

## CLLC Victorian Periodical Corpus

### Some Statistics: as at February 2008

- number of articles: 182
- number of authors: 21 (14 men and 7 women)
- number of words: 1,765,454

### Background of Corpus

Some years ago Dr. Ellen Jordan – a social historian with a special interest in the Victorian era – came to the CLLC with an attribution problem. She was interested to know if the computational stylistics techniques being developed at the Centre could determine the probability of correctness of her "strong hunch" that Anne Mozley had written a number of anonymous articles of interest to her. She was told that she would need to build up a corpus of well attributed articles by comparable authors – similar in date and genre to her "mystery" articles. Ellen decided that she would choose articles written by well regarded female journalists in "high class" literary journals around the 1850s and 60s.

Thus, the corpus began to take shape. The first articles were transcribed from photocopies Ellen had taken from the periodical Journals themselves. The authors, apart from Anne Mozley, were Frances Power Cobbe, George Eliot, Harriet Martineau, Margaret Oliphant and Elizabeth (Lady Eastlake) Rigby.

Ellen was then approached by Eileen Curran to test what Eileen suspected were Wellesley misattributions of the two Scottish writers John Stuart Blackie and John Hill Burton, both of whom were born in 1809 in Edinburgh, attended the same College and wrote for *Tait's Edinburgh Review*. In order to do this testing, it was necessary to begin adding male authors to the corpus. The authors added at this stage were those whose articles could be downloaded as electronic texts from the online Gutenberg collection.

The Centre's research assistant, Alexis Antonia, had assisted Ellen in building this initial corpus, and found that she very much enjoyed working with these periodical articles. When the opportunity arose for working on a Research Higher Degree, Alexis chose a topic which would involve using and expanding the existing Victorian Periodical Corpus.

### Acquisition of Electronic Texts for the Corpus

Setting the boundaries and obtaining the texts:

- (i) the fifty year period from 1830 to 1880 was chosen as encompassing what has been called "the golden years" of periodical writing. These were also the years during which the great anonymity versus signature debate was enacted. Anonymity was the norm at the beginning of this period, but gradually towards the end of it, signature was more in evidence. Though most of the articles in the corpus were anonymous at the time of publication, they have all been fairly reliably attributed and quite a few were subsequently republished in authorial collections of essays and writings.

(ii) the articles were all taken either from one of the major quarterlies or from one of a number of well-regarded monthlies.

**Quarterlies**

Edinburgh Review  
 Quarterly Review  
 Westminster Review  
 Bentley's Quarterly Review  
 National Review

**Monthlies**

Blackwoods Edinburgh Review  
 Cornhill Magazine  
 Fortnightly Review (which became monthly)  
 Frasers Magazine  
 MacMillans Magazine  
 Tait's Edinburgh Review

(iii) authors were chosen with the idea of getting as wide a range as possible of writers who were known to have been writing for our listed periodical journals during our fifty year period.

**Men**

Walter Bagehot (1826-77)  
 John Stuart Blackie (1809-1895)  
 John Hill Burton (1809-1881)  
 Thomas Carlyle (1795-1881)  
 Lord Robert Cecil (1830-1903)  
 John Wilson Croker (1780-1857)  
 James Anthony Froude (1819-1894)  
 William Rathbone Greg (1809-1881)  
 Abraham Hayward (1801-1884)  
 Thomas Henry Huxley (1825-1895)  
 Charles Kingsley (1819-1875)  
 George Henry Lewes (1817-1878)  
 Thomas Babington Macaulay (1800-1859)  
 Sir Leslie Stephen (1832-1904)

**Women**

Frances Power Cobbe (1822-1904)  
 Caroline Frances Cornwallis (1786-1858)  
 George Eliot (1819-1880)  
 Harriet Martineau (1802-1876)  
 Anne Mozley (1809-1891)  
 Margaret Oliphant (1828-1897)  
 Elizabeth Lady Eastlake Rigby (1809-1893)

(iv) the following methods were used to obtain the electronic texts of Victorian Periodical articles in the Corpus:

1. The first method used was to transcribe the text onto the computer from a photocopy of the journal article. This method was used subsequently for a number of MacMillans articles which the librarian in the University's Rare Books Collections photocopied for us.
2. The second method was to find public domain electronic texts available in online collections. Of these, only Gutenberg allowed downloading of text in editable form.
3. Other online collections such as the ILEJ (Bodley, Oxford) site for *Blackwoods* provided only photo image copies of texts; these could be printed for subsequent scanning or could be transcribed from the photo image into editable electronic text form.
4. Newcastle University has a number of Victorian periodical journals (*Westminster*, *Edinburgh*, *Frasers*, *Tait's* and an incomplete *MacMillans*) available on microfilm. Microfilm printouts were obtained from these for many articles; most of these

were transcribed onto the computer; occasionally a microfilm printout was considered clear enough to permit scanning.

5. Where published editions of periodical articles existed in authorial collections of writings, these were photocopied and scanned. Sometimes, if the photocopy was not suitable for scanning, the article was transcribed.
6. For journals, such as the *Quarterly Review* and the *Fortnightly*, which the Newcastle Library doesn't hold, Inter-Library loan requests obtained the file in TIFF format. These articles were printed and either transcribed or scanned depending on the quality of the copy.

### **Editing of the electronic texts in preparation for the Centre's counting programs.**

- Each text was subjected to careful proof reading, since both transcription and scanning can produce some unexpected errors.
- Quotations (that is to say, text included in the article which does not belong to the author) were identified and removed from the count.
- Foreign words which were longer than two or three words and which were not part of the syntax of the sentence were identified and removed from the count.
- Portions of text where the author assumes a persona for illustrative or dramatic purposes, or where he or she feels obliged to use inverted commas to signal his adoption for the moment of a special way of phrasing something, were identified but left in the count (the idea being that it is still his/her voice).
- Words which are used by some authors as a single compound and by others as two separate words were identified and united using TEI format. (for example, `<reg orig="can not"> cannot </reg>`) The various compounds of *any*, *some*, *every*, and *no* with *one*, *thing*, *how*, *where* ... were united in this fashion.
- Negative forms such as *don't* and *can't* (which are not common in Victorian writing anyway) were left untouched.
- Occasionally an article included tables of statistics and so on. These were generally omitted.