

Quantitative Dramenanalyse: Ein explorativer Zugang zu Gesundheit in literarischen Texten

Gesundheit verstehen – Digital Humanities im Dialog

Nils Reiter

Center for Reflected Text Analytics,
Institut für Maschinelle Sprachverarbeitung,
Universität Stuttgart

30.04.2019

Einführung

Quantitative Drama Analytics (QuaDramA)

▶ Ziele

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

Quantitative Drama Analytics (QuaDramA)

- ▶ Ziele
 - ▶ Analyse konstitutiver Eigenschaften von Figurentypen
 - ▶ (Automatische) Erkennung von Figurentypen
 - ▶ Entwicklung von Figurentypen dramahistorisch und innerhalb des Dramas
 - ▶ Interaktion(stypen) zwischen Figuren(typen)
- ▶ Neuere Deutsche Literaturwissenschaft und Computerlinguistik
 - ▶ Marcus Willand (Uni Heidelberg, früher Stuttgart)
 - ▶ Zwei Doktoranden (Benjamin Krautter, Janis Pagel)
 - ▶ Förderung durch VolkswagenStiftung

Quantitative Drama Analytics (QuaDramA)

- ▶ Ziele
 - ▶ Analyse konstitutiver Eigenschaften von Figurentypen
 - ▶ (Automatische) Erkennung von Figurentypen
 - ▶ Entwicklung von Figurentypen dramahistorisch und innerhalb des Dramas
 - ▶ Interaktion(stypen) zwischen Figuren(typen)
- ▶ Neuere Deutsche Literaturwissenschaft und Computerlinguistik
 - ▶ Marcus Willand (Uni Heidelberg, früher Stuttgart)
 - ▶ Zwei Doktoranden (Benjamin Krautter, Janis Pagel)
 - ▶ Förderung durch VolkswagenStiftung
- ▶ Kooperation und Interaktion auf Augenhöhe
 - ▶ Getrennte Forschungsworkflows mit definierter Schnittstelle
 - ▶ Keine GUIs
 - ▶ Literaturwissenschaftler programmieren selbst

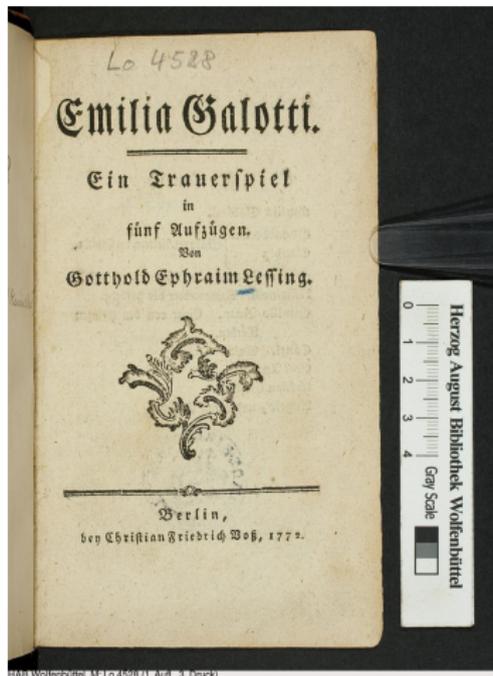
Quantitative Drama Analytics

Das Drama

- ▶ Textvorlage für Theaterstücke
 - ▶ Sprecheranweisungen, Akt-/Szenengrenzen
 - ▶ Figurenrede
 - ▶ Bühnen-/Regieanweisungen
- ▶ Komödie: Am Ende wird geheiratet
- ▶ Tragödie: Am Ende wird gestorben
- ▶ Untergattungen

Quantitative Drama Analytics

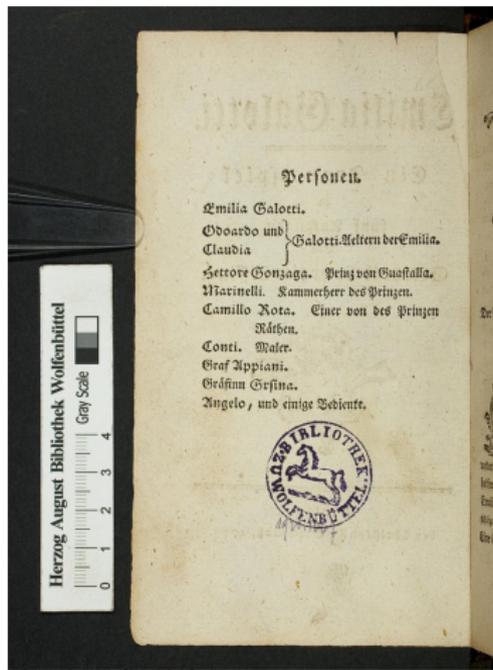
Das Drama



Lessing: Emilia Galotti (1772). Titelblatt.
(Deutsches Textarchiv)

Quantitative Drama Analytics

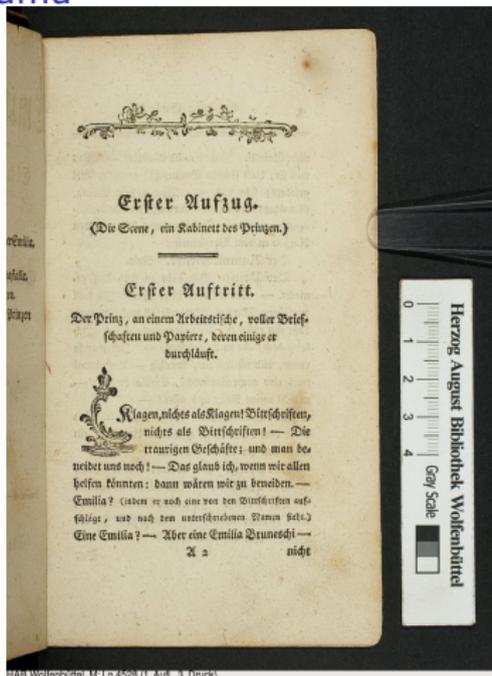
Das Drama



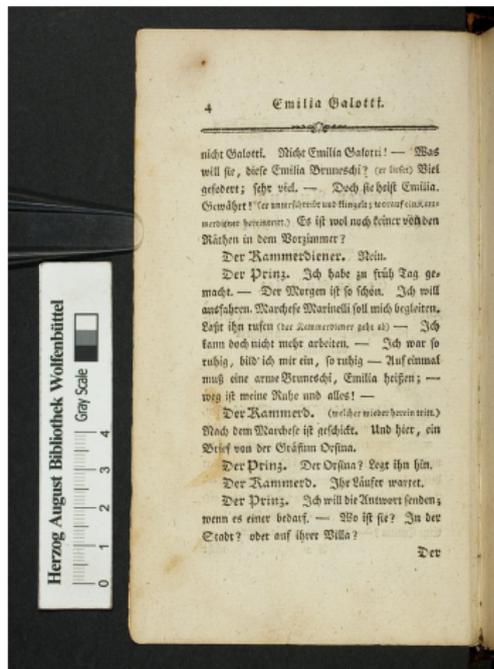
Lessing: Emilia Galotti (1772). Dramatis Personae.
(Deutsches Textarchiv)

Quantitative Drama Analytics

Das Drama



HAB Wolfenbüttel, M. Lo 4528 (1. Aufl., 3. Druck)



HAB Wolfenbüttel, M. Lo 4528 (1. Aufl., 3. Druck)

Lessing: Emilia Galotti (1772). Text.
(Deutsches Textarchiv)

Quantitative Drama Analytics

Beispiele

- ▶ Gotthold Ephraim Lessing (1729-1781)
 - ▶ Miss Sara Sampson (1755)
 - ▶ Emilia Galotti (1772)
- ▶ Bürgerliche Trauerspiele (= Tragödien)
 - ▶ “Bürgerliche Hochkultur”
- ▶ Ähnliche Figurentypen
 - ▶ Tochter
 - ▶ Väter
 - ▶ Liebhaber
 - ▶ ...

Quantitative Drama Analytics: Korpus

- ▶ Korpus: ca. 500 deutschsprachige Dramen zwischen 1730 und 1924
- ▶ TextGrid Repository, GerDraCor
- ▶ Aufbereitet in XML (= “maschinenlesbar”)

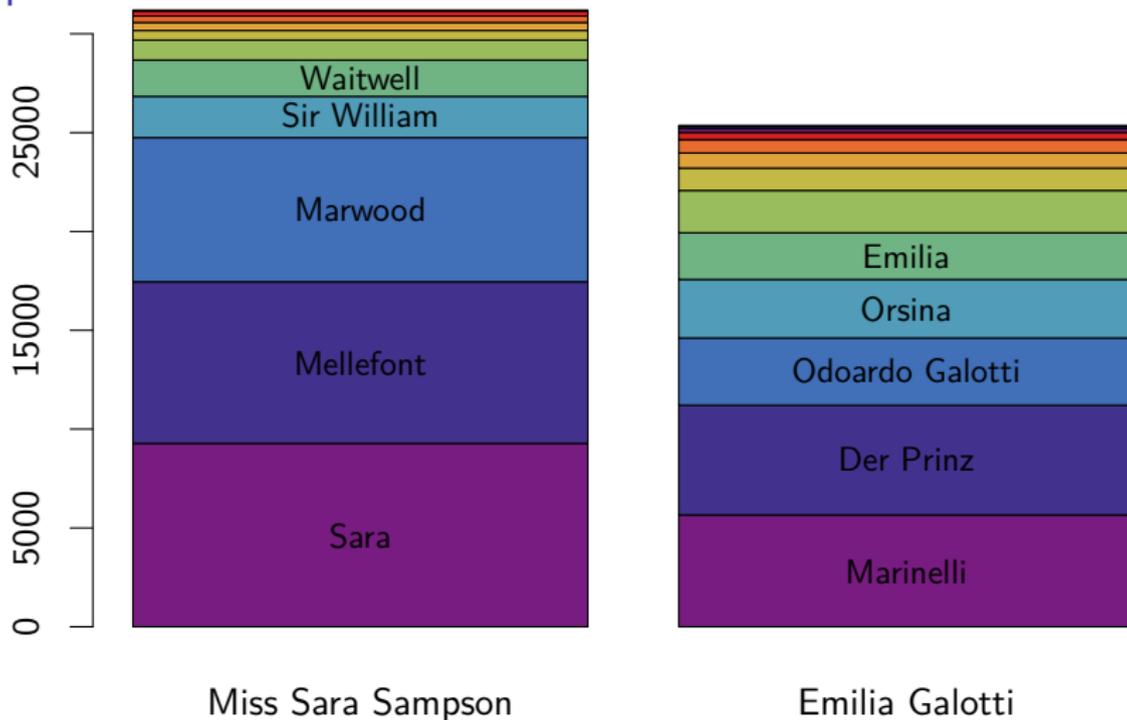
Quantitative Drama Analytics: Korpus

Beispiel XML

```
...
<head>Erster Auftritt</head>
<stage>
  <hi>Fräulein Amalie. Fräulein Karoline.</hi>
</stage>
<sp who="#fräulein_amalie">
  <speaker>FRÄULEIN AMALIE.</speaker>
  <p>Ach, liebe Schwester! wenn doch nur heute unsrer Frau
    Muhme keine Verhinderung in den Weg käme; daß sie das
    Testament machen könnte!</p>
</sp>
<sp who="#fräulein_karoline">
  <speaker>FRÄULEIN KAROLINE.</speaker>
  <l>Warum denn das?</l>
</sp>
...
```

