

Quantitative Dramenanalyse

Vortrag, Workshop, Diskussion

Nils Reiter

Institut für Digital Humanities, Universität zu Köln

07.01.20

Plan für heute

- ▶ Einführung
 - ▶ QuaDramA
 - ▶ Mixed methods
 - ▶ Operationalisierung
- ▶ Fallstudien
 - ▶ Figurenpräsenz
 - ▶ Gattungen
 - ▶ Redesemantiken
- ▶ Lab session
 - ▶ Programmiersprache R, Programmierumgebung RStudio
 - ▶ Bibliothek DramaAnalysis

QuaDramA, das Projekt

Ziele

QuaDramA = Quantitative Drama Analytics

- ▶ Analyse konstitutiver Eigenschaften von Figurentypen
- ▶ (Automatische) Erkennung von Figurentypen
- ▶ Entwicklung von Figurentypen dramanhistorisch und innerhalb des Dramas
- ▶ Interaktion(stypen) zwischen Figuren(typen)

Methodische und inhaltliche Ziele!

QuaDramA, das Projekt

Eckdaten

- ▶ Neuere Deutsche Literaturwissenschaft und Computerlinguistik
 - ▶ Computerlinguistik: Modellierung/Operationalisierung/Automatisierung menschlicher Sprache mit dem Computer
 - ▶ 'Flaggschiff'-Anwendung: Maschinelle Übersetzung

QuaDramA, das Projekt

Eckdaten

- ▶ Neuere Deutsche Literaturwissenschaft und Computerlinguistik
 - ▶ Computerlinguistik: Modellierung/Operationalisierung/Automatisierung menschlicher Sprache mit dem Computer
 - ▶ 'Flaggschiff'-Anwendung: Maschinelle Übersetzung
- ▶ Projektleiter
 - ▶ Marcus Willand (Uni Heidelberg, früher Stuttgart)
 - ▶ Nils Reiter (Unis Köln und Stuttgart)
- ▶ Zwei Doktoranden
 - ▶ Benjamin Krautter, NDL
 - ▶ Janis Pagel, CL

QuaDramA, das Projekt

Eckdaten

- ▶ Neuere Deutsche Literaturwissenschaft und Computerlinguistik
 - ▶ Computerlinguistik: Modellierung/Operationalisierung/Automatisierung menschlicher Sprache mit dem Computer
 - ▶ 'Flaggschiff'-Anwendung: Maschinelle Übersetzung
- ▶ Projektleiter
 - ▶ Marcus Willand (Uni Heidelberg, früher Stuttgart)
 - ▶ Nils Reiter (Unis Köln und Stuttgart)
- ▶ Zwei Doktoranden
 - ▶ Benjamin Krautter, NDL
 - ▶ Janis Pagel, CL
- ▶ Förderung durch VolkswagenStiftung seit 04/2017

QuaDramA, das Projekt

„Mixed methods“ in den Geisteswissenschaften?

Ziel ist die Förderung von „Hybrid-Projekten,“ in denen qualitativ-hermeneutische und digitale Ansätze integrativ kombiniert werden zu einem gemeinsamen, innovativen und für das Fach signifikanten Forschungsvorhaben.

Anhand einer spannenden wissenschaftlichen Frage soll die Verknüpfung und das Zusammenwirken von nicht-digitalen und digitalen Verfahren konzipiert werden. Der geplante Erkenntnisgewinn ist für beide Bereiche darzustellen. [...]

Ausschreibungstext

QuaDramA, das Projekt

Konsequenzen für uns

Kooperation und Interaktion auf Augenhöhe

- ▶ Getrennte Forschungsworkflows mit definierter Schnittstelle
- ▶ Keine GUIs: Literaturwissenschaftler programmieren selbst
- ▶ Enge Kooperation (= ständiger Austausch)
- ▶ Ergebnisse: Je eine Promotion in NDL und CL
 - ▶ Beide mit DH-Anteilen

QuaDramA, das Projekt

Ziele

Überprüfung von Hypothesen über dramatische Figuren im großen Maßstab

- ▶ Figurentypen
 - ▶ Nach Geschlecht, Handlung, Status, Alter, 'stock character' ...
 - ▶ Wie wird ein 'zärtlicher Vater' als zärtlicher Vater markiert?

QuaDramA, das Projekt

Ziele

Überprüfung von Hypothesen über dramatische Figuren im großen Maßstab

- ▶ Figurentypen
 - ▶ Nach Geschlecht, Handlung, Status, Alter, 'stock character' ...
 - ▶ Wie wird ein 'zärtlicher Vater' als zärtlicher Vater markiert?
- ▶ Relationen zwischen Figuren(typen)
 - ▶ Welche Gemeinsamkeiten gibt es zwischen Figuren(typen)?
 - ▶ Was sagen Figuren(typen) übereinander?

