnen:

- Nutzung von sprach- und textbasierten Sammlungen geschriebener, gesprochener Sprache als Quellen für literarische und nicht literarische Primärtexte (z. B. DeReKo, DGD, DTA, TextGrid Repository)
- Nutzung von lexikalischen Ressourcen (z. B. DWDS)
- Nutzung von Analyseplattformen und Anwendung von Werkzeugen zum empirischen Arbeiten mit Korpora sowie zur Visualisierung der Ergebnisse von Recherchen (z. B. KorAP, TextGrid, WebLicht, TextGrid Repository)
- Verwendung von bestehenden Lehr- und Schulungsmaterialien zum Selbststudium oder Einsatz in der Lehre

Im Folgenden wird auf die einzelnen Angebote der drei Anwendungsbereiche näher eingegangen und deren Mehrwert für die digitale Lehre erläutert.

Das Deutsche Referenzkorpus und KorAP

Das Deutsche Referenzkorpus – DeReKo⁵ – ist mit aktuell über 46,9 Milliarden Wörtern die weltweit größte linguistisch motivierte Sammlung elektronischer Korpora von geschriebenen deutschsprachigen Texten der Gegenwart und der neueren Vergangenheit. DeReKo wird am Leibniz-Institut für Deutsche Sprache (IDS) in Mannheim gepflegt und kontinuierlich ausgebaut (vgl. Kupietz et al., 2018). In der Germanistik eignet sich das Korpus in der Lehre zur Beantwortung linguistischer Fragestellungen, zum Finden von Belegen, zur Entdeckung von beispielsweise semantischen oder syntaktischen Mustern und zu systematischen statistischen Auswertungen. Das Referenzkorpus bietet

⁵ DeReKo, URL: <https://www1.ids-mannheim.de/kl/projekte/korpora/> (Zugriffsdatum: 27.10.2020).

somit eine empirische Grundlage insbesondere für die quantitative Analyse der deutschen Gegenwartssprache in ihrer Verwendung und Entwicklung. Über die Korpusanalyseplattform KorAP⁶ können Studierende eigenständig große Datenbestände nach linguistischen Phänomenen aus den Bereichen Semantik, Grammatik, Textlinguistik oder Diskurs systematisch durchsuchen, deren Sprachgebrauch in verschiedenen Textgenres analysieren und sich zu den Suchergebnissen Annotationen wie beispielsweise zu Wortformen oder syntaktischer Struktur anzeigen lassen (siehe Beispiele in Kupietz et al., 2019). Die KorAP-Benutzeroberfläche ist intuitiv verständlich und ermöglicht auch ohne technische Vorkenntnisse und ohne eine zeitaufwendige Einarbeitung die Durchführung einer Korpusrecherche, Datenauswertung und -visualisierung. Der Einsatz von DeReKo und KorAP in der Lehre kann Studierende dabei unterstützen, sich mit den Grundlagen des methodisch validen Umgangs mit Korpusdaten im Bereich der Sprachwissenschaft und insbesondere der empirisch germanistischen Forschung auseinanderzusetzen und vertraut zu machen.

Datenbank für Gesprochenes Deutsch

Die Datenbank für gesprochenes Deutsch – DGD⁷ – des Leibniz-Instituts für Deutsche Sprache (IDS) ist ein kontinuierlich aktualisiertes Korpusmanagementsystem, das eine umfangreiche Sammlung von binnen- und auslandsdeutschen Varietäten und Dialekten mit Tonaufnahmen, Transkripten, Videoaufnahmen und Zusatzmaterialien aus verbalen Interaktionen in verschiedenen sozialen und situativen Kontexten umfasst (vgl. z. B. Schmidt et al., 2019). In der germanistischen Lehre eignet sich die Datenbank, die gegenwärtig 34 Korpora (Stand 08.10.2020) umfasst, besonders um Varietäten des Deutschen (Dialekte, Umgangssprachen und das gesprochene Standarddeutsch) sowie verbale Interaktionen in verschiedenen sozialen und situativen Kontexten aufzuzeigen und zu analysieren. Eine weitere Einsatzmöglichkeit ist die Verwendung von DGD-Sprachdaten als Datenquelle im DaF- und DaZ-Unterricht, indem Studierenden authentische Texte in informellen und alltäglichen Gesprächssituationen angeboten werden. Als Fragmente authentischer Alltagssituationen können sie den Sprachgebrauch lebhaft sowie anschaulich wiedergeben und somit zu einer höheren Lernmotivation beitragen. Da die Daten sowohl in sprachlicher als auch schriftlicher Form permanent zur Verfügung stehen und nach Belegen und Metadaten durchsuchbar sind, können sie je nach Thema und Relevanz für Lernende gezielt ausgewählt werden.

