(One Year After) A Question of Style: individual voices and corporate identity in the *Edinburgh Review*, 1814-1820

Dr Francesca Benatti and Dr David King, The Open University
Introduction

A Question of Style in brief
• Winner, 2016 RSVP Field Development Grant
  • (thank you!!!)

• Research question:
  • Did the *Edinburgh Review* create a “transauthorial discourse” (Klancher) that hid individual authorial voices behind an impersonal corporate style?
INTRODUCTION
A Question of Style

• When last we saw our heroes in July 2017…

• David presented at RSVP 2017 in Freiburg on
  • corpus selection
  • OCR correction

• Francesca presented at SHARP 2017 (here in Victoria) and BARS 2017 in York on
  • authorship
  • computational criticism

• Where are we now?
Corpus creation

The composition and rationale of our corpus
CORPUS
Size, composition, rationale

• Driven by research goal:
  • Analyse the style of ER
  • Compare it to style of QR

• Driven by prior research on Christabel review:
  • Focus on literature, history, travel

• Emergent research questions:
  • Does multiple authorship hide style? (6 multi-authored articles)
  • Are reviews of the same text similar to one another? (10 pairs of reviews)
  • Does the genre of the text being reviewed influence the style of the review? (3 genres)
OCR correction

Post- Optical Character Recognition processing
OCR

Post-OCR processing

- **Aim:** Develop semi-automated process for OCR correction
- **Problems:** OCR errors too inconsistent for automation
  - Individual spelling choices
    - Publick
  - Regional identities
    - Perswaded
  - Political perspectives
    - Breslaw/Breslau
  - Language transformation
    - Shakspear, Shakspeare, Shakespear, Shakespeare
- **Solution:** David reviewed all automated corrections and “spelling mistakes” against the digitised source image
Quotations

Or, what is in a review?
QUOTATIONS
Or, what is in a review?
QUOTATIONS
Or, what is in a review?

Quotations % Max and Min

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Quote % Max</th>
<th>Quote % Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERCorp</td>
<td>79%</td>
<td>1%</td>
</tr>
<tr>
<td>QRCorp</td>
<td>65%</td>
<td>11%</td>
</tr>
</tbody>
</table>
Or, what is in a review?

• The presence of quotations is a problem for our chosen analytical methods

• e.g. we want to analyse Francis Jeffrey’s style (reviewer), not Walter Scott’s (reviewed)

• Solution:
  • David marked quotations using TEI XML <quote> element
  • Then we removed quotations using XSL transformation

• This reduces the total size of the corpora to 512,702 words
Analysis

Stylometry and corpus stylistics
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> (MFW)</td>
<td>Still <strong>working out</strong> how to identify with computational methods</td>
</tr>
</tbody>
</table>

* as defined by Sarah Allison at DH2016, Stylistics workshop, 12 July 2016
• Style as Fingerprint:
  • Stylometry

• Style as signature:
  • Corpus stylistics
    • Keywords
    • N-grams
    • Clusters
Stylometry in brief

- 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

- Generally concerned with authorship attribution but increasingly used to study authorship more broadly

- Burrows’ s Delta method implemented by Eder, Rybicki and Kestemont’s Stylo software package

- Improved method Cosine Delta developed by University of Würzburg

- Based on analysis of most frequent words
Stylometry: by genre (using Stylo (Eder, Rybicki, Kestemont); Cosine Delta; 300 MFW)

