

CHAPTER SEVEN

A CORPUS OF LATE MODERN BRITISH AND AMERICAN ENGLISH PROSE (COLMOBAENG)¹

TERESA FANEGO

UNIVERSITY OF SANTIAGO DE COMPOSTELA

1. Introduction

Until the turn of the millennium, the structure and use of English in the Late Modern English period (LME; 1700-1900) had received less scholarly attention than that of other periods, perhaps, as Kytö, Rydén & Smitterberg suggest, as a result of “the view that the language of this period was not an interesting topic for historical research, where the ‘antiquity’ of the English language has often been in focus” (2006:1). In the past ten years, however, the eighteenth and nineteenth centuries have begun to generate the attention they deserve; witness, for example, collections of studies such as that of Kytö, Rydén & Smitterberg, above, linguistic introductions to the period such as Görlach (1999, 2001), Beal (2004) and Tieken-Boon Van Ostade (2009), and the successful conference series on Late Modern English, the first held in Edinburgh in 2001, the second in Vigo in 2004, the third in Leiden in 2007, and the fourth in Sheffield in May 2010.

New developments in corpus linguistics have also enabled researchers to offer a fuller account of LME than previously possible. When David

¹ For generous financial support I am grateful to the Autonomous Government of Galicia (grants CN2011/011 and CN2012/012) and the Spanish Ministry for Economy and Competitiveness and the European Regional Development Fund (grants HUM2007-60706 and FF12011-26693-C02-01).

Denison wrote his survey of the syntax of LME, “there was little machine-readable corpus material dated between *c.* 1700 and *c.* 1960” (Denison 1998: 94), so he had to construct a 100,000-word corpus of private letters written between 1861 and 1918 (Denison 1994) to supplement ARCHER (at that time a preliminary and incomplete version), the corpus compiled by Douglas Biber and Edward Finegan to trace the development of historical English registers from 1650 to 1990 (cf. Biber, Finegan & Atkinson 1994). Since then, additional LME language data has been made available in electronic collections such as the one-million word CONCE (Corpus of Nineteenth-Century English; see Kytö, Rudanko & Smitterberg 2000), and CLMET (Corpus of Late Modern English Texts), a database consisting of almost ten million words from the period 1710–1920, compiled by Hendrik De Smet on the basis of texts drawn from Project Gutenberg resources and the Oxford Text Archives (see De Smet 2005). To these could be added the somewhat earlier COPC (Century of Prose Corpus), 500,000 words of British prose texts dating back to 1680–1780 (see Milic 1995).²

However, most of the historical corpora mentioned above comprise only British English. An exception is ARCHER, which divides the eighteenth and nineteenth centuries into four subperiods, namely 1700–1749, 1750–1799, 1800–1849 and 1850–1899; of these, the first and third include only British English, the second and fourth both British and American English, the word count for each subperiod and dialect coming to around 180,000 words (that is, a total of 720,000 words of BrE and 360,000 words of AmE). It is clear, therefore, that there is still a scarcity of historical corpora offering parallel samples of the two major standard varieties of English, and it was partly to provide for this need that the machine-readable corpus described in the present chapter was produced.

In what follows, I will discuss the structure (section 2) of A Corpus of Late Modern British and American English Prose (COLMOBAENG), and some of its research possibilities and limitations (section 3).

² Since December 2010 the 400-million-word Corpus of Historical American English (COHA), compiled by Mark Davies, has also been available; it comprises material from the early 1800s to the present time. Cf. <http://corpus.byu.edu/coha/>

2. Corpus Make-up

2.1. Period, Dialect and Genre Division

COLMOBAENG (1700-1879) is a 1,170,000 word database comprising texts drawn from both printed and electronic sources; the former are listed in the references at the end of the chapter, the latter include the Century of Prose Corpus, Project Gutenberg, the Online Books Page at the University of Pennsylvania, the Electronic Text Center at the University of Virginia, and Chadwyck-Healey’s collections of Early American Fiction, Eighteenth-Century Fiction and Nineteenth-Century Fiction. At present, the structure of the corpus is as shown in Tables 1-3.

