Quantitative Drama Analytics

Heidelberg Computational Humanities Summer School

July 15, 2019

https://quadrama.github.io
Outline

Part 1: Theory and overview

Introduction (Marcus)
  ● QuaDramA: People
  ● QuaDramA: Research

Operationalisation: A key challenge in Computational Literary Studies (Benjamin)

Use cases in quantitative drama analysis
  ● Utterances (Janis)
  ● Semantics of character speech (Benjamin)
  ● Network analysis (Nils)

Part 2: Practice and lab session

Introduction (Nils)
  ● R Basics
  ● The R package DramaAnalysis

Lab session (Benjamin/Janis/Marcus/Nils)
Work by yourselves on the data, using a corpus of English or German plays

Wrap-Up (Nils)
Observations, insights, follow-ups, questions

quadrama.github.io/blog/2019/03/08/quadrama-tutorial.en
Introduction - QuaDramA: People

Dr. Nils Reiter

Janis Pagel

PhD

CL

PI

LS

Benjamin Krautter

Dr. Marcus Willand
Introduction - QuaDramA: People

Students (Theses):

Sonja Eberhardt
*Hebbel's fathers and children from a psychoanalytical perspective* [Working title]

Christiane Schneider
“You whistle / how the beak has grown for you” - *The Scaramutza character in the German-speaking drama*

Dominik Wabersich
*Especially Typical. For the understanding of the Faustfigure around 1800*

Nathalie Wiedmer
Automatic extraction and quantitative analysis of relations of dramatic characters

Nina Wittmann
*Character types in Expressionist drama* [Working title]
Introduction - QuaDramA: People

Students (Assistant):  

Annotation / (Meta) Data  
Sonja Eberhard, Nathalie Schürmann, Martin Glasbrenner, Ivan Vidakovic, Laura Friedrichsohn  
(Anja Braun, Dominik Wabersich, Martin Kuhn, Annika Haag, Anja Schmelzle, Tanja Preuß, Nina Stark, Alexander Frank)

Technical Support  
Tim Strohmayer
Introduction - QuaDramA: Research

Overall goal:

Test hypotheses about dramatic characters on large corpora using NLP methods
Introduction - QuaDramA: Research

Overall goal:

Test hypotheses about dramatic characters on large corpora using NLP methods

Hypothesis:
A proposed explanation for a phenomenon.
Introduction - QuaDramA: Research

Overall goal:

Test hypotheses about dramatic characters on large corpora using NLP methods

Hypothesis:
A proposed explanation for a phenomenon.

Scientific hypothesis:
Requires that one can test it.
Introduction - QuaDramA: Research

Overall goal:

Test hypotheses about dramatic characters on large corpora using NLP methods

Hypothesis:
A proposed explanation for a phenomenon.

Scientific hypothesis:
Requires that one can test it.

Hypothesis in literary studies?
Introduction - QuaDramA: Research

Overall goal:

Test hypotheses about dramatic characters on large corpora using NLP methods

Hypothesis:
A proposed explanation for a phenomenon.

Scientific hypothesis:
Requires that one can test it.

Hypothesis in literary studies?

The play is brilliant because god made its author brilliant
vs.

The play is brilliant because its characters speak in a very unique way
Introduction - QuaDramA: Research

Overall goal:

Test hypotheses about dramatic characters on large corpora using NLP methods

Hypothesis:
A proposed explanation for a phenomenon.

Scientific hypothesis:
Requires that one can test it.

Hypothesis in literary studies?

The play is brilliant because god made its author brilliant
vs.
The play is brilliant because its characters speak in a very unique way

aesthetics
structuralism
Introduction - QuaDramA: Research

Research objects:

Dramatic Texts
= Plays
≠ Prose
≠ Poetry

EIGENSCHAFTEN FIKT LIT → DLS vs xy

Operationalisierung als Grundherausforderung
Introduction - QuaDramA: Research

Research objects:

> 465 german dramatic texts (soon some more)
  Textgrid / DLINA (Frank Fischer, Peer Trilcke et al.)

