

forTEXT

Literatur digital erforschen

DraCor: Drama Corpora Project

Textsammlungen · 24. Juni 2019 · Jan Horstmann

1. Kurzbeschreibung

DraCor bietet für eine Vielzahl von deutschsprachigen, russischen, spanischen, altgriechischen oder auch für Shakespeare-Dramen den zielgenauen Zugriff auf bestimmte Textuntermengen (wie etwa der gesprochene Text pro Figur; nur der Nebentext, nur Texte weiblicher Figuren etc.). DraCor generiert zudem automatisch Netzwerke (vgl. → [Netzwerkanalyse](#)), die figürliche Kopräsenzen anzeigen. Die Netzwerkdaten können Sie im CSV- oder → [Gephi](#)-kompatiblen GEXF-Format herunterladen und weiterverarbeiten. Eine API bietet zudem zahlreiche Möglichkeiten und zusätzliche Formate zur Weiterverarbeitung der vereinheitlichten Metadaten.

German Drama Corpus		Greek Drama Corpus		Russian Drama Corpus	
Number of plays	472	Number of plays	39	Number of plays	163
person + personGrp	10,683 (male: 7326, female: 2286)	person + personGrp	438 (male: 77, female: 38)	person + personGrp	2,861 (male: 2011, female: 658)
text	8,300,957 tokens	text	321,599 tokens	text	1,781,108 tokens
sp	328,623 (7,927,345 tokens)	sp	15,693 (320,968 tokens)	sp	89,963 (1,687,200 tokens)
stage	159,852 (1,012,330 tokens)	stage	11 (20 tokens)	stage	38,485 (165,220 tokens)
Last update	19.6.2019, 21:42:28	Last update	22.6.2019, 08:00:19	Last update	19.6.2019, 21:29:46
Shakespeare Drama Corpus		Spanish Drama Corpus			
Number of plays	37	Number of plays	25		
person + personGrp	1,433	person + personGrp	567		
text	908,286 tokens	text	444,647 tokens		
sp	31,066 (876,744 tokens)	sp	23,601 (388,883 tokens)		
stage	10,450 (41,230 tokens)	stage	8,200 (72,719 tokens)		
Last update	23.6.2019, 23:40:19	Last update	19.6.2019, 21:16:59		

Abb. 1: Startseite von DraCor mit den fünf Korpora: deutsche, griechische, russische, spanische und Shakespeare-Dramen

Steckbrief

- <https://dracor.org>
- herausgegeben durch das Centre for Digital Humanities an der HSE Moskau und die Universität Potsdam; entstanden aus dem mittlerweile eingestellten [DLINA](#)-Projekt und → [Ezlinavis](#); derzeit noch in der Alphaphase online (offizielles Erscheinungsdatum steht noch nicht fest)
- derzeit 472 deutschsprachige, 155 russische, 39 griechische, 25 spanische und 37 Shakespeare-Dramen mit insg. knapp 12 Mio. Tokens, demnächst erscheint ein schwedisches Dramenkorporus,

mehr ist in Planung

- strukturelle (Akte, Szenen, Auftritte, Szenenbeschreibungen etc.), semantische (z. B. Gender der jeweiligen Figuren) und weitere Metadaten: Autor, Entstehungsdatum, Veröffentlichung und Premiere des Stücks, Figurennamen, [Link](#) zur Textquelle
- zwischen 1730–1930 erschienene vollständige Dramentexte (keine Fragmente), insb. übernommen vom → [TextGrid Repository](#)

2. Anwendungsbeispiel

Sie untersuchen in einem Forschungsprojekt Figurenkonstellationen in Dramen des Sturm und Drang und interessieren sich für das Verhältnis von weiblichen und männlichen Figuren. Mit DraCor erhalten Sie mit wenigen Klicks [automatisch](#) Figurennetzwerke zu den Dramen und können diese miteinander vergleichen. Sie sehen, welche Figuren in einzelnen Akten gemeinsam auftreten und wie groß der Anteil einer jeden Figur am gesprochenen Text ist. Die Texte können Sie per API in vielerlei Formen (bspw. Text pro Figur inkl. Genderangabe, nur Nebentexte/Regieanweisungen, nur gesprochene Texte etc.) als [TXT](#)-, [RDF](#)-, [CSV](#)-, [GEXF](#)- oder [JSON](#)-Datei herunterladen und weiterverwenden. 4

