

klemmHA

esteeschwarz

2022-09-18

Contents

1. einleitung	1
2. klemm	2
2.1 bio	2
2.2 stück	2
2.3 verortung, rezeption	2
3. dokumentation aufbereitung	2
3.1. transkription	2
3.2 TEI-auszeichnung	2
4. some visualisations	4
5. REF:	8

1. einleitung

precondition: Im Seminar “Digitale Dramenanalyse” des SS2022 (Frank Fischer, FUB) war den Teilnehmern freigestellt, sich während des Semester einem Stück aus der DraCor Einakterdatenbank (Çakir and Fischer 2022) & (einakter and Klemm 2022) zu widmen. Das sollte den Prozess umfassen der Transkription und Auszeichnung nach TEI Standard. Es wurden 9 Stücke selbst gewählt und, teilweise im Team, daran gearbeitet. Über den Zwischenstand wurden Referate zu Stück und Autor sowie evtl. schon während des Seminars kennengelernter Analysen vorgestellt. Ich habe Christian Gottlieb Klemms “Der Besuch” ausgesucht, bin bei der Auswahl aber nur nach Sprache **Deutsch** und einem allein zu bewältigenden Umfang gegangen. Ich werde in dieser Arbeit kurz dokumentieren, wie die Aufgaben umgesetzt wurden, welche Methoden dabei zum Einsatz kamen und einen Ausblick auf literaturwissenschaftliche Analysen geben, die mit den gelernten Verfahren möglich sind.

2. klemm

2.1 bio

Klemm wurde 1736 in Schwarzenberg/Erzgebirge geboren und lebte bis 1802 in Leipzig, Jena, Frankfurt, ab 1759 in Wien. Dort wurde er Herausgeber verschiedener Wochenblätter (Die Welt, Der österreichische Patriot, Beyträge zum deutschen Theater, Wiener Allerlei) in denen neben theaterkritischen und -theoretischen Beiträgen teilweise auch seine Stücke erschienen. (Killy 2009)

2.2 stück

2.3 verortung, rezeption

3. dokumentation aufbereitung

3.1. transkription

Als Textgrundlage des im folgenden *Rohtext* genannten Materials dienten die von google books öffentlich zur Verfügung gestellten scans der Erstausgabe von 1765 (A1765_Q) (google and Klemm 1765), die grösztenteils auf den in der Österreichischen Nationalbibliothek schon vorhandenen Digitalisierungen (ÖNB and Klemm 1765) basieren. Diese wurden auf wikisource und hiermit auf einer platform zur kollaborativen Erstellung von u.a. Transkriptionen (vom Seminarleiter) weiterhin zugänglich gemacht und durchliefen nun den Prozess von Transkription und zweifacher unabhängiger Korrektur (wikisource and Klemm 1765). Zur Transkription diente mir im ersten Stadium eine zu 90% verlässliche OCR (optical character recognition) mittels TRANSKRIBUS (read-coop 2022). Benutzt wurde das frei verfügbare OCR-model “ONB_Newseye_GT_M1+”, das zufriedenstellende Ergebnisse lieferte – allerdings nur mit der Ausgabe von 1767 (A1767_Q), die ich als scan in der Digitalen Sammlung der Münchner Staatsbibliothek (MDZ and Klemm 1767) gefunden hatte. Da sich diese, wie ich bald feststellte, doch in einigen Punkten von der Erstausgabe unterschied, wurde der Rohtext auf Wikisource also :nur unter Zuhilfenahme: dieser OCR-Fassung erstellt, d.h. im Einzelnen:

1. OCR (A1767_Q)
 1. Korrektur desselben noch in TRANSKRIBUS selbst
 2. Resultat Rohtext (A1767_R)
2. Abgleich A1767_R mit A1765_Q in der Bearbeitungsmaske von Wikisource
 1. Korrektur bzw. Anpassung des auf diese Weise erstellten Rohtext der Erstausgabe (A1765_R)
 2. upload des Rohtextes
3. zwei Korrekturdurchgänge
 1. die von weiteren Personen auf Wikisource durchgeführt wurden
 2. dabei wurde A1765_R bereits mit Szenenüberschriften und Sprechermarkierungen versehen

