MIKE KESTEMONT (UNIVERSITY OF ANTWERP)
GUNTHER MARTENS (GHENT UNIVERSITY)
THORSTEN RIES (GHENT UNIVERSITY)

A CHALLENGE FOR STYLOMETRY AND AUTHORSHIP ATTRIBUTION METHODS: GOETHE’S CONTRIBUTIONS TO THE FRANKFURTER GELEHRTE ANZEIGEN 1772/73
The Problem

- J. W. Goethe was editor-in-chief of the *Frankfurter Gelehrte Anzeigen* 1772 (J. C. Deinet bought the *Frankfurter Gelehrtenzeitung* in 1771)
- *Frankfurter gelehrte Anzeigen* (FgA) became a flagship of the “Sturm und Drang” movement in this year because of its editors Goethe, Johann Heinrich Merck and Johann Georg Schlosser and contributors such as Johann Gottfried Herder.
- Goethe most likely also wrote for the FgA in 1773.
- The “Rezensionen” (i.e. articles) of the FgA were published anonymously and have often been redacted by the editors, some have even been written collaboratively.
The Problem

- 900 pages of anonymous journal text, authorship of a lot of the nearly 400 Rezensionen by around 40 contributors of 1-7pp length is unclear or authorship attribution relies on shaky hermeneutic arguments.
- It is in many cases unclear which have been penned by Goethe, which ones were heavily redacted by him, and which ones have been collaborative “protocol reviews”.
- His self-attribution of some FgA-Rezensionen in his self-edited edition of his works is regarded as unreliable (1772/73: 35).
- The majority of texts have not been tested systematically at all.
The Project

• Computational stylometrics and authorship attribution
  • Burrows’s *Delta*
  • Mike Kestemont (stylo, imposters method).
• Check all Rezensionen in the FgA 1772/73 with computational stylometric methods to verify whether Goethe wrote them or not.
• The advantages
  • statistical method to detect the stylistic footprint
    • tested on large corpora and was trained on large corpora by the authors.
    • Imposters Method: very good accuracy
• We hope to attribute new texts to Goethe, correct previous false positives, a new scientific foundation to previous correct authorship attributions (poorly tested, small data basis).
Early Approaches

- Max Morris and Hermann Bräuning-Octavio endowed large parts of their academic life to this authorship attribution question, gathering all philological evidence, producing several 800 pp thick monographs. Otto Trieloff and Wilhelm Scherer joined the conversation.

- Rather vague notions of style and thematic preference, hermeneutic arguments, recurrence of opinions and topics, individual spelling characteristics (“warrlich”, “Shäckespear”). Prose rhythm (Karl Marbe, 1904, 1912, without success).

- Statistical, stylometrical, stylistic approaches have been tried at small scale, on small samples with a very limited basis and methodological foundation.
Early Approaches

Specific for Goethe?

“Schäckespear”
Statistical and Stylistic Approaches

- Bräuning-Octavio 1966: set of ‘typical features’ of Goethe’s style; language rhythm and melody, favourite expressions, rhetorical features such as (vague) specifics of exclamation, questions, address, double negation, accumulation and enumeration, anaphora, parenthesis, typical Rezensionen beginning, Goethe’s grammar during the “Werther Periode”, sentences omitting verb, parallelisms, inversion, emphatic sentence endings, latin quotes etc.

- The results - beyond the direct philological proof found - remained vague.

- But Bräuning-Octavio already worked on a prototype of stylometrics, as his private archive collection in the Archive of the Technische Universität Darmstadt shows
In Bräuning-Octavio's archive

"Statistik der Füllwörter in den FGA"

Statistics of the expletives in the FgA


24,2: Statistik der Füllwörter in den FGA 1772 mit Vorarbeiten ohne die notwendige (kostspielige) elektronische Auswertung. Dazu ein Karteikasten.


24,3: Katalog der Herzogin Caroline von Pfalz-Zweibrücken. Xerokopie. - Katalog der Bibliothek Ludwigs IX., Xeroxnebst Materialien.
Statistical and Stylistic Approaches


Statistical and Stylistic Approaches

**Herbert Sparmann 1970:** Tried to distinguish Goethe from Merck by the frequency of the use of the definite article, finding Merck uses the **definite article** 40% more frequent than Goethe.

**Very small corpus,** taken from FgA!
Statistical and Stylistic Approaches


- A profile categorising **frequency of word function**: nouns, adjectives, and lexicon variation; analysis of words in 1st, 2nd, last position in the sentence.

- First **computational approach**! Using LDVLIB by R. Drewek, an early textstatistical processor about which you hardly find anything but mentions in books online.

- Very **small corpus**, taken from the FgA!
FgA Challenges for Computational and Stylometric Approaches

- **Corpus acquisition**: OCR – German Fraktur of the 18th century, specific training of engines needed. Consequently, our corpus was partly ‘dirty’.
- **Length** of the Rezensionen varies between 1 page to 7 pages, the shortness of samples may be a problem for authorship attribution.
- Goethe has - in his role of an editor-in-chief - certainly **redacted** some or many Rezensionen by others.
- Corpus: the co-editors have **not written as much as Goethe**
- Corpus: Goethe’s **style might have changed** over the years (from a literary perspective for sure, from a statistical perspective, we don’t know)
We took four examples from the FgA for a test drive (blind test):