3. Diskussion

3.1 Kann ich DraCor für wissenschaftliche Arbeiten nutzen?

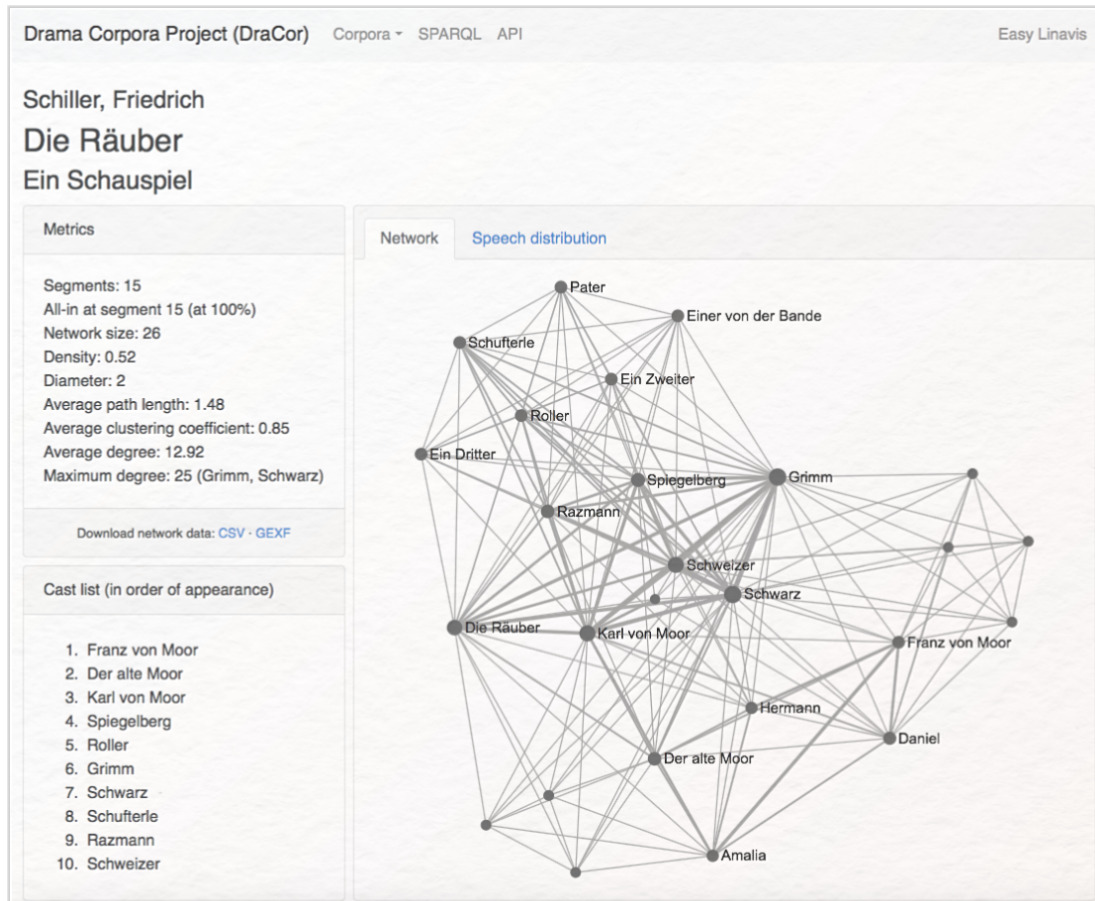
Ja. Den deutschen Texten liegt insbesondere das → [TextGrid Repository](#) mit dem dort erarbeiteten [TEI-XML](#) zugrunde, das korrigiert und angereichert wurde. Zudem werden Texte von Gutenberg-DE und Wikisource integriert und formal angeglichen. Außerdem werden zukünftig weitere Dramen aus dem → [Deutschen Textarchiv \(DTA\)](#) integriert. Für die zusätzlichen Dramen aus Gutenberg-DE und Wikisource ist das Aufnahmekriterium, dass es eine bekannte Druckvorlage gibt, die als Vergleichsgrundlage herangezogen wird. Die russischen Dramen stammen aus der Wikisource, der Russian Virtual Library, der Online Library of Alexei Komarov und der Maxim Moshkov's Library, die Shakespeare-Dramen aus der Folger Shakespeare Library, die spanischen Dramen aus der Biblioteca Electrónica Textual del Teatro en Español (BETTE), die altgriechischen aus der Perseus Digital Library. Eigene wissenschaftliche Anwendungen von DraCor dokumentiert das dlina-Team [hier](#). Es werden nur Texte in DraCor aufgenommen, die erstens keine Fragmente und zweitens in den Jahren 1730–1930 erschienen sind. Dabei möchte man eher ein repräsentatives als ein riesiges Korpus aufbauen, das Ziel sind etwa 1000 deutschsprachige und mehr als 500 russische Stücke. Zusätzlich zum Shakespearekorpus sind derzeit zwei weitere autorspezifische Korpora zu Henrik Ibsen und Ludvig Holberg geplant. 5