British English	American English
BrE1 1700-1726	—
BrE2 1732-1757	AmE2 1732-1759
BrE3 1761-1797	AmE3 1774-1804
BrE4 1850-1879	AmE4 1851-1879

Table 1. Range of periods and dialects represented

BrE1 (1700-1726)	BrE2 (1732-1757)	BrE3 (1761-1797)	BrE4 (1850-1879)
nonfiction: 76,000 w 26 texts 22 authors	nonfiction: 76,000 w 27 texts 24 authors	nonfiction: 76,000 w 26 texts 22 authors	nonfiction: 76,000 w 15 texts 12 authors
fiction: 124,000 w 11 texts 8 authors	fiction: 124,000 w 8 texts 7 authors	fiction: 124,000 w 8 texts 8 authors	fiction: 124,000 w 8 texts 8 authors
Total words: 200,000	Total words: 200,000	Total words: 200,000	Total words: 200,000

Table 2. Range of texts from British English: fiction vs nonfiction (word [w] totals)

AmE2 (1732-1759)	AmE3 (1774-1804)	AmE4 (1851-1879)
nonfiction: 50,000 w 7 texts 5 authors	nonfiction: 40,000 w 7 texts 5 authors	nonfiction: 76,000 w 16 texts 12 authors
	fiction: 80,000 w 4 texts 4 authors	fiction: 124,000 w 10 texts 8 authors
Total words: 50,000	Total words: 120,000	Total words: 200,000

Table 3. Range of texts from American English: fiction vs nonfiction (word [w] totals)

The nonfiction samples in the first three subperiods of British English (BrE1-3) have been taken from Milic's COPC, and this has influenced the choice of nonfiction texts in the rest of the database. The COPC is intended to constitute "an inventory of the daily language of the literate members of English society" of the eighteenth century (Milic 1995: 329). Therefore, in the interest of comparability, no attempt was made to include in BrE4 (1850-1879) or in the several subperiods of American English records of town, court and church proceedings, or the unstudied writing of uneducated individuals, since none of these would have an equivalent in the COPC.

A second criterion guiding the selection of the American English texts was the exclusion of authors born outside the United States, though exceptions were made in the case of Charles Thomson (1729-1824), whose family migrated from Ireland in 1739, Hugh Henry Brackenridge (1748-1816), born near Campbelltown, Scotland, but brought up in Pennsylvania from the age of five, Alexander Hamilton (1757-1804), who at the age of fifteen moved to New York from the Caribbean island of Nevis, and John William Draper (1811-1882). Draper was born in Liverpool, England, and came to the United States in 1832; he graduated from the medical department of the University of Pennsylvania in 1836 and in 1839 became Professor of Chemistry at the University of the City of New York. It seems reasonable to assume that by 1874, when the *History of the Conflict between Religion and Science* – the only work by Draper included in COLMOBAENG – was published, his writing style would have been representative of standard American usage.

Considerable efforts were also made to leave out all American works printed in Great Britain, since, as noted by Kytö & Rissanen (1983: 476), “there is always the possibility of British interference on the printer’s side.” However, in view of the scarcity of materials published in America during AmE2 (1732-1759) I finally opted for the inclusion of two texts printed in London, namely John Bartram’s *Observations* (1751) and Charles Thomson’s *An Enquiry* (1759).

2.2. Contents: Fiction Texts

Unless otherwise indicated, all extracts have around 20,000 words and have been taken from Chadwyck-Healey’s electronic text collections. Dates of publication are given after each title; birth-dates in brackets.

Subperiod BrE1 (1700-1726):

- Anonymous. *The Adventures of Lindamira, a Lady of Quality*, 1702. [2,000 words]. COPC
- Barker, Jane (1675-1743). *The Amours of Bosvil and Galesia*, 1719. [14,000 words]
- Barker, Jane (1675-1743). *Exilius: or, The Banish’d Roman*, 1719. [6,300 words]
- Davys, Mary (1674-1732). *Familiar Letters Betwixt a Gentleman and a Lady*, 1725. [13,110 words]
- Davys, Mary (1674-1732). *The Lady’s Tale*, 1725. [7,240 words]
- Defoe, Daniel (1659?-1731). *The Fortunes and Misfortunes of the Famous Moll Flanders*, 1722.
- Haywood, Eliza Fowler (1693?-1756). *Fantomina: or, Love in a Maze*, 1725. [11,900 words]
- Haywood, Eliza Fowler (1693?-1756). *The Fatal Secret: or, Constancy in Distress*, 1725. [9,100 words]
- Hearne, Mary (dates unknown). *Lover’s Week*, 1718. [2,000 words]. COPC
- Manley, Mary de la Rivière (1663-1724). *The Secret History of Queen Zarah and the Zarazians*, 1705.
- Swift, Jonathan (1667-1745). *Gulliver’s Travels*, 1726.