Quantitative Drama Analytics

Beispiele



Figuren und ihre Redeanteile in zwei Lessing-Dramen

Quantitative Drama Analytics

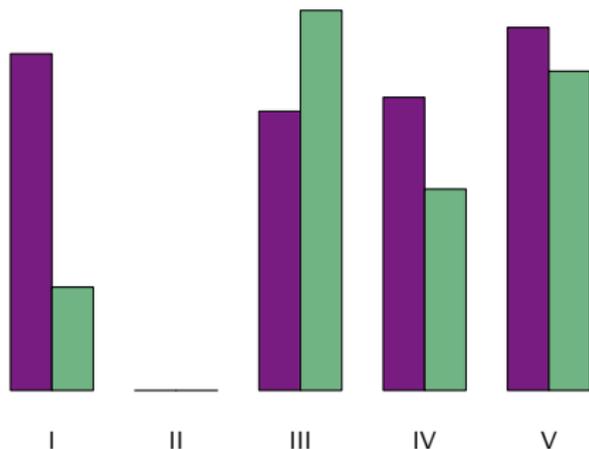
Beispiele



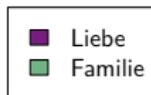
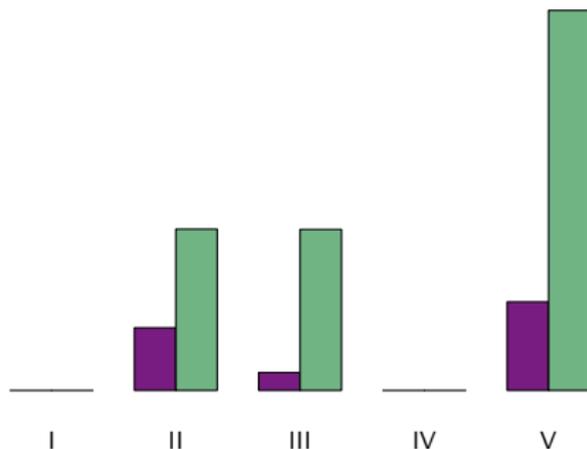
Quantitative Drama Analytics

Beispiele

Rede von Sara



Rede von Emilia



Gesundheit in Dramen

Wortfeldanalyse

- ▶ Wortfeld: Sammlung thematisch ähnlicher Wörter
 - ▶ QuaDramA-Arbeiten: Religion, Liebe, Krieg, Familie, Ratio

Wortfeldanalyse

- ▶ Wortfeld: Sammlung thematisch ähnlicher Wörter
 - ▶ QuaDramA-Arbeiten: Religion, Liebe, Krieg, Familie, Ratio
- ▶ Franz Dornseiff (1934): 'Der deutsche Wortschatz nach Sachgruppen'
 - ▶ Uwe Quasthoff (2004): 8., erweiterte Auflage
- ▶ Drei relevante Bereiche:
 - ▶ 2.23: Gesundheit (102 Wörter)
 - ▶ 2.33: Krankheit (451 Wörter)
 - ▶ 2.35: Behandlung (378 Wörter)

Wortfeldanalyse

Zufällige 20 Wörter je Wortfeld

| Gesundheit | Krankheit | Behandlung |
|--|---|--|
| gesundheitsamt, impfausweis, rüstig, strotzen, gesundheitsschutz, blühen, auf dem damm sein, wohlgemut, stabil, schaffenskraft, zustand, untersuchungsergebnis, wohlauf, gesundheitszentrum, nichts zugestoßen, gesundheitsexperte, ertüchtigen, gesundheitsgefährdend, gesundheitspolitik, unversehrt | übertragbar, feuermal, weicher schanker, krankheit des bewegungsapparates, glatze, schluckauf, anfall, lungentuberkulose, heruntergekommen, blutarmut, kröpf, melancholie, diabetes, hexenschuss, pflegebedürftiger, tripper, aidskranker, ast, grippe, kotzen | schmerzstillend, betriebsarzt, krankenpfleger, seebad, militärhospital, wieder wohlauf sein, heilmethode, poliklinik, orthopäde, heilpraktiker, krankenhaus, unschädlich, rezeptblock, physiotherapie, hygiene, blutsenkung, medizin, universitätsklinik, anästhesist, ärztehaus |

Wortfeldanalyse

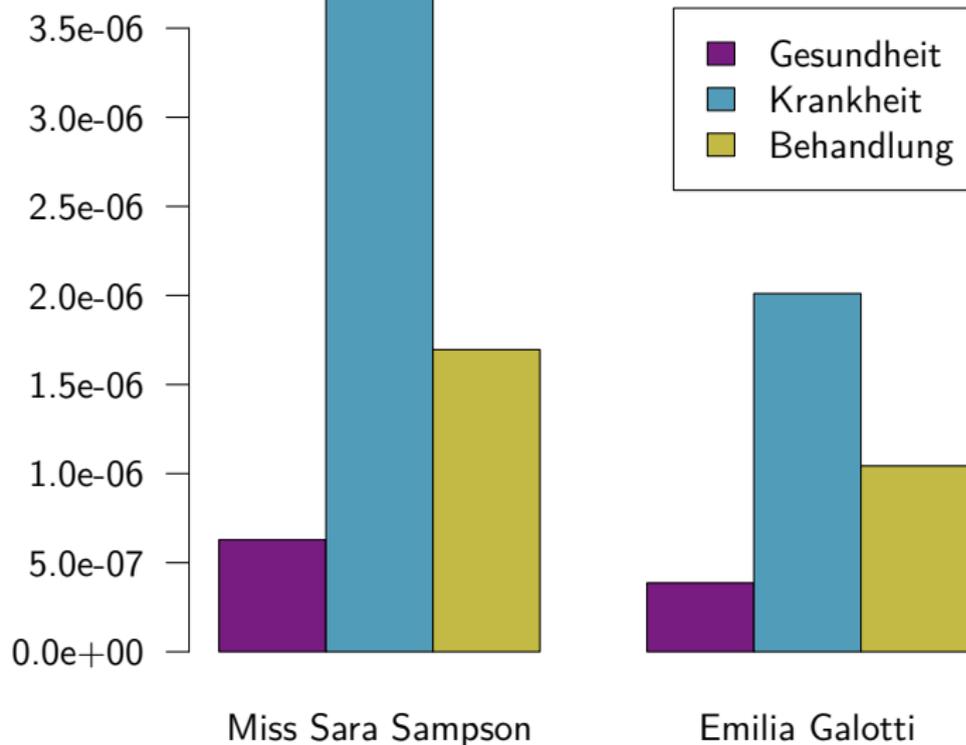
Berechnung

$$s(t, w) = |t \cap w| \times \frac{1}{|w|} \times \frac{1}{|t|}$$

- ▶ Automatische Lemmatisierung
- ▶ Anteil der Wörter aus dem Wortfeld, die im Text/in der Rede einer Figur vorkommen
- ▶ Normalisierung
 - ▶ Redeanteil der Figur (Figuren die viel reden haben mehr Gelegenheit Wörter zu verwenden)
 - ▶ Größe des Wortfeldes (Große Wortfelder haben mehr Chancen dranzukommen)
- ▶ Caveats
 - ▶ Mehrwortausdrücke
 - ▶ Ambige/moderne Wörter

