THE CORPUS OF EARLY ENGLISH CORRESPONDENCE EXTENSION (CEECE)

THE CORPUS

The Corpora of Early English Correspondence (CEEC) are diachronic corpora of personal letters designed for historical sociolinguistics, compiled at the University of Helsinki (see CoRD entry). CEEC consists of three separate corpora, which together span the years c. 1400–1800 and contain over 5 million words (Nevala & Nurmi forthcoming). This poster introduces the Late Modern English part: CEEC Extension (CEECE).

CEECE covers the long eighteenth century, starting from 1680 (where the original CEEC cuts off) and ending in 1800 (Table 1 and Fig. 1; see Laitinen 2002).

Table 1. CEECE in numbers by gender.

<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th>Women</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informants</td>
<td>214</td>
<td>94</td>
<td>308</td>
</tr>
<tr>
<td>Letters</td>
<td>3,681</td>
<td>1,242</td>
<td>4,923</td>
</tr>
<tr>
<td>Words</td>
<td>1,622</td>
<td>0.6m</td>
<td>2.22m</td>
</tr>
</tbody>
</table>

18th-century England saw rapid growth in literacy. However, as CEECE is compiled from printed (original-spelling) editions, it reflects “the edited truth”: gentry remain well represented, but there are sorely few editions of Late Modern English letters by writers from the lower social ranks (Table 2, Fig. 2).

Table 2. Social ranks of informants.

<table>
<thead>
<tr>
<th>Social Rank</th>
<th>Men</th>
<th>Women</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Royalty</td>
<td>10</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td>Nobility</td>
<td>18</td>
<td>8</td>
<td>26</td>
</tr>
<tr>
<td>Gentry</td>
<td>55</td>
<td>26</td>
<td>81</td>
</tr>
<tr>
<td>Clergy</td>
<td>46</td>
<td>10</td>
<td>56</td>
</tr>
<tr>
<td>Professionals</td>
<td>41</td>
<td>12</td>
<td>53</td>
</tr>
<tr>
<td>Merchants</td>
<td>16</td>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td>Other non-gentry</td>
<td>28</td>
<td>12</td>
<td>40</td>
</tr>
</tbody>
</table>

The scarcity of East Anglian informants in CEECE is another reflection of our sources. Yet changes in English society are visible in the geographical coverage of CEECE. Increased regional mobility can make establishing domiciles difficult. But many lived in the capital: London tripled in size 1650–1800, and the proportion of the population living in London doubled from 6% to 12%.

REGIONS

The CEECE corpora were compiled at the Research Unit for Variation, Contacts and Change in English (VARIENG), Department of Modern Languages, University of Helsinki. The compilers of CEECE are Terttu Nevalainen (leader), Helena Raumolin-Brunberg, Samuli Kaislaniemi, Mikko Hakala, Minna Nevala, Arja Nurmi, Minna Palander-Collin, Tanja Säily, and Anni Sairio.

Since the publication of PCEEC in 2006 (see CoRD), publishers have become increasingly reluctant to grant permissions, or demand inordinate fees while insisting on unrealistic restrictions on the use of texts. We are looking for an ARCHER-like solution for making CEECE publicly available; at present access is only on-site. Studies using CEECE continue to come out at a steady pace (see bibliography in CoRD entry). These four posters are the first presentations of new work done on the CEECE to be published in Nevalainen et al. (forthcoming).

Table 3. Regional distribution of informants.

<table>
<thead>
<tr>
<th>Region</th>
<th>Men</th>
<th>Women</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Court</td>
<td>5%</td>
<td>4%</td>
<td>5%</td>
</tr>
<tr>
<td>London</td>
<td>23%</td>
<td>27%</td>
<td>24%</td>
</tr>
<tr>
<td>East Anglia</td>
<td>2%</td>
<td>0%</td>
<td>1%</td>
</tr>
<tr>
<td>Home Counties</td>
<td>16%</td>
<td>24%</td>
<td>19%</td>
</tr>
<tr>
<td>North England</td>
<td>17%</td>
<td>3%</td>
<td>13%</td>
</tr>
<tr>
<td>Other areas</td>
<td>37%</td>
<td>41%</td>
<td>38%</td>
</tr>
</tbody>
</table>

Table 4. Comparison of informants in CEEC and CEECE: social ranks.

<table>
<thead>
<tr>
<th>Social Rank</th>
<th>CEEC Men</th>
<th>CEEC Women</th>
<th>Total</th>
<th>CEECE Men</th>
<th>CEECE Women</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Royalty</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nobility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gentry</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clergy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professionals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Merchants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other non-gentry</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. CEECE words over time.

Figure 2. Map of regions covered by CEECE.

Figure 2. Proportion of words from the highest (Royalty & Nobility) and lowest (Other non-gentry) social ranks in CEEC and CEECE. *First two periods in CEEC combined.

CEECE VS. CEECE

In terms of socio-regional representativeness, CEECE is not quite as well balanced as CEEC. But CEECE is ‘thicker’ than the original CEEC: it contains over twice as much material measured over time or per informant. CEECE also contains proportionately more letters from women and from the middle and lower social ranks than CEEC (Tables 1, 4).

CURRENT STATE

COMPILERS

The CEEC corpora were compiled at the Research Unit for Variation, Contacts and Change in English (VARIENG), Department of Modern Languages, University of