

# Corpus of Modern Scottish Writing (CMSW)

*Wendy Anderson, Jenny Bann, David Beavan*

Department of English Language

University of Glasgow

*W.Anderson@englang.arts.gla.ac.uk; J.Bann@englang.arts.gla.ac.uk; D.Beavan@englang.arts.gla.ac.uk*

## **Abstract**

This poster describes the online Corpus of Modern Scottish Writing (1700-1945), being created at the University of Glasgow. The corpus fills the chronological gap between the Helsinki Corpus of Older Scots (1375-1700) and the Scottish Corpus of Texts & Speech (1945-present). The period covered by CMSW is an important time in the history of Scotland and Scots. It begins with the last stages of the standardisation of written English and the onset of the ‘Vernacular Revival’ in literary Scots. Out of the interaction between Broad Scots and written Standard English, the hybrid prestige variety of today’s Scottish English is said to emerge: CMSW will allow researchers to substantiate this claim, among many others. Once complete, CMSW will contain at least 4 million words of text, with accompanying metadata, covering a range of genres, including personal writing, administrative prose, verse and drama, and the writings of language commentators.

## **1. Introduction**

A new diachronic corpus, the Corpus of Modern Scottish Writing, is being created by a team led by Professor John Corbett in the Department of English Language at the University of Glasgow. The project is grant-funded by the Arts and Humanities Research Council.<sup>i</sup> Once complete, CMSW will be able to be browsed, searched and analysed through a freely available online interface.<sup>ii</sup> Chronologically, CMSW fills the gap between the Helsinki Corpus of Older Scots (1450-1700) (see Meurman-Solin 1995), and the Scottish Corpus of Texts & Speech (1945-present) (University of Glasgow, 2007).<sup>iii</sup> It will therefore soon be possible to analyse the Scots language across a period of five and a half centuries.

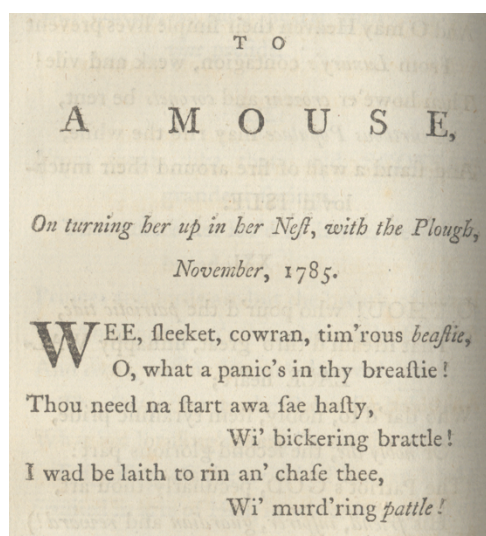
### **1.1. The period 1700-1945 in the history of Scots**

The period covered by CMSW is an important time in the history of Scotland and Scots. It begins with the last stages of the standardisation of written English and the onset of the Vernacular Revival in literary Scots that produced writers like Robert Burns. Language use in Scotland in the modern period, conventionally dated from 1700, can be described as a continuum with Standard English at one end, and social and regional varieties of Broad Scots at the other. Writers vary their performance along that continuum, to a greater or lesser extent, depending on their social background and the context of writing. Out of the interaction between Broad Scots and written Standard English, the hybrid prestige variety of today’s Scottish English is said to emerge. However, there is comparatively little study of how this happened, beyond some detailed analysis of the evidence of spelling reformers of the eighteenth century, mainly in relation to changes in pronunciation of the period. As a searchable digital archive of Scottish writing from this key period, CMSW lays the foundations for a new account of language development in Scotland, and will allow researchers to substantiate claims previously made.

## 2. Description of the corpus

Once complete, CMSW will contain at least 4 million words of text, with accompanying demographic and textual metadata, covering a range of genres, including administrative, expository, instructional, imaginative and religious prose, personal writing, verse and drama, and journalism. Genre categories were selected in order to make the corpus broadly comparable to the Helsinki Corpus of Older Scots, and therefore facilitate extensive diachronic research. CMSW will also feature a category of writing by contemporary commentators on the Scots language, the so-called 'orthoepists'. As with its predecessor project, the Scottish Corpus of Texts & Speech, CMSW will be able to be searched and analysed through integrated tools. We anticipate that it will attract a wide range of users, from linguists of various persuasions interested in analysing the corpus as an integral resource, to historians drawing on the resource as a source of digital images and carefully transcribed text.