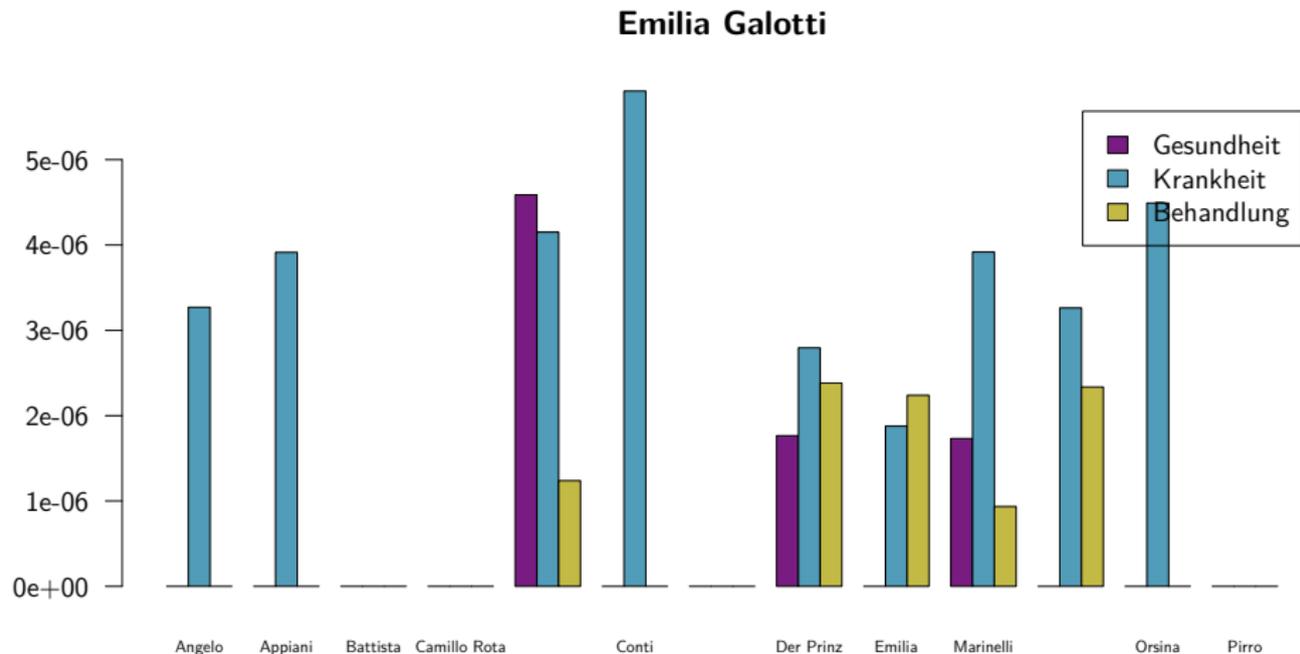
Wortfeldanalyse: Gesundheit

Globalsicht



Wortfeldanalyse: Gesundheit

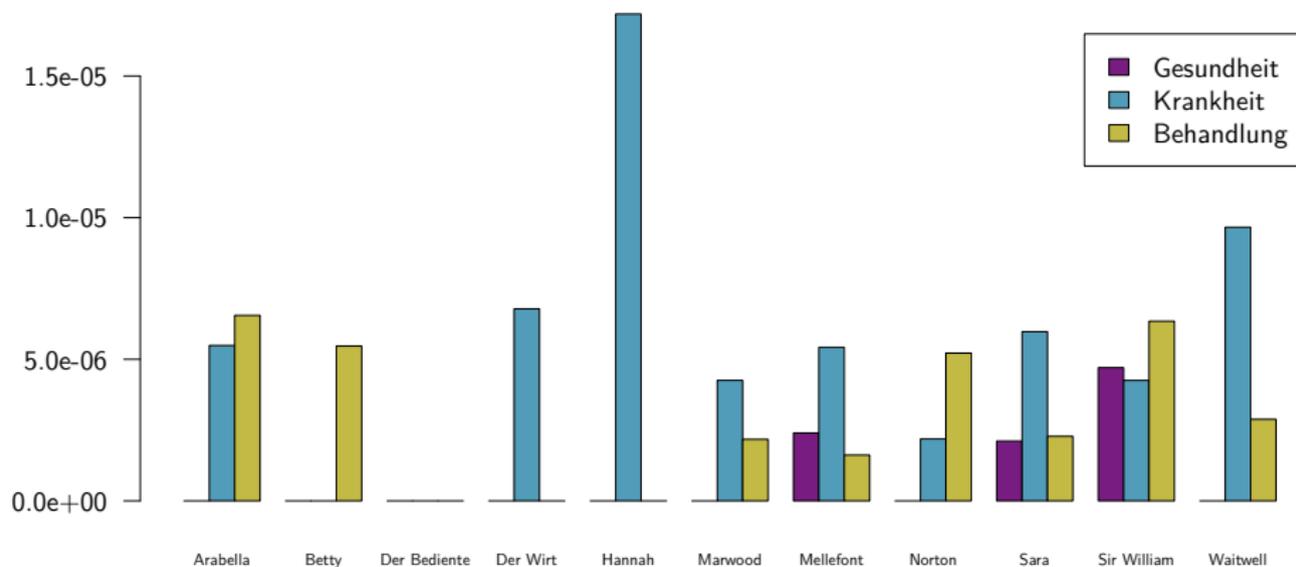
Emilia Galotti



Wortfeldanalyse: Gesundheit

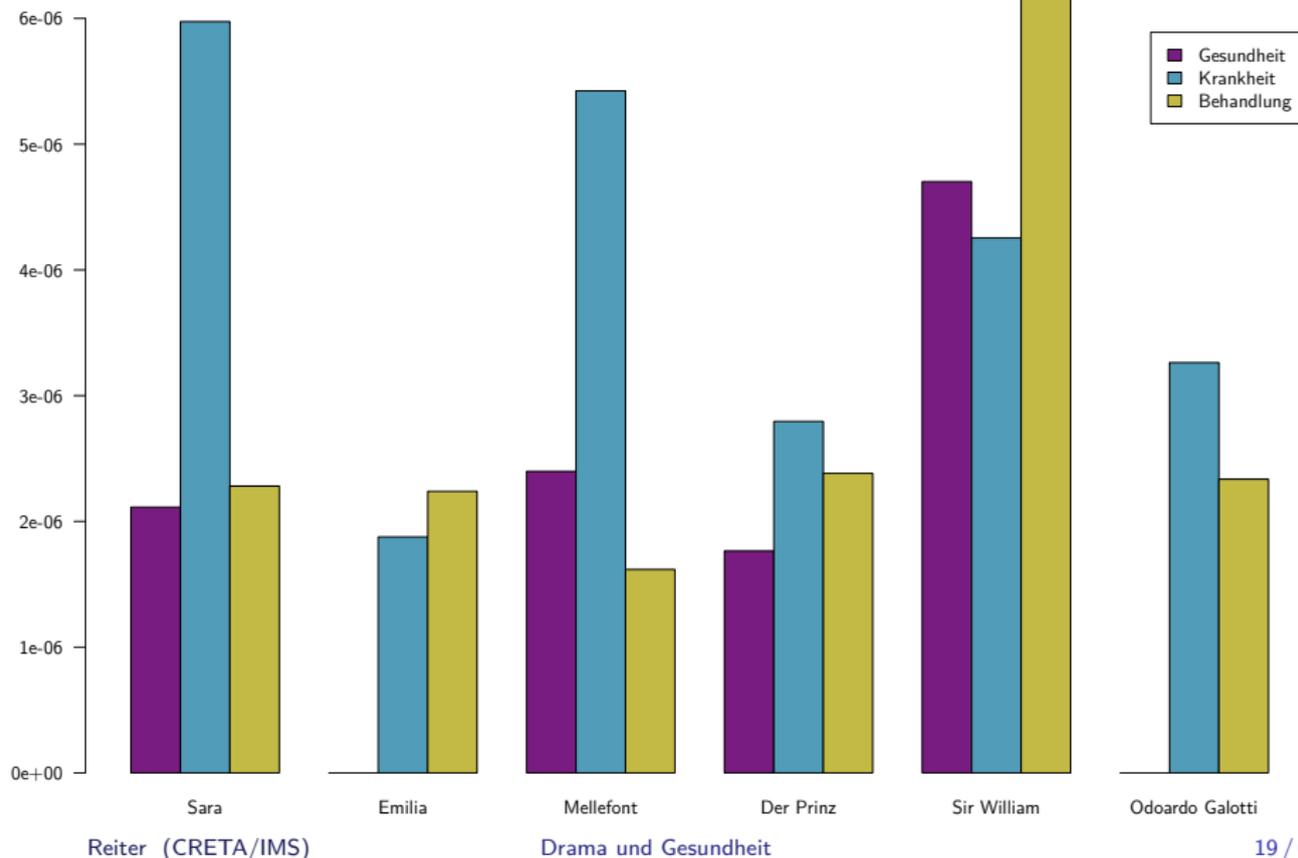
Miss Sara Sampson

Miss Sara Sampson



Wortfeldanalyse: Gesundheit

Figurentypen im Vergleich



Wortfeldanalyse: Gesundheit

Welche Wörter?

| Typ | Wortfeld | Emilia Galotti | Miss Sara Sampson |
|-----------|------------|---|--|
| Töchter | Behandlung | pflügen, bessern | betäuben, heilen, beruhigen, betäubung, helfen, pflügen, bessern, gesund |
| Verführer | Krankheit | schwermut, liegen, elend, leiden, liegen, liegen, elend | elend, schwach, elend, elend, gefährlich, ansteckend, gefährlich, fall, elend, liegen, fall, ernst, übel, übel, schmerz, schmerz, ohnmacht, erbrechen, ohnmacht, schmerz |
| Väter | Krankheit | tödlich, schmerz, fall, fall, liegen | elend, gefährlich, schwach, schwach |
| Väter | Gesundheit | | kräftig |

Wortfeldanalyse: Gesundheit

Welche Wörter?

| Typ | Wortfeld | Emilia Galotti | Miss Sara Sampson |
|-----------|------------|---|--|
| Töchter | Behandlung | pflügen, bessern | betäuben, heilen, beruhigen, betäubung, helfen, pflügen, bessern, gesund |
| Verführer | Krankheit | schwermut, liegen, elend, leiden, liegen, liegen, elend | elend, schwach, elend, elend, gefährlich, ansteckend, gefährlich, fall, elend, liegen, fall, ernst, übel, übel, schmerz, schmerz, ohnmacht, erbrechen, ohnmacht, schmerz |
| Väter | Krankheit | tödlich, schmerz, fall, fall, liegen | elend, gefährlich, schwach, schwach |
| Väter | Gesundheit | | kräftig |

Gesundheitswörter werden (hier) metaphorisch gebraucht ('Liebeskrank')