⁶ KorAP, URL: <https://korap.ids-mannheim.de/> (Zugriffsdatum: 27.10.2020).

⁷ DGD, URL: https://dgd.ids-mannheim.de/dgd/pragdb.dgd_extern.welcome (Zugriffsdatum: 27.10.2020).

Deutsches Textarchiv

Das Deutsche Textarchiv – DTA⁸ – ist ein von der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW) erstelltes, disziplinen- und gattungsübergreifendes digitales Volltextkorpus deutschsprachiger Texte aus dem frühen 16. bis zum frühen 20. Jahrhundert (vgl. z. B. Geyken et al., 2018). Die Sammlung umfasst diverse Textsorten wie Zeitung, Gebrauchsliteratur, Belletristik, verschiedene wissenschaftliche Disziplinen von gedruckten und handschriftlichen Vorlagen, die sprachhistorische Entwicklungen der deutschen Sprache seit dem Ende der frühneuhochdeutschen Sprachperiode abbilden. Die Texte des DTA-Volltextkorpus liegen in diversen Dateiformaten – XML (TEI P5), HTML, Text, TCF (tokenisiert, serialisiert, lemmatisiert, normalisiert) – vor und sind frei zugänglich. In der Lehre eignet sich das DTA somit als Quelle für zitierfähige Primärtexte und ermöglicht zudem das Zusammenstellen individueller Korpora, die darüber hinaus auch mit etablierten Forschungswerkzeugen analysiert werden können.

Digitales Wörterbuch der deutschen Sprache

Das Digitale Wörterbuch der deutschen Sprache – DWDS⁹ – ist ein von der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW) gepflegtes digitales Wortinformationssystem zur Auskunft über den deutschen Wortschatz. Die in die Rechercheplattform des DWDS eingebundenen Referenzkorpora beinhalten Texte, die insgesamt einen Zeitraum von ca. 1600 bis heute abdecken (vgl. z. B. Geyken et al., 2017). Mit über 555.000 Einträgen in verschiedenen Wörterbüchern und über 25 Milliarden Belegen (Stand 6.10.2020) aus historischen und gegenwartssprachlichen Textkorpora kann das DWDS vielseitig für Recherchen eingesetzt werden. So können diverse Bedeutungen von Begriffen, deren Etymologie, Anwendungsbeispiele, Angaben über die Worthäufigkeit und Wortverlaufskurven sowie typische Verbindungen angezeigt und zeitliche Verläufe statistisch ausgewertet werden.

⁸ DTA, URL: <http://www.deutschestextarchiv.de/> (Zugriffsdatum: 27.10.2020).

⁹ DWDS, URL: <https://www.dwds.de/> (Zugriffsdatum: 27.10.2020).