QuaDramA, das Projekt

Ziele

Überprüfung von Hypothesen über dramatische Figuren im großen Maßstab

- ▶ Figurentypen
 - ▶ Nach Geschlecht, Handlung, Status, Alter, 'stock character' ...
 - ▶ Wie wird ein 'zärtlicher Vater' als zärtlicher Vater markiert?
- ▶ Relationen zwischen Figuren(typen)
 - ▶ Welche Gemeinsamkeiten gibt es zwischen Figuren(typen)?
 - ▶ Was sagen Figuren(typen) übereinander?
- ▶ Entwicklung von Figuren(typen)
 - ▶ Innerhalb eines Dramas
 - ▶ Dramenhistorisch

QuaDramA, das Projekt

Computerlinguistische Ziele

- ▶ Strukturierte Textsorte
 - ▶ Dramen sind kein unstrukturierter Wortstrom

QuaDramA, das Projekt

Computerlinguistische Ziele

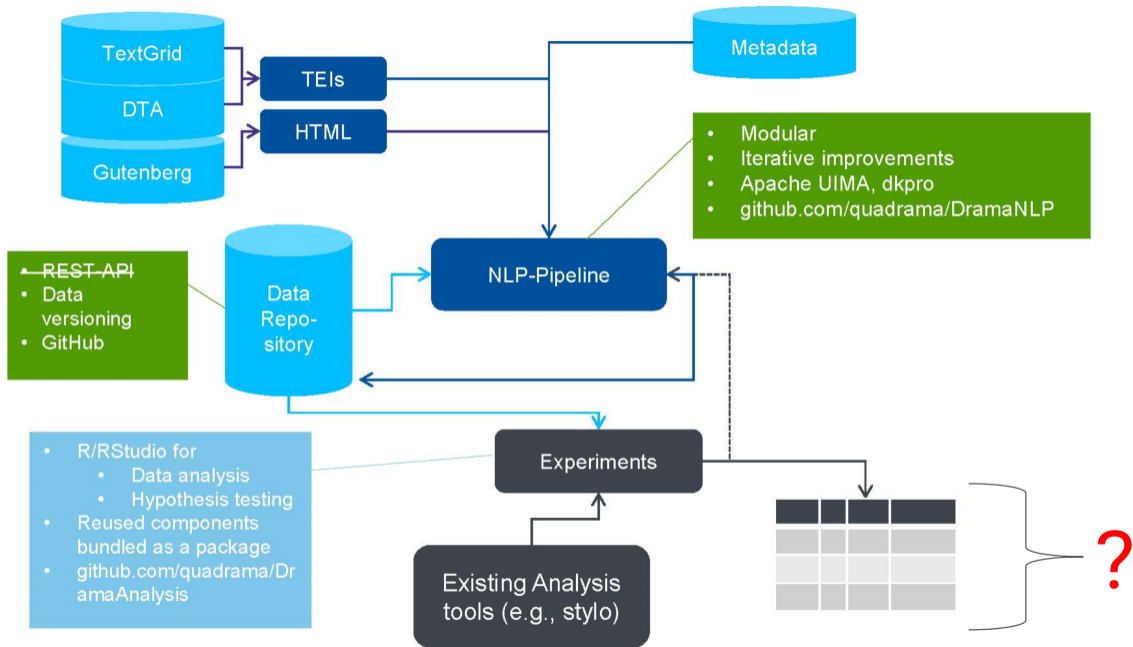
- ▶ Strukturierte Textsorte
 - ▶ Dramen sind kein unstrukturierter Wortstrom
- ▶ Domain adaptation
 - ▶ Texte sind literarisch, fiktional, historisch, ...
 - ▶ Einfluss auf alle meiste Ebenen der Textanalyse
 - ▶ Wortarten, Syntax, Semantik, ...

QuaDramA, das Projekt

Computerlinguistische Ziele

- ▶ Strukturierte Textsorte
 - ▶ Dramen sind kein unstrukturierter Wortstrom
- ▶ Domain adaptation
 - ▶ Texte sind literarisch, fiktional, historisch, ...
 - ▶ Einfluss auf alle meiste Ebenen der Textanalyse
 - ▶ Wortarten, Syntax, Semantik, ...
- ▶ Koreferenzresolution (Figurenreferenzen)
 - ▶ Eigennamen
 - ▶ Pronomen
 - ▶ Nominalphrasen

DER PRINZ. So wett' ich, Conti, daß es des Künstlers eigene Gebieterin ist. – *Indem der Maler das Bild umwendet.* Was seh' ich? Ihr Werk, Conti? oder das Werk meiner Phantasie? – Emilia Galotti!



Operationalisierung

Wie misst man ein theoretisches Konzept?

The concept of length is therefore fixed when the operations by which length is fixed are fixed: that is, the concept of length involves as much and nothing more than the set of operations by which length is determined.

Bridgman, 1927

Operationalisierung

Wie misst man ein theoretisches Konzept?

The concept of length is therefore fixed when the operations by which length is fixed are fixed: that is, the concept of length involves as much and nothing more than the set of operations by which length is determined.

Bridgman, 1927

- ▶ Definition durch Operationalisierung

Operationalisierung

Wie misst man ein theoretisches Konzept?

The concept of length is therefore fixed when the operations by which length is fixed are fixed: that is, the concept of length involves as much and nothing more than the set of operations by which length is determined.

Bridgman, 1927

- ▶ Definition durch Operationalisierung
 - ▶ Literaturwissenschaftliche Konzepte typischerweise nicht so definiert
- ▶ Direkte Operationalisierung: Konzept so (um)definieren dass es messbar wird

Operationalisierung

Wie misst man ein theoretisches Konzept?

