RSVP Corpus: Foscolo Supplement, 2018

The **RSVP Corpus: Foscolo Supplement, 2018**, is an output of the *Foscolo* study, a follow on project to the original *A Question of Style* project.

Aims and purpose

The original A Question of Style project created a corpus of 780,000 words published in the *Edinburgh Review* and *Quarterly Review* during 1814–1820, with the aim of gaining new insights on the writing and editorial practices of the *Edinburgh Review* during the editorship of Francis Jeffrey. The *Edinburgh Review* was the main literary journal in early nineteenth-century Britain. It reviewed the most significant publications in literature, history, politics and science and its contributors included some of the most prominent figures of the age. Authors believed that a review from the *Edinburgh Review* could make or break the fortunes of a publication.

This supplementary corpus has been created to support a study into the contributions of Italian poet Ugo Foscolo to these periodicals in 1818-1819. Foscolo broke with both reviews in 1820, allegedly because of the poor quality of the translations of his articles from Italian and French into English, executed by, among others, James Mackintosh, Francis Palgrave and Francis Jeffrey. Additionally, Foscolo objected to the cuts and interpolations the two journals operated, accusing them both of "tampering with articles".

This follow on study aimed to investigate if Foscolo's claims were a convenient fiction or a fact that can be tested empirically. The corpus is our raw data to allow an assessment of Digital Humanities methodologies, such as stylometry and corpus stylistics, and if they can help confirm or deny Foscolo's claims.

Funding

The A Question of Style project was funded by the inaugural Field Development Grant by the Research Society for Victorian Periodicals from the bequest of the Eileen Curran estate,.

This follow on study is the unfunded work of the two researchers.

Personnel

The *A Question of Style* and *Foscolo* projects were led by Dr Francesca Benatti, francesca.benatti@open.ac.uk, Digital Humanities Scholar at the Open University, in collaboration with Dr David King, david.king@open.ac.uk, Visiting Fellow at the Open University, during 2018-2019.

Use

The collection comprises five articles encoded in standard TEI XML. No TEI extensions are used.

The files can be processed with any XML or text tool.

Articles

The corpus comprises all five articles authored by Foscolo that appeared in the *Edinburgh Review* and the *Quarterly Review*:

- *Parga*: ER VOLUME 32, OCTOBER, 1819, pages 263-293 – Google Books link
- *Pius VI*: ER VOLUME 31, MARCH, 1819, pages 271-295 – Google Books link
- *Dante part 1*: ER VOLUME 29, FEBRUARY, 1818, pages 453-474 – Google Books link
- *Dante part 2*: ER VOLUME 30, SEPTEMBER, 1818, pages 317-351 – Google Books link
- *Narrative Poems*: QR VOLUME 21, April 1819, pages 486-556 – Google Books link

Files

The base set of five files are, with three versions of each file:

- Foscolo-Jeffrey_Parga_ER_32_1819
- Foscolo-Jeffrey_Pius_VI_ER_31_1819
- Foscolo-Mackintosh_Dante_part01_ER_29_1818
- Foscolo-Wallace_Dante_part02_ER_30_1818
- Foscolo-Palgrave_Narrative_Poems_QR_21_42_1819

The filenames are based on:

- original author, always Foscolo for this corpus
- translator
- short title of article
- periodical - either ER for *Edinburgh Review* or QR for *Quarterly Review*
- volume number
- issue number if QR
- year of review's publication
- suffix:
  - _dirty.txt

- _clean.txt
- .xml

The three file versions are:

- _dirty.txt, individual article text manually extracted from full text downloads from Google Books
- _clean.txt, cleaned versions of the dirty text files
- .xml, encoded versions of the clean text files

Simple statistics

Using the unix `wc` command the statistics for the clean text versions of the articles are:

Article	Lines	Words	Characters
Foscolo-Jeffrey_Parga_ER_32_1819	105	14525	84386
Foscolo-Jeffrey_Pius_VI_ER_31_1819	93	11130	65449
Foscolo-Mackintosh_Dante_part01_ER_29_1818	329	9447	53984
Foscolo-Wallace_Dante_part02_ER_30_1818	439	15686	91423
Foscolo-Palgrave_Narrative_Poems_QR_21_42_1819	1096	32053	185471
Totals	2062	82841	480713

Data suitability

The study worked solely on PDFs available in Google Books, and the accompanying OCR texts. The study did not scan the documents nor OCR the resulting images.

The corpus is primarily intended to support stylometric analysis of the writings of Ugo Foscolo in the context of early nineteenth-century literature. The corpus contents are also relevant to researchers interested in political and social history.

Owing to curatorial decisions relevant to the *A Question of Style* project, especially those relating to the handling of punctuation, this corpus is not suitable for examining critical editions of the original works.

A secondary purpose of the corpus is to support research into curation of OCR processed texts. Typically, the OCR process is imperfect, especially on older texts. The contents of this corpus provide both source, dirty texts and their clean, curated equivalents.

Curation

The articles are curated as described in the TEI header in each file. As such, the files are digital artefacts derived from the original published article.

Reuse

The corpus is freely available for reuse.

The corpus' DOI is 10.21954/ou.rd.7472210.

The original RSVP curated corpus and its source text companion corpus may also be downloaded from the *A Question of Style* project's online data site.

The Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) Licence applies all three corpora.

Contact

For more information on the corpus, please email Digital Humanities at the Open University, fass-digital-humanities@open.ac.uk,

or write to:

Digital Humanities
The Open University
Walton Hall
Milton Keynes
MK7 6AA
United Kingdom