

The Corpus of Early English Correspondence Sampler (CEECS)

Arja Nurmi
University of Helsinki

1 Introduction

The compilation of the *Corpus of Early English Correspondence* (CEEC)¹ began in September 1993, and while the full corpus is still not ready for publication we are able to release a part of it. This part is called the *Corpus of Early English Correspondence Sampler* (CEECS). The corpora are compiled by the Sociolinguistics and Language History Project Team at the Department of English, University of Helsinki. The team includes Terttu Nevalainen, Helena Raumolin-Brunberg, Jukka Keränen, Minna Nevala, Arja Nurmi and Minna Palander-Collin.

The creation of the corpus proved necessary in order to provide suitable material for the application of sociolinguistic methods to the historical study of English (more on the theoretical background is found in Nevalainen and Raumolin-Brunberg 1994 and 1996, pilot studies using the CEEC and exploring the methodology are collected in Nevalainen and Raumolin-Brunberg (eds) 1996).

Our work has been financed by the Academy of Finland (1993–95) and the University of Helsinki (1996–98). Since the project funding is now over, the development of the full CEEC has temporarily been halted, waiting for clearing of copyright and final corrections, but as an interim measure we decided to offer the parts of the corpus no longer in copyright for the general use of researchers. This resulted in the CEECS presented in this article and available on the new ICAME CD-ROM.

2 Description of the corpus

The *Corpus of Early English Correspondence Sampler* (CEECS) reflects the same compilation principles as the full *Corpus of Early English Correspondence* (CEEC). The aim has been to reach as broad a social representativeness as possible, though several obstacles immediately present themselves. The two most

obvious ones, the low level of literacy in late medieval and early modern England and the haphazard preservation of letters, are shared by all historical research. A more unexpected problem is the penchant, particularly of 19th-century editors, to edit only the letters of historically important people, and ones describing important historical events. Editors often disregarded family letters concerning everyday life, which would serve as better material for historical sociolinguistics.

Table 1: The composition of the *Corpus of Early English Correspondence Sampler* (CEECS): subdivision, collections included, years covered and word counts of the samples used (see Appendix for the bibliographical information)

	Collection	Years	Word count
CEECS		1418–1680	450085
	<i>CEECS1</i>	1418–1638	246055
	Original 1	1418–1529	23176
	Stonor	1424–1483	38006
	Marchall	1440–1476	4834
	Shillingford	1447–1448	13527
	Plumpton	1461–1550	36530
	Rerum	1483–1509	5915
	Original 2	1520–1586	16879
	Hutton	1566–1638	25319
	Leycester	1585–1586	67786
	Royal 1	1585–1596	14083
	<i>CEECS2</i>	1580–1680	204030
	Original 3	1580–1665	9948
	Henslowe	1600–1610	551
	Royal 2	1612–1614	227
	Cornwallis	1613–1644	61603
	Cosin	1617–1669	37853
	Harley	1625–1666	24915
	WeSa	1632–1642	4320
	Charles	1634–1678	2964
	Wharton	1642	8068
	Hamilton	1648–1650	1091
	Jones	1651–1660	33877
	Basire	1651–1666	7068
	Tixall	1656–1680	11545

In selecting the letters we have aimed at the greatest possible authenticity, choosing autograph letters whenever possible and looking for editions which not only produce original spelling, but also explain their editorial principles as explicitly as possible. While we have striven for authenticity, the poor quality of existing material from certain periods, regions, and social strata has forced us occasionally to accept also copies, scribal letters and doubtful cases in order to ensure as wide a social representativeness as possible. A case in point are the Plumpton letters, edited from early sixteenth-century copies, but representing the otherwise scarce language of fifteenth- and early sixteenth-century Northern England. The text of the Plumpton letters comes from the old edition, but the dating follows the more reliable modern one (Kirby 1996). As an extreme measure we have edited a handful of letters ourselves to fill in gaps in the social representation (for a list of editions used see Appendix).