The concept of length is therefore fixed when the operations by which length is fixed are fixed: that is, the concept of length involves as much and nothing more than the set of operations by which length is determined.

Bridgman, 1927

- ▶ Definition durch Operationalisierung
 - ▶ Literaturwissenschaftliche Konzepte typischerweise nicht so definiert
- ▶ Direkte Operationalisierung: Konzept so (um)definieren dass es messbar wird
- ▶ Approximative Operationalisierung: Mehrere messbare Phänomene die mit Zielphänomenen zu tun haben
 - ▶ 'instrumental variables' (Sack, 2011)

Operationalisierung

Wie misst man ein theoretisches Konzept?

The concept of length is therefore fixed when the operations by which length is fixed are fixed: that is, the concept of length involves as much and nothing more than the set of operations by which length is determined.

Bridgman, 1927

- ▶ Definition durch Operationalisierung
 - ▶ Literaturwissenschaftliche Konzepte typischerweise nicht so definiert
- ▶ Direkte Operationalisierung: Konzept so (um)definieren dass es messbar wird
- ▶ Approximative Operationalisierung: Mehrere messbare Phänomene die mit Zielphänomenen zu tun haben
 - ▶ 'instrumental variables' (Sack, 2011)
- ▶ Operationalisierung durch Annotation: Annotationsrichtlinien Form von Operationalisierung

Operationalisierung

- ▶ Schlüsselherausforderung in den *computational literary studies*
- ▶ Nur operationalisierte Konzepte können quantitativ untersucht werden

Operationalisierung

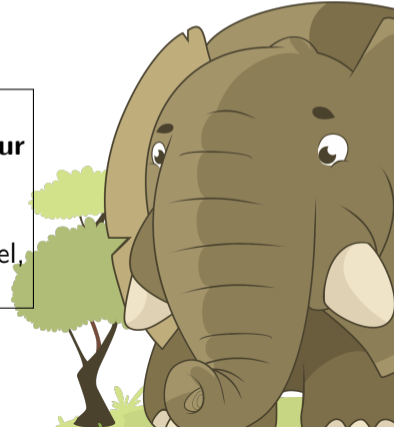
- ▶ Schlüsselherausforderung in den *computational literary studies*
- ▶ Nur operationalisierte Konzepte können quantitativ untersucht werden

Workshop:

Vom Phänomen zur Analyse – ein CRETA-Workshop zur reflektierten Operationalisierung in den DH

02./03. März, DHd 2020, Paderborn

Nora Ketschik, Benjamin Krautter, Sandra Murr, Janis Pagel,
Nils Reiter



Section 2

Emilia vs. Sara

Einleitung

- ▶ Gotthold Ephraim Lessing (1729-1781)
 - ▶ Miss Sara Sampson (1755)
 - ▶ Emilia Galotti (1772)
- ▶ Bürgerliche Trauerspiele
- ▶ Figurentypen
 - ▶ Töchter (Emilia, Sara)
 - ▶ Liebhaber (Prinz, Mellefont)
 - ▶ Väter (Odoardo, William)