ANALYSIS

I. Hunt Not Surrey ER 27 1818
H. Croker Hobhouse Letters OR 14 28 1816
L. Hibbert Coleridge Literary Life ER 58 1817
M. Coleridge Lyassington ER 27 1816
L. Croker Keats Enquiry OR 19 3 1818
L. Croker Hunt Memoir OR 14 28 1816
L. Gifford Hazlitt-Shakespeare Q 16 26 1818
H. Brougham Phillips Speeches ER 29 1817
L. ColeridgeJ T Hazlitt Table OR 17 33 1817
L. Macaulay Wrigg Answer OR 28 25 1815
L. Gifford Hazlitt-Shakespeare OR 22 43 1819
L. Moore Boy Fathers ER 24 1814
L. Moore Hunt/Poems ER 23 1814
L. Moore Coleridge Christian ER 22 1816
T. Smith Heude Voyage ER 32 1819
L. Moore French Novels ER 34 1818
L. Croker-Nelson Morgan France OR 17 33 1817
L. Croker Shelley Frankenstein OR 18 26 1817
L. Croker Scott Anhjuary ER 15 25 1815
M. Coleridge Godwin Drama ER 17 8 1817
L. ColeridgeJ T Hemans Poems OR 24 47 1820
T. Brougham Forsyth Italy ER 22 1814
L. Lamb Wordsworth White Doe OR 14 27 1815
L. Scott Lord of the Isles OR 30 20 1810
T. Smith Travellers in America ER 31 1818
H. Smith Ireland ER 34 1820
H. Brougham Restorning Poland ER 23 1814
T. Cockburn Julian's Papers OR 26 1816
H. Brougham Nelson Letters ER 24 1814
H. Brougham Croydon Defense ER 25 1815
T. Brougham Shepard Paris ER 23 1814
H. Brougham Franklin Correspondence ER 28 18
L. Brougham Junius Letters ER 29 1817
T. Brougham Columbus ER 21 1810
Brougham Park Journey ER 22 1815
T. Brougham Simple Germany ER 22 1814
H. Brougham Melanges ER 30 1819
H. Moore Jonson Travels Brougham's Turkey ER 25 1818
H. Brougham Memoir ER 32 1819
T. Italy Labumia Campaign in Russia OR
H. Allen Warden St Helen ER 27 1818
H. Payler Lisbon Campaign in Russia ER 2
H. Smith Geographical Memoir ER 30 1819
H. Patmore Goethe part ER 20 1817
H. Patmore Goethe part ER 20 1816
P. Patmore Northern Poetry ER 20 1816
H. Allen Hallam Middle Age ER 35 1818
H. Smith Burghardt Nubia ER 34 1820
H. Smith Bowditch Assemblies ER 32 1819
T. Playfair Bright Travels in Hungary ER 31 1818
T. Birmi Bumford American Researches D
T. Barrow Burghardt Nubia OR 22 44 1820
G. Southern Lewis-Clarke American
T. Smith Quem Botany Bay ER 34 1820
L. MacRae Dowd Slavery ER 22 1
L. MacRae Rogers Poems ER 22 1815
L. Mackintosh Godwin Milton ER 20 1815
L. MacRae Stoddart Novels ER 26 1816
L. Hadlil Schlegel Drama ER 26 1816
H. MacRae Seminary Literature ER 26 1815
T. Scott Maturin Women ER 30 1819
T. Scott Maturin Emma ER 14 27 1815
L. Scott Enquiry-Scotic Stoddard OR 16 32 1
L. Gifford-Onford Cloane Poems OR 23 45 1820
T. Scott Byron Childs Hours IV OR 19
T. Scott Byron Childs Hours III OR 16
T. Wilson Wilsons Haplo OR 16 1816
L. Jeffrey Wortworth White Doe ER 25 1815
T. Jeffrey Wortworth Excursion ER 24 1816
T. Jeffrey Scott of the Isle ER 24 1815
T. Scott Montgomery Fool OR 11 21 1814
L. Scott Hunt Rumin ER 20 1815
T. Jeffrey Keats Endymion ER 34 1820
T. Jeffrey Hazlitt Shakespeare ER 25 1817
T. Jeffrey Elgarworth Tales ER 26 1817
T. Jeffrey Wilson City Plague ER 20 1816
L. Jeffrey Rogers Human Life ER 31 1819
T. Jeffrey Moore Lalla Rookh ER 29 1817
T. Jeffrey Byron Manifold ER 29 1817
T. Jeffrey Byron Consuel Bride ER 22 1814
Stylometry: by journal (using Stylo (Eder, Rybicki, Kestemont); Cosine Delta; 900 MFW)
Stylometry using Stylo (Eder, Rybicki, Kestemont); Cosine Delta; 300 MFW
Keywords

• Keywords are words that are significantly more frequent in the target text/corpus than in a chosen reference corpus

• Target corpus:
  • ERCorp
  • QRCorp

• Reference corpus:
  • 1780-1850 part of CLMET (Corpus of Late Modern English Texts; created by Hendrik de Smet)
  • 99 texts
  • 11 million words

• Used to represent (written) language norm for the period
• How do ERCorp and QRCorp differ from this baseline?
## ANALYSIS

Keywords: top 10 keywords, ERCorp and QRCorp

<table>
<thead>
<tr>
<th>Rank</th>
<th>ERCorp keyword</th>
<th>QRCorp keyword</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>of</td>
<td>the</td>
</tr>
<tr>
<td>2</td>
<td>the</td>
<td>of</td>
</tr>
<tr>
<td>3</td>
<td>cannot</td>
<td>cannot</td>
</tr>
<tr>
<td>4</td>
<td>we</td>
<td>we</td>
</tr>
<tr>
<td>5</td>
<td>author</td>
<td>which</td>
</tr>
<tr>
<td>6</td>
<td>poetry</td>
<td>author</td>
</tr>
<tr>
<td>7</td>
<td>which</td>
<td>river</td>
</tr>
<tr>
<td>8</td>
<td>his</td>
<td>readers</td>
</tr>
<tr>
<td>9</td>
<td>is</td>
<td>poem</td>
</tr>
<tr>
<td>10</td>
<td>poetical</td>
<td>poetry</td>
</tr>
<tr>
<td>Keyword classification</td>
<td>ERCorp</td>
<td>QRCorp</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>Texts, authors, readers</td>
<td>author poetry poetical readers literature</td>
<td>story character poem description poet</td>
</tr>
<tr>
<td></td>
<td>narrative poet metres reader character</td>
<td></td>
</tr>
<tr>
<td>Reviewing</td>
<td>genius style diction antient modern</td>
<td>interesting great original romantic imagination</td>
</tr>
<tr>
<td></td>
<td>peculiar genius criticism interesting taste</td>
<td></td>
</tr>
<tr>
<td>Grammatical</td>
<td>of the we ... most</td>
<td>of the we</td>
</tr>
<tr>
<td>Verbs</td>
<td>cannot</td>
<td>cannot</td>
</tr>
</tbody>
</table>
# CONCORDANCES