Even though the corpus is based on editions, we have found it a reliable tool for the study of morphology and syntax, as well as pragmatics. Despite its limited size, it can be useful in some types of semantic and textual research, but not in the study of orthography. Particularly the older editions (ie the ones included in the CEECS) cannot be relied upon in questions of spelling, as the editors' priorities were often not linguistic but historical. Even in the case of newer editions – and those by language scholars – the influence of nearly illegible manuscripts, the result of decay and bad handwriting, may make these a less than reliable source for studies of orthography.

Table 2 illustrates the differences between the full letter corpus and the sampler part now published. The sampler may not be quite representative of the whole CEEC, as the criterion for selection was corpus external and depended on such a random variable as the date of the editor's death; the editions used for CEECS are out of copyright because all the editors have died more than 70 years ago. The only exceptions to the rule are ourselves; Jukka Keränen has edited the sample now made available of the Henslowe collection and Jukka Keränen, Terttu Nevalainen and Arja Nurmi have re-edited the letters in the Marchall collection, as Lyell's (1934) version proved too unreliable.

Table 2: The *Corpus of Early English Correspondence* (CEEC) and the *Corpus of Early English Correspondence Sampler* (CEECS) compared

	CEEC	CEECS
size (words)	2.7 million	450,085
time span	1417–1681	1418–1680
number of collections	96	23
number of letters	ca 6000	1147
number of writers	777	194
% women of writers	20%	23%

Although the CEECS is less representative than the CEEC, the sampler corpus resembles the full CEEC in most respects. The temporal division is skewed in the same way with most material in the latter end of the corpus, and the proportion of women is approximately the same. There is a slight overrepresentation of royalty, as these were the letters the majority of early editors were most eager to publish, and some of our best collections of non-gentry writers (Cely, Johnson and Marescoe) are still in copyright.

3 Coding

The CEECS contains two levels of coding: textual and parameter coding. The text level coding has been adopted from the Helsinki Corpus (see Kytö 1996 for details) with the exception of line format: as letter editions have not preserved the original lineation of the manuscripts, we have seen no need to preserve the line format of our selected editions. The parameter coding is very much a stripped version of the Helsinki Corpus: only <B ‘name of text file’, <Q ‘text identifier’, <A ‘author’ and <P ‘page’ have been retained.²

Due to the different nature of the two corpora the text identifier contains somewhat different information from that of the Helsinki Corpus. As the CEECS is based on letter collections, the first three characters given in the CEECS code identify the collection in question. That is followed by date (or an estimated decade in some cases; for details see Nurmi 1998) and writer code. The code <Q OR3 1600? FDEVEREUX> tells us that the example in question comes from the collection Original 3 (index of collection codes is included in the manual), that it has presumably been written in the year 1600 and that the writer is F. Devereux. The codes are explicated in the manual index of writer codes, but the ‘author’ parameter can also be used for this purpose: <A

FRANCES DEVEREUX> tells us that the writer is a woman called Frances Devereux, and the editorial heading of the letter itself [] [LETTER CCXXXVII. THE COUNTESS OF ESSEX TO MR SECRETARY CECIL.] [] tells us that she is the Countess of Essex.

Our letter writers are identified by first and last name rather than by title, as during the period covered by our corpus there are several holders of some individual titles (eg Duke of Norfolk), and several titles for some individuals (eg William Cecil, later Sir William Cecil and then Lord Burghley). This also simplifies the use of reference works, such as the *Dictionary of National Biography*, which follow a similar policy.