~ 1740 - 1930

- canonical / non-canonical
- Tragedy / Comedy
  - subgenres

Corpus: Plays per Decade
**Introduction - QuaDramA: Research**

**structural drama research: non-digital past**

**CONFIGURATION MATRICES**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dirne</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Soldat</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Stubenmädchen</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Junger Herr</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Junge Frau</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ehemann</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Süßes Mädel</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Dichter</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Schauspielerin</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Graf</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Pfister: Das Drama (2001[1977]), S. 239 Schnitzlers *Round Dance*
Introduction - QuaDramA: Research

structural drama research: digital present

NETWORK-GRAPHS

https://dracor.org/ger/schnitzler-reigen
Introduction - QuaDramA: Research

structural drama research: digital present
Speech-based analysis
Introduction - QuaDramA: Research

Focus: Dramatic characters
3 Dimensions

● Character types
  ○ By gender, action, social class, stock character
  ○ How are fathers marked as being ‘tender’?

● Relations between characters
  ○ What are shared topics/emotions?
  ○ What do characters say about each other?

● Character type development
  ○ How do topics change?
  ○ How are relations changing?
Operationalization: A key challenge in Computational Literary Studies
Operationalization

Question: How to measure a theoretical concept?

- P. W. Bridgman: *Logic of Modern Physics* (1927)
  Example:
  “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.”

- Franco Moretti: “Operationalizing”: *or, the function of measurement in modern literary theory* (2013)
  “[...] In our case: from the concepts of literary theory, through some form of quantification, to literary texts.”
Operationalization

Question: How to measure a theoretical concept in literary studies?

- An adequate operationalization is important for the acceptance of the results in literary studies
- How to: direct operationalization and/or approximate operationalization

Some potential problems:

- Focus of the theoretical frame: author, text, reader?
- Different approaches: a concept as an analytic tool for text analysis or text interpretation to sharpen the concept
- Concepts range from being rather formalistic (e.g., narratology) to (deliberately) vague
- Oftentimes, literary concepts cannot be measured directly
Operationalization

Direct operationalization:
- Size of the stage personnel (how many characters?)
- Number of acts and/or scenes
- When are characters introduced?
- How do characters appear and disappear on stage? (*liaison de scène*)
- ...

Most of the times, these questions are relevant for the comparison within a large(r) text corpus, e.g., for studies in literary history.
Operationalization

Approximate Operationalization:

- Is it possible to operationalize vague concepts? How can we measure concepts that were not established with quantification in mind?
- Possible concepts:
  - Who is the protagonist of a literary text?
  - What are the themes/topics characters talk about?
  - Is it possible to distinguish different character types? (lover, father, villain,...)
  - Is it possible to determine the genre of a given text?
  - ...

- Poetological concepts: e.g., the concept of natural language, a character development (e.g., from Katilina to Brutus) or a specific aesthetic effect for the reader/spectator
Operationalization

Approximate Operationalization - **natural language**:

- Approach: split the concept in individual aspects that are easier to measure; combine the individual aspects to get an approximation
- criteria/Instrumental variables (Graham Sack: Simulating Plot, 2011)
  - sentence length
  - utterance length
  - use of punctuation (e.g., dashes)
  - use of vocabulary (type/token ratio)
  - ...

Operationalization through annotation

- **Goal**: try to sharpen the meaning/sense of a concept through manual annotation
- **Approach**: textual work on individual cases
- **Needs** annotation guidelines that are continuously updated: Guideline as applicable operationalization of the concept
- **Application** can result in a training corpus, e.g., for machine learning (second operationalization)
Use Cases in Quantitative Drama Analysis
Use Cases in Quantitative Drama Analysis

Detecting Protagonists
Protagonists

Need to make literary view operationalizable
Protagonists

Need to make literary view operationalizable

Need for clear guidelines for annotation
Protagonists

Need to make literary view operationalizable

Need for clear guidelines for annotation

Our definition:

1. Figure causes or solves the central dramatic conflict
2. Either actively or passively
Features

Come up with good features/characteristics that capture what it means/entails to be a protagonist
Group discussion

What could be reasonable features to characterize and classify protagonists in plays?