Subperiod BrE2 (1732-1757):

- Cleland, John (1709-1798). *Memoirs of a Woman of Pleasure*, 1749.
- Coventry, Francis (died 1754). *Pompey the Little*, 1752. [2,000 words]. COPC
- Fielding, Henry (1707-1754). *A Journey from this World to the Next*, 1743.
- Fielding, Sarah (1710-1768). *The Adventures of David Simple*, 1744.
- Fielding, Sarah (1710-1768). *The Governess*, 1749. [2,000 words]. COPC
- Haywood, Eliza Fowler (1693?-1756). *The History of Miss Betsy Thoughtless*, 1751.
- Richardson, Samuel (1689-1761). *The History of Sir Charles Grandison*, 1754.
- Smollett, Tobias (1721-1777). *The Adventures of Roderick Random*, 1748.

Subperiod BrE3 (1761-1797):

- Burney, Frances (1752-1840). *Evelina, or, a Young Lady's Entrance into the World*, 1778.
- Godwin, William (1756-1836). *Things as They Are; or, The Adventure of Caleb Williams*, 1794.
- Holcroft, Thomas (1745-1809). *The Adventures of Hugh Trevor*, 1794-1797.
- Jenner, Charles (dates unknown). *The Placid Man*, 1770. [2,000 words]. COPC
- Johnstone, Charles (c1719-c1800). *Chrysal or, The Adventures of a Guinea*, 1761. [2,000 words]. COPC
- Mackenzie, Henry (1745-1831). *The Man of Feeling*, 1771.
- Reeve, Clara (1729-1807). *The Old English Baron: A Gothic Story*, 1780.
- Scott, Sarah Robinson (1723-1795). *The History of Sir George Ellison*, 1766.

Subperiod BrE4 (1850-1879):

- Blackmore, Richard Doddridge (1825-1900). *Lorna Doone: A Romance of Exmoor*, 1869.
- Braddon, Mary Elizabeth (1837-1915). *Lady Audley's Secret*, 1862. [2,000 words]
- Dickens, Charles (1812-1870). *Great Expectations*, 1861.
- Gaskell, Elizabeth Cleghorn (1810-1865). *North and South*, 1855. [2,000 words]
- Le Fanu, Joseph Sheridan (1814-1873). *Uncle Silas: A Tale of Bartram-*

Haugh, 1864.

Linton, Elizabeth Lynn (1822-1898). *The True History of Joshua Davidson*, 1872.

Oliphant, Mrs. (Margaret) (1828-1897). *Salem Chapel: Chronicles of Carlingford*, 1863.

Wood, Mrs. Henry (1814-1887). *East Lynne*, 1861.

Subperiod AmE3 (1774-1804):

Belknap, Jeremy (1744-1798). *The Foresters, an American Tale*, 1792.

Brackenridge, Hugh Henry (1748-1816). *Modern Chivalry: Containing the Adventures of a Captain and Teague O'Regan, His Servant*, 1804.

Foster, Hannah Webster (1759-1840). *The Coquette; or, The History of Eliza Wharton*, 1797.

Tyler, Royall (1757-1826). *The Algerine Captive; or, The Life and Adventures of Doctor Updike Underhill*, 1797.

Subperiod AmE4 (1851-1879):

Alcott, Louise May (1832-1888). *Little Women*, 1869. [2,060 words]
UVirg

Bennett, Emerson (1822-1905). *Ellen Norbury; or, The Adventures of an Orphan*, 1855. [2,080 words]

Davis, Rebecca Harding (1831-1910). *Margret Howth: A Story of Today*, 1862. UVirg

Hawthorne, Nathaniel (1804-1864). *The House of the Seven Gables*, 1851.

Higginson, Thomas Wentworth (1823-1911). *Malbone: An Oldport Romance*, 1869.