Das Projekt ist nicht primär als Repositorium für Volltexte gedacht, auch wenn Sie die gesamte Textsammlung mit einer Zeile in der [Commandline](#) downloaden können; eine Anleitung dazu finden Sie [hier](#) am Ende von „Corpus Description“. Ein Gesamtdownload der jeweiligen [XML-Datei](#) ist mit Rechtsklick auf „TEI version“ zwar möglich, in Textsammlungen wie → [TextGrid Repository](#) (das in DraCor zu jedem Drama verlinkt ist) oder dem → [Deutschen Textarchiv \(DTA\)](#) jedoch etwas benutzerfreundlicher gestaltet. Das Ziel ist vielmehr, einen Ausgangspunkt für verschiedene digitale Projekte bereitzustellen, für die man sich die Daten in der jeweils präferierten Form über die API-Abfrage herunterladen und weiterverarbeiten kann. Gemäß dieses Grundsatzes sind alle Daten auf DraCor für jeden zugänglich. Bei Bedarf ist es zudem möglich, die gesamte Plattform lokal zu installieren (s. [hier](#)). Projekte, die auf die DraCor-Daten zugreifen und andere Netzwerke erstellen als DraCor selbst sind z. B. [Shiny DraCor](#) oder [webweb](#) (Sie haben hier weitere Möglichkeiten, die Korpora zu explorieren bzw. die Visualisierungen zu beeinflussen). 6

Strukturelle Metadaten wie auch in das TEI übernommene Seitenumbrüche ermöglichen die Orientierung in den Texten. Die Metadaten sind auf vereinheitlichtem Niveau, im Shakespeare-Korpus fehlen bislang lediglich die Genderannotationen für Figuren. Geplant ist hierbei die Möglichkeit, dass Benutzer*innen weitere Annotationen über GitHub beisteuern können.

7

3.2 Wie benutzerfreundlich ist die Arbeit mit DraCor?



8

Abb. 2: Netzwerk für Schillers Drama *Die Räuber* (1781) in DraCor

Die gesuchten Texte finden Sie in DraCor sehr schnell (s. Abschnitt 4 dieses Beitrags). Ein Klick auf den Titel des jeweiligen Dramas öffnet das entsprechende Netzwerk (vgl. Abb. 2), in das Sie beliebig hineinzoomen können. Um zu verdeutlichen, dass eine Netzwerkvisualisierung immer eine Interpretation von Daten ist und z. B. die Position der Figurennamen auf der Fläche keine Rolle spielt (im Gegensatz zu ihren Verbindungen), sieht das jeweilige Netzwerk zu einem Stück bei jedem Aufruf etwas anders aus (vgl. → [Textvisualisierung](#)).

