

Open Research Online

The Open University's repository of research publications and other research outputs

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

Conference or Workshop Item

How to cite:

Benatti, Francesca and King, David (2016). A Question of Style: individual voices and corporate identity in the Edinburgh Review, 1814-1820. In: Digital Humanities Congress 2016, 8-10 Sep 2016, University of Sheffield, UK.

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.

oro.open.ac.uk

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*

Style as fingerprint

Unconscious elements
in the way we write

(e.g. Van Halteren et al.
"Existence of a human
stylome." (2005))

Reflected by use of **Most
Frequent Words**

Style as signature

Conscious choice of
words, sentences, tone

(e.g. Van Dalen-Oskam
[Riddle of Literary Quality](#)
project)

Still **unsure** how to
identify with stylometry

* 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.”

John Burrows, “Delta”, *Literary and Linguistic Computing* 17.3, 2002

Delta continued

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

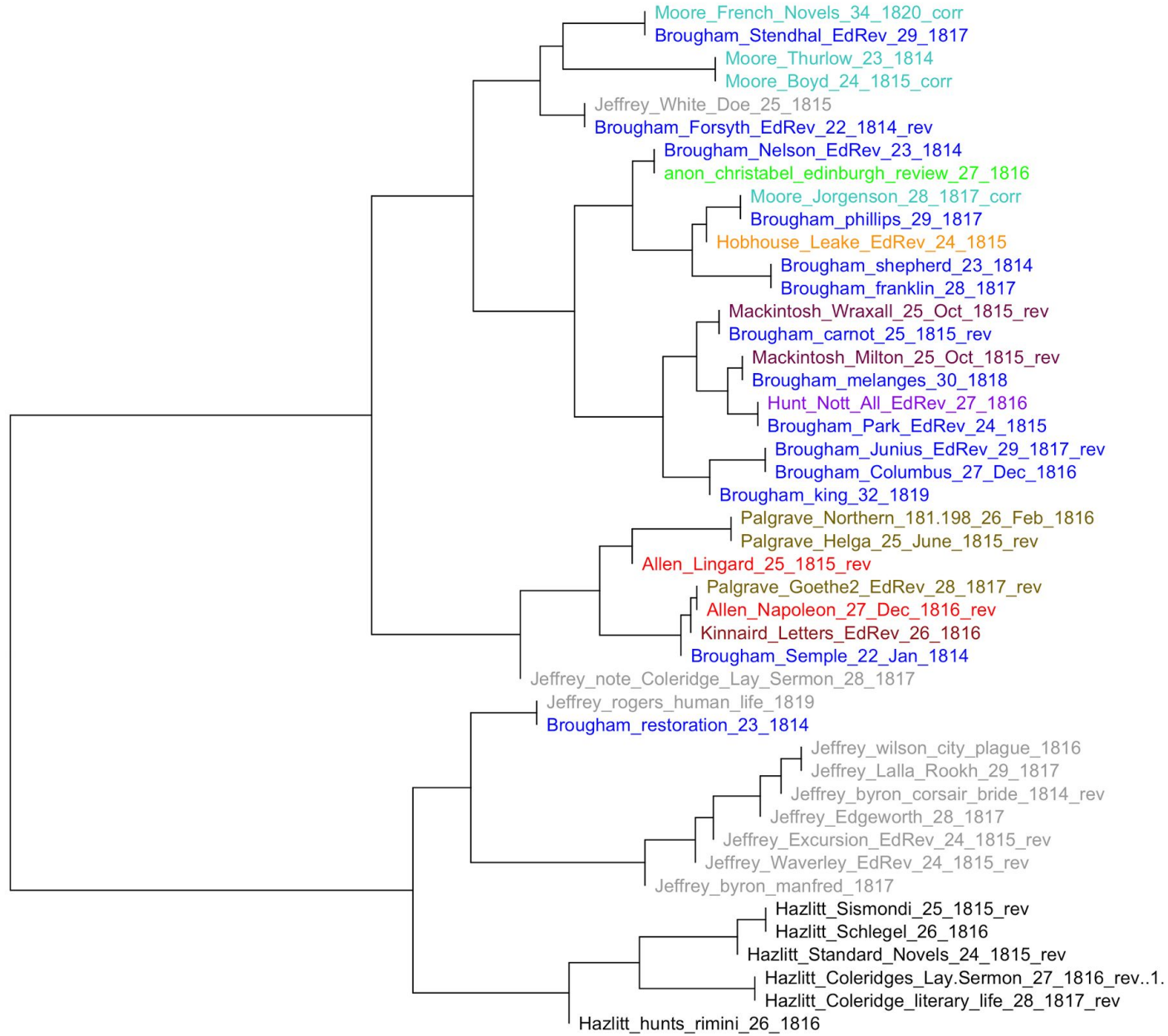
All authors use Most Frequent Words differently