Our earlier articles have discussed the possibility of coding some of the sociolinguistic background information we have gathered into the corpus. This background information can be divided into two groups: 1) facts and 2) interpretations (Nevalainen and Raumolin-Brunberg 1996: 53; Raumolin-Brunberg 1997). Some of the facts we have gathered (name of letter writer and year of letter) have been included, but most have been left out for three reasons. Firstly, our background data is in a separate database and cannot be easily incorporated into the corpus as such. Secondly, manually coding any information into each of the 1,147 letters of the CEECS is beyond our resources at the moment. Finally, most of the incontestable facts (names, dates, careers) are easily available for corpus users in standard reference works such as the *Dictionary of National Biography* or the *British Biographical Index* (not to forget the editions themselves), largely thanks to the 19th-century editors' preference for editing letters by important people. On the other hand, our interpretations (social rank and mobility, domicile, etc) are still preliminary classifications and need testing as to their usefulness and reliability.

4 Types of letter included

It is important to note that the CEEC and CEECS consist of personal letters (discussed in more detail by Nevalainen and Raumolin-Brunberg 1996),³ and private letters are only a part of this category. Personal letters here mean letters written by an identifiable individual to another identifiable individual (some letters written by a collective, such as the Privy Council, have been included in the CEEC as comparative material, but only one such letter is found in CEECS). The most typical example of truly private letters are family letters, as in example [1]:

- [1] <Q COR 1623? NA2BACON>
<A NATHANIEL BACON>
<P 83>
[] [LVI. NATHANIEL BACON TO JANE LADY CORNWALLIS.\] }]
Sweet Hart,
 Hauinge this opportunitie by M=r= Chittock, I thought these my letters
myght come to yo=e= hands before yo=e= retourne, being assured you
would be very glad to heare of o=e= estate here in the contry. For my
self, therefore, you may vnderstand that I
<P 84>
am reasonably well, but cannot quite my self of my distempers, although
they be very small. [...] O=e= childeren ar well; & little Nick hath cast
his cote, and seemeth metamorphosed into a grasshopper. Jane is a very
modest mayden, & is wholly taken vpp w=th= trauailinge by her self,
w=ch= she perfourmeth very hansomely, & wil be ready to runn at yo=e=
comand when you retourne. Thus w=th= my best love & prayers I leaue,
resting alwaies & onely
 Yo=es= Nath. Bacon.
[1622-3.\]
 To his best respected friend the Lady Cornewalleys, at her lodginge
ouer agaynst York Howse, at the signe of the Stirrop.

Perhaps untypically, here it is the husband who remains at home with the children while the wife is at Court, but otherwise the content is very characteristic of family letters (due to restrictions of space some passages in the middle have been left out). Another typical type of letter is the business letter, as in [2]:

- [2] <Q HEN 1600S WFAWNTE>
<A WILLIAM FAWNTE>
<P F83>
[] [F.83. WILLIAM FAWNTE TO EDWARD ALLEYN\] }]
M=r=. Allin mey Loue remembered I vnderstoode bey a man which
came with too Beares from the gardeyne that you haue a deseyre to bey~
one of mey Boles. I haue three westerne boles at this teyme but I haue
had verey ell loeck with them for one of them hath lost his horne to the
queyck that I think that hee will neuer bee to feyght a gayne that is mey
ould star of the west hee was a verey esey bol and mey Bol, Bevis hee
hath lost one of his eyes but I think if you had him hee would do you
more hurt then good for I protest I think hee would ether throo vp your

dodges in to the loftes oreles dingout theare braynes a gen~st the grates
so that I think hee is not for your turne [...]

Here one organiser of bear-baiting and other animal fights writes to another in a very businesslike manner about buying and selling animals. Most business letters do not of course have such gruesome content; letters dealing with the buying and selling of wool and other commodities are far more common. Merchants' letters are underrepresented in the CEECS, as mentioned above, because many of the best collections are still in copyright.