- Number of uttered tokens
- ...
- ...
- ...
- ...
- ...
Our features

- Features that we have experimented with
  - Uttered tokens
Our features

- Features that we have experimented with
  - Uttered tokens
  - Stage presence (active and passive)
Our features

- Features that we have experimented with
  - Uttered tokens
  - Stage presence (active and passive)
  - Centrality measures
    - (Weighted) Degree
    - Eigenvalue centrality
    - Betweenness
    - Closeness
Our features

- Features that we have experimented with
  - Uttered tokens
  - Stage presence (active and passive)
  - Centrality measures
    - (Weighted) Degree
    - Eigenvalue centrality
    - Betweenness
    - Closeness
  - Topic modeling
Our features

- Features that we have experimented with
  - Uttered tokens
  - Stage presence (active and passive)
  - Centrality measures
    - (Weighted) Degree
    - Eigenvalue centrality
    - Betweenness
    - Closeness
  - Topic modeling
  - Appears in last act?
Our features

● Features that we have experimented with
  ○ Uttered tokens
  ○ Stage presence (active and passive)
  ○ Centrality measures
    ■ (Weighted) Degree
    ■ Eigenvalue centrality
    ■ Betweenness
    ■ Closeness
  ○ Topic modeling
  ○ Appears in last act?
  ○ Genre/Epoch
Classification

Next step:

Train model to decide: protagonist or not-protagonist (based on features)
Feature importance

Results on 34 plays with 171 protagonists
Shapley analysis

Emilia Galotti
Use Cases in Quantitative Drama Analysis

Word field analysis - Kleist’s plays
Word field analysis

How can we measure prevalent themes in the character speech of German plays (1750-1830) that are interpretable?

- Dictionary based approach, created by a domain expert; currently 5 dictionaries:
  - Love (91 words)
  - Family (73 words)
  - Reason (108 words)
  - War (112 words)
  - Religion (57 words)
  - others are wip (politics, economy, …)
Word field analysis

- adore, adored, adoring, sincere, sincerity, covetousness, desire, eagerness, mourn, relationship, burning, ceremony, copulation, copulate, honest, feeling, sentiment, angel, delight, feeling, spark, beloved, lover, spouse, wife, luck, marriage, heart, wedding, homage, cuddle, kiss, passion, love, gracious, graciousness, caress, lip, lust, rose, beauty, soul, sensual, sensuality, marriage, marry, tender, tenderness, ceremony, lovest, [...]


Word field analysis - single character

![Graph showing word frequency analysis]
Word field analysis - Kleist’s *Familie Schroffenstein*
Word field analysis - Kleist's plays

Die Familie Schroffenstein
Amphitryon
Der zerbrote Krug
Penthesilea
Das Käthchen von Heilbronn
Die Hermannsschlacht
Prinz Friedrich von Homburg

- Love
- War
- Family
- Reason
- Religion
Word field analysis - Kleist’s plays
Word field analysis - Kleist’s plays
Word field analysis - Kleist’s plays
Use Cases in Quantitative Drama Analysis

Relations and Character Presence
Configuration

- Represents the character stage presence
  - Right: Romeo & Juliet, by scene
- Extension in two directions:
  - Copresence networks
  - Passive presence

<table>
<thead>
<tr>
<th>character</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benvolio</td>
<td>TRUE</td>
<td>TRUE</td>
<td>FALSE</td>
<td>TRUE</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>Tybalt</td>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>TRUE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Capulet</td>
<td>TRUE</td>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>TRUE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Lady Capulet</td>
<td>TRUE</td>
<td>FALSE</td>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Montague</td>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Escalus</td>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Romeo</td>
<td>TRUE</td>
<td>TRUE</td>
<td>FALSE</td>
<td>TRUE</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>Paris</td>
<td>FALSE</td>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Nurse</td>
<td>FALSE</td>
<td>FALSE</td>
<td>TRUE</td>
<td>FALSE</td>
<td>TRUE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Juliet</td>
<td>FALSE</td>
<td>FALSE</td>
<td>TRUE</td>
<td>FALSE</td>
<td>TRUE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Mercutio</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>TRUE</td>
<td>FALSE</td>
<td>TRUE</td>
</tr>
<tr>
<td>Friar Lawrence</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
</tbody>
</table>
# Configuration → Copresence Network

<table>
<thead>
<tr>
<th>Character</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Romeo</td>
<td>✅</td>
<td>✅</td>
<td>✗</td>
<td>✅</td>
<td>✗</td>
</tr>
<tr>
<td>Juliet</td>
<td>✗</td>
<td>✅</td>
<td>✅</td>
<td>✅</td>
<td>✅</td>
</tr>
<tr>
<td>Paris</td>
<td>✗</td>
<td>✗</td>
<td>✅</td>
<td>✗</td>
<td>✗</td>
</tr>
</tbody>
</table>