Melville, Herman (1819-1891). *I and my Chimney*, 1856. [5,480 words]
UVirg

Melville, Herman (1819-1891). *Bartleby, the Scrivener*, 1856. [14,540 words]

Spofford, Harriet Prescott (1835-1921). *In a Cellar*, 1859. [8,250 words]

Spofford, Harriet Prescott (1835-1921). *Dark Ways*, 1863. [11,940 words]
UVirg

Stowe, Harriet Beecher (1811-1896). *Uncle Tom's Cabin, or, Life among the Lowly*, 1852.

2.3. Contents: Nonfiction Texts

Subperiod BrE1 (1700-1726):

8 extracts of 5,000 words each from Part A of the COPC:³

Addison, Joseph (1672-1719). *Spectator* nos. 13, 61, 119, 125 (1710-1711).

Addison, Joseph (1672-1719). *Guardian* no. 100 (1713), *Freeholder* nos. 2, 4, 40 (1715-1716).

Berkeley, George (1685-1753). *Essay Toward a New Theory of Vision* (1709).

Berkeley, George (1685-1753). *Treatise Concerning Principles of Human Knowledge* (1710).

Defoe, Daniel (1659?-1731). *An Appeal to Honour and Justice* (1715).

Defoe, Daniel (1659?-1731). *A Tour through the Whole Island of Great Britain* (1724).

Steele, Richard (1672-1729). *Christian Hero* (1705), *Tatler* nos. 61, 115 (1710).

Steele, Richard (1672-1729). *Spectator* nos. 11, 49, 109, 113 (1710-1711).

18 extracts of 2,000 words each from Part B of the COPC:

Atterbury, Francis (1663-1732). *English Advice* (1714).

Baron, Samuel (dates unknown). *Kingdom of Tonqueen* (1700).

Bell, John (1691-1780). *St. Petersburg to Peking* (1719).

Blackmore, Richard (1654-1729). *Essays* (1716).

Boyer, Abel (1667?-1729). *Memoirs of Sir William Temple* (1714).

Clarke, Samuel (1675-1729). *Letter to Dodwell* (1706).

Collier, Arthur (1680-1732). *Clavis Universalis* (1713).

Evelyn, John (1620-1706). *Memoirs for his Grandson* (1704).

The Free-Thinker (1720).

Johnson, Richard (died 1721). *Grammatical Commentaries* (1706).

Kennet, White (1660-1728). *History of England* (1706).

King, William (1663-1712). *Journey to London* (1705).

Lockhart, George (1673-1731). *Memoirs of Scotland* (1714).

Mead, Richard (1673-1754). *Account of Poisons* (1702).

Monthly Miscellany (1707).

Ray, John (1627-1705). *Philosophical Letters* (1718).⁴

³ The COPC consists of two parts: Part A, which comprises 5,000-word extracts from “the foremost prose writers of the period” (Milic 1995: 329), and Part B, which comprises 2,000-word extracts from the works of secondary writers.

Rowe, Nicholas (1674-1718). *Life of Shakespear* (1709).
 Ward, John (dates unknown). *Young Mathematician's Guide* (1719).

Subperiod BrE2 (1732-1757):

9 extracts of 5,000 words each from Part A of the COPC:

Bolingbroke, Henry St. John (1678-1751). *Patriot King* (1738).
 Chesterfield, Philip Dormer Stanhope, Earl of (1694-1773). *Letters* (1745-1748).
 Chesterfield, Philip Dormer Stanhope, Earl of (1694-1773). *The World* nos. 24, 49, 98 (1753-1754).
 Fielding, Henry (1707-1754). *Joseph Andrews* (1742).
 Fielding, Henry (1707-1754). *Covent Garden Journal* nos. 3-6, 10 (1752).
 Hume, David (1711-1776). *Treatise of Human Nature* (1739).
 Hume, David (1711-1776). *Essays: Of Eloquence; Of Some Remarkable Customs* (1741-1742).
 Johnson, Samuel (1709-1784). *Rambler* nos. 122, 187 (1751). [2,500 words]
 Smith, Adam (1723-1790). *History of Astronomy* (1750). [2,500 words]

18 extracts of 2,000 words each from Part B of the COPC:

Bower, Archibald (1686-1766). *History of the Popes* (1749).
 Bradley, James (1693-1762). *On Nutation* (1748).
 Cibber, Colley (1671-1757). *Occasional Letter to Pope* (1744).
Daily Advertiser (1741).
 Egmont, Percival (1683-1748). *Diary* (1739).
 Hales, Stephen (1677-1761). *Philosophical Experiments* (1739).
 Hartley, David (1705-1757). *Observations on Man* (1749).
 Holmes, John (dates unknown). *Art of Rhetoric* (1739).
 Lediard, Thomas (1685-1743). *Life of Marlborough* (1736).
London Magazine (1736).
 Luxborough, Henrietta (1699?-1756). *Letters to Shenstone* (1748).
 Melmoth, William (1710?-1799). *Letters of Fitzosborne* (1742).
 Neal, Daniel (1678-1743). *History of the Puritans* (1732).