Aktive Präsenz

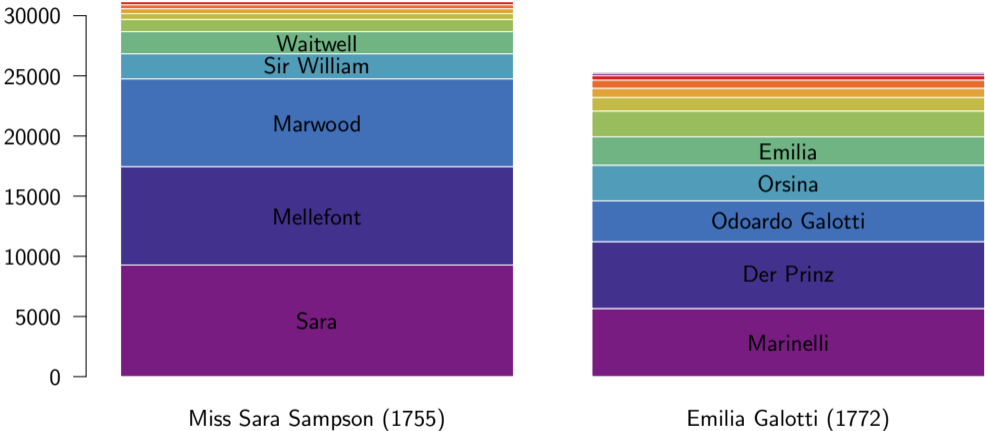


Abbildung: Redeanteile der Figuren in Lessing-Dramen

Aktive Präsenz

Nach Akt

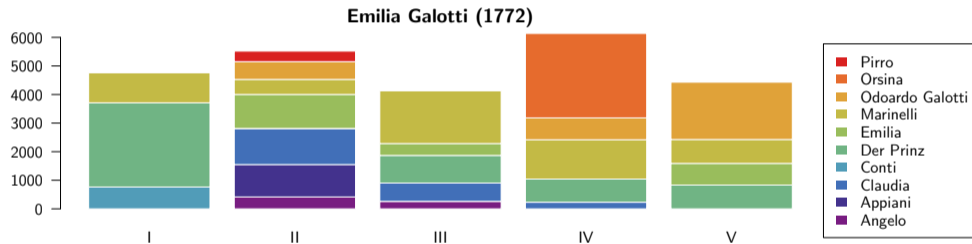


Abbildung: Redeanteile der Figuren in Akten

Aktive Präsenz

Nach Akt

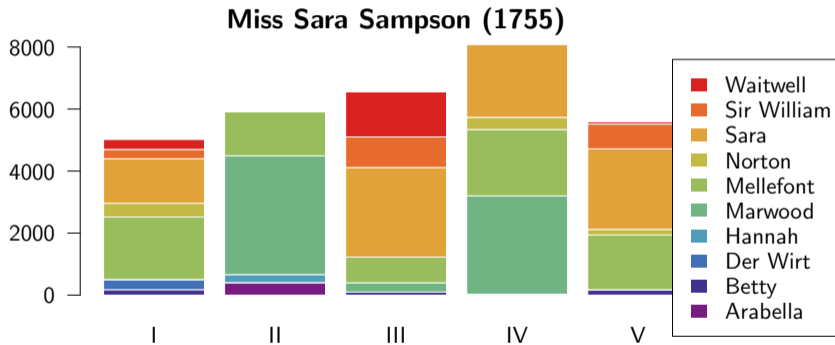


Abbildung: Redeanteile der Figuren in Akten

Aktive Präsenz

Sprechereignisse



Abbildung: Sprechereignisse einzelner Figuren im Verlauf der Dramen

Präsenz

- ▶ Aktive Präsenz: Figur ist auf Bühne
- ▶ Passive Präsenz: Figur ist nicht auf Bühne, aber anderweitig präsent
 - ▶ Rede anderer Figuren

Präsenz

- ▶ Aktive Präsenz: Figur ist auf Bühne
- ▶ Passive Präsenz: Figur ist nicht auf Bühne, aber anderweitig präsent
 - ▶ Rede anderer Figuren

Koreferenzannotationen

- ▶ Basierend auf manuellen Annotationen
- ▶ Erzeugt im QuaDramA-Projekt durch Hilfskräfte
- ▶ Automatische Auflösung ist in Arbeit ...

Passive Präsenz

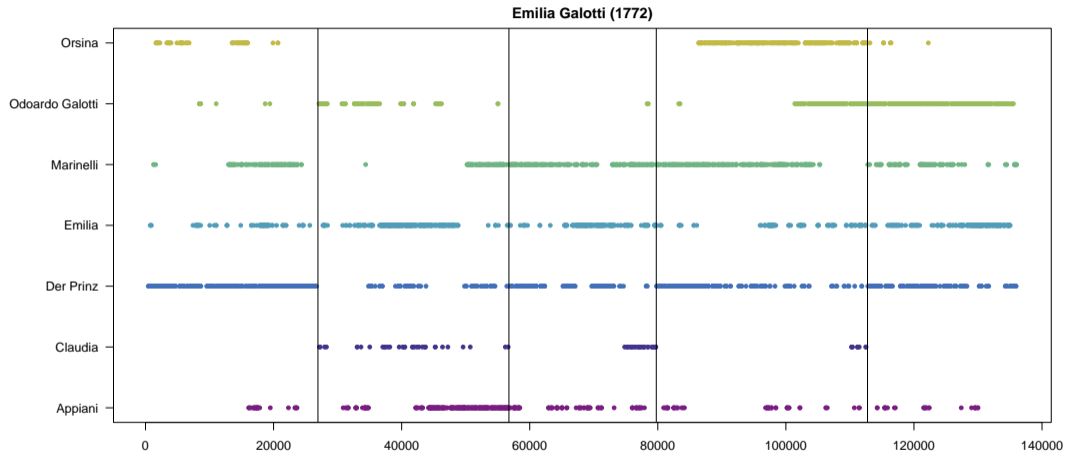


Abbildung: Erwähnungen einzelner Figuren im Verlauf des Dramas

Passive Präsenz

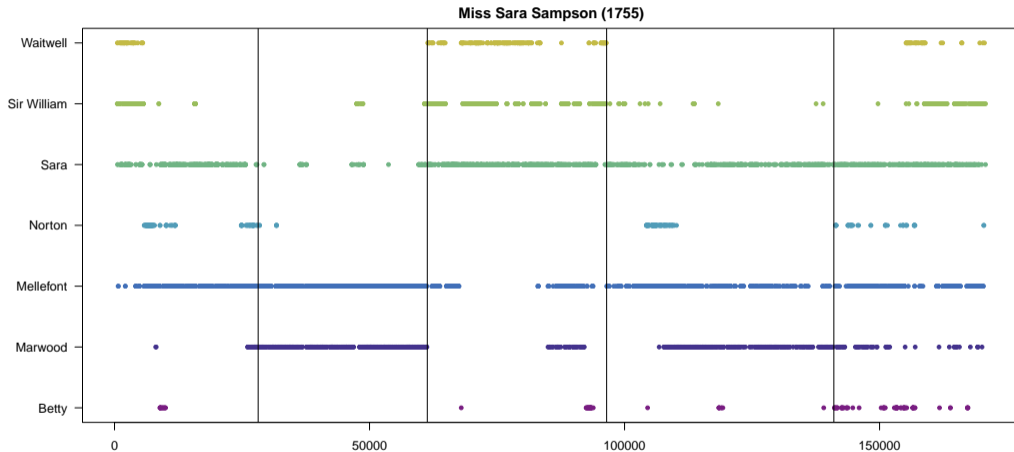


Abbildung: Erwähnungen einzelner Figuren im Verlauf des Dramas

Präsenz im Korpus

- ▶ Bühnenpräsenz kann nicht nur als 'aktive Präsenz' verstanden werden
- ▶ Vollständige Koreferenzannotation: Hohe Anforderungen!