Underpinned by solid mathematical and linguistic foundations

Delta - example

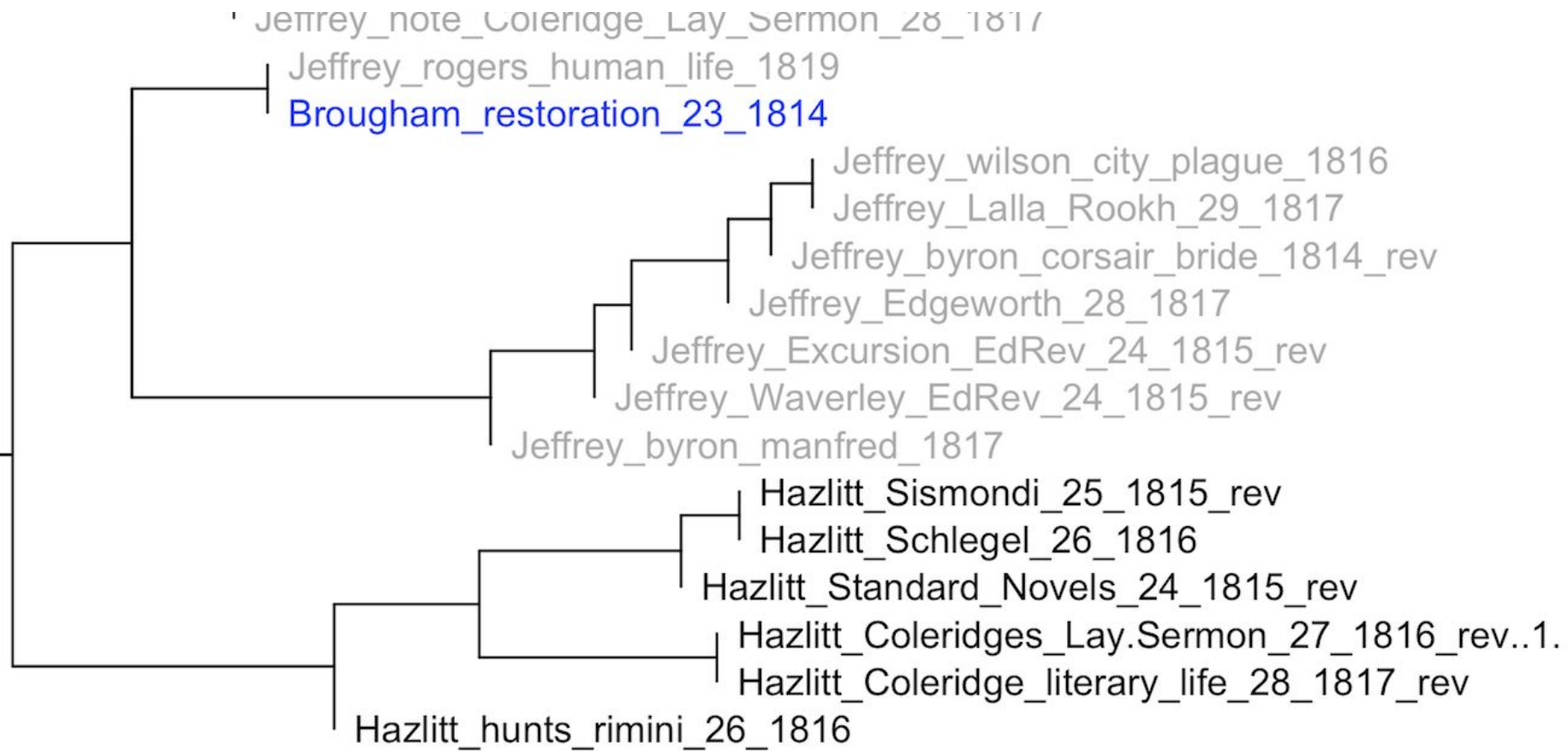
| Word | Moore | Coleridge | Godwin | Southey |
|------|-------|-----------|--------|---------|
| the | 7.71 | 6.4 | 6.9 | 7.69 |
| of | 5.85 | 5.06 | 4.49 | 3.54 |
| and | 2.83 | 3.95 | 3.52 | 3.15 |
| to | 2.97 | 3.04 | 3.01 | 3.11 |