3.2 TEI-auszeichnung

Der grosze Teil der weiteren Arbeit bestand in der ersten vereinfachten Auszeichnung nach TEI-Standard. Ich habe hierzu ein umfassendes script in der Programmiersprache R (“RStudio IDE” 2022) entworfen, das in A1765_R mittels der Suche mit regulären Ausdrücken (z.B.(regexr and Klemm 2022)) Szenenwechsel, Sprecherbezeichnungen und Regieanweisungen nach einigem Aufwand zu finden vermochte¹ um diese

¹dieser Aufwand wäre um einiges geringer ausgefallen, wenn ich mich, statt mit dem Text im Format .txt (plaintext) zu arbeiten, mit der bereits basis-HTML formatierten (durch Fett/Kursivdruck zb. unterschiedene Passagen für Sprecher, Regieanweisungen und Szenenwechsel) bedient hätte.

schließlich in HTML-tags zu verpacken: d.h. die einzelnen verschiedenen Elemente des Textes nach einem Standard zu markieren, der uns vorlag und bereits als Schema zur Validierung im Editor verwendet werden konnte (drama and dracor 2022). Es handelte sich dabei noch nicht um die finale, schemagerechte Version. Diese wurde weiterhin vor der Integration des Stückes in das dracor-Korpus (Fischer et al. 2019) vom Seminarleiter auf Grundlage der vereinfachten Version erstellt.

3.2.1.beispiel der auszeichnung:

```
library(rvest)
library(stringi)
src<-"https://de.wikisource.org/wiki/Der_Besuch_(Klemm)"
dta1<-read_html(src)
xpathkl<-"//*[@id="mw-content-text"]/div[1]/div[2] '
#xpath copied from browser developer tools (safari)
#html_nodes(dta1,xpath = xpathkl)
txt<-html_nodes(dta1,xpath = xpathkl) %>%html_text()
regx1<-intToUtf8(926)
repl1<-"@begin"
txtm1<-gsub(regx1,repl1,txt,perl = T)
txt<-txtm1
# txt<-ground(txt)
repl1<-'<div type="scene"><head>\\1 \\2</head><stage>'
regx1<-"(Erster|Zweyter|Dritter|Vierter|Fünfter|Sechster|Siebenter|Achter|Neunter) (Auftritt.)"
#stri_extract_all_regex(txtm,regx3)
m<-gregexec(regx1,txt,perl = T)
#regmatches(txt,m)
txtm1<-gsub(regx1,repl1,txt,perl = T)
#wks.
cat(stri_sub(txtm1,1,500),"[...]" )

##
## [@begin] Der österreichische Patriot.
## Drey und vier und siebenzigstes Stück.
## Freytags den 14. Junii. 1765.
## Der
## Besuch,
## Ein Lustspiel
## in einem Aufzuge.
##
## Personen.
## Celimene, Finette, Erast, Chlorinde, Cydalise, Damis.
##
## <div type="scene"><head>Erster Auftritt.</head><stage>
## Celimene, Finette.
## Celimene.
## Siehst du nun, wie behutsam man mit den Mannspersonen
## umgehen muß? Wie lange man sie für
## rechtschaffen, für zärtlich, für beständig halten
## kann? und auf einmal ist der Verräther da. Alle
## sind Verrät [...]
```