Präsenz im Korpus

- ▶ Bühnenpräsenz kann nicht nur als 'aktive Präsenz' verstanden werden
- ▶ Vollständige Koreferenzannotation: Hohe Anforderungen!
- ▶ Netzwerkanalytische Ansätze bei Dramen arbeiten mit Szenen

Präsenz im Korpus

- ▶ Bühnenpräsenz kann nicht nur als 'aktive Präsenz' verstanden werden
- ▶ Vollständige Koreferenzannotation: Hohe Anforderungen!
- ▶ Netzwerkanalytische Ansätze bei Dramen arbeiten mit Szenen

Definitionen

Anteil der Szenen in denen Figur c spricht ($[0; 1]$)

$$active(c) = \frac{|\{s | c \text{ spricht in } s\}|}{scenes(d)}$$

Präsenz im Korpus

- ▶ Bühnenpräsenz kann nicht nur als 'aktive Präsenz' verstanden werden
- ▶ Vollständige Koreferenzannotation: Hohe Anforderungen!
- ▶ Netzwerkanalytische Ansätze bei Dramen arbeiten mit Szenen

Definitionen

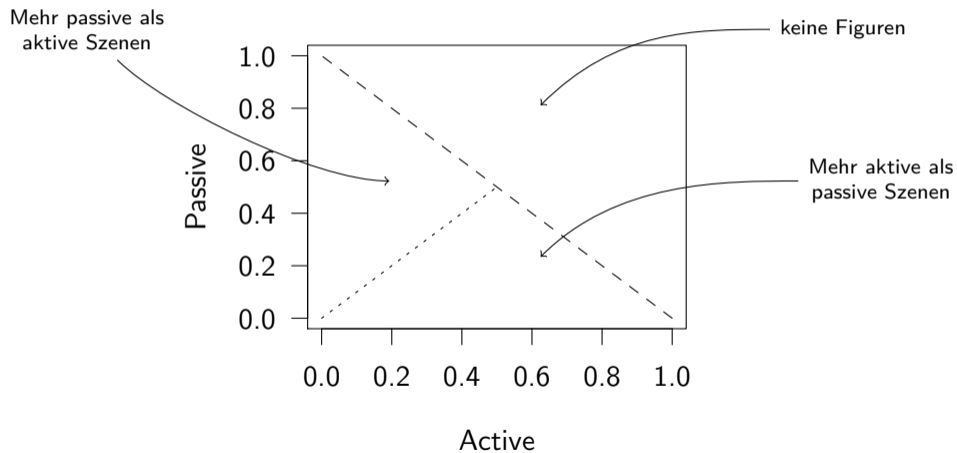
Anteil der Szenen in denen Figur c spricht ($[0; 1]$)

$$active(c) = \frac{|\{s | c \text{ spricht in } s\}|}{scenes(d)}$$

Anteil der Szenen, in denen Figur c erwähnt wird und *nicht* selbst spricht ($[0; 1]$)

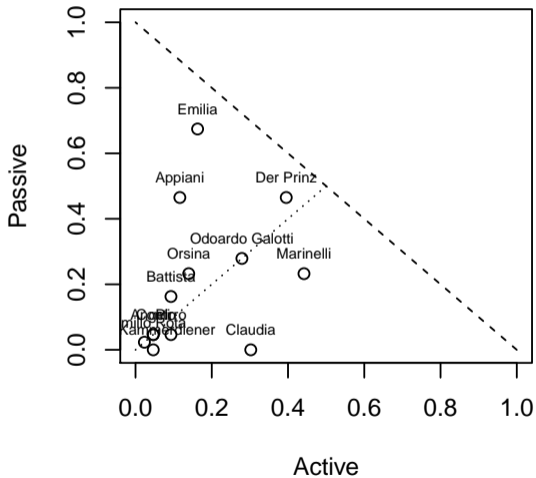
$$passive(c) = \frac{|\{s | c \text{ wird erwähnt in } s \wedge \neg c \text{ spricht in } s\}|}{scenes(d)}$$

Präsenz als Abbildung

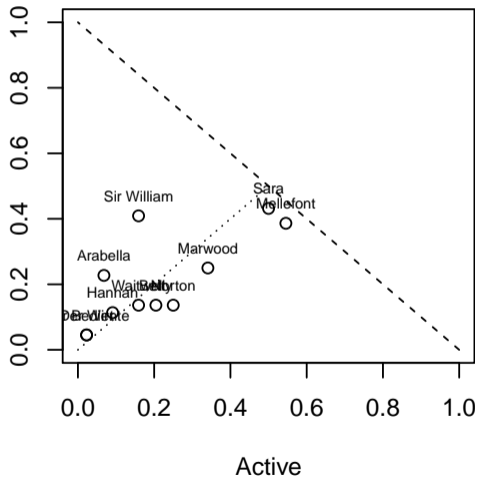


Präsenz

Emilia Galotti (1772)