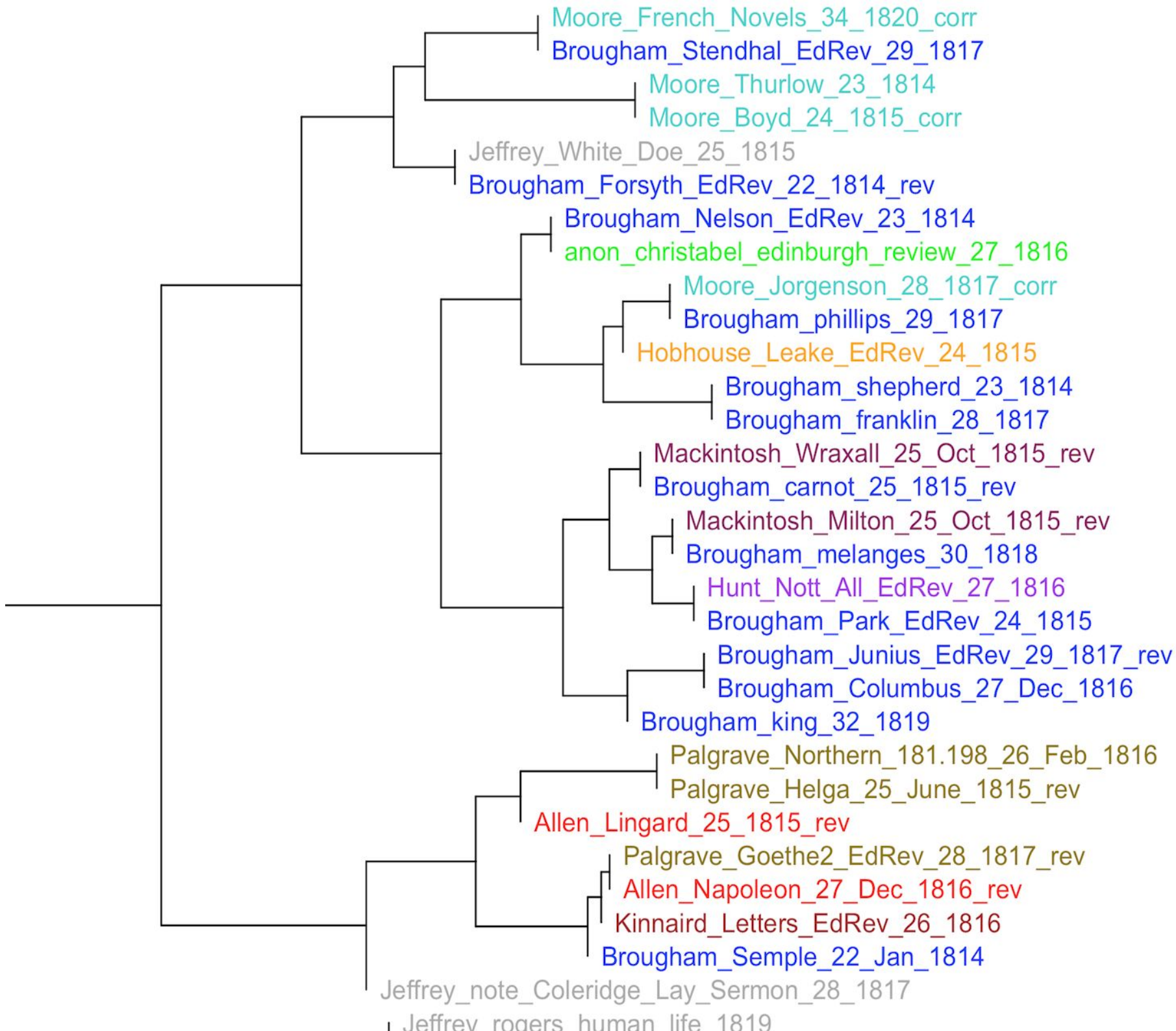
BARS Cluster Analysis



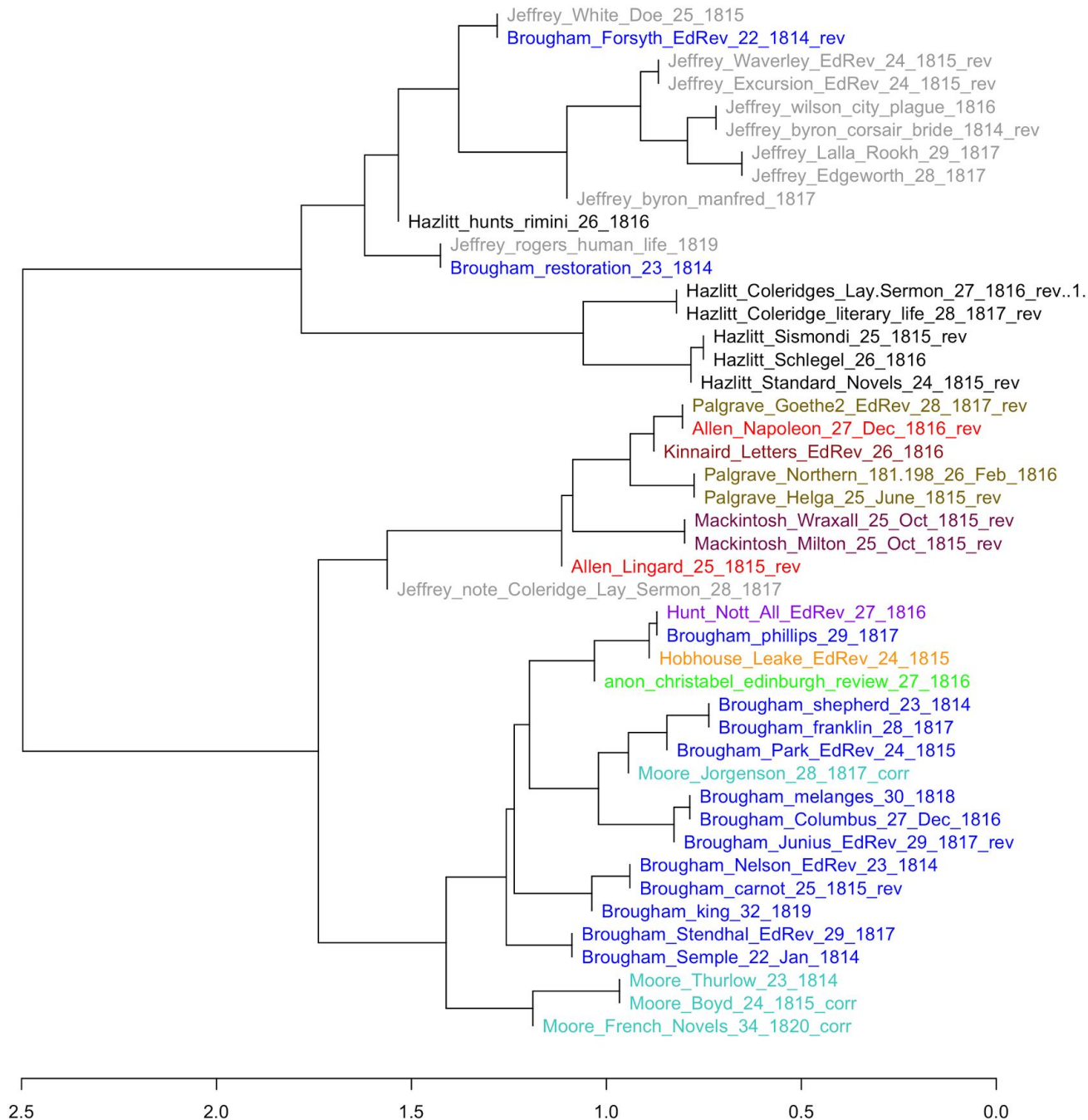
3.0 2.5 2.0 1.5 1.0 0.5 0.0

100 MFW Cullted @ 0%
Pronouns deleted Classic Delta distance





BARS Cluster Analysis



200 MFW Culled @ 70%