9

Klicken Sie oben auf den Link „API“, gelangen Sie zu den zahlreichen Möglichkeiten, mehrere oder einzelne Texte (bzw. auch Textteile) in unterschiedlichen Formen und Formaten zu exportieren (verschiedene Metadaten, TEI, Figuren, verschiedene Listen). Diese Funktionen können eine erhebliche Zeitersparnis bedeuten, wenn Sie an bestimmten Aspekten der Dramentexte interessiert sind.

10

In den Grundfunktionen (Textsuche, Netzwerkgenerierung, Download von Netzwerkdaten als CSV-Datei oder im → [Gephi](#)-kompatiblen GEXF) funktioniert DraCor bereits sehr intuitiv. Die Arbeit mit der API-Seite (vgl. Abb. 3) kann zu Beginn eine Herausforderung darstellen, wird mit der Zeit aber handhabbar.

11

Parameters

GET /corpora/{corpusname}/play/{playname}/spoken-text Get spoken text of a play (excluding stage directions)

Cancel

Name	Description
corpusname * required string (path)	ger
playname * required string (path)	schiller-die-raeuber
gender string (query)	FEMALE

Execute

Responses

Code	Description	Links
200	Returns plain text document text/plain Controls Accept header. Example Value Schema string	No links

Abb. 3: Export der gesprochenen Texte weiblicher Figuren (Amalia) in Schillers *Die Räuber* (1781)

Auf der GitHub-Seite von DraCor gibt es für das deutschsprachige Dramenkorpus ein **Wiki**, das **13** Notizen zu einzelnen Dramen enthält (z. B.: „Zwischen Franz und Karl wird wegen IV/2 eine Relation hergestellt. Aber eigentlich begegnen sie sich gar nicht“ für *Die Räuber*) sowie eine Dokumentation der Veränderungen im Zuge des Imports von TextGrid bereitstellt. Über GitHub bzw. die API-Seite („contact the developer“) können Sie die Herausgeber von DraCor zudem per Email kontaktieren.

Für Texte, die Sie nicht in DraCor finden, können Sie sehr leicht selbst ein entsprechendes **14** Netzwerk aufbauen. Das Tool → **Ezlinavis** (bzw. Easy Linavis) existierte bereits vor DraCor und ist auf der DraCor-Seite oben rechts verlinkt. Es bietet die Möglichkeit, Netzwerke zügig und gänzlich ohne technische Vorkenntnisse zu erstellen.

4. Wie funktioniert die Textsuche in DraCor?

Auf der Startseite entscheiden Sie sich für eine der angebotenen Textsammlungen, die Sie mit **15** einem Klick öffnen können. Dort haben Sie die Möglichkeit, die aufgeführten Dramen per Eingabe im Suchfeld oben zu suchen oder nach Autor*in, Titel, Netzwerkgröße, Erscheinungsjahr (generiert aus Entstehungsjahr, Premierenjahr und Druckjahr) oder Textquelle neu zu sortieren, indem Sie auf den entsprechenden Reiter klicken (s. Abb. 4). Die Suche einzelner Texte über die Suchleiste funktioniert sehr schnell. Ein Klick auf den Titel des jeweiligen Dramas öffnet das entsprechende Netzwerk.