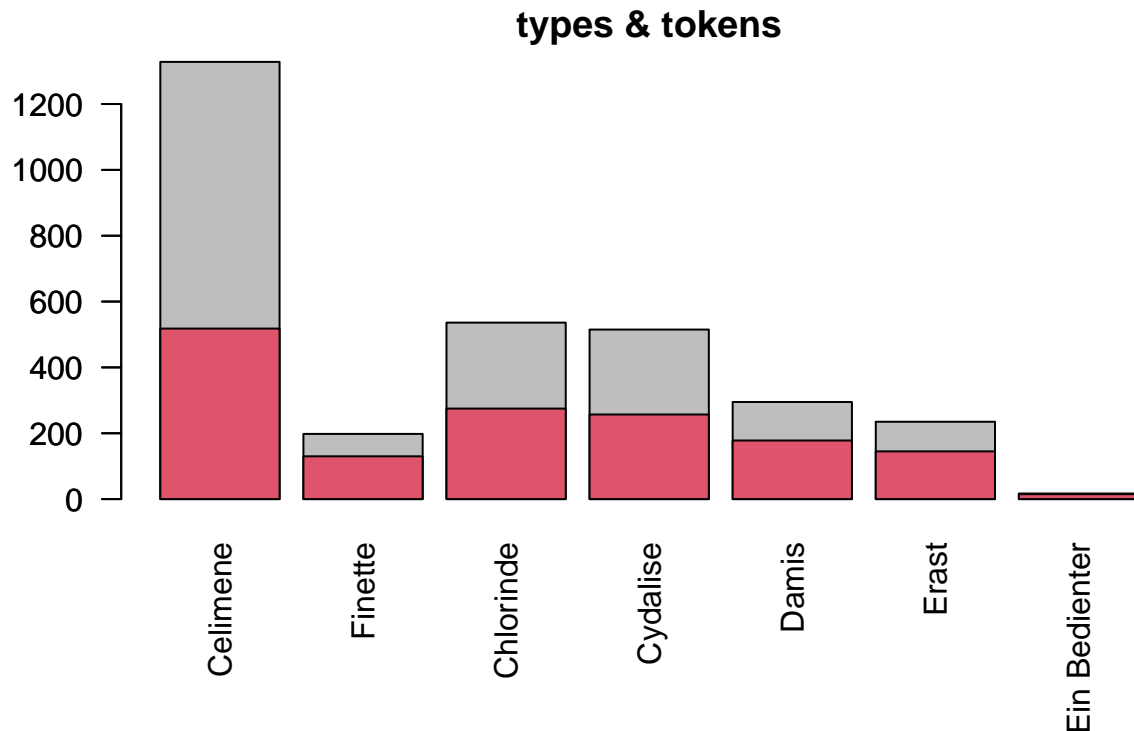
3.2.2 finales TEI

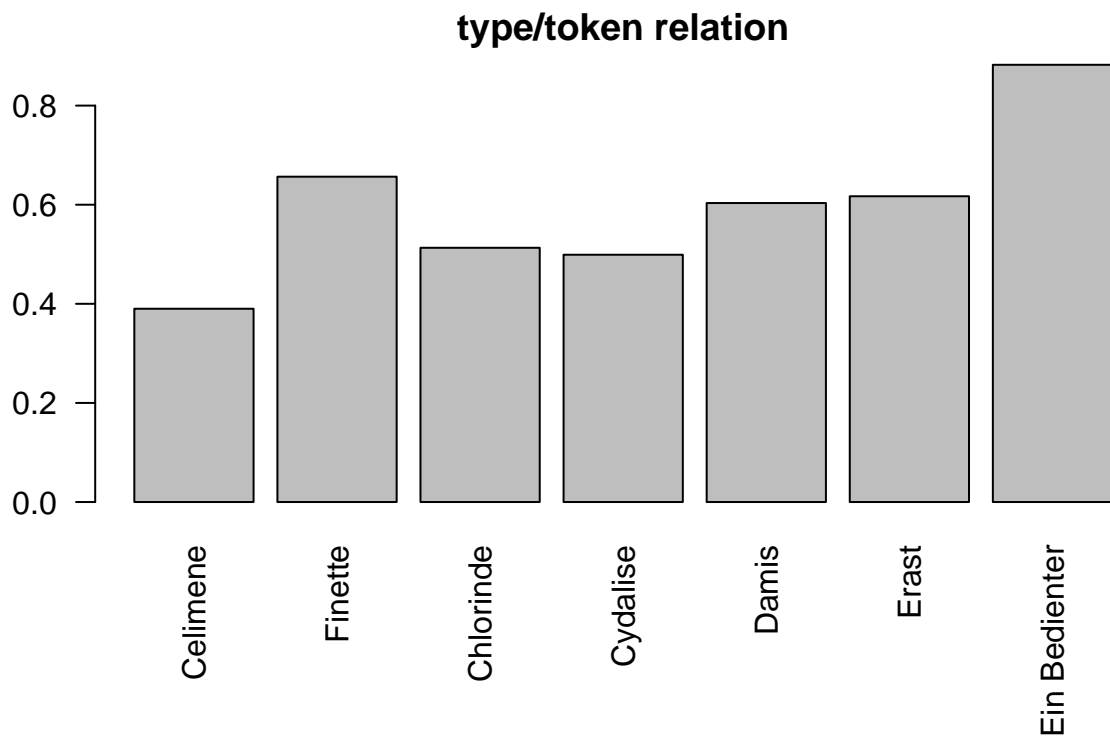
Ziel der Bearbeitung war, das Stück auf DraCor für weitere Analysen in standardisierten Formaten (TEI, RDF, JSON sowie TXT und GRAPHML/GEFX Netzwerkdaten) zur Verfügung stellen zu können. Hierzu war es nötig, zuerst die Textgrundlage und erste Auszeichnung herzustellen. Mit Verfügbarkeit des Dramas im schemagerechten TEI Austauschformat, das im Bereich der digital humanities oder der Wissenschaft seit Jahren als Standard etabliert ist (TEI 1987), wenn es um die Bereitstellung/ Katalogisierung, Archivierung bzw. eben den Austausch von Daten oder Metadaten geht, können nun weitreichende Analysen, vergleichende Studien und weiteres angestellt werden. Ich werde im Folgenden einen kleinen Ausschnitt dessen vorstellen, was möglich, aber in jedem Fall erweiterbar ist.

