In Search of the Voice of the Edinburgh Review

Conference or Workshop Item

How to cite:

For guidance on citations see FAQs.

© 2016 Open University

Version: Version of Record

Link(s) to article on publisher’s website:
http://www.bars.ac.uk/

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.
In Search of the Voice of the Edinburgh Review

Francesca Benatti and David King
The Open University
Introduction

Did the *Edinburgh Review* create a “transauthorial discourse” (Klancher 1987) that hid the voices of individual contributors behind a corporate style?
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
“Many interesting things cannot be counted, but many others can.”

John Burrows
Why use stylometry?

Perception of authorial “voice” is quite subjective
• e.g. Duncan Wu (2007)

Computer-aided analysis can supplement humanistic research
“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.”

John Burrows
Delta continued

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

We all use Most Frequent Words differently

Underpinned by solid mathematical and linguistic foundations
BARS
Cluster Analysis

Moore_French_Novels_34_1820_corr
Brougham_Stendhal_EdRev_29_1817
Moore_Thurlow_23_1814
Moore_Boyd_24_1815_corr
Jeffrey_White_Doe_25_1815
Brougham_Forsyth_EdRev_22_1814_rev
Brougham_Nelson_EdRev_23_1814
anon_christabel_edinburgh_review_27_1816
Moore_Jorgenson_28_1817_corr
Brougham_phillips_29_1817
Hobhouse_Leake_EdRev_24_1815
Brougham_shepherd_23_1814
Brougham_franklin_28_1817
Mackintosh_Wraxall_25_Oct_1815_rev
Brougham_carnot_25_1815_rev
Mackintosh_Milton_25_Oct_1815_rev
Brougham_melanges_30_1818
Hunt_Nolt_All_EdRev_27_1816
Brougham_Park_EdRev_24_1815
Brougham_Junius_EdRev_29_1817_rev
Brougham_Columbus_27_Dec_1816
Brougham_king_32_1819
Palgrave_Northern_181.198_26_Feb_1816
Palgrave_Helga_25_June_1815_rev
Allen_Lingard_25_1815_rev
Palgrave_Goethe2_EdRev_28_1817_rev
Allen_Napoleon_27_Dec_1816_rev
Kinnaid_Letters_EdRev_26_1816
Brougham_Semple_22_Jan_1814
Jeffrey_ncte_Coleridge_Lay_Sermon_28_1817
Jeffrey_rogers_human_life_1819
Brougham_restoration_23_1814

Jeffrey_wilson_city_plague_1816
Jeffrey_Lalla_Rookh_29_1817
Jeffrey_byronCorsair_bride_1814_rev
Jeffrey_Edgworth_28_1817
Jeffrey_Historical_EdRev_24_1815_rev
Jeffrey_Waverley_EdRev_24_1815_rev
Jeffrey_byron_manfred_1817
Hazlitt_Sismondi_25_1815_rev
Hazlitt_Schlegel_26_1816
Hazlitt_Standard_Novels_24_1815_rev
Hazlitt_Coleridge_Lay_Sermon_27_1816_rev
Hazlitt_Coleridge_literary_life_28_1817_rev

100 MFW Culled @ 0%
Pronouns deleted Classic Delta distance
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

MFW

Frequent words

Choose what to *include* in the analysis

TF:IDF

Significant words

Choose what to *exclude* from the analysis
Evaluation

Distinctions between authors weak but present

How noticeable are they to a human reader?

Do Jeffrey’s editorial interventions erase individual voices?

Spot patterns that deserve close reading
Future work

Extend corpus:
  • Quarterly Review
  • CLMET

AntConc:
  • keywords (words statistically more frequent)
  • methods from corpus linguistics

Funding from RSVP to extend research (Jan-Oct 2017)