# Towards Tool Criticism: Complementing Manual with Computational Literary Analyses

Melina L. Jander

July 9-12, 2019

*DH 2019: Complexities*

**eTRAP**
Electronic Text Reuse Acquisition Project
GÖTTINGEN CENTRE FOR
DIGITAL HUMANITIES

GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

# TABLE OF CONTENTS

# Introduction

"Tool criticism is the critical inquiry of knowledge technologies used for research purposes. It reviews the qualities of the tool in light of the research activities and reflects on how the tool [...] affects the research process and output." (Van Es et al., 2018)

## RESEARCH OBJECTIVES

1. Approach tool criticism through two prominent applications: Topic Modelling (TM) and Sentiment Analysis (SA).

2. Use TM and SA tools on a self-designed corpus of literary relevance: dystopian novels of the 19th and 20th century.

# Research Data

| Languages | Novels | Tokens | Types | Token-type ratio |
|---|---|---|---|---|
| American English | 39 | 3,167,702 | 136,954 | 23.1 |
| British English | 35 | 2,660,983 | 112,012 | 23.8 |
| German | 28 | 1,872,969 | 98,497 | 19.1 |
| TOTAL | 102 | 7,701,654 | 480,651 | 16.1 |

Table 1: Overview of the data-set.

The novels were published between 1836 and 1979.
→ Time span of 143 years.

| Novel | Language | Year | Tokens | Types | Token-type ratio |
|---|---|---|---|---|---|
| Zamyatin: *We* | AE | 1924 | 63,161 | 6,377 | 9.9 |
| Kafka: *Der Prozeß* | GER | 1925 | 71,835 | 7,866 | 9.1 |
| Huxley: *Brave New World* | BE | 1932 | 64,130 | 8,231 | 7.8 |
| Orwell: *Nineteen Eighty-Four* | BE | 1949 | 104,205 | 9,215 | 11.4 |
| Bradbury: *Fahrenheit 451* | AE | 1953 | 45,857 | 5,173 | 8.9 |

Table 2: "Canonical" dystopian novels.

A well specified canon of dystopian novels does not exist. The genre accumulates many other literary kinds:

Science Fiction, Utopias, Feminist Fiction, Post-apocalyptic Fiction, etc.

# Experiments

**Topic Modelling**:

- MALLET (McCallum, 2002)
- Stanford Topic Modeling Toolbox (Ramage et a., 2009)

"MALLET is a Java-based package for statistical natural language processing, document classification, clustering, topic modeling, information extraction, and other machine learning applications to text." (McCallum 2002)

# MALLET

## Experiment 1

1. Research data: Cleaned English part of the corpus (.txt files).
2. Settings: MALLET implementations.
3. Number of topics: Ranging between 10 and 100.

## Experiment 2

1. Research data: Cleaned English part of the corpus (.txt files).
2. Settings: MALLET implementations and a modified stop word list.
   → Personal names.
3. Number of topics: 30.

# Stanford Topic Modeling Toolbox

"The Stanford Topic Modeling Toolbox (TMT) brings topic modeling tools to social scientists and others who wish to perform analysis on datasets that have a substantial textual component." (Ramage & Rosen 2009)

# Stanford Topic Modeling Toolbox

## Experiment 1

1. Research data: Cleaned English part of the corpus (.csv files).
2. Settings: Can be changed individually → No stop word list **but** 30 most common terms are filtered out.
3. Number of topics: Ranging between 10 and 100.

## Experiment 2

1. Research data: Cleaned English part of the corpus (.csv files).
2. Settings: Remained the same.
3. Number of topics: 30.

**Sentiment Analysis**:

- Stanford Core NLP Sentiment Annotator (Socher et al., 2013)
- Berlin Affective Word List - Reloaded (Võ et al., 2009)

# STANFORD SENTIMENT ANNOTATOR

- Recursive Neural Tensor Network: Prediction accuracy of fine-grained sentiment labels reaches 80.7%.
- Training data: Stanford Sentiment Treebank.
  $\rightarrow$ Includes labels for 215,154 phrases in the parse trees of 11,855 sentences.

## STANFORD SENTIMENT ANNOTATOR

**Experiment**

1. Research data: Cleaned English part of the corpus (.txt files).
2. Settings: Stanford Core NLP Sentiment Annotator implementations.

   → Rating system:
   Very negative - Negative - Neutral - Positive - Very positive

# BERLIN AFFECTIVE WORD LIST RELOADED

- A list of more than 2,900 German words.
  $\rightarrow$ Represents negative, neutral and positive affective valences.

- Annotation: 200 psychology students.
- Categories:
  - Emotional valence
  - Arousal valence
  - Imageability valence

# BERLIN AFFECTIVE WORD LIST RELOADED

**Experiment**

1. Research data: Cleaned German part of the corpus (.txt files).
2. Settings: A dictionary-based algorithm.
   It scans the novels for the terms present in the BAWL-R and analyses their different valences (Roth-Kleyer, 2018).

   → Rating system:
   Depends on the analysed valence.