Miss Sara Sampson (1755)



Section 3

Dramatic Genres

Dramatic Genres

Plays by Kleist and literary studies

Die Familie Schroffenstein	Der zerbrochene Krug	Penthesilea	Das Käthchen von Heilbronn	Die Hermannsschlacht	Prinz Friedrich von Homburg	Amphitryon	Robert Guiskard
Ein Trauerspiel in fünf Aufzügen	Ein Lustspiel	Ein Trauerspiel	Ein großes historisches Ritterschauspiel	Ein Drama	Ein Schauspiel	Ein Lustspiel nach Molière	Fragment aus dem Trauerspiel
Tragödie	Lustspiel	Drama				Lustspiel	
	Lustspiel			politisches Drama			
	Lustspiel		Drama	Tendenzdrama	Schauspiel		
Trauerspiel	Lustspiel	Schauspiel	Schauspiel	Arminius-Drama	Schauspiel	Komödie	
Drama	Komödie	Tragödie	Drama	Drama	Drama	Lustspiel	Drama-Fragment
Drama	Lustspiel	Tragödie (eng.)	historisches Ritterschauspiel	Drama	Schauspiel (eng.)	Lustspiel	Dramenfragment
Tragödie	Komödie	Tragödie	historisches Ritterschauspiel (eng.)	Schauspiel (eng.)	Schauspiel (eng.)	Komödie (eng.)	
Schauspiel (eng.)	Komödie	Trauerspiel	historisches Ritterschauspiel	Drama	Schauspiel	Drama (Kreutzer bezweifelt, dass sich dieses Stück zuordnen lässt)	Tragödie
		Trauerspiel	Drama (Lü: Gattungszuschreibung schwierig, eher: eine Liebesgeschichte)			Drama	
Tragödie	Lustspiel	Trauerspiel		Tendenzstück	Schauspiel	"Jugendwo zwischen Tragödie und Komödie", S. 226	Tragödie
	Lustspiel	Trauerspiel	Schauspiel		Schauspiel		
Drama	Lustspiel	Trauerspiel	Ritterschauspiel	Schauspiel	Schauspiel	Tragikkomödie	Tragödie
Drama	Komödie	Drama der Affekte	Ritter- bzw. Sagenstück	(vaterländische) Tendenzstück	Drama	Lustspiel	Trauerspiel
Drama	Lustspiel	Trauerspiel	Schauspiel	Drama	Schauspiel	Lustspiel	Fragment
Drama	Lustspiel		Stück mit Zügen vom Lustspiel bzw. derben Volksstück, S. 239	Drama	Stück	Tragikkomödie	Tragödie
Trauerspiel	Lustspiel						Tragödie
Tragödie	Lustspiel	Tragödie	Drama	Drama	Drama	Lustspiel	
						Tragikkomödie	
Tragödie	Komödie					Tragikkomödie	Tragödie

Dramatic Genres

Stylometry 101

- Word frequencies as vectors

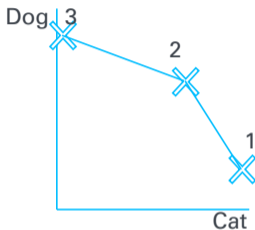
Text	Dog	Cat	Mouse	...
1	15	1	2	
2	17	13	2	
3	0	25	11	
...				

Dramatic Genres

Stylometry 101

- Word frequencies as vectors

Text	Dog	Cat	Mouse	...
1	15	1	2	
2	17	13	2	
3	0	25	11	
...				



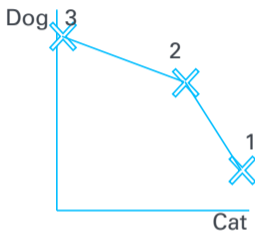
- Text distance \leftrightarrow spatial distance
 - Vector space with n dimensions

Dramatic Genres

Stylometry 101

- Word frequencies as vectors

Text	Dog	Cat	Mouse	...
1	15	1	2	
2	17	13	2	
3	0	25	11	
...				



- Text distance \leftrightarrow spatial distance
 - Vector space with n dimensions
- Authorship attribution: Function words!

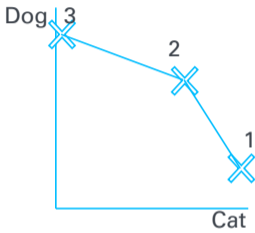
Dramatic Genres

Stylometry 101

- Word frequencies as vectors

Text	Dog	Cat	Mouse	...
1	15	1	2	
2	17	13	2	
3	0	25	11	
...				