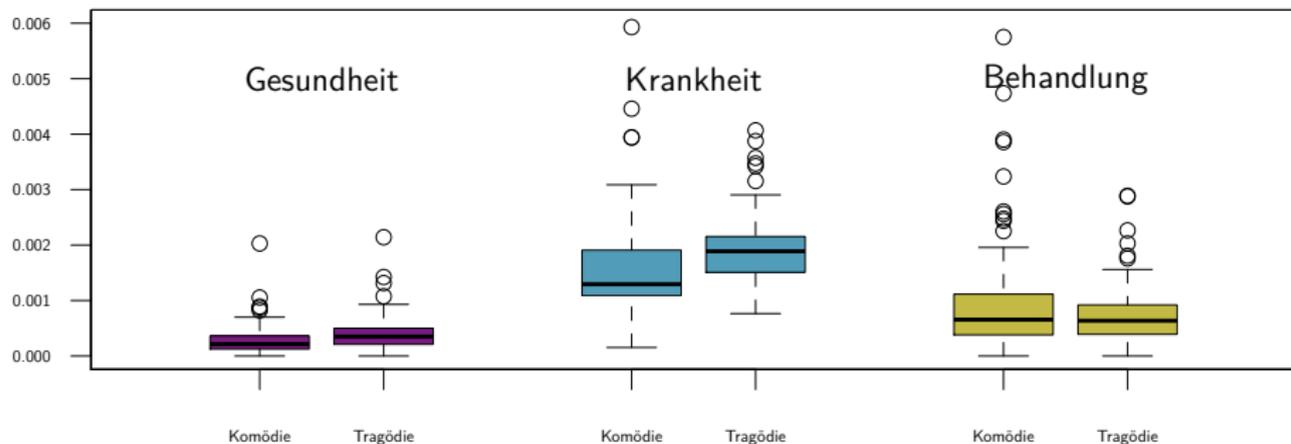
Zooming out

Zooming out

- ▶ Dramenweise Analyse auf größerem Korpus
 - ▶ 76 Tragödien, 81 Komödien

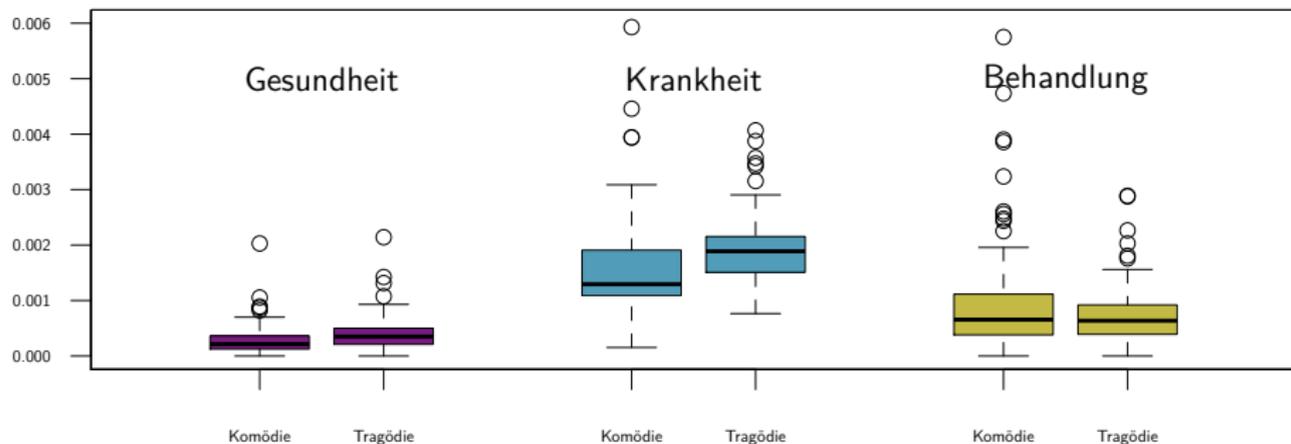
Zooming out

- ▶ Dramenweise Analyse auf größerem Korpus
 - ▶ 76 Tragödien, 81 Komödien



Zooming out

- ▶ Dramenweise Analyse auf größerem Korpus
 - ▶ 76 Tragödien, 81 Komödien



- ▶ Größere Streuung in Komödien

Welche Wörter?

Benutzte Wörter aus dem Wortfeld 'Gesundheit'

| Komödie | Tragödie |
|---|--|
| frisch, gesund, gesundheit, munter, zustand, heil, kräftig, blühen, derb, verhüten, vorbeugen, wohlsein, gedeihen, fidel, unversehrt, wohlauf, rüstig, verfassung, strotzen, unverwüstlich, wohlergehen, stählen, unbeschädigt, zäh, abgehärtet, abhärten, habitus, robust, stabil, wohlbefinden, wohlbehalten | heil, frisch, gesund, blühen, kräftig, zustand, gesundheit, munter, verhüten, gedeihen, derb, stählen, rüstig, verfassung, wohlgemut, wohlsein, zäh, unversehrt, vorbeugen, fidel, konstitution, strotzen, wohlergehen, befund, wohlbehalten, habitus, wohlauf, abhärten, gesundheitszustand, kerngesund, unbeschädigt, untersuchungsergebnis, unverwüstlich, wohlbefinden |

Welche Wörter?

Benutzte Wörter aus dem Wortfeld 'Krankheit'

| Komödie | Tragödie |
|--|---|
| <p>liegen, mal, übel, ernst, fall, leiden, schmerz, krank, gefährlich, krankheit, elend, schwach, schwäche, ohnmacht, bedenklich, schädlich, buckel, pest, anfall, floh, patient, kopfweh, schwindel, zucker, gebrechen, erbrechen, fieber, schwindsucht, anstecken, husten, lahm, krüppel, matt, melancholie, reizung, giftig, schwermut, leidend, schielen, begleiterscheinung, krampf, kränklich, sepsis, siech, tödlich, unheilbar, ausschlag, erblich, hinken, lahmen, schwächlich, verderblich, verstopfung, aufstoßen, bruch, gebrechlich, notfall, schlucken, schmarotzer, blindheit, kraftlos, kranker, krebs, krisis, migräne, schleichend, störung, ungesund, ansteckend, erkranken, gicht, hypochonder, kränkeln, laus, rheumatismus, schnupfen, star, tumor, würmer, ast, bettlägerig, bösartig, cholera, hinfällig, höcker, kahlkopf, klumpfuß, kotzen, krächzen, muttermal, platte, tränen, übelkeit, affektion, angegriffen, bandwurm,</p> | <p>liegen, mal, schmerz, ernst, leiden, elend, fall, schwach, krank, übel, gefährlich, fieber, krankheit, ohnmacht, schwäche, pest, matt, tödlich, bedenklich, giftig, floh, schwindel, verderblich, gebrechen, anfall, leidend, erbrechen, melancholie, anstecken, störung, hinken, karzinom, krebs, schleichend, buckel, kraftlos, krampf, lahm, schielen, ast, blindheit, krächzen, krüppel, schlucken, bruch, erkranken, schädlich, erblich, scharlach, unheilbar, bösartig, gebrechlich, geschwür, glatze, schwermut, ansteckend, ausschlag, blinder, cholera, notfall, patient, rückfall, schwächlich, seuche, symptom, zucker, aussatz, fiebern, kopfweh, kranker, lahmen, siechtum, ungesund, gicht, heruntergekommen, hinfällig, husten, laus, nervosität, platte, schmarotzer, unwohl, asthma, eiter, erkrankung, gelbsucht, knochenbruch, kränklich, reizung, rekonvaleszent, schlapp, star, taubheit,</p> |

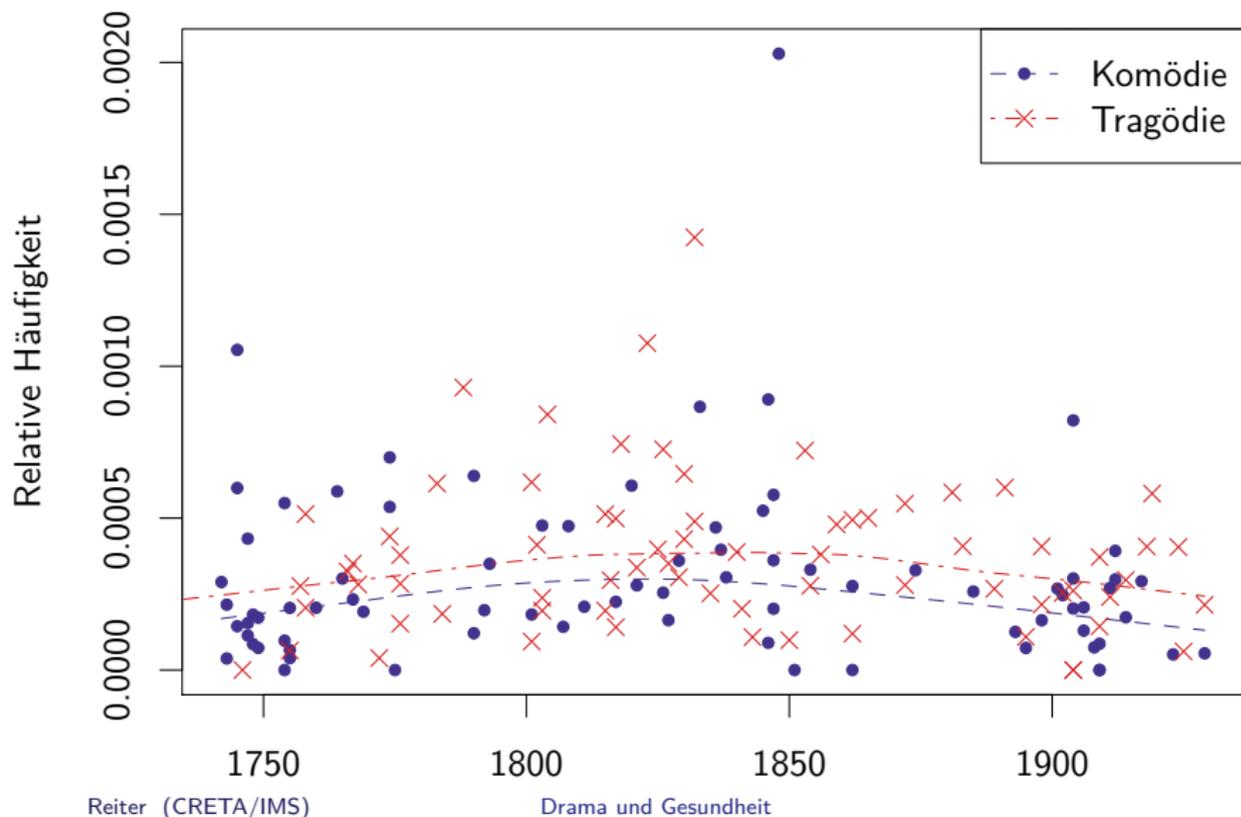
Welche Wörter?

Benutzte Wörter aus dem Wortfeld 'Behandlung'

| Komödie | Tragödie |
|--|--|
| <p>schwester, helfen, doktor, gesund, pflegen, arzt, bessern, behandeln, beruhigen, bedienen, untersuchung, heilen, kurieren, eingeben, harmlos, rezept, verschreiben, kur, medizin, anschlagen, behandlung, vorbeugen, ärztlich, erleichtern, stützen, krankenhaus, wohltätig, betäuben, eingriff, mediziner, pflaster, beschwichtigen, diät, genesung, heilsam, operation, operieren, praktizieren, spital, klinik, pflege, diagnose, förderlich, mildern, verjüngen, verordnen, attest, binde, chirurg, einschläfern, hospital, unschädlich, aufleben, badeort, bestrahlen, dermatologie, genesen, hygiene, quacksalber, sanatorium, schröpfen, seeluft, zuträglich, augenarzt, badekur, betäubung, chirurgie, dermatologe, einlauf, geburtshilfe, hebamme, internist, krankenschwester, kurort, lazarett, meditation, pfleger, physiologe, quarantäne, röntgen, schmerzstillend, therapie, watte, wickel, wiederherstellen</p> | <p>schwester, helfen, doktor, arzt, pflegen, heilen, gesund, beruhigen, bessern, stützen, behandeln, betäuben, untersuchung, bedienen, betäubung, verjüngen, heilsam, erleichtern, eingeben, mildern, binde, genesung, pflege, kur, wohltätig, genesen, harmlos, pflaster, verschreiben, krankenhaus, medizin, verordnen, ärztlich, eingriff, kräftigen, vorbeugen, anschlagen, behandlung, pfleger, unschädlich, beschwichtigen, spital, assistenzarzt, attest, aufleben, befund, diagnose, förderlich, rezept, wickel, wiederherstellen, zuträglich, diät, einschläfern, injektion, lazarett, sanatorium, schröpfen, seebad, aufbauend, bestrahlen, betreuen, erstarken, geburtshelfer, gynäkologe, hausarzt, hebamme, heilkunst, impfen, internist, krankenschwester, küren, kurieren, massage, mediziner, obduktion, operation, physiologe, psychiater, quacksalber, seeluft, tierarzt, untersuchungsergebnis, watte</p> |