⁴ Ray's collected letters were published posthumously in 1718. Milic (1995: 330) points out that because "editors and printers... had during the eighteenth century substantial latitude in altering an author's words, it is safe to assume that the language of the selection as represented in these posthumously published letters is that of the publication date and not that of writing, which we cannot accurately determine in any case."

North, Roger (1651?-1734). *Life of Francis North* (1742).⁵
 Perry, Charles (1698-1780). *View of the Levant* (1743).
 Richardson, Jonathan (1665-1745). *Notes on Milton* (1734).
 Walpole, Robert (1676-1745). *Speech* (1734).
 Wood, Robert (1716-1771). *Ruins of Balbec* (1757).

Subperiod BrE3 (1761-1797):

8 extracts of 5,000 words each from Part A of the COPC:

Boswell, James (1740-1795). *Hypochondriack* nos. 1, 21, 27 (1777-1779).
 Boswell, James (1740-1795). *Selections from the Life of Johnson* (1791).
 Gibbon, Edward (1737-1794). *History*, Chapter XXII (1781).
 Gibbon, Edward (1737-1794). *Memoirs* (1790).
 Johnson, Samuel (1709-1784). *Preface to Shakespeare* (1765), *Some Thoughts on the Falkland's Islands* (1771).
 Johnson, Samuel (1709-1784). *Journey to the Western Islands* (1775), *Life of Thomson* (1781).
 Walpole, Horace (1717-1797). *Historic Doubts about Richard III* (1767).
 Walpole, Horace (1717-1797). *Life of Mr. Thomas Baker* (1778).

18 extracts of 2,000 words each from Part B of the COPC:

Aikin, John (1747-1822). *Natural History and Poetry* (1777).
 Chandler, Robert (1738-1810). *Ionian Antiquities* (1769).
 Chelsum, James (1738-1801). *Remarks on Gibbon* (1778).
 Dalrymple, John (1726-1810). *Memoirs of Great Britain* (1771).
 Garrick, David (1717-1779). *Letters* (1771).
 Granger, James (1723-1776). *Biographical History* (1769).
 Gregory, John (1724-1773). *Legacy to his Daughters* (1774).
 Jones, William (1726-1800). *Natural Philosophy* (1762).
 Langhorne, John (1735-1779). *Theodosius and Constantia* (1764).
 Law, Edmund (1703-1787). *Life of Locke* (1777).
 Massey, William (1691-1764?). *Origin of Letters* (1763).
 Osborn, Sarah (1721-1773). *Letters* (1766).
 Pasley, Thomas (1734-1808). *Private Sea Journals* (1778).
 Priestley, Joseph (1733-1804). *Principles of Government* (1768).
 Tytler, William (1711-1792). *Mary Queen of Scots* (1760).
Universal Museum (1762).
 Watson, Richard (1737-1816). *Chemistry* (1771).

⁵ The *Life of Francis North* was published eight years after Roger North's death, by his son Montagu.

Westminster Magazine (1776).

Subperiod BrE4 (1850-1879):