- Text distance \leftrightarrow spatial distance
 - Vector space with n dimensions
- Authorship attribution: Function words!
- Green: Tragedies, Blue: Comedies



Dramatic Genres

Content differences

- 178 dramatic texts
- Classification based on self-description
 - “Trauerspiel” → Tragedy, “Lustspiel” → Comedy
- Extraction of
 - lemmas of
 - nouns, verbs, adjectives of
 - 10 most talkative figures per play
- Minimal frequency 50
- Content words with significantly different frequencies
 - log-likelihood, 0.01

	Texts	Figures	Tokens
Comedy	86	783	1,535,665
Tragedy	92	879	1,797,430

Dramatic Genres

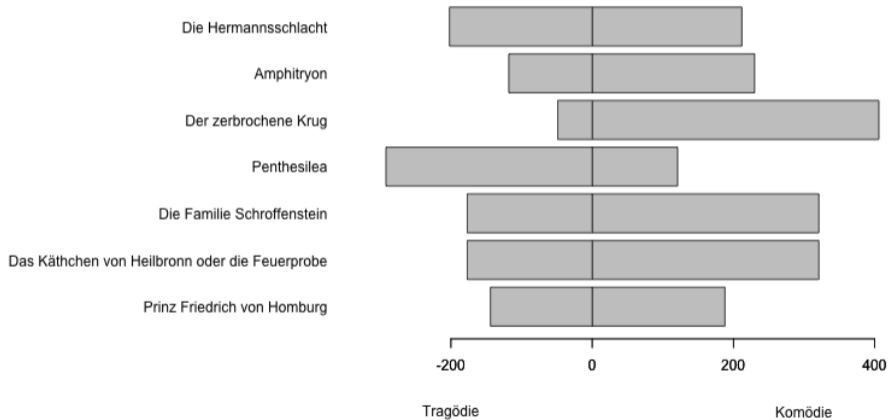
Content differences

Comedy	Tragedy
fräulein, professor, hübsch, kriegen, frau, geschwind, bekommen, wahrhaftig, anfangen, richtig, leute, gesellschaft, gelegenheit, gewiß, zufrieden, geld, ehrlich, reisen, fein, manne, dumm, versichern, woche, person, nit, bezahlen, vergnügen, ehe, sache, fehler, kömmen, gnädig, hochzeit, uhr, besuch, angenehm, kerl, antworten, ansehen, brav, herr, essen, annehmen, henker, holen, onkel, mühe, bey, zimmer, buch, ...	haupt, könig, fürst, dunkel, blut, schauen, volk, brust, haß, frieden, sturm, ruhm, grab, sieg, erde, wild, meer, tod, erheben, tot, retten, hölle, fern, stern, träne, tor, regen, rache, macht, schatten, feind, bild, schlacht, kraft, gunst, lippe, eign, zittern, licht, stürzen, eignen, stehn, tief, traum, drängen, sonne, staub, dringen, fürchterlich, zorn, ...

- 50 tokens per genre, sorted by frequency difference

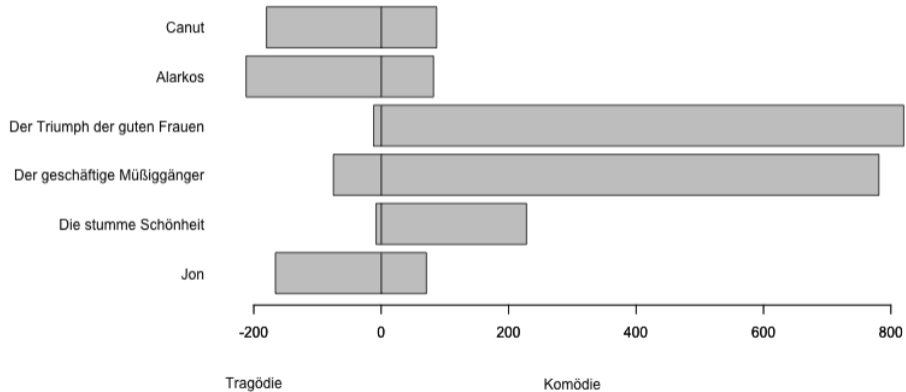
Dramatic Genres

Kleist: Comedy or Tragedy?



Dramatic Genres

Schlegel: Tragödie oder Komödie?



Dramatic Genres

Conclusions

- Dramatic genres can be distinguished by their use of words
- About 100 words are enough to get good classification results

Dramatic Genres

Conclusions

- Dramatic genres can be distinguished by their use of words
- About 100 words are enough to get good classification results

- Is that surprising?

Dramatic Genres

Conclusions

- Dramatic genres can be distinguished by their use of words
- About 100 words are enough to get good classification results

- Is that surprising?
- No
 - The genres are clearly different
 - Different endings (wedding vs. death) need to invoke different vocabulary

Dramatic Genres

Conclusions

- Dramatic genres can be distinguished by their use of words
- About 100 words are enough to get good classification results

- Is that surprising?
- No
 - The genres are clearly different
 - Different endings (wedding vs. death) need to invoke different vocabulary
- Yes
 - It's surprisingly easy
 - 100 words without context

Section 4

Figurenrede analysieren

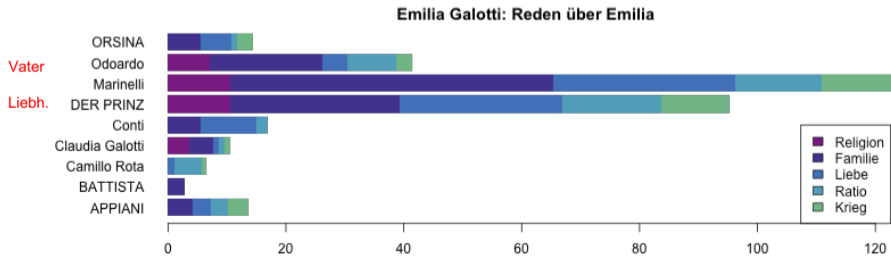
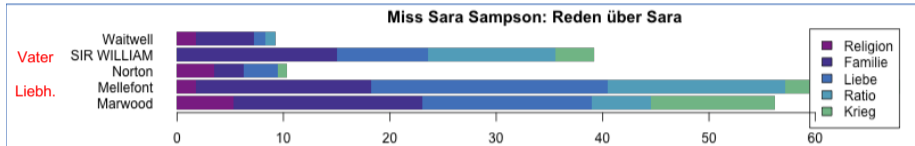
Thematische Analyse der Figurenrede