Perhaps the more typical opposite pole to private letters are official letters written by a government official in his official capacity on official business. Example [3] shows instructions from Sir Francis Walsingham, Elizabeth I's spymaster to Robert Dudley, Earl of Leicester (the corpus follows the old-fashioned spelling 'Leycester' used by the edition) during the latter's government of the Low Countries. Dudley was having a hard time trying to please the Queen's ever changing opinions about what should be done:

[3] <Q LEY 1586 FWALSINGHAM>
<A FRANCIS WALSYNGHAM>
<P 221>
{ } [LETTER LXXXI. MR. SECRETARY WALSYNGHAM TO THE
EARL OF LEYCESTER. 11TH APRIL 1586.\] { }
<P 222>

My very good lord, as I have alwaies thowght, sythence your lordships first entrie into the charge you now howld, the assystance of sir William Pelham most necessarye for your lordship, so have I just cause, wayghing the late mutiny happened at Utreck by a bande perteyning unto coroneel Norryce, to thinke the removing of the one as necessarye as the placying of the other. I see some reason to dowbt that the grownde of the seyd coronells caryag of himself towardses your lordship grew by practyce from hence. The nurishing of factyon at home and abroade is thowght here the best coorse of pollecye, but the myschefe yt wyll breed I feare wyll prove irreparable. [...]

One particular subtype of letters are the pleas to high officials, usually begging for release from prison, return of property and other vital favours. In the following, the Countess of Essex pleads to William Cecil, Lord Burghley, Chief Minister of Queen Elizabeth, to prevent the execution of her husband, the Earl of

Essex (Elizabeth's favourite). Despite the touching plea Essex was beheaded in 1601.

- [4] <Q OR3 1600? FDEVEREUX>
<A FRANCES DEVEREUX>
<P 55>
{ } [LETTER CCXXXVII. THE COUNTESS OF ESSEX TO MR SECRETARY CECIL.] }]
<P 56>
Sir
Although the awnswere I receved from you two daies since gave mee small incouragement to flatter my self that anie importunity I could make should bee able to appease the scandall you had conceaved to bee geven you by my unfortunat husband: yet hade it not pleased God to powre uppon mee one affliction
<P 57>
after an other, and to add to the immesurable sorrowes of my harte so violent a sicknes as I am not able of my self to stur out of my bed, I had presented unto your vew the image of the importunante Widow mentioned in the Scriptur, and had never ceased to pester you with my complaines till you had afforded me some assurance that, whatsoever respects might dehorte you from so much as wishinge my husband's good, yet that an afflicted and wofull lady should not wholly loose her labor, or returne desperate of such comforts at the last yeare you so honorably ministred unto mee in a great affliction though differinge from this in qualitie. [...]

Conclusion

In publishing the *Corpus of Early English Correspondence Sampler* (CEECS) we hope to give a preview of what is to come later, with the eventual publication of the full corpus. Although relatively small in size, we hope CEECS will prove a useful tool for the linguistic community.

Notes

1. CEEC is pronounced /si:k/ and CEECS is pronounced /si:ks/. The earlier name of the corpus, the *Helsinki Corpus of Early English Correspondence* (HEEC) was changed because it proved too misleading.

2. Keränen (1998: 33) describes these parameters in a slightly different form. In the final editing of CEECS the information in the text identifier was modified to suit the publication format of the corpus, and the author parameter was changed from X to A in accordance with the usage of the Helsinki Corpus.
3. For a definition of letter, see eg Kohonen (1997).