- Arnold, Matthew (1822-1888). *The Study of Celtic Literature*, 1867. [5,000 words] GUT
- Arnold, Matthew (1822-1888). *Culture and Anarchy: An Essay in Political and Social Criticism*, 1869. [5,080 words] GUT
- Bagehot, Walter (1826-1877). *Physics and Politics or Thoughts on the Application of the Principles of 'Natural Selection' and 'Inheritance' to Political Society*, 1869. [5,050 words] GUT
- Borrow, George (1803-1881). *Wild Wales: Its People, Language and Scenery*, 1862. [5,300 words] GUT
- Burton, Richard F. (1821-1890). *The Land of Midian (Revisited)*, 1879. [5,180 words] GUT
- Carlyle, Thomas (1795-1881). *Latter-Day Pamphlets*, 1850. [5,060 words] GUT
- Carlyle, Thomas (1795-1881). *The Life of John Sterling*, 1851. [5,100 words] GUT
- Darwin, Charles Robert (1809-1882). *On the Origin of Species by Natural Selection*, 1859. [5,070 words] GUT
- Private Letters by John Richard Green and Lord and Lady Amberley*, 1861 and 1872. [5,000 words] Denison
- Meredith, George (1828-1909). *Letters Written to the 'Morning Post' from the Seat of War in Italy*, 1866. [5,110 words] GUT
- Meredith, George (1828-1909). *On the Idea of Comedy and of the Uses of the Comic Spirit*, 1877. [5,210 words] GUT
- Mill, John Stuart (1806-1873). *Considerations on Representative Government*, 1861. [5,080 words] GUT
- Mill, John Stuart (1806-1873). "The Contest in America", *Fraser's Magazine*, 1862. [5,080 words] GUT
- Morris, William (1834-1896). *Hopes and Fears for Art*, 1877-1879. [5,080 words] GUT
- Ruskin, John (1819-1900). *Sesame and Lilies*, 1865. [5,060 words] GUT

Subperiod AmE2 (1732-1759):

- Bartram, John (1699-1777). *Observations on the Inhabitants, Climate, Soil, Rivers, Productions, Animals, and Other Matters Worthy of Notice Made by Mr. John Bartram, in his Travels from Pensilvania to Onondago, Oswego and the Lake Ontario, in Canada*, 1751. [11,510 words] UPenn

- Edwards, Jonathan (1703-1758). *Letter to Rev. Dr. Benjamin Colman*, 1735. [4,510 words]
- Edwards, Jonathan (1703-1758). *Freedom of the Will*, 1754. [10,970 words]
- Franklin, Benjamin (1706-1790). *Poor Richard's Almanack*, 1733-1758. [5,900 words] UPenn
- Franklin, Benjamin (1706-1790). *Proposals Relating to the Education of Youth in Pensilvania*, 1749. [2,483 words] UPenn
- Thomson, Charles (1729-1824). *An Enquiry into the Causes of the Alienation of the Delaware and Shawanese Indians from the British Interest*, 1759. [10,135 words] UPenn
- Woolman, John (1720-1772). *Some Considerations on the Keeping of Negroes*, 1754. [4,731 words]

Subperiod AmE3 (1774-1804):

- Franklin, Benjamin (1706-1790). *The Autobiography. Part Three*, 1788-1789. [10,100 words] UVirg
- Hamilton, Alexander (1757-1804). *The Federalist Papers*, 1787-1788. [7,700 words] UVirg
- Hamilton, Alexander (1757-1804). *Letters to Angelica Schuyler Church*, 1789-1801. [2,300 words] UVirg
- Jay, John (1745-1829). *The Federalist Papers*, 1787-1788. [5,200 words] UVirg
- Jefferson, Thomas (1743-1826). *Notes on the State of Virginia*, 1781-1782. [5,000 words] UVirg
- Jefferson, Thomas (1743-1826). *Letters*, 1797-1799. [5,260 words] UVirg
- Woolman, John (1720-1772). *Journal*, 1774. [Chapter XI, written in 1772; 5,000 words] UVirg

Subperiod AmE4 (1851-1879):

- Abbott, John S. C. (1805-1877). *David Crockett: His Life and Adventures*, 1874. [5,020 words] UVirg
- Draper, John William (1811-1882). *History of the Conflict between Religion and Science*, 1881[1874]. [5,120 words] UVirg
- Emerson, Ralph Waldo (1803-1882). *The Conduct of Life*, 1860. [6,010 words] UVirg
- Emerson, Ralph Waldo (1803-1882). *Thoreau*, 1862. [4,544 words]
- Fiske, John (1842-1901). *Spain and the Netherlands*, 1868. [5,040 words] UVirg