- ▶ Wörterbuchbasiert
- ▶ Thematische Wörterbücher für Familie, Liebe, Krieg, Ratio, Religion
- ▶ Automatische Lemmatisierung mit mate-tools (Björkelund et al., 2010)
- ▶ Analyse der Figurenrede in Szenen
 - ▶ in denen Titelfiguren nicht selbst sprechen, und
 - ▶ in denen die Titelfiguren namentlich erwähnt werden

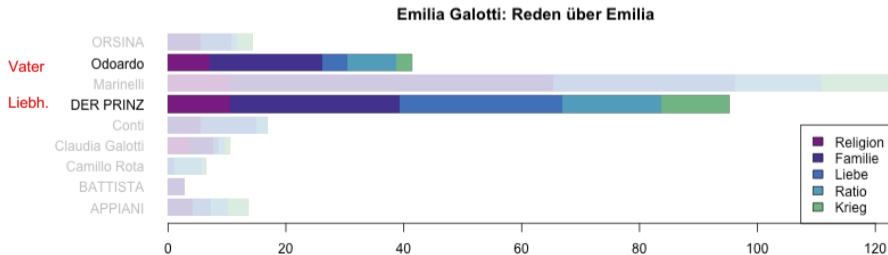
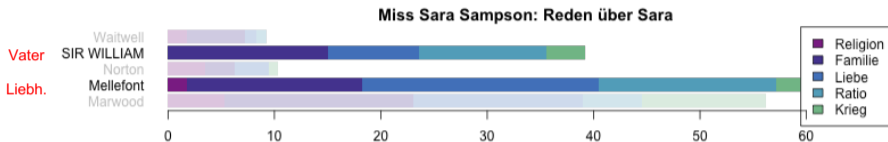
Wortfeld: Liebe

adorieren, anbeten, Angebetete, aufrichtig, Aufrichtigkeit, begehren, Begierde, begieren, beweinen, Beziehung, brennen, Ceremonie, Copulation, copulieren, ehrlich, empfinden, Empfindsamkeit, Engel, Entzücken, fühlen, Funke, Gefühl, geliebt, Geliebte, Geliebter, Gemahl, Gemahlin, Glück, Heirat, Heirath, Herz, Herzen, Hochzeit, huldigen, kuscheln, Kuss, Kuß, küssen, Leidenschaft, Liebe, lieben, liebenswürdig, Liebenswürdigkeit, Liebhaber, Liebhaberin, liebkosen, Liebkosung, Lippe, Lust, lüstern, Rose, Schönheit, Seele, sehnen, Sehnsucht, Sinn, sinnlich, Sinnlichkeit, streicheln, Trauung, treu, Umarmung, Unschuld, verehren, Verehrung, vergöttern, Vergötterung, Verlangen, verlangen, vermählen, Vermählung, verzehren, zärtlich, Zärtlichkeit, Zeremonie, Abschiedskuß, Brust, Busen, geliebt, Geliebte, Geliebtste, Jüngling, Liebchen, liebeglühend, Liebende, Liebesfest, Liebliche, Liebste, Liebster, Liebstes, Treue, vergöttern, Verlieben, Wollust

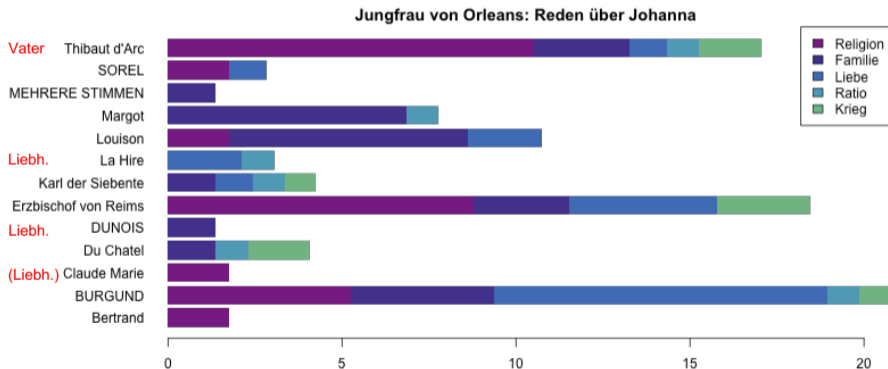
Reden über ...



Reden über ...



Reden über ...



Section 5

Fazit

Fazit

- ▶ Strukturanalysen sind bei Dramen aufschlussreich
- ▶ Textanalyse profitiert von Struktureigenschaften der Dramen
- ▶ Operationalisierung: Wie messen wir eigentlich was?
 - ▶ Kernbereich unserer Arbeit
 - ▶ Gute Strategien erfordern Kenntnisse aus Domäne und Methodik
 - ▶ Validierung von Operationalisierung über Annotationen oder Introspektion