A Question of Style: individual voices and corporate identity in the Edinburgh Review, 1814-1820

Conference or Workshop Item

How to cite:


For guidance on citations see FAQs.

© 2016 The Open University

Version: Version of Record

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online’s data policy on reuse of materials please consult the policies page.
A Question of Style: individual voices and corporate identity in the *Edinburgh Review*, 1814-1820

Francesca Benatti and David King
The Open University
Research question

Did the *Edinburgh Review* create a “transauthorial discourse” (Klancher 1987) that hid the voices of individual contributors behind a corporate style?

Funded by the Research Society for Victorian Periodicals Field Development Grant (January-October 2017)
The *Edinburgh Review*

Most influential periodical in early 19th C.

Edited by Francis Jeffrey, who could make alterations to any article

All articles published anonymously
Existing Corpus

*Edinburgh Review:*
- 45 articles
- 10 authors and one anonymous article
- 269,622 ‘words’

Preparation:
1. OCR with manual curation
2. TEI manual mark-up
3. attention to quotations
Stylometry

The study of how hidden stylistic traits can be measured through statistical methods to trace an author's voice

Made better known by John Burrows in his 2001 Busa Award lectures and beyond

Perception of authorial “voice” is quite subjective

• e.g. Duncan Wu (Introduction, *New Writings of William Hazlitt*, 2007)
Two interpretations of style*

<table>
<thead>
<tr>
<th>Style as fingerprint</th>
<th>Style as signature</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unconscious</strong> elements in the way we write</td>
<td><strong>Conscious</strong> choice of words, sentences, tone</td>
</tr>
<tr>
<td>(e.g. Van Halteren et al. &quot;Existence of a human stylome.&quot; (2005))</td>
<td>(e.g. Van Dalen-Oskam <em>Riddle of Literary Quality</em> project)</td>
</tr>
<tr>
<td>Reflected by use of <strong>Most Frequent Words</strong></td>
<td>Still <strong>unsure</strong> how to identify with stylometry</td>
</tr>
</tbody>
</table>

* as defined by Sarah Allison at DH2016, Stylistics workshop, 12 July 2016
Signature - possible routes

Van Dalen-Oskam
- vocabulary richness?
- word length?
- sentence length?

Allison
- medium-frequency words?
- words used vs. words avoided?
Fingerprint - Delta method

“Delta is the mean of the absolute differences between the z-scores for a set of word-variables in a given text-group and the z-scores for the same set of word-variables in a target text.”

Delta continued

Delta works on the Most Frequent Words present in a given set of texts

All authors use Most Frequent Words differently

Underpinned by solid mathematical and linguistic foundations
## Delta - example

<table>
<thead>
<tr>
<th>Word</th>
<th>Moore</th>
<th>Coleridge</th>
<th>Godwin</th>
<th>Southey</th>
</tr>
</thead>
<tbody>
<tr>
<td>the</td>
<td>7.71</td>
<td>6.4</td>
<td>6.9</td>
<td>7.69</td>
</tr>
<tr>
<td>of</td>
<td>5.85</td>
<td>5.06</td>
<td>4.49</td>
<td>3.54</td>
</tr>
<tr>
<td>and</td>
<td>2.83</td>
<td>3.95</td>
<td>3.52</td>
<td>3.15</td>
</tr>
<tr>
<td>to</td>
<td>2.97</td>
<td>3.04</td>
<td>3.01</td>
<td>3.11</td>
</tr>
</tbody>
</table>
Data exploration with multidimensional scaling — spot the cluster
False clusters

Female pronouns

- Moore_French_Novels_34_1820_corr 36%
- Jeffrey_Edgeworth_28_1817 33%
- anon_christabel_edinburgh_review_27_1816 32%
- Jeffrey_Lalla_Rookh_29_1817 23%
- Brougham_melanges_30_1818 21%

...and 10 texts contained no female pronouns at all
Increasing rigour

With clustering techniques that
• rely on random seeding, the results depend too heavily on the random starting point
• have parameters, the results depend too heavily on those parameters

Therefore, applied
• both *agglomerative* (hierarchy) and *partition* (kmeans) clustering techniques
• drilled down through two feature sets initially (lexical, POS), and later a third (tf:idf)
Two weak clusters emerge
## MFW vs TF:IDF

Both attempt to remove the influence of content over style in the analysis.

<table>
<thead>
<tr>
<th>MFW</th>
<th>TF:IDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequent words</td>
<td>Significant words</td>
</tr>
<tr>
<td>Choose what to <em>include</em> in the analysis</td>
<td>Choose what to <em>exclude</em> from the analysis</td>
</tr>
<tr>
<td>Unconscious style?</td>
<td>Conscious style?</td>
</tr>
</tbody>
</table>
Future work

Extend corpus:
  • Python toolset to assist
    • OCR correction
    • TEI markup

Further methods:
  • corpus stylistics
  • Burrows’ Zeta and Iota