TextGrid

TextGrid¹⁰ ist ein 2006 ins Leben gerufenes Verbundprojekt, das aus zehn institutionellen und universitären Partnern besteht. Neben dem TextGrid Repository¹¹ – einem Langzeitarchiv für geisteswissenschaftliche Forschungsdaten – bietet es mit dem TextGrid Laboratory¹² eine virtuelle Forschungsumgebung, die einen kollaborativen Forschungsprozess – insbesondere bei der Erstellung digitaler Editionen – durch Open-Source-Werkzeuge und Open-Source-Services unterstützt. Über die Digitale Bibliothek¹³ bietet das TextGrid Repository Zugriff auf einen stetig wachsenden Bestand von gegenwärtig rund 100.000 zitierfähigen deutschsprachigen Primärtexten aus verschiedenen literarischen Genres und Epochen vom Anfang des Buchdrucks bis hin zu den ersten Jahrzehnten des 20. Jahrhunderts. Die Texte können in unterschiedlichen Ausgabeformaten (z. B. XML, ePub, PDF) kostenlos heruntergeladen werden und eignen sich für den Einsatz in der Germanistik, der vergleichenden Literaturwissenschaft, der Philosophie und den Kulturwissenschaften.

WebLicht

Mit WebLicht¹⁴, einem Werkzeug zur Verarbeitung natürlicher Sprache, stellt CLARIAH-DE Ressourcen für Forschende und Lehrende aus der Linguistik zur Verfügung, die bei der Vermittlung sprachwissenschaftlicher Lehrinhalte zur Tokenisierung, Lemmatisierung, Part-of-Speech-Tagging (PoS-Tagging), Syntaxanalyse (Konstituenz und Dependenz), Named Entity Recognition (NER) und Geotagging von digitalen Texten eingesetzt werden können.

CLARIAH-DE Tutorial Finder

Für die Anwendung digitaler Methoden, Forschungsdaten und Dienste sind zahlreiche Lehr- und Schulungsmaterialien für einen niederschweligen Einstieg vorhanden. Einen Überblick ermöglicht der sich im Aufbau befindende *Tutorial Finder*¹⁵, der geprüfte Materialien aus nationalen und internationalen Quellen sowie Eigenentwicklungen erfasst. Momentan beinhaltet der *Tutorial Finder* sechs Sammlungen mit 558 Lehrangeboten (Stand: 27.10.2020) und wird sukzessiv erweitert. Vor-

¹⁰ TextGrid, URL: <https://textgrid.de/web/guest> (Zugriffsdatum: 27.10.2020).

¹¹ TextGrid Repository, URL: <https://textgridrep.org/> (Zugriffsdatum: 27.10.2020).

¹² TextGrid Laboratory, URL: <https://textgrid.de/download> (Zugriffsdatum: 27.10.2020).

¹³ Digitale Bibliothek von TextGrid, URL: <https://textgrid.de/digitale-bibliothek> (Zugriffsdatum: 27.10.2020).

¹⁴ WebLicht, URL: https://weblight.sfs.uni-tuebingen.de/weblightwiki/index.php/Main_Page (Zugriffsdatum: 27.10.2020).

¹⁵ Tutorial Finder, URL: <https://teaching.clariah.de> (Zugriffsdatum: 27.10.2020).

schläge für weitere Sammlungen und/oder eigenentwickelte Lehr- und Schulungsmaterialien können gerne beim *Helpdesk*¹⁶ eingereicht werden. Der *Tutorial Finder* hostet die Lehr- und Schulungsmaterialien jedoch nicht selbst, sondern ermöglicht einen Zugriff über Metadaten und Volltexte, sodass die integrierten Materialien ihre zugrunde liegende Lizenz behalten. Nach ihrer Aufnahme in den *Tutorial Finder* stehen die Schulungs- und Lehrmaterialien einem breiten Nutzerkreis der Geisteswissenschaften und ihrer benachbarten Disziplinen zur Verfügung.