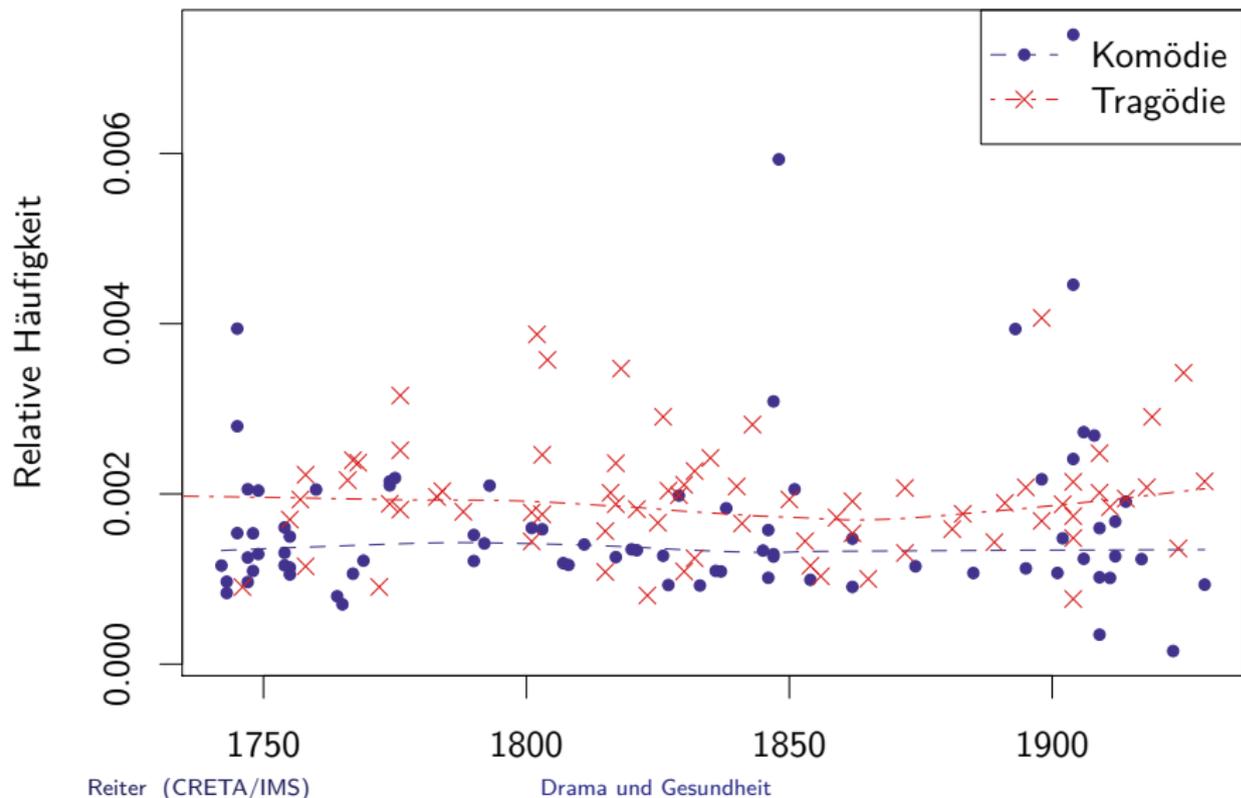
Diachrone Entwicklung

Gesundheit



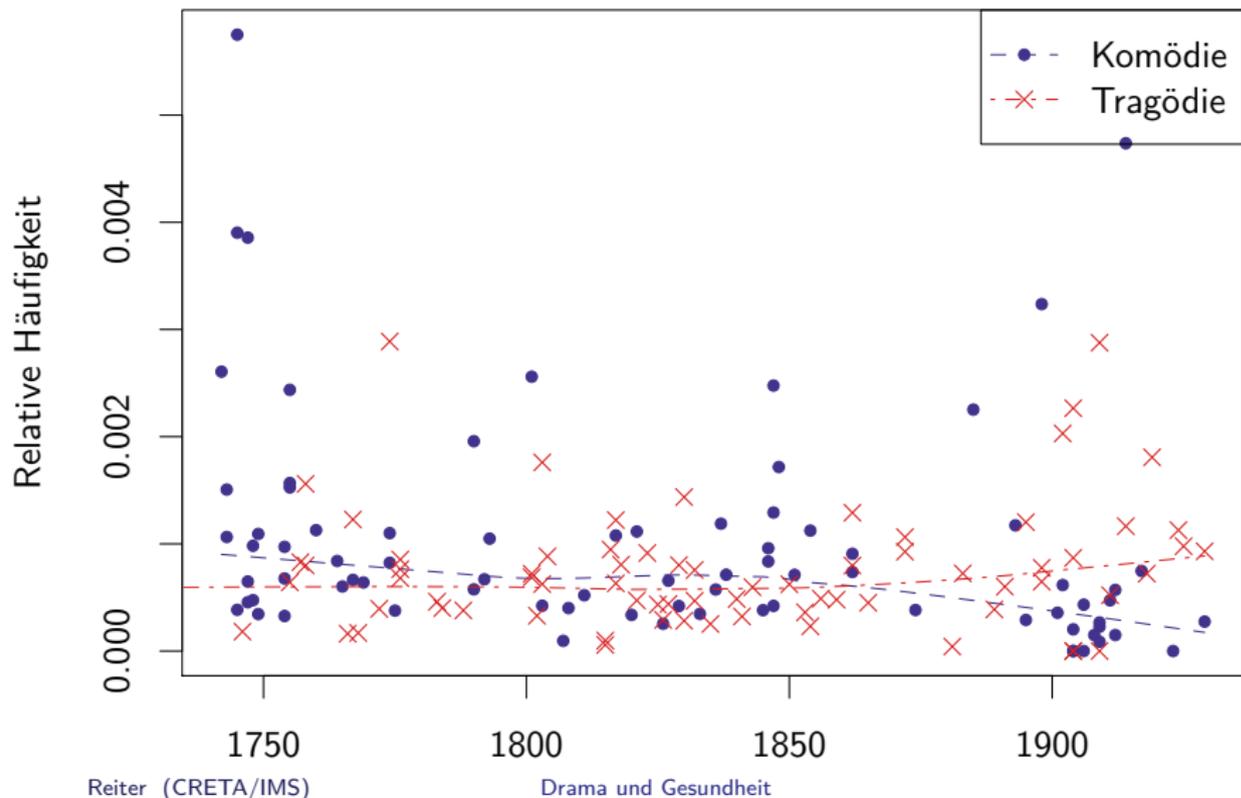
Diachrone Entwicklung

Krankheit

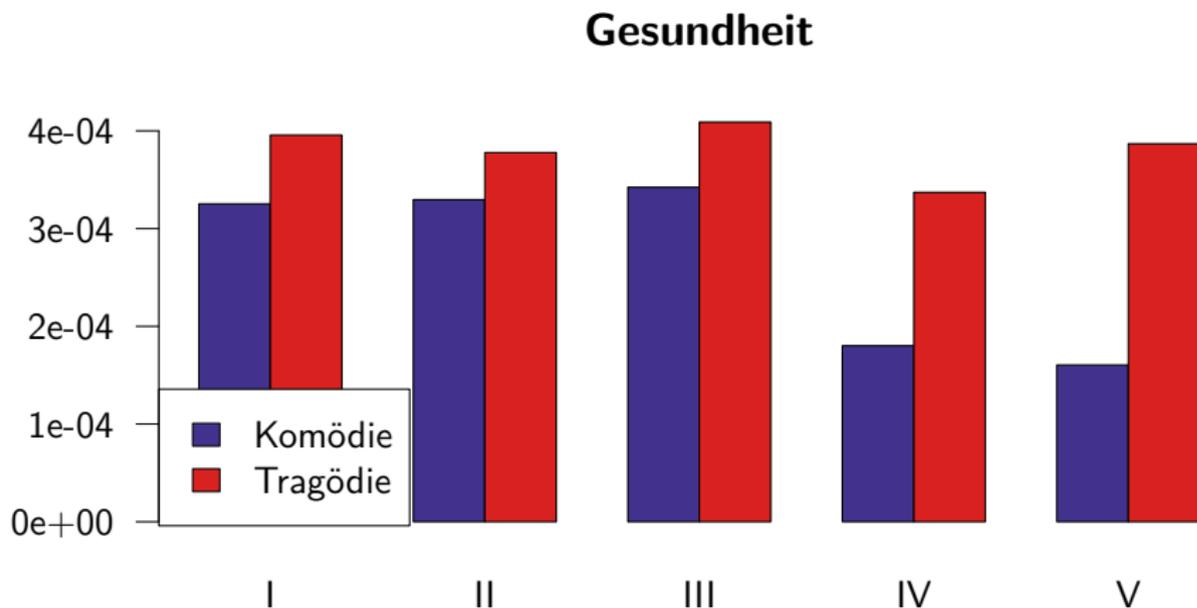


Diachrone Entwicklung

Behandlung



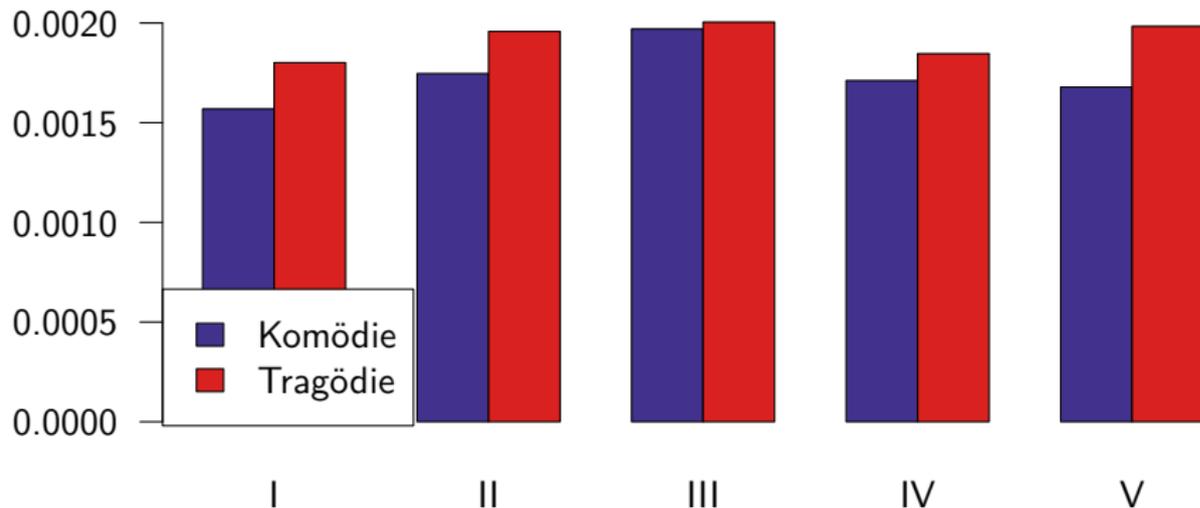
Wo im Drama?