- Fiske, John (1842-1901). *Athenian and American Life*, 1873. [5,130 words] UVirg
- Holmes, Oliver Wendell (1809-1894). *The Autocrat of the Breakfast-Table*, 1858. [5,050 words] GUT
- Holmes, Oliver Wendell (1809-1894). *Medical Libraries*, 1878. [5,120 words] GUT
- Motley, John Lothrop (1814-1877). *The Life and Death of John of Barneveld, Advocate of Holland with a View of the Primary Causes and Movements of the Thirty Years' War*, 1874. [5,040 words] GUT
- Several authors. A group of private letters, 1861-1867. [about 3,800 words] UVirg
- Spooner, Lysander (1808-1887). *No Treason: The Constitution of No Authority*, 1870. [5,020 words] UVirg
- Stowe, Harriet Beecher (1811-1896). *The Education of Freedmen*, 1879. [5,100 words] UVirg
- Thoreau, Henry David (1817-1862). *The Correspondence of Henry David Thoreau*, 1851-1862. [Three letters from the years 1851 and 1862; 1,260 words.]
- Thoreau, Henry David (1817-1862). *Walden, or, Life in the Woods*, 1854. [5,040 words] UVirg
- Twain, Mark (1835-1910). *Innocents Abroad*, 1869. [5,000 words] UVirg
- Whitman, Walt (1819-1892). *Memoranda during the War*, 1875-1876. [5,100 words] UVirg

3. Research Possibilities and Limitations

The corpus is too small for lexicographic purposes, and is also unsuitable for the study of phenomena related to spelling or punctuation. It should also be borne in mind that texts retrieved from internet sources often lack exact bibliographical information, so that it is not clear who the editors were, nor the degree of editorial intervention. Furthermore, Milic (1995: 330-331) acknowledges that he opted for “normalizing” the spelling, punctuation and capitalization practices of the texts in the COPC.

A second limitation is that, as noted in section 2, the corpus is biased towards texts written by literate members of English and American society in the eighteenth and nineteenth centuries, and includes neither speech-related genres such as town and court depositions, where speech is taken down as either direct or indirect speech, nor drama, in which dialogues are written to be spoken. Many of the fiction extracts included in the corpus contain dialogue written to represent speech, but in general COLMOBAENG is not adequate to address questions having to do with the natural spoken

language of the period. By contrast, in terms of Biber's multi-dimensional approach to register variation, discourse dimensions such as narrative vs. non-narrative concerns (Dimension 2; see Biber 1988: 109, 136) and informational vs. involved production (Dimension 1; see Biber 1988: 107, 128) are well represented in the corpus and might be explored by selecting samples from the appropriate registers (e.g. fictional and historical prose, philosophical writings, academic prose, etc.).

COLMOBAENG was compiled with the specific aim of investigating syntactic development within the Late Modern period, and thus far has been used successfully to explore changes in the English system of sentential complementation, such as the competition between infinitives and gerunds as preverbal and extraposed subjects (e.g. "*Inviting / To invite the twins* was a bad mistake"; see Fanego 2007: 186-219, 2010), and the specialization of gerunds to encode object clauses with "non-controlled" implicit subjects (e.g. "A visit to London entails *bringing gifts for friends*") as opposed to infinitival object clauses with controlled implicit subjects (e.g. "I want *to see Mary*"; see Fanego 2007: 178-186). Other areas of grammar that were also undergoing change in LME and which would thus constitute suitable research topics include the *be Ving* progressive, the passive and other impersonalizing strategies, and the variation between subjunctives, indicatives and modals in certain types of clauses (e.g. "... if the temperature *be increased / was increased / should be increased*").

Finally, in terms of size the corpus is comparable to traditional small historical corpora such as the Helsinki Corpus (551,000 words of Early Modern English; see Kytö 1996), CONCE (1,000,000 words of nineteenth-century English) and ARCHER (1,700,000 words for the time span 1650-1990). It can be used most successfully to supplement the latter two, in particular ARCHER, since this, as already noted, comprises only a few relatively small samples of eighteenth- and nineteenth-century British and American English. Unlike ARCHER, however, which can only be consulted in situ at one of the fourteen departments belonging to the ARCHER Consortium (see López-Couso & Méndez-Naya, this volume), COLMOBAENG is freely available and can be obtained by simply contacting the author of this chapter (teresa.fanego@usc.es).