### 2.1. Example texts



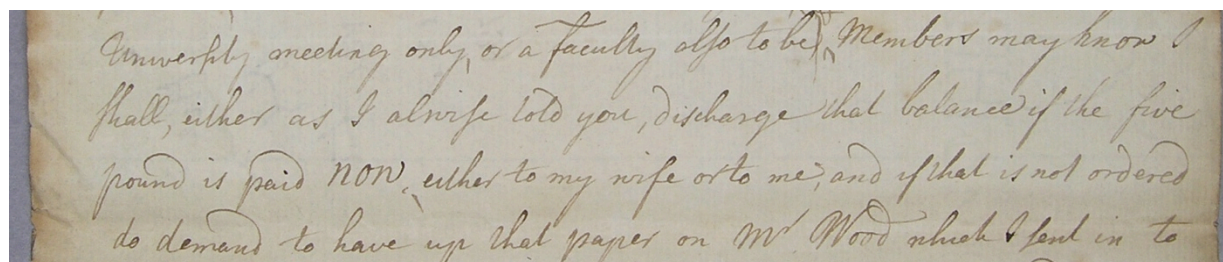
**Robert Burns**

*Poems, Chiefly in the Scottish Dialect, 1786*

University of Glasgow Library, Special Collections

In 2009, the 250<sup>th</sup> anniversary of the birth of the Scottish poet Robert Burns, CMSW made available the full text and images of the first, Kilmarnock, edition of Burns' *Poems Chiefly in the Scottish Dialect* (1786). With an initial print-run of only 612 copies, it sold out within a month. Burns, who was just 27 at the time, had been planning to emigrate to the West Indies, but following the success of the book he stayed in Scotland and moved from his home in Ayrshire to Edinburgh, the capital.

**Professor Andrew Rosse, Chair of Humanity, University of Glasgow**  
**Letter to Professor Robert Simson, Clerk of Senate 16 August 1735**  
University of Glasgow Archives

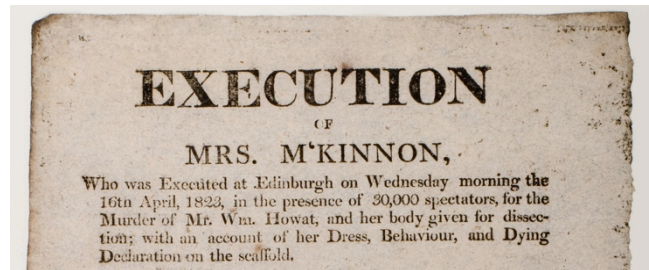


Andrew Rosse was owed £5 in additional salary by the University. In this letter he demands the matter be brought up at a meeting and threatens to take further measures if it remains unpaid. Unfortunately, we do not have a response, although we know that Rosse resigned his position in this same year on condition that the university would continue to pay his salary for the rest of his life. CMSW will contain a number of documents from Glasgow University archives, including personal correspondence from the 1715 Jacobite rebellion, proceedings of university disciplinary trials from the 18th century, and legal correspondence concerning land rights.

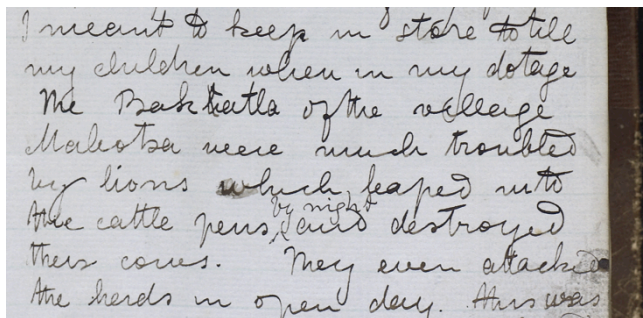
Legal writing in particular contains a high number of Scots words presented in a formal register, since Scots law remained distinct from English law after the 1707 Act of Union.

**Broadside announcing the execution of Mrs Mary McKinnon**  
**Printed by John Muir, Glasgow, 1823**  
**University of Glasgow Library, Special Collections**

The broadside relates the details of the execution of Mrs Mary McKinnon, who stabbed a writer's clerk to death during an intoxicated squabble at a brothel she ran. The article describes the crowd as 'unusually great', estimating that 30,000 spectators turned out to watch the hanging of this woman 'elegantly dressed in a black silk gown'. Broadside were cheaply produced and widely distributed to members of the public interested in the scandalous details of such crimes. While CMSW will contain an amount of newspaper material from the 18th century onwards, broadsides were an important form of early journalism.