Nur Fünfkfater (24/55), durchschnittliche relative Häufigkeit von Wörtern aus dem Wortfeld je Akt

Wo im Drama?

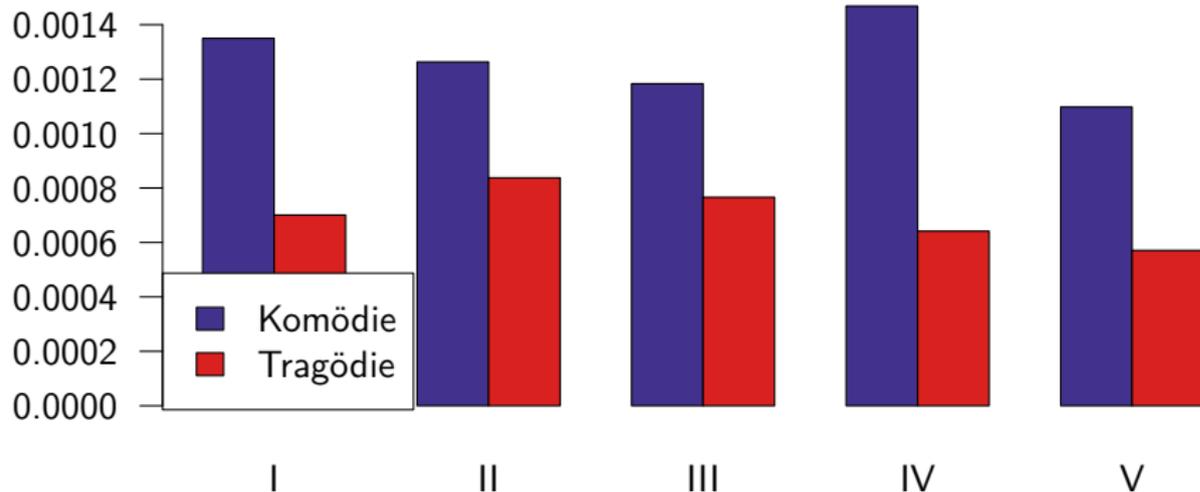
Krankheit



Nur Fünkfaktor (24/55), durchschnittliche relative Häufigkeit von Wörtern aus dem Wortfeld je Akt

Wo im Drama?

Behandlung



Nur Fünkfaktor (24/55), durchschnittliche relative Häufigkeit von Wörtern aus dem Wortfeld je Akt

Einzeltexte

| Wortfeld | Autor | Titel | Datum |
|------------|---|--|-------|
| Gesundheit | Goethe, Johann Wolfgang | Faust. Der Tragödie zweiter Teil | 1832 |
| Krankheit | Hartleben, Otto Erich | Rosenmontag | 1898 |
| Behandlung | Gottsched, Luise Adelgunde Victorie | Das Testament | 1745 |

Goethe, Johann Wolfgang: Faust. Der Tragödie zweiter Teil (1832)

“Das gesündeste Drama!”

##

| | | | | |
|----|--------|------|---------|--------|
| ## | frisch | heil | kräftig | gesund |
|----|--------|------|---------|--------|

| | | | | |
|----|----|----|----|---|
| ## | 34 | 23 | 16 | 7 |
|----|----|----|----|---|

| | | | | |
|----|--------|--------|------|-----------|
| ## | munter | blühen | derb | wohlgemut |
|----|--------|--------|------|-----------|

| | | | | |
|----|---|---|---|---|
| ## | 5 | 4 | 4 | 4 |
|----|---|---|---|---|

| | | | | |
|----|--------|----------|------------|---------|
| ## | rüstig | gedeihen | unversehrt | zustand |
|----|--------|----------|------------|---------|

| | | | | |
|----|---|---|---|---|
| ## | 2 | 1 | 1 | 1 |
|----|---|---|---|---|

Goethe, Johann Wolfgang: Faust. Der Tragödie zweiter Teil (1832)

*Des Lebens Pulse schlagen frisch lebendig,
Wie eine Pfirsche frisch und voller Saft!
Nun zum Peneios frisch hinab!*

Goethe, Johann Wolfgang: Faust. Der Tragödie zweiter Teil (1832)

*Des Lebens Pulse schlagen frisch lebendig,
Wie eine Pfirsche frisch und voller Saft!
Nun zum Peneios frisch hinab!*

*Heil dem Wasser! Heil dem Feuer!
Heil dem seltnen Abenteuer!*

Goethe, Johann Wolfgang: Faust. Der Tragödie zweiter Teil (1832)

*Des Lebens Pulse schlagen frisch lebendig,
Wie eine Pfirsche frisch und voller Saft!
Nun zum Peneios frisch hinab!*

*Heil dem Wasser! Heil dem Feuer!
Heil dem seltnen Abenteuer!*

- ▶ Oft nicht mit “gesundheitlicher Bedeutung” verwendet

Krankheit

Hartleben, Otto Erich: Rosenmontag (1898)

| | | | |
|----|----------------|------------|---------|
| ## | | | |
| ## | mal | liegen | ernst |
| ## | 79 | 18 | 9 |
| ## | gefährlich | fall | krank |
| ## | 5 | 4 | 3 |
| ## | übel | elend | leiden |
| ## | 3 | 2 | 2 |
| ## | anstecken | bedenklich | bruch |
| ## | 1 | 1 | 1 |
| ## | krankheit | ohnmacht | pest |
| ## | 1 | 1 | 1 |
| ## | rekonvaleszent | schmerz | schwach |
| ## | 1 | 1 | 1 |
| ## | typhus | | |
| ## | 1 | | |

Behandlung

Gottsched, Luise Adelgunde Victorie: Das Testament (1745)

| | | | |
|----|-----------|--------------|----------|
| ## | | | |
| ## | doktor | schwester | gesund |
| ## | 71 | 39 | 13 |
| ## | arzt | helfen | eingeben |
| ## | 9 | 8 | 4 |
| ## | kurieren | bedienen | bessern |
| ## | 4 | 2 | 2 |
| ## | genesung | medizin | rezept |
| ## | 2 | 2 | 2 |
| ## | beruhigen | heilen | hospital |
| ## | 1 | 1 | 1 |
| ## | pflegen | praktizieren | |
| ## | 1 | 1 | |

Behandlung

Gottsched, Luise Adelgunde Victorie: Das Testament (1745)

Personen dieses Lustspiels

Frau Oberstin von Tiefenborn, eine alte reiche Witwe

Fräulein Karoline

Fräulein Amalie, ihrer Schwester Kinder

Herr von Kaltenbrunn

Herr Hauptmann von Wagehals

Herr Dr Hippokras, ein Arzt

Herr Landrat von Ziegendorf

Der Kammerherr von Ziegendorf

Herr von Kreuzweg

Ein Notarius

Heinrich, ein Lakai

Themen erkennen

Methodenentwicklung in NLP/CLS

- ▶ Ziel: Automatische Erkennung linguistischer Phänomene (z.B. Wortarten)
- ▶ Ansätze
 - ▶ Regelbasierte Systeme: Computerlinguist_innen schreiben Programme mit linguistischen Regeln (außer Mode)
 - ▶ Maschinelle Lernverfahren: Computermodelle werden auf annotierten Daten trainiert und lernen selbständig
 - ▶ z.B., dass Großschreibung und Wortart zusammenhängen
- ▶ Computerlinguistik ist eine Methodendisziplin
 - ▶ Entwickelte Verfahren müssen *evaluiert* (= systematisch getestet) werden
 - ▶ Annotierte Daten!
- ▶ Goldstandard: Vertrauenswürdig annotierte Daten
 - ▶ Training von maschinellen Lernverfahren
 - ▶ Evaluation von Systemen (gleich welcher Art)

Villa Vigoni



DFG Symposium 2017: Digitale
Literaturwissenschaft

(nicht im Bild: Comer See)

Nils Reiter, Marcus Willand. *What are they talking about? A Systematic Exploration of Theme Identification Methods for Character Speech in Dramatic Texts*

Themen erkennen

- ▶ Verschiedene Methoden Themen zu erkennen
 - ▶ Wörterbuch-basiert (Dornseiff, WordNet, GermaNet)
 - ▶ Topic Modelling
- ▶ Kein Goldstandard
 - ▶ “Worum geht es in diesem Text(abschnitt)?” ist eine interpretative Frage (insb. bei literarischen Texten)
 - ▶ Propositionaler Gehalt (X begrüßt Y) nur die halbe Miete
 - ▶ Viele Textabschnitte enthalten keinen propositionalen Gehalt
 - ▶ Methodische Herausforderung
 - ▶ Goldstandard auch mit viel Geld nicht herstellbar

Themen erkennen

- ▶ Verschiedene Methoden Themen zu erkennen
 - ▶ Wörterbuch-basiert (Dornseiff, WordNet, GermaNet)
 - ▶ Topic Modelling
- ▶ Kein Goldstandard
 - ▶ “Worum geht es in diesem Text(abschnitt)?” ist eine interpretative Frage (insb. bei literarischen Texten)
 - ▶ Propositionaler Gehalt (X begrüßt Y) nur die halbe Miete
 - ▶ Viele Textabschnitte enthalten keinen propositionalen Gehalt
 - ▶ Methodische Herausforderung
 - ▶ Goldstandard auch mit viel Geld nicht herstellbar
- ▶ Aber: Zugang zu kooperativen Domänenexpert_innen
 - ▶ Wissen über literarische Texte

Können wir trotzdem Methodenevaluation (u. -entwicklung) betreiben?
Wenn ja, wie?