Author ▼	Title	Network Size	Year (normalized)	Source
Alberti, Konrad pnd:116009926	Brot! Ein soziales Schauspiel in fünf Akten Wikidata: Q51370104	49	1888 🗨️ 1888 📅 1888	TextGrid Repository (TEI version)
Alberti, Konrad pnd:116009926	Im Suff Naturalistische Spital-Katastrophe in zwei Vorgängen und einem Nachgang Wikidata: Q51370930	14	1890 📅 1890	TextGrid Repository (TEI version)
Anzengruber, Ludwig pnd:11850357X	Das vierte Gebot Volksstück in vier Akten Wikidata: Q1171707	32	1877 🗨️ 1877 📅 1878	TextGrid Repository (TEI version)
Anzengruber, Ludwig pnd:11850357X	Der Wissenswurm Bauernkomödie mit Gesang in drei Akten Wikidata: Q1194083	15	1874 🗨️ 1874 📅 1874	TextGrid Repository (TEI version)
Anzengruber, Ludwig pnd:11850357X	Der Meineidbauer Volksstück mit Gesang in drei Akten Wikidata: Q1195757	22	1871 🗨️ 1871 📅 1871	TextGrid Repository (TEI version)

Abb. 4: Textsuche im German Drama Corpus von DraCor

Um sich die von Ihnen präferierte Form des Textes (bspw. die weibliche Figurenrede) downloaden zu können, gehen Sie auf die API-Seite, klicken (in diesem Fall bei „Get spoken text of a play (excluding stage directions)“) auf „GET“ und dann auf „Try it out“ und füllen die entsprechende Suchmaske aus. Die für eine erfolgreiche API-Abfrage benötigte Form des „corpusname“ sowie des „playname“ folgt einem Schema, das Sie in jeder URL der Dramennetzwerke wiederfinden können. Das German Drama Corpus hat bspw. das Kürzel „ger“, Schillers *Räuber* das Kürzel „schiller-die-raeuber“, woraus sich die URL dracor.org/ger/schiller-die-raeuber ergibt. Nach einem Klick auf „Execute“ erhalten Sie die angeforderten Daten zum Download.

17

5. Nachweise und weiterführende Literatur

DraCor: <https://dracor.org/>.

GitHub-Seite von GerDraCor: <https://github.com/dracor-org/gerdracor>.

Fischer, Frank, Tatyana Orlova, German Palchikov, Irina Pavlova, Daniil Skorinkin und Natasha Tyshkevich (2017): *Introducing RusDraCor. A TEI-encoded Russian Drama Corpus for the Digital Literary Studies*. <http://dlina.github.io/presentations/2017-spb/#/> [Zugriff: 15. Mai 2019].

Fischer, Frank, Ingo Börner, Mathias Göbel, Angelika Hecht, Christopher Kittel, Carsten Milling und Peer Trilcke (2019): „Programmable Corpora – Die digitale Literaturwissenschaft zwischen Forschung und Infrastruktur am Beispiel von DraCor“. In: *DHd 2019. Digital Humanities: multimedial & multimodal. Konferenzabstracts*, 194–197. DOI: [10.5281/zenodo.2596095](https://doi.org/10.5281/zenodo.2596095).

Skorinkin, Daniil, Frank Fischer und German Palchikov (2018): „Building a corpus for the quantitative research of russian drama: composition, structure, case studies“. In: *Computational Linguistics and*

Intellectual Technologies: Proceedings of the International Conference "Dialogue 2018". <http://www.dialog-21.ru/media/4332/skorinkind.pdf> [Zugriff: 15. Mai 2019].

Trilcke, Peer, Frank Fischer, Mathias Göbel und Dario Kampkaspar (2016): „Theatre Plays as ‘Small Worlds’? Network Data on the History and Typology of German Drama, 1730–1930“. In: *Digital Humanities 2016: Conference Abstracts*. Jagiellonian University & Pedagogical University, Kraków, 385–387. <http://dh2016.adho.org/abstracts/360> [Zugriff: 15. Mai 2019].

Verwandte Beiträge

Ressourcen

[Deutsches Textarchiv \(DTA\)](#)

[TextGrid Repository](#)

Routinen

[Netzwerkanalyse](#)

[Netzwerkanalyse mit Gephi](#)

[Textvisualisierung](#)

Tools

[Ezlinavis](#)

[Gephi](#)