**David Livingstone, Explorer and Missionary**  
**Draft of Missionary Travels and Researches in South Africa, 1850**  
**John Murray Archive, National Library of Scotland**

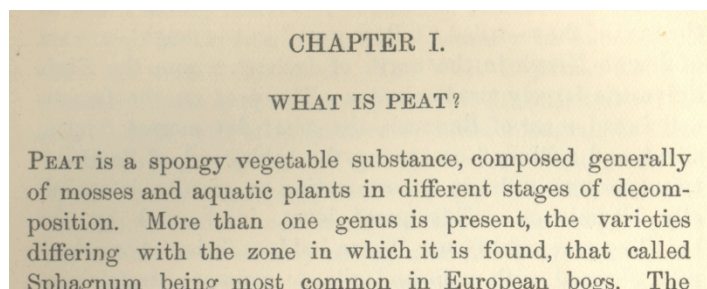


In his own hand, Livingstone here recounts several stories from his time in Africa, including this one which, he says, 'I meant to keep in store to tell my children when in my dotage. The Bakatla of the village Mabotsa were much troubled by lions which leaped into the cattle pens by night and destroyed their cows.' CMSW will contain a number of letters and letter-books

from the John Murray Archive, which covers seven generations of the Murray family's publishing empire.

**W A Kerr**  
**Peat and its Products, 1905**  
**University of Glasgow Library**

The book is an illustrated treatise on peat and its products as a national source of wealth. Peat, the 'sombre genius of the moor', has multiple uses as fuel, manure and building material, and possesses health-giving properties!



**2.2. Orthoepist material**

One novel aim for CMSW is the inclusion, alongside the main corpus, of commentaries on Modern Scots by orthoepists, or language commentators, from the eighteenth to the twentieth

centuries, such as James Murray, John Callander, and Sir James Wilson. The availability of a series of observations on the Scots language alongside samples of its usage will be a powerful aid to orthographic research.

Hugh Mitchell's *Scotticisms, Vulgar Anglicisms* (1793), for example, was written for the instruction of Scottish schoolchildren. Working from the principle that '[v]ulgar words and phrases must be known to be so before they can be avoided', Mitchell lists the most common of these alongside their standard English equivalent: the Scottish 'Indeed no' becomes 'No, indeed', 'plenish' becomes 'furnish', and 'What's your will?' becomes 'What would you have?'

### **2.3. Integrated tools**

The Corpus of Modern Scottish Writing will provide an integrated suite of tools for linguistic analysis, including maps, word frequency and key word in context concordances. Texts will also be downloadable in a number of formats, for use with researchers' own stand-alone software packages. Expert users will be able to take advantage of a suite of advanced features, offering the possibility of selecting documents to form their own sub-corpora. These in turn can be downloaded in a variety of standards-based formats, such as TEI (Text Encoding Initiative), allowing offline analysis and integration with other corpora and data sets. Casual and more novice users who may be unable, or unwilling to install specific software packages are supported through the integrated tools. These tools are designed for ease of use, while aiding browsing and navigation, and provide the opportunity to showcase the corpus to new research communities. Recent visualisation innovations, such as collocate clouds, provide a way of summarising the entire resource, identifying patterns in language otherwise difficult to identify.

### **3. Research possibilities**

Within the project time-frame, the linguistic research aims to account for the structures of Modern Scots orthography, with a view to enhancing automatic identification of spelling and dialectal variants. This groundwork will lead to further linguistic analysis, for example of the relationship between orthography and phonology, to what extent Modern Scots orthography is lexicalised in different phases within the period, and how particular orthographies are motivated by stylistic or philological intentions on the part of the author.

CMSW will be able to be used to investigate various types of linguistic research question. For example:

To what extent does the language of writers like Robert Burns and others in the Vernacular Revival have a lasting impact today?

What has been the nature of the relationship between Scots and English over the period 1700-1945?

How do the prescriptions of the orthoepists (or language commentators) compare with actual written usage of the same period?

In what genres of Scottish Standard English can lexical items from Broad Scots be found?

To what extent can a more detailed description of the structure of modern Scots orthography allow us to automate the identification of Scots spelling variants?

### **References**

Meurman-Solin, A. (1995). "A new tool: the Helsinki Corpus of Older Scots (1450-1700)." *ICAME Journal*, 19, 49-62.

- 
- <sup>i</sup> CMSW is funded from 2007-2010 by the Arts and Humanities Research Council in the UK, [www.ahrc.ac.uk](http://www.ahrc.ac.uk).
- <sup>ii</sup> Information about CMSW and the initial dataset can be found at [www.scottishcorpus.ac.uk/cmsw](http://www.scottishcorpus.ac.uk/cmsw). Once complete, the corpus will also be made available through this site.
- <sup>iii</sup> For information about the Helsinki Corpus of Older Scots, among other diachronic corpora, see ICAME, <http://icame.uib.no/>. The Scottish Corpus of Texts & Speech can be freely searched online at [www.scottishcorpus.ac.uk](http://www.scottishcorpus.ac.uk).