## Genius

<table>
<thead>
<tr>
<th>GENIUS</th>
<th>ER</th>
<th>QR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>189</td>
<td>56</td>
</tr>
<tr>
<td>Range</td>
<td>37/61</td>
<td>14/24</td>
</tr>
<tr>
<td>Authors</td>
<td>14</td>
<td>8</td>
</tr>
<tr>
<td>Highest user</td>
<td>Mackintosh (54)</td>
<td>Scott (33)</td>
</tr>
</tbody>
</table>
**ANALYSIS**

**N-grams**

- Repeated sequences of n number of words
- Repeated patterns are repeated because they are associated with the communicative needs of a certain register (Mahlberg)
- 4-grams and above have an associated semantic prosody (Fischer-Starcke)
- Semantic prosody: attitudinal element, motivation
# ANALYSIS

## 4-grams

<table>
<thead>
<tr>
<th>Rank</th>
<th>ERCorp</th>
<th>QR Corp</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>one of the most</td>
<td>the greater part of</td>
</tr>
<tr>
<td>2</td>
<td>the greater part of</td>
<td>one of the most</td>
</tr>
<tr>
<td>3</td>
<td>a great deal of</td>
<td>it is impossible to</td>
</tr>
<tr>
<td>4</td>
<td>as one of the</td>
<td>for the most part</td>
</tr>
<tr>
<td>5</td>
<td>we have no doubt</td>
<td>greater part of the</td>
</tr>
<tr>
<td>6</td>
<td>for the most part</td>
<td>it is plain that</td>
</tr>
<tr>
<td>7</td>
<td>it is impossible to</td>
<td>as much as possible</td>
</tr>
<tr>
<td>8</td>
<td>a great part of</td>
<td>in a great measure</td>
</tr>
<tr>
<td>9</td>
<td>is by no means</td>
<td>it is clear that</td>
</tr>
<tr>
<td>10</td>
<td>a good deal of</td>
<td>it was impossible to</td>
</tr>
</tbody>
</table>
ANALYSIS
4-grams

• Out of the 200 most frequent 4-grams in ERCorp and QRCorp
  • There are more types of 4-grams marking strong judgement in ERCorp
  • 4-gram types marking weak judgement are similar in number
  • ERCorp strong judgement 4-grams occur in multiple authors

• Fewer shared expressions of judgement in QRCorp
  • A marker of house style?

<table>
<thead>
<tr>
<th>Type</th>
<th>ERCorp</th>
<th>QRCorp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong judgement</td>
<td>30</td>
<td>14</td>
</tr>
<tr>
<td>Weak judgement</td>
<td>22</td>
<td>19</td>
</tr>
</tbody>
</table>
Clusters

- Certain clusters seem more characteristic of ERCorp e.g.
  - “have no doubt”
  - “great deal”
- N.B. CLMET is 30 times the size of ERCorp

<table>
<thead>
<tr>
<th>Cluster</th>
<th>ERCorp</th>
<th>QRCorp</th>
<th>CLMET</th>
</tr>
</thead>
<tbody>
<tr>
<td>have no doubt</td>
<td>19</td>
<td>1</td>
<td>58</td>
</tr>
<tr>
<td>great deal</td>
<td>35</td>
<td>2</td>
<td>694</td>
</tr>
</tbody>
</table>
Conclusion
CONCLUSION

And next steps

• Some traces of “house style”
  • Relatively faint through stylometry (“fingerprint” level)
  • More perceptible through corpus stylistics (signature level)

• Influence of genre of text being reviewed on its stylistic fingerprint
CONCLUSION
And next steps

• Corpus stylistics with:
  • Part of speech tags
  • Lemmatisation
  • Author-based corpora
    • Jeffrey, Brougham, Croker, Scott

• Stylometry with:
  • Character n-grams
  • Positive vs. negative reviews

• Assessment of the benefits of curation:
  • Keeping quotations
  • Using “raw” OCR
“Many interesting things cannot be counted, but many others can.”

–John Burrows
THANK YOU!

Download our corpus from ORDO (search for RSVP)
doi: 10.21954/ou.rd.6850865

Follow us on
http://www.open.ac.uk/blogs/styleproject/
@rhymesontheroad
@dh_ou