References

- British biographical index*. 1990. Ed. by David Bank and Anthony Esposito. Vol. 1–4. Glasgow University. Humanities Reference Systems Unit.
- Dictionary of national biography*. 1885–1912. Ed. by Leslie Stephen and Sidney Lee. Vol. 1–63. London.
- Hickey, Raymond, Kytö, M., Lancashire I. and Matti Rissanen (eds). 1997. *Tracing the trail of time. Proceedings from the Second Diachronic Corpora Workshop*. Language and Computers: Studies in Practical Linguistics 18. Amsterdam & Atlanta, GA: Rodopi.
- Keränen, Jukka. 1998. The Corpus of Early English Correspondence: Progress report. In A. Renouf (ed) *Explorations in corpus linguistics*. Language and Computers: Studies in Practical Linguistics 23. Amsterdam & Atlanta, GA: Rodopi, 29–37.
- Kirby, Joan (ed). 1996. *The Plumpton letters and papers*. Camden Fifth Series 8. London: Cambridge University Press for the Royal Historical Society.
- Kohonen, Thomas. 1997. Toward a theoretical foundation of ‘text type’ in diachronic corpora: Investigations with the Helsinki Corpus. In R. Hickey et al (eds), 185–195.
- Kytö, Merja (comp). 1996. *Manual to the diachronic part of the Helsinki Corpus of English Texts. Coding conventions and lists of source texts*. 3rd edition. Department of English, University of Helsinki.
- Lyell, Laetitia (ed). 1934. *A mediaeval post-bag*. London: Jonathan Cape.
- Nevalainen, Terttu. 1997. Ongoing work on the Corpus of Early English Correspondence. In R. Hickey et al (eds), 81–90.
- Nevalainen, Terttu and Helena Raumolin-Brunberg. 1994. Sociolinguistics and language history: The Helsinki Corpus of Early English Correspondence. *Hermes, Journal of Linguistics* 13:135–143.
- Nevalainen, Terttu and Helena Raumolin-Brunberg. 1996. The Corpus of Early English Correspondence. In T. Nevalainen and H. Raumolin-Brunberg (eds), 39–54.

- Nevalainen, Terttu and Helena Raumolin-Brunberg (eds). 1996. *Sociolinguistics and language history. Studies based on the Corpus of Early English Correspondence*. Language and Computers: Studies in Practical Linguistics 15. Amsterdam & Atlanta, GA: Rodopi.
- Nurmi, Arja (ed). 1998. *Manual for the Corpus of Early English Correspondence Sampler CEECS*. Department of English. University of Helsinki. Available at <http://www.hit.uib.no/icame/ceecs/index.htm>.
- Raumolin-Brunberg, Helena. 1997. Incorporating sociolinguistic information into a diachronic corpus of English. In R. Hickey et al (eds), 105–117.

**Appendix: Letter collections included in the CEECS
(for sample size see Table 2)**

Basire

The correspondence of Isaac Basire, D.D. Archdeacon of Northumberland and prebendary of Durham, in the reigns of Charles I. and Charles II. with a memoir of his life. Ed. by Darnell, William Nicholas. London: John Murray. 1831.

Charles

Sample 1: *Five letters of King Charles II*. Ed. by the Marquis of Bristol. Camden Miscellany 5. Camden First Series 87. 1864/1968.

Sample 2: *Letters of the Council to Sir Thomas Lake, relating to the proceedings of Sir Edward Coke at Oatlands*. Ed. by Gardiner, Samuel Rawson. Camden Miscellany 5. Camden First Series 87. 1864/1968.

Cornwallis

The private correspondence of Jane Lady Cornwallis; 1613–1644. Ed. by Griffin, Richard, Baron Braybrooke. London: S. & J. Bentley, Wilson, & Fley. 1842.

Cosin

The correspondence of John Cosin, D.D. Lord Bishop of Durham: Together with other papers illustrative of his life and times. Parts I–II. Ed. by Ornsby, George. Publications of the Surtees Society 52, 55. 1868, 1870.

Hamilton

The Hamilton papers: Being selections from original letters in the possession of his grace the Duke of Hamilton and Brandon, relating to the years 1638–1650. Ed. by Gardiner, Samuel Rawson. Camden New Series 27. 1880.

Harley

Letters of the Lady Brilliana Harley, wife of Sir Robert Harley, of Brampton Bryan, Knight of the Bath. Ed. by Lewis, Thomas Taylor. Camden First Series 58. 1854.

Henslowe

Sample 1: “Forgeries and one-eyed bulls: Editorial questions in corpus work”. Ed. by Keränen, Jukka. *Neuphilologische Mitteilungen* 99/2: 217–226. Helsinki. 1998.

Hutton

The correspondence of Dr. Matthew Hutton, Archbishop of York. With a selection from the letters, etc. of Sir Timothy Hutton, Knt., his son; and Matthew Hutton, Esq., his grandson. Ed. by Raine, James. Publications of the Surtees Society 17. 1843.