Literaturwissenschaftliches Wissen

- ▶ Literaturwissenschaft erzeugt qualitatives/vages Wissen
 - ▶ Komplex kontextabhängig
 - ▶ Nicht direkt quantifizierbar
 - ▶ Nicht empirisch
 - ▶ Entwickelt über Jahrhunderte

Literaturwissenschaftliches Wissen

- ▶ Literaturwissenschaft erzeugt qualitatives/vages Wissen
 - ▶ Komplex kontextabhängig
 - ▶ Nicht direkt quantifizierbar
 - ▶ Nicht empirisch
 - ▶ Entwickelt über Jahrhunderte

- ▶ Wissen über Gattungen
 - ▶ Komödien: Figuren heiraten – Tragödien: Figuren sterben
 - ▶ Ausnahmen sind möglich!
 - ▶ Keine Aussagen über jeden einzelnen Text

Literaturwissenschaftliches Wissen

- ▶ Literaturwissenschaft erzeugt qualitatives/vages Wissen
 - ▶ Komplex kontextabhängig
 - ▶ Nicht direkt quantifizierbar
 - ▶ Nicht empirisch
 - ▶ Entwickelt über Jahrhunderte

- ▶ Wissen über Gattungen
 - ▶ Komödien: Figuren heiraten – Tragödien: Figuren sterben
 - ▶ Ausnahmen sind möglich!
 - ▶ Keine Aussagen über jeden einzelnen Text

- ▶ Generische Aussagen: Vögel können fliegen

Idee

- ▶ Postulat: Erwartete Beziehung zwischen Themen in Figurenrede und einer messbaren Eigenschaft
- ▶ Gesamtschau: Mehrere Postulate gemeinsam betrachten

Idee

- ▶ Postulat: Erwartete Beziehung zwischen Themen in Figurenrede und einer messbaren Eigenschaft
- ▶ Gesamtschau: Mehrere Postulate gemeinsam betrachten
- ▶ Wie viele Dramen verhalten sich wie erwartet bei Methode X?
 - ▶ Messung für verschiedene Anteile
- ▶ Testen von verschiedenen Methoden der Themenerkennung

Idee

- ▶ Postulat: Erwartete Beziehung zwischen Themen in Figurenrede und einer messbaren Eigenschaft
- ▶ Gesamtschau: Mehrere Postulate gemeinsam betrachten
- ▶ Wie viele Dramen verhalten sich wie erwartet bei Methode X?
 - ▶ Messung für verschiedene Anteile
- ▶ Testen von verschiedenen Methoden der Themenerkennung
- ▶ Verwandt mit
 - ▶ Downstream evaluation
 - ▶ Distant supervision

Postulate

1. Geteilte Bühnenpräsenz führt zu ähnlichen Themen
 - ▶ Weil Figuren interagieren
2. Mehr Figurenaustausch führt zu mehr unterschiedlichen Themen über Segmentgrenzen hinweg
3. Tragödien und Komödien weisen unterschiedliche Themen am Ende auf
 - ▶ Sterben/heiraten sollte sich in Themen niederschlagen
4. Themen in der Rede von Protagonist_innen ändern sich über den dritten Akt
 - ▶ Der Wendepunkt im Drama ist im dritten Akt
 - ▶ (nicht in dieser Präsentation)

Annahmen

- ▶ Themenzuweisungen können als Vektor repräsentiert werden
- ▶ Arbiträre Textsegmente auf Themen untersuchbar
- ▶ Dramatische Textstruktur

Annahmen

- ▶ Themenzuweisungen können als Vektor repräsentiert werden
- ▶ Arbiträre Textsegmente auf Themen untersuchbar
- ▶ Dramatische Textstruktur

- ▶ Methoden der Themenerkennung
 - ▶ Wortfelder (manuell): Liebe, Familie, Ratio, Krieg, Religion (Domänenwissen)
 - ▶ Wortfelder (semi-automatisch erweitert)
 - ▶ GermaNet Wurzelebene
 - ▶ Vier Varianten von topic modelling
 - ▶ Nach Figur vs. nach Stück
 - ▶ 20 vs. 100 topics

Postulat 1: Geteilte Bühnenpräsenz → ähnliche Themen

Operationalisierung

- ▶ Gemessen über Figurenpaare

$$C_{INV}(f_1, f_2) \in [0; 1]$$

Invertierte, relative Häufigkeit an Szenen, in denen ein Figurenpaar kopräsent sind

- ▶ Ähnlichkeit der thematischen Profile: Euklidische Distanz zwischen thematischen Vektoren
- ▶ Korrelation zwischen Distanz und C_{INV} quantifiziert als Pearson-Korrelation

Beispiel: Geteilte Bühnenpräsenz

| ↓ Figur / Szene→ | 1 | 2 | 3 | 4 |
|------------------|---|---|---|---|
| A | x | x | | |
| B | x | x | x | |
| C | | x | | x |

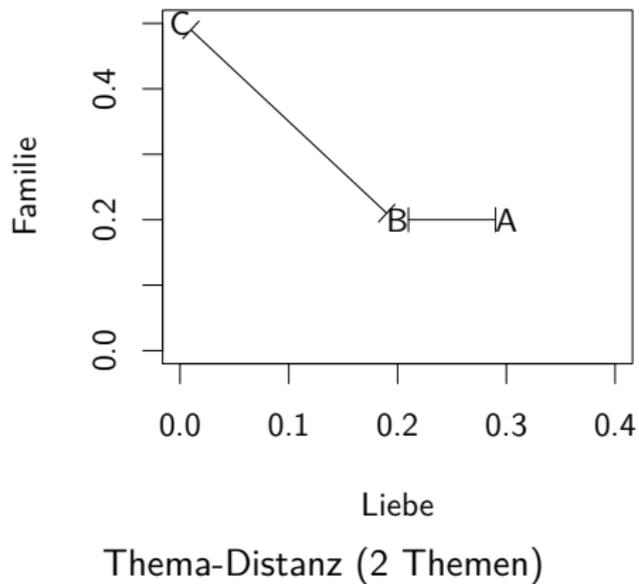
$$C_{INV}(A, B) = 1 - \frac{2}{4} = \frac{1}{2} \quad C_{INV}(B, C) = 1 - \frac{1}{4} = \frac{3}{4}$$

Beispiel: Ähnliche Themen

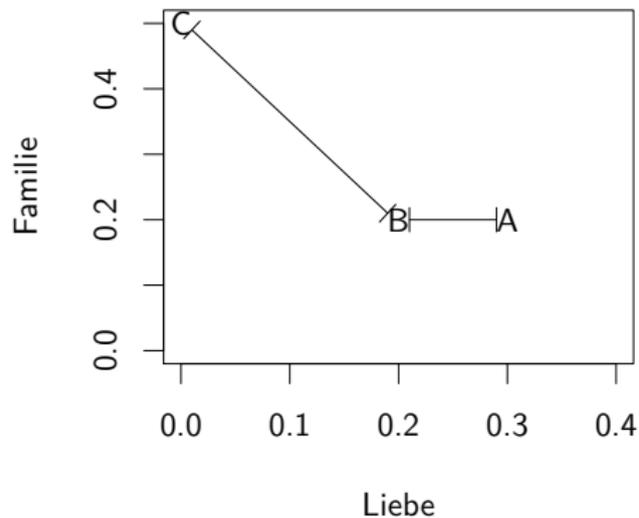
Thematische Profile der Figuren A, B und C

| | Liebe | Familie | Krieg | Religion | Ratio |
|---|-------|---------|-------|----------|-------|
| A | 3e-01 | 0.2 | 3e-01 | 0.0 | 0.0 |
| B | 2e-01 | 0.2 | 1e-04 | 0.1 | 0.1 |
| C | 1e-04 | 0.5 | 1e-04 | 0.2 | 0.2 |

Beispiel: Ähnliche Themen



Beispiel: Ähnliche Themen

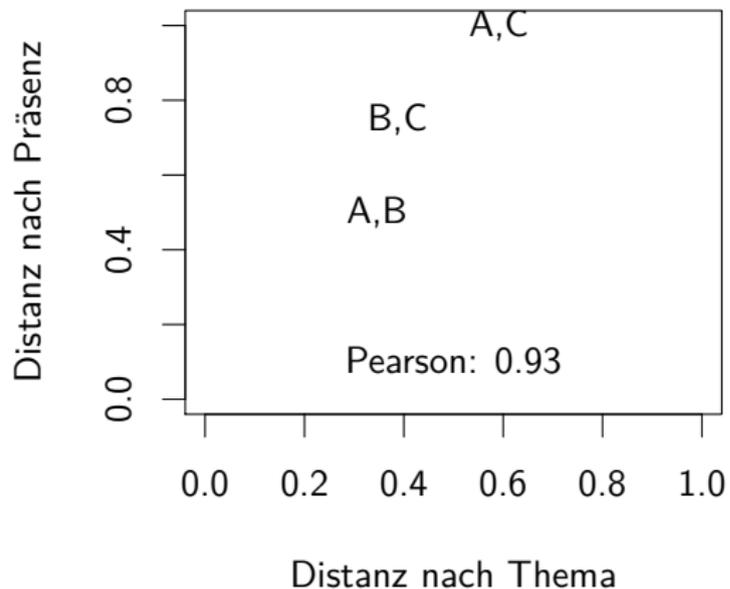


Thema-Distanz (2 Themen)

Thema-Distanz (5 Themen)

| | A | B | C |
|---|------|------|------|
| A | 0.00 | 0.35 | 0.59 |
| B | 0.35 | 0.00 | 0.39 |
| C | 0.59 | 0.39 | 0.00 |