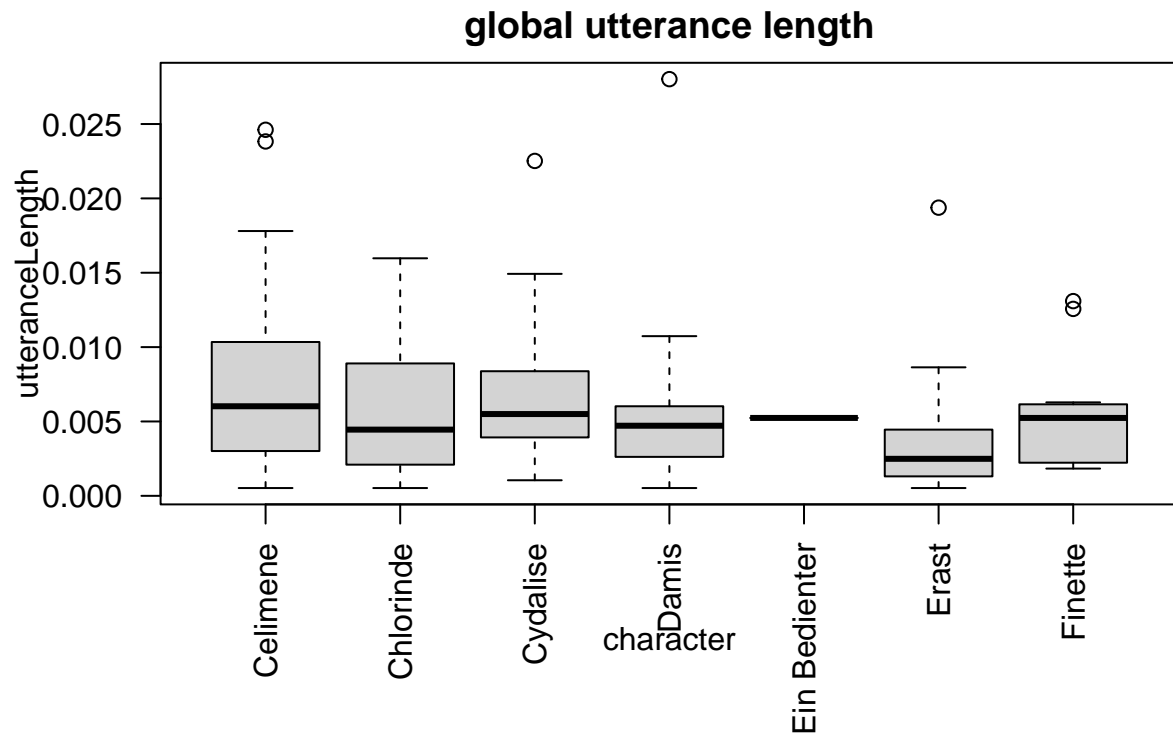
4. some visualisations

basics...

```
## load TEI from: https://dracor.org/api/corpora/ger/play/klemm-der-besuch/tei
```



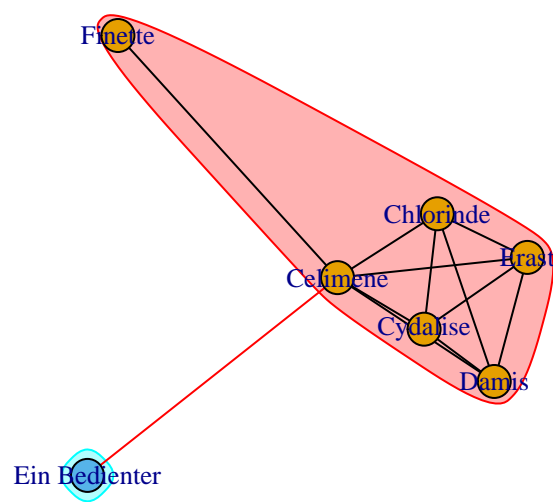




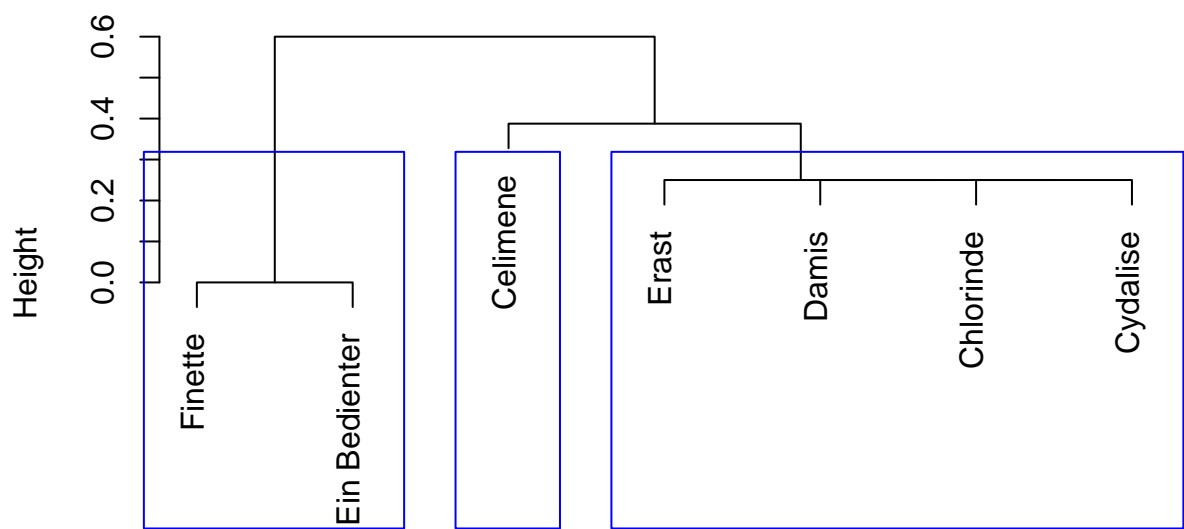
4.1.1 further...