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

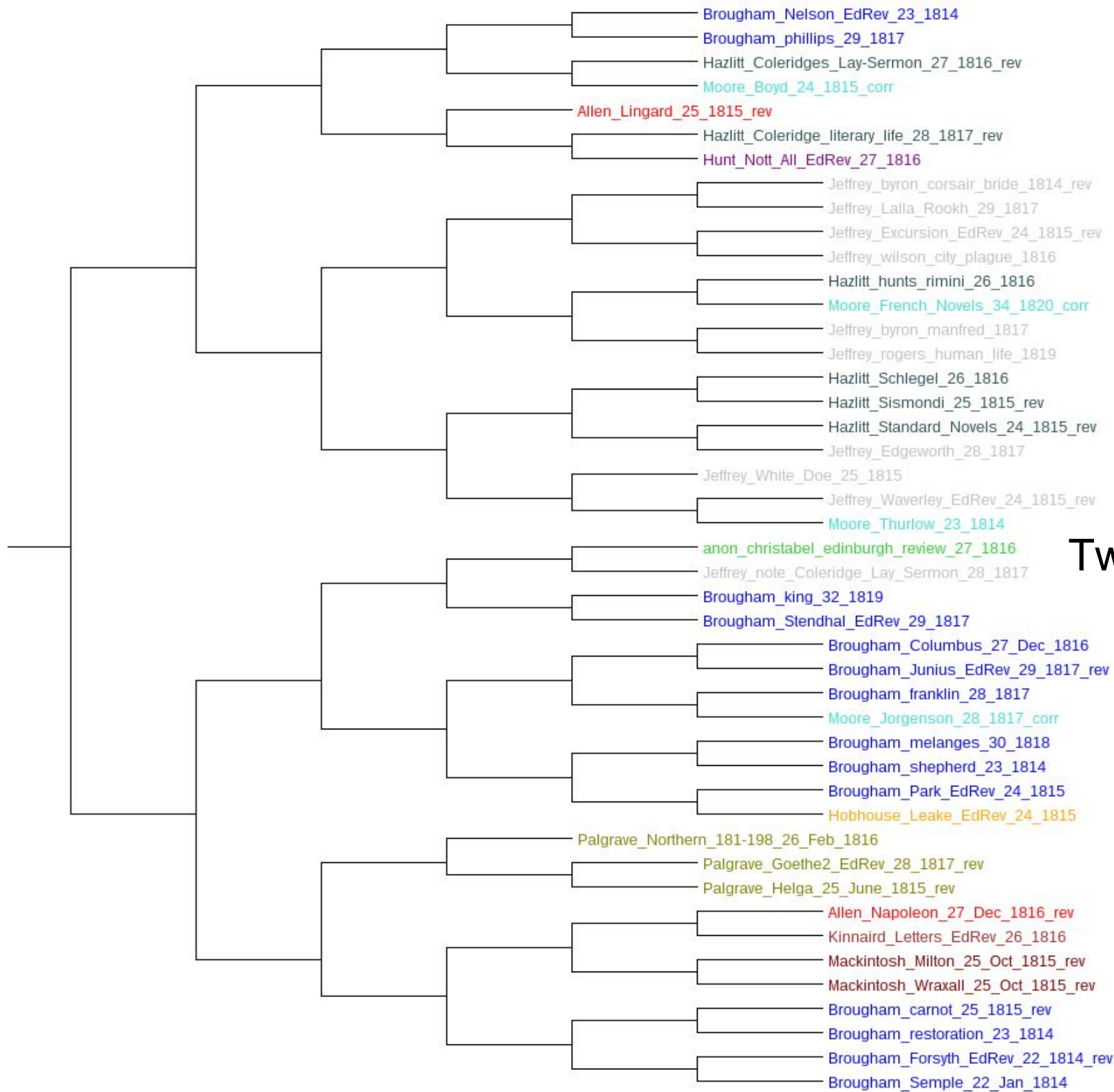
Increasing rigour

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

Unconscious style?

TF:IDF

Significant words

Choose what to *exclude*
from the analysis

Conscious style?

Future work

Extend corpus:

- Python toolset to assist
 - OCR correction
 - TEI markup

Further methods:

- corpus stylistics
- Burrows' Zeta and Iota

Digital Humanities at the Open University
The Open University
Walton Hall
Milton Keynes
MK7 6AA

Arts-digital-humanities@open.ac.uk
www.open.ac.uk