# Outputs

# MALLET

| | Terms | Potential Topic Labels |
|---|---|---|
| 1 | rubashov party krug winston cell o'brien gletkin round ivanov voice big rubashov's revolution war one's tele-screen political history started newspeak | Nineteen Eighty-Four; Darkness at Noon |
| 2 | alfred arctor barris fred knight hermann thought german von women donna luckman bob hess hitler men germany house girl god | Swastika Night; A Scanner Darkly |
| 3 | alvin donald norman diaspar siegmund hilvar urbmon lys chad city beninia sugaiguntung khedron people elihu building jason jeserac human computer | The City and the Stars; The World Inside; Stand on Zanzibar |
| 4 | dream orr haber bernard poole savage lenina boz shrimp round lottie hanson she'd milly director he'd birdie george dreams linda | The Lathe of Heaven; Brave New World; 334 |
| 5 | francis brother abbot god priest father percy thon abbey holy rome church lord monk paulo leibowitz order dom paused saint | A Canticle for Leibowitz; Christianity related terms |

Table 3: The 5 most meaningful topics out of 30.

| | Terms | Potential Topic Labels |
|---|---|---|
| 1 | party voice stood head knew felt turned door asked years cell answer heard light sat new left wall saw revolution | Nineteen Eighty-Four; Verbs in past tense |
| 2 | alfred knight women german von knights hitler god germany really boy germans english book course christian nazi shall years joseph | Swastika Night; Human life |
| 3 | didnt youre thats new course donald ive wasnt hed want guess hes mr mind norman id theres home hell let | Stand on Zanzibar; Contractions [LING.] |
| 4 | dr says savage bernard voice suddenly course young head silence shot room words round hand mr director world hands linda | Brave New World; Human life |
| 5 | brother francis god father priest new abbey rome church world tell monk lord order holy thats saint paused ill began | A Canticle for Leibowitz; Christianity related terms |

Table 4: 5 topics out of 30 that match the MALLET topics.

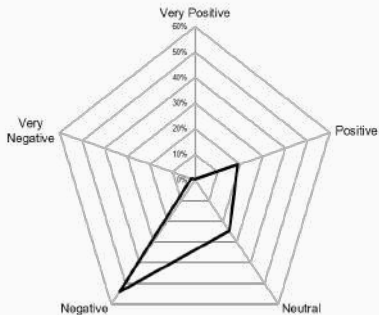# STANFORD SENTIMENT ANNOTATOR

American English:

British English:



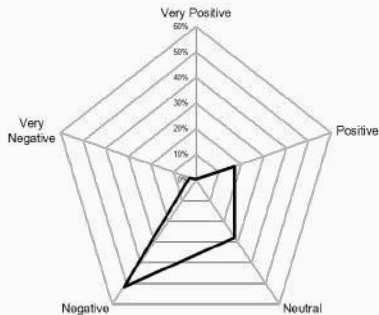Figure 1: Sentiment distribution in the American part of the corpus.



Figure 2: Sentiment distribution in the British part of the corpus.

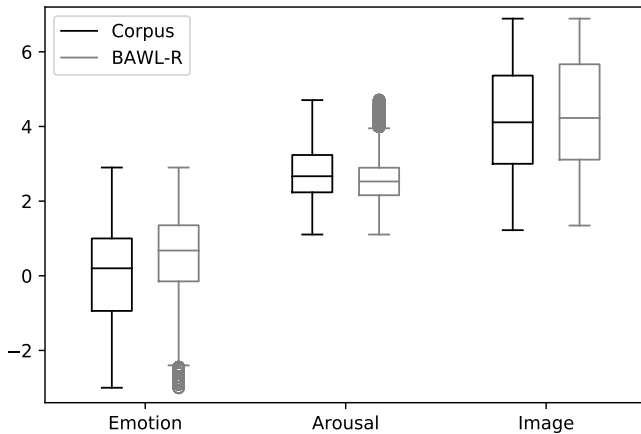Figure 3: Comparing the values of **German dystopian novels** and the BAWL-R dictionary.

# Conclusions

**Computational analysis:**

- TM tools can represent certain novels of the research corpus.

**Manual analysis:**

- Core theme of DNs: The world's gone awry.

**Concerns:**

- Modifying the stop word list.
- Labelling the topics.

→ **To sum up:** TM reveals only rudimental insights about our corpus of DNs.

# SENTIMENT ANALYSIS

**Computational analysis:**

- DNs carry primarily negative sentiments.
- Extreme emotions are rare.

**Manual analysis:**

- DNs are pessimistic literature.
- Sarcasm, irony and cynicism are common stylistic devices.

**Concerns:**

- Multiple meanings of text: The context should not be ignored.
- Interpretations and emotions change according to a person's background and personality.
- SA tools cannot detect implicit meanings.

→ **Question:** How can we analyse sentiments in literary texts holistically?

**Hypothesis**: The sentiments a text carries depend on the reader's cultural and social background.

→ **Experiment:** Sentiment Annotation

1. Choose a 'canonical' dystopian novel: *Brave New World* (Huxley, 1932).
2. Use Amazon Mechanical Turk to globally reach out to people for annotating key paragraphs of the novel.
3. Evaluate the test persons' demographic information.

**Thank you!**

The theme this presentation is based on is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License. Changes to the theme are the work of eTRAP.