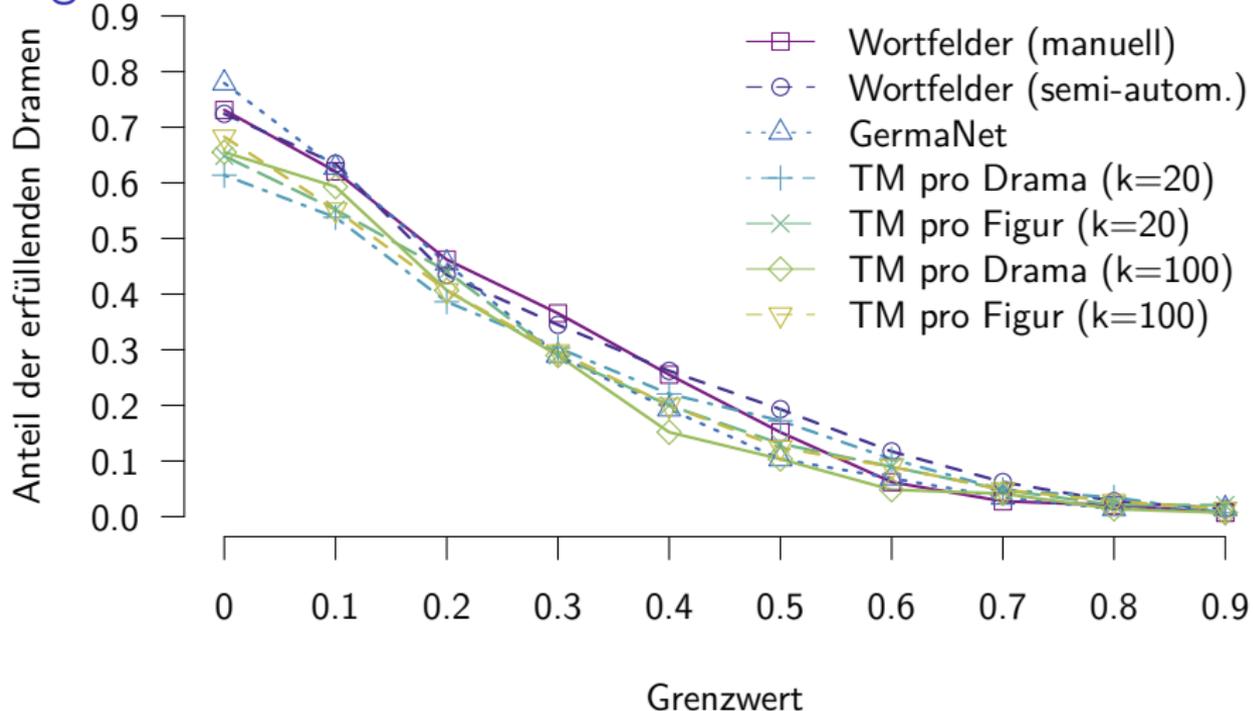
Beispiel: Ähnliche Themen



Operationalisierung

1. Berechne je Figurenpaar
 - ▶ Distanz nach Kopräsenz (C_{INV})
 - ▶ Distanz nach thematischem Profil
2. Berechne Korrelation der Distanzen für alle Paare in einem Drama
3. Bestimme wie viele Dramen über welchem Grenzwert liegen

Ergebnisse



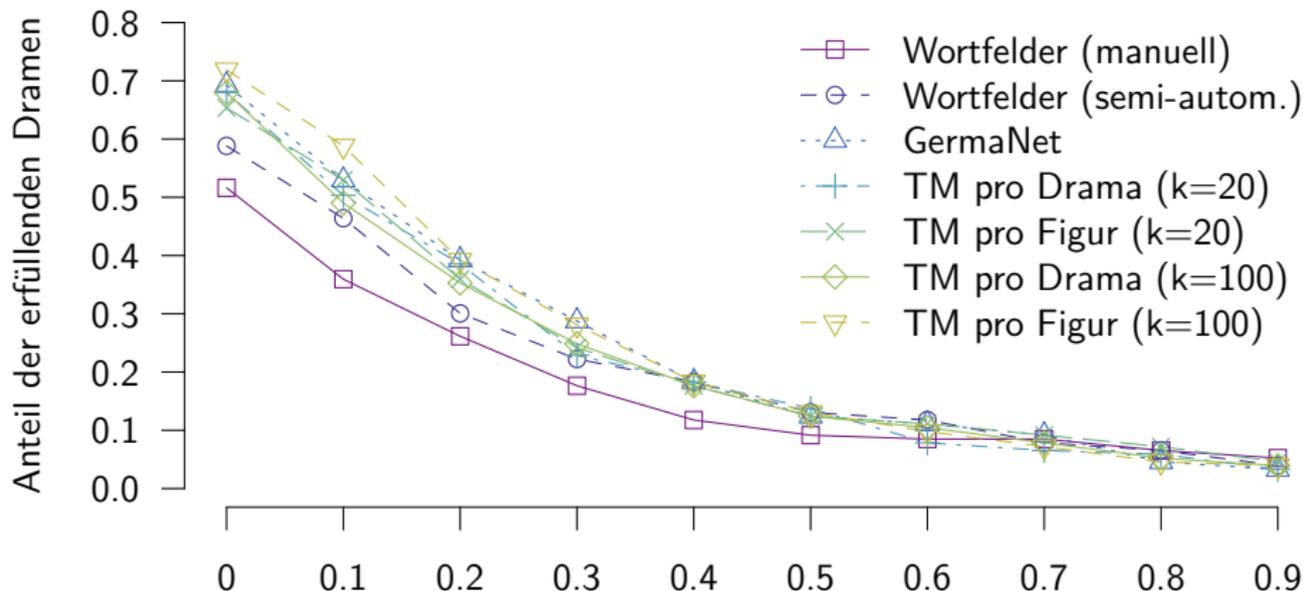
- ▶ Bandbreite über 20 Prozentpunkte
- ▶ TM schlechter für die meisten Grenzwerte, Wortfelder und GermaNet sehr ähnlich

Postulat 2: Mehr Figurenaustausch → unterschiedlichere Themen

Operationalisierung

- ▶ Figurenaustausch (Trilcke et al., 2017): $r_i = \frac{C_i \Delta C_{i+1}}{C_i \cup C_{i+1}}$
 - ▶ Anteil der ausgetauschten Figuren an den Figuren, die in zwei benachbarten Szenen auftreten
 - ▶ Alle Figuren bleiben: $r_i = 0$
 - ▶ Alle Figuren wechseln: $r_i = 1$
- ▶ Thematische Distanz: Wie vorher, Euklidische Distanz der Themen-Vektoren über die Figurenrede in Szene i und $i + 1$
 - ▶ Alle Figuren zusammen
- ▶ Korrelation

Ergebnisse



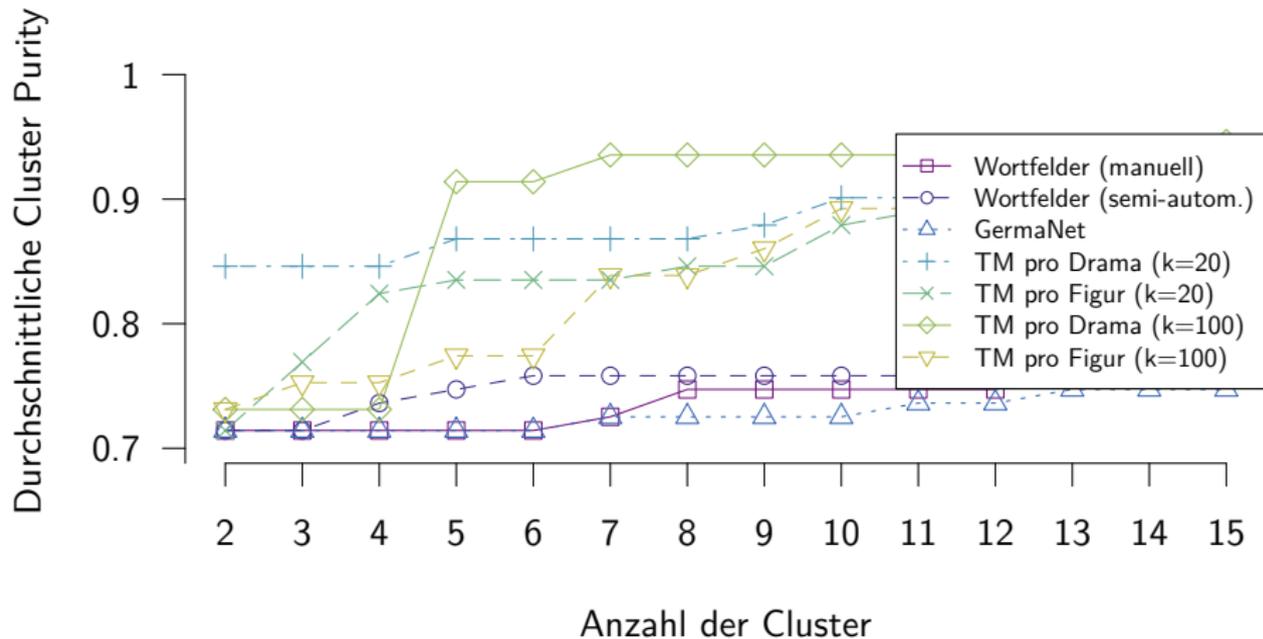
- ▶ Insgesamt niedrigeres Niveau an erfüllenden Dramen
- ▶ Manuell erstellte Wortfelder deutlich schlechter
- ▶ Wortfelder automatisch erweitern führt hier zu Verbesserung

Großgattung → Unterschiedliche Themen am Ende

Operationalisierung

- ▶ Ende: Letzte Szene
- ▶ Clustering der Enden in verschiedenen Dramen basierend auf Themen
- ▶ Cluster-Zahl nicht festgelegt, wg. Untergattungen
- ▶ Messung der Cluster-Purity
 - ▶ Jedes Cluster wird der am häufigsten vorkommenden Gattung zugewiesen
 - ▶ Anteil der korrekt zugewiesenen Dramen

Ergebnisse



- ▶ Topic modelling gut für Gattungen
- ▶ Unterschiede zwischen zwei Gattungen werden gut per TM erkannt
- ▶ Mehr topics helfen für mehr Untergattungen

Schluss

Gesundheit in Dramen

- ▶ Um 1870 kommen Behandlungswörter in Komödien nicht mehr vor (in unserem Korpus)
- ▶ Gesundheitswörter kommen in Akt IV und V von Komödien seltener vor

Gesundheit in Dramen

- ▶ Um 1870 kommen Behandlungswörter in Komödien nicht mehr vor (in unserem Korpus)
- ▶ Gesundheitswörter kommen in Akt IV und V von Komödien seltener vor
- ▶ Folgeuntersuchungen
 - ▶ Wortfelder für Liebeswahn, Hysterie, Leidenschaft
 - ▶ Dornseiff zu allgemein und modern
 - ▶ Normvorstellungen 18./19. Jahrhundert
 - ▶ Tugendhaft = gesund, Leidenschaftlich = krank
 - ▶ Schlägt sich Freud-Rezeption (z.B. bei Schnitzler) in den Dramen nieder?

Themenerkennung

- ▶ Keine klare, eindeutig beste Methode (nicht zu erwarten)
- ▶ Topic modelling gut für Gattungsunterscheidungen geeignet
 - ▶ Hier: Keine Annahmen über Interpretierbarkeit der Gattungen
- ▶ Integration von Domänenwissen
 - ▶ Wortfeld-Definition: Nicht das Optimum

Meta-Methodik

- ▶ Erlaubt Benutzung von unscharfem Wissen für Methodenentwicklung
 - ▶ Keine “ground truth”
- ▶ To do
 - ▶ Mehr Variablenkontrollieren (z.B. Textlänge, Anzahl der Figuren)
 - ▶ Operationalisierung der Postulate nicht alternativlos
 - ▶ Validierung der Operationalisierungen durch synthetische Daten
 - ▶ Negativ-Postulate womöglich leichter zu finden
 - ▶ Interpretierbarkeit als Evaluationskriterium

Fragen?

Vielen Dank für's Zuhören!