vacat

4.2 network



Cluster Dendrogram



hclust(*, "average")
7

networks graph cf. (Poggel 2022),(Franceschet 2022)

4.3 sources comparison (LERA)

LERA output cf. (Pöckelmann et al. 2022)

5. REF:

- Çakir, Dilan Canan, and Frank Fischer. 2022. “Dramatische Metadaten - Die Datenbank Deutschsprachiger Einakter 1740–1850,” March. <https://zenodo.org/record/6327977>.
- drama, and dracor. 2022. “Dracor Xml Model.” http://www.tei-c.org/release/xml/tei/custom/schema/relaxng/tei_drama.rng.
- einakter, and Christian Gottlob Klemm. 2022. “Einakter: Klemm, Christian Gottlob: Der Besuch.” <https://einakter.dracor.org/klemm-der-besuch>.
- Fischer, Frank, Ingo Börner, Mathias Göbel, Angelika Hechtel, Christopher Kittel, Carsten Milling, and Peer Trilcke. 2019. “Programmable Corpora: Introducing DraCor, an Infrastructure for the Research on European Drama,” July. <https://zenodo.org/record/4284002>.
- Franceschet, Massimo. 2022. “Communities in Igraph.” <https://users.dimi.uniud.it/~massimo.franceschet/R/communities.html>.
- google, and Christian Gottlob Klemm. 1765. *Klemm: Der Österreichische Patriot: eine Wochenschrift* / *google books*. Gedruckt u. zu finden bey Georg Ludwig Schulz. https://books.google.de/books?id=P2nXk3Ib9LkC&pg=PA577&redir_esc=y#v=onepage&q&f=false.
- Killy, Walter. 2009. *Killy Literaturlexikon, Band 6 Huh - Kräf*. De Gruyter. <https://doi.org/10.1515/9783110213942>.
- MDZ, and Christian Gottlob Klemm. 1767. “Klemm, Christian Gottlob: Christian Gottlieb Klemm’s Beyträge Zum Deutschen Theater. 1, Bild 122 von 620 | MDZ.” <https://www.digitale-sammlungen.de/de/view/bsb10112352?page=122,123>.
- ÖNB, and Christian Gottlob Klemm. 1765. “Klemm: Der Österreichische Patriot: Eine Wochenschrift. ÖNB.” https://digital.onb.ac.at/OnbViewer/viewer.faces?doc=ABO_%2BZ184248300.
- Pöckelmann, Marcus, André Medek, Jörg Ritter, and Paul Molitor. 2022. “LERA - an Interactive Platform for Synoptical Representations of Multiple Text Witnesses.” *Digital Scholarship in the Humanities (DSH)*. <https://lera.uzi.uni-halle.de/about?lang=de>.
- Poggel, Lisa. 2022. “10. Sitzung: Netzwerkanalyse.” https://lipogg.github.io/r-tutorium-website/tutorium10_11072022_netzwerke.html.
- read-coop. 2022. “Transkribus.” <https://readcoop.eu/transkribus/>.
- regexr, and Christian Gottlob Klemm. 2022. “Regexr: Learn, Build, & Test RegEx.” <https://regexr.com/6mr4b>.
- “RStudio IDE.” 2022. <https://www.rstudio.com/products/rstudio/download/>.
- TEI. 1987. “History – TEI: Text Encoding Initiative.” <https://tei-c.org/about/history/>.
- wikisource, and Christian Gottlob Klemm. 1765. “Der Besuch (Klemm) – Wikisource.” [https://de.wikisource.org/wiki/Der_Besuch_\(Klemm\)](https://de.wikisource.org/wiki/Der_Besuch_(Klemm)).