Fazit und Ausblick

Dieser Beitrag zeigt exemplarisch, wie die digitale Forschungsinfrastruktur CLARIAH-DE Lehrende und Studierende in der digitalen Lehre bei der Vermittlung der Lehrinhalte im Bereich der Sprach- und Literaturwissenschaft unterstützen kann. Die Bereitstellung von Forschungsdaten und Diensten innerhalb einer Forschungsinfrastruktur ermöglicht Forschenden, Lehrenden und Studierenden einen räumlich und zeitlich uneingeschränkten Zugang zu den Angeboten, der gerade für die asynchrone Lehre einen großen Vorteil mit sich bringt. Zudem eignen sich bereits geprüfte Online-Tutorials für den Einsatz in der Lehre oder zum vertiefenden Selbststudium. In Diskussionen mit potenziellen Nutzenden stellte sich heraus, dass die in CLARIAH-DE enthaltenen Angebote gegenwärtig noch sehr zurückhaltend eingesetzt werden. So bestehen beispielsweise derzeit die Bedenken, dass die Einarbeitung in den Umgang mit den digitalen Ressourcen und Diensten zu viel Zeit in Anspruch nimmt oder fundierte technische Vorkenntnisse vorausgesetzt werden. Wir hoffen, mit unserem Beitrag zeigen zu können, dass mögliche Vorkenntnisse abhängig vom jeweiligen Anwendungsbereich sehr unterschiedlich ausfallen und die angestrebten Lernziele häufig an den Wissenshintergrund der Lehrenden und Studierenden angepasst werden können. Die Angebote der Forschungsinfrastruktur CLARIAH-DE werden von Geisteswissenschaftler*innen mit dem Ziel konzipiert, die Forschung und Vernetzung in den Geisteswissenschaften zu unterstützen. Aus diesem Grund sind das Feedback und die Zusammenarbeit mit den Forschenden und Lehrenden, die die angebotenen Ressourcen in der Praxis testen und nutzen, eine wertvolle Ressource für die Selbstreflexion und die Verbesserung unserer eigenen Angebote.

Bibliographie

Geyken, Alexander / Barbaresi, Adrien / Didakowski, Jörg / Jurish, Bryan / Wiegand, Frank / Lemnitzer, Lothar: Die Korpusplattform des „Digitalen Wörterbuchs der deutschen Sprache“ (DWDS). In: Zeitschrift für germanistische Linguistik 45,2 (2017), S. 327–344.

¹⁶ CLARIAH-DE Helpdesk, URL <https://www.clariah.de/support> (Zugriffsdatum: 27.10.2020).

- Geyken, Alexander / Boenig, Matthias / Haaf, Susanne / Jurish, Bryan / Thomas, Christian / Wiegand, Frank: Das Deutsche Textarchiv als Forschungsplattform für historische Daten in CLARIN. In: Lobin, Henning / Schneider, Roman / Witt, Andreas (Hg.): Digitale Infrastrukturen für die germanistische Forschung. (Germanistische Sprachwissenschaft um 2020,6) Berlin, New York: De Gruyter 2018, S. 219–248.
- Kupietz, Marc / Diewald, Nils / Margaretha, Eliza / Bodmer Mory, Franck / Stallkamp, Helge / Harders, Peter: Neues von KorAP. In: Eichinger, Ludwig M. / Plewnia, Albrecht (Hg.): Neues vom heutigen Deutsch: Empirisch – methodisch – theoretisch. (Jahrbuch des Instituts für Deutsche Sprache 2018) Berlin, New York: De Gruyter 2019, S. 345–350.
- Kupietz, Marc / Längen, Harald / Kamocki, Paweł / Witt, Andreas: The German reference corpus DeReKo: New developments – new opportunities. In: Calzolari, Nicoletta / Choukri, Khalid / Cieri, Christopher / Declerck, Thierry / Goggi, Sara Sara / Hasida, Koiti / Isahara, Hitoshi / Maegaard, Bente / Mariani, Joseph / Mazo, Hélène / Moreno, Asuncion / Odijk, Jan / Piperidis, Stelios / Tokunaga, Takenobu (Hg.): Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018), 7-12 May 2018, Miyazaki, Japan. Paris, France: European language resources association (ELRA), S. 4353–4360.
- Schmidt, Thomas / Gasch, Joachim / Kaiser, Julia: DGD – die Datenbank für Gesprochenes Deutsch. In: Eichinger, Ludwig M. / Plewnia, Albrecht (Hg.): Neues vom heutigen Deutsch: Empirisch – methodisch – theoretisch. (Jahrbuch des Instituts für Deutsche Sprache 2018). Berlin, New York: De Gruyter 2019, S. 351–354.