<table>
<thead>
<tr>
<th>Title</th>
<th>Length in pp</th>
<th>Goethe self-attributed?</th>
<th>Haenelt</th>
<th>Kestemont, Martens, Ries.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cymbelline, ein Trauerspiel, nach einem von Schäckespear erfundnen Stoffe.</td>
<td>2</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empfindsame Reisen durch Deutschland von S. 2ter Theil. Bey Zimmermann ...</td>
<td>4</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Essays on song-writing: with a collection of such Englisch Songs, as are most eminent for poetical merit. [...]</td>
<td>5</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Die schönen Künste in ihrem Ursprung, ihrer wahren Natur und besten Anwendung, betrachtet von J. G. Sulzer. 1772</td>
<td>7</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Authorship attribution

Anonymous
document X

[Stamatatos 2009]
Represent documents in bag-of-words table

<table>
<thead>
<tr>
<th>Item 1</th>
<th>Item 2</th>
<th>Item 3</th>
<th>Item 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Document 1</td>
<td>10</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>Document 2</td>
<td>2</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>Document 3</td>
<td>7</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Document 4</td>
<td>12</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

“Vocabulary” of features

Find ‘nearest neighbor’ using a distance metric

[Argamon 2008; Burrows 2002; Evert et al. 2017; Sebastiani 2002]
Authorship verification

Anonymous document X

Or somebody else…

Author A

Author B

Author C
Authorship verification

Anonymous document X

yes / no

Author A

yes / no

Author B

yes / no

Author C

Anonymous

document X
Imposters approach

[Remote Figure]

Vocabulary

[Remote Figure]

[Author A]

Anonymous document X

Imposters “pool”

[Remote Figure]

[Koppel & Winter 2014]
Repeat e.g. 100 times

Random selection
e.g. 50%

Vocabulary

Anonymous
document X

Author A

Closest?

Imposters “pool”

Random selection
e.g. 100 imposters

[Koppel & Winter 2014]
• Bootstrapped or stochastic process: $n$ samples in two dimensions
• Single verification score: e.g. 15/100 vs 87/100
• Apply threshold: e.g. $\geq 25$ -> attribute
• Intuition: documents by same author are similar across random samples from the vocabulary and more similar than other random selections of texts
• Good results in competitions (e.g. PAN)
Imposter selection

• Main difficulty: come with good pool of imposters (cf. police line-up)
• As similar as possible to test and train texts, in terms of genre, date, etc.
• But not too similar either…
• Rezensionen from the Frankfurter Gelehrten Anzeigen would be ideal, but problematic because all anonymous…
• Restricted to Goethe (target author) vs. Herder & Schlosser (imposters)
Calibration: development results

- Segment to shortest test sample size (=2,102 words)
- Set apart development set (20% of documents)
- Verify authorship of development set
- Evaluate accuracy (and F1-score) of verifications
- Sampling clearly helps, across both dimensions

<table>
<thead>
<tr>
<th></th>
<th>baseline</th>
<th>+ features (50%)</th>
<th>+ features (50%) + imposters (250)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>92.48</td>
<td>95.01</td>
<td>98.02</td>
</tr>
<tr>
<td>F1-score (macro)</td>
<td>92.34</td>
<td>94.77</td>
<td>97.83</td>
</tr>
</tbody>
</table>
Test

• Apply calibrated system to:
  • verified Herder (7)
  • verified Goethe (4)
  • unverified texts (4)
• Optimal thresholds:
  • >= 25: Goethe
  • <= 16: Herder
• Test scores: .51 (solid “yes”) to 0.1 (solid “no”); two in between (.41, .31)
• Send email to Thorsten… (Unbiased)
## FgA Test Case(s)

- **Summary of test results:**

<table>
<thead>
<tr>
<th>Title</th>
<th>Haenelt</th>
<th>Kestemont, Martens, Ries</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Cymbelline […]</td>
<td>Highest probability: <strong>Herder</strong></td>
<td>borderline case, wouldn't bet on it, but it does reach the attribution threshold in this experiment.</td>
<td>Same result: it is unclear. It is reasonable to assume that there are features of all three authors in here: collaboration or redaction by multiple authors. The result seems to corroborate some Goethe impact here.</td>
</tr>
<tr>
<td>Features: 2xMerck, 2xHerder, 1xGoethe</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Empfindsame Reisen durch Deutschland […]</td>
<td>Highest probability: <strong>Goethe</strong></td>
<td>really big chance that it's <strong>Goethe</strong></td>
<td>Same result: Even clearer here than Haenelt's, where the non-Goethe feature was the vocab composition and distribution, which is one of the most style-determining feature in her matrix.</td>
</tr>
<tr>
<td>Features: 4xGoethe, 1xHerder</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Essays on song-writing: […]</td>
<td><strong>Herder</strong></td>
<td>very <strong>unlikely that it's Goethe</strong></td>
<td>Same result.</td>
</tr>
<tr>
<td>(Haenelt is sure)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Features: 5x Herder</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Die schönen Künste in ihrem Ursprung, […]</td>
<td><strong>Goethe</strong></td>
<td><strong>reasonable chance that it's Goethe</strong></td>
<td>Almost same result.</td>
</tr>
<tr>
<td>(Haenelt is sure)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Features: 5x Goethe</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Selected references

- Evert et al., Understanding and explaining Delta measures for authorship attribution, *DSH* [2017].
- Kestemont et al., Authenticating the writings of Julius Caesar. *ESWA* (2016).
- Koppel & Winter, Determining if Two Documents are by the Same Author, *JASIST* (2014).