References

Primary sources

COLMOBAENG comprises materials from the following sources:

a) electronic databases:

COPC	Century of Prose Corpus (Milic 1995).
Denison	A Corpus of Late Modern English Prose (Denison 1994).
EAF	Early American Fiction, Chadwyck-Healey.
ECF	Eighteenth-Century Fiction, Chadwyck-Healey.
GUT	Project Gutenberg.
NCF	Nineteenth-Century Fiction, Chadwyck-Healey.
UPenn	Online Books Page, University of Pennsylvania.
UVirg	Electronic Text Center, University of Virginia.

b) printed texts:

- Edwards, Jonathan. 1735. *Letter to Rev. Dr. Benjamin Colman*. In Nina Baym (ed.), *The Norton Anthology of American Literature*, Vol. I: *Literature to 1820*. New York: W.W. Norton & Company, 2003.
- . 1754. *Freedom of the Will*. In Paul Ramsey (ed.), *The Works of Jonathan Edwards*, Vol. I. New Haven: Yale University Press, 1957.
- Emerson, Ralph Waldo. 1862. *Thoreau*. In Richard Poirier (ed.), *The Oxford Authors: Ralph Waldo Emerson*. Oxford: Oxford University Press, 1990.
- Thoreau, Henry David. 1851-1862. In Walter Harding & Carl Bode (eds.), *The Correspondence of Henry David Thoreau*. New York: New York University Press, 1958.
- Woolman, John. 1754. *Some Considerations on the Keeping of Negroes*. In Nina Baym (ed.), *The Norton Anthology of American Literature*, Vol. I: *Literature to 1820*. New York: W.W. Norton & Company, 2003.

Secondary sources

- Beal, Joan C. 2004. *English in Modern Times 1700-1945*. London: Arnold.
- Biber, Douglas. 1988. *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- Biber, Douglas, Edward Finegan & Dwight Atkinson. 1994. ARCHER

- and its challenges: Compiling and exploring A Representative Corpus of Historical English Registers. In Udo Fries, Peter Schneider & Gunnel Tottie (eds.), *Creating and Using English Language Corpora. Papers from the 14th International Conference on English Language Research on Computerized Corpora, Zurich 1993*, 1-13. Amsterdam: Rodopi.
- Denison, David. 1994. A Corpus of Late Modern English Prose. In Merja Kytö, Matti Rissanen & Susan Wright (eds.), *Corpora across the Centuries*, 7-16. Amsterdam: Rodopi.
- . 1998. Syntax. In Suzanne Romaine (ed.), *The Cambridge History of the English Language*, Vol. 2: 1776-1997, 92-329. Cambridge: Cambridge University Press.
- De Smet, Hendrik. 2005. A Corpus of Late Modern English Texts. *ICAME Journal* 29: 69–82.
- Fanego, Teresa. 2007. Drift and the development of sentential complements in British and American English from 1700 to the present day. In Javier Pérez-Guerra, Dolores González-Álvarez, Jorge L. Bueno-Alonso & Esperanza Rama-Martínez (eds.), *'Of Varying Language and Opposing Creed': New Insights into Late Modern English*, 161-235. Bern, etc.: Peter Lang.
- . 2010. Variation in sentential complements in eighteenth- and nineteenth-century English: A processing-based explanation. In Raymond Hickey (ed.), *Eighteenth-century English: Ideology and Change*, 200-220. Cambridge: Cambridge University Press.
- Görlach, Manfred. 1999. *English in Nineteenth-Century England. An Introduction*. Cambridge: Cambridge University Press.
- . 2001. *Eighteenth-Century English*. Heidelberg: Universitätsverlag C. Winter.
- Kytö, Merja. 1996 [1991]. *Manual to the Diachronic Part of the Helsinki Corpus of English Texts: Coding Conventions and Lists of Source Texts*. 3rd ed. Helsinki: Department of English, University of Helsinki.
- Kytö, Merja & Matti Rissanen. 1983. The syntactic study of Early American English. The variationist at the mercy of his corpus. *Neophilologische Mitteilungen* 84: 470-490.
- Kytö, Merja, Juhani Rudanko & Erik Smitterberg. 2000. Building a bridge between the present and the past: A Corpus of 19th-century English. *ICAME Journal* 24: 85–97.
- Kytö, Merja, Mats Rydén & Erik Smitterberg, eds. 2006. *Nineteenth-Century English. Stability and Change*. Cambridge: Cambridge University Press.

- Milic, Louis T. 1995. The Century of Prose Corpus: A half-million word historical data base. *Computers and the Humanities* 29: 327-337.
- Tieken-Boon Van Ostade, Ingrid. 2009. *An Introduction to Late Modern English*. Edinburgh: Edinburgh University Press.