Jones

Inedited letters of Cromwell, Colonel Jones, Bradshaw and other regicides. Ed. by Mayer, Joseph. Transactions of the Historic Society of Lancashire and Cheshire New Series 1. 1861.

Leycester

Correspondence of Robert Dudley, Earl of Leycester, during his government of the Low Countries, in the years 1585 and 1586. Ed. by Bruce, John. Camden First Series 27. 1844.

Marchall

Letters from the Marchall correspondence. Ed. by Keränen, Jukka, Terttu Nevalainen and Arja Nurmi. Public Record Office, SC1. 1996.

Original 1

Original letters, illustrative of English History; including numerous royal letters: from autographs in the British Museum, and one or two other collections. Vol I. 2nd edition. Ed. by Ellis, Henry. London: Harding, Triphook, and Lepard. 1825.

Original 2

Original letters, illustrative of English history; including numerous royal letters: from autographs in the British Museum, and one or two other collections. Vol II. 2nd edition. Ed. by Ellis, Henry. London: Harding, Triphook, and Lepard. 1825.

Original 3

Original letters, illustrative of English history; including numerous royal letters: From autographs in the British Museum, and one or two other collections. Vol III. 2nd edition. Ed. by Ellis, Henry. London: Harding, Triphook, and Lepard. 1825.

Plumpton

Plumpton correspondence. A series of letters, chiefly domestick, written in the reigns of Edward IV. Richard III. Henry VII. and Henry VIII. Ed. by Stapleton, Thomas. Camden First Series 4. 1839.

Rerum

Rerum Britannicarum medii aevi scriptores. Or chronicles and memorials of Great Britain and Ireland during the Middle Ages. Letters and papers illustrative of the reigns of Richard III. & Henry VII. Vol. I–II. Ed. by Gairdner, James. London: Longman, Green, Longman, and Roberts. 1861, 1863.

Royal 1

Letters of Queen Elizabeth and King James VI. of Scotland; some of them printed from originals in the possession of the Rev. Edward Ryder, and others from a MS. which formerly belonged to Sir Peter Thompson, Kt. Ed. by Bruce, John. Camden First Series 46. 1849.

Royal 2

Sample 1: *Letters to King James the Sixth from the Queen, Prince Henry, Prince Charles, the Princess Elizabeth and her Husband Frederick King of Bohemia, and from their Son Prince Frederick Henry.* Ed. by Walker, Sir Patrick and Alexander Macdonald. Edinburgh: The Maitland Club. 1835.

Shillingford

Letters and papers of John Shillingford, Mayor of Exeter 1447–50. Ed. by Moore, Stuart A. Camden New Series 2. 1871/1965.

Stonor

Sample 1: *The Stonor letters and papers, 1290–1483.* Vols. I–II. Ed. by Kingsford, Charles Lethbridge. Camden Third Series 29, 30. 1919.

Sample 2: *Supplementary Stonor letters and papers (1314–1482).* Ed. by Kingsford, Charles Lethbridge. Camden Miscellany 13. Camden Third Series 34. 1924 [referred to as Vol. 3 in page number coding in the corpus].

Tixall

Tixall letters; or the correspondence of the Aston family, and their friends, during the seventeenth century. Vol II. Ed. by Clifford, Arthur. London: Longman, Hurst, Rees, Orme, and Brown and Edinburgh: Archibald Constable and Co, and John Ballantyne and Co. 1815.

WeSa

Sample 1: *Four letters of Lord Wentworth, afterwards Earl of Strafford, with a poem on his illness.* Ed. by Gardiner, Samuel Rawson. Camden Miscellany 8. Camden New Series 31. 1883/1965.

Sample 2: *Papers relating to the delinquency of Lord Savile, 1642–1646.* Ed. by Cartwright, James J. Camden Miscellany 8. Camden New Series 31. 1883/1965.