1: Configuration
Configuration → Copresence Network

1: Configuration

<table>
<thead>
<tr>
<th>Character</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Romeo</td>
<td>✅</td>
<td>✅</td>
<td>✗</td>
<td>✅</td>
<td>✗</td>
</tr>
<tr>
<td>Juliet</td>
<td>✗</td>
<td>✅</td>
<td>✅</td>
<td>✅</td>
<td>✅</td>
</tr>
<tr>
<td>Paris</td>
<td>✗</td>
<td>✗</td>
<td>✅</td>
<td>✗</td>
<td>✗</td>
</tr>
</tbody>
</table>

2: Adjacency matrix

<table>
<thead>
<tr>
<th></th>
<th>Romeo</th>
<th>Juliet</th>
<th>Paris</th>
</tr>
</thead>
<tbody>
<tr>
<td>Romeo</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Juliet</td>
<td>2</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Paris</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
**Configuration → Copresence Network**

1: Configuration

<table>
<thead>
<tr>
<th>Character</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Romeo</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>Juliet</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Paris</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
</tr>
</tbody>
</table>

2: Adjacency matrix

<table>
<thead>
<tr>
<th></th>
<th>Romeo</th>
<th>Juliet</th>
<th>Paris</th>
</tr>
</thead>
<tbody>
<tr>
<td>Romeo</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Juliet</td>
<td>2</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Paris</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

3: Copresence network
Copresence Network

Network metrics

- Avg. degree, centrality, densely connected clusters, …

Trilcke et al. (2015)
Copresence Network

Network metrics

- Avg. degree, centrality, densely connected clusters, …
- Drama-historic developments

Trilcke et al. (2015)
Copresence Network

Network metrics

- Avg. degree, centrality, densely connected clusters, …
- Drama-historic developments
- Character relations
- Visualisation

Trilcke et al. (2015)
Configuration and Presence

- Configuration represents the character stage presence
- Active presence: Character on stage
  - Misleading

Act I, Scene 1.

Prince. Complaints; nothing but complaints! Petitions; nothing but petitions! [...] Emilia? (opening a petition, and looking at the signature.) An Emilia? Yes - but an Emilia Bruneschi - not Galotti. Not Emilia Galotti. What does she want, this Emilia Bruneschi? (Reads) She asks much--too much. But her name is
Configuration and Presence

- Configuration represents the character stage presence
- Active presence: Character on stage
  - Misleading
- Passive presence: Character not on stage, but talked about
- Coreference resolution
  - Proper names
  - Pronouns
  - Nominal phrases

We are still working on that!

Act I, Scene 1.

Prince. Complaints; nothing but complaints! Petitions; nothing but petitions! [...] Emilia? (opening a petition, and looking at the signature.) An Emilia? Yes - but an Emilia Bruneschi - not Galotti. Not Emilia Galotti. What does she want, this Emilia Bruneschi? (Reads) She asks much--too much. But her name is Emilia. It is granted (signs the paper, and rings).
## Active and Passive Presence

<table>
<thead>
<tr>
<th>Character</th>
<th>Active scenes</th>
<th>Passive scenes</th>
<th>Presence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appiani</td>
<td>5</td>
<td>25</td>
<td>-0.465</td>
</tr>
<tr>
<td>Claudia</td>
<td>13</td>
<td>12</td>
<td>0.023</td>
</tr>
<tr>
<td>Der Prinz</td>
<td>14</td>
<td>26</td>
<td>-0.279</td>
</tr>
<tr>
<td>Emilia</td>
<td>7</td>
<td>34</td>
<td>-0.627</td>
</tr>
<tr>
<td>Marinelli</td>
<td>19</td>
<td>23</td>
<td>-0.093</td>
</tr>
<tr>
<td>Odoardo</td>
<td>12</td>
<td>20</td>
<td>-0.186</td>
</tr>
<tr>
<td>Orsina</td>
<td>6</td>
<td>16</td>
<td>-0.232</td>
</tr>
</tbody>
</table>
Presence of Non-Characters

Coreference annotation also gives us a grasp on objects and abstract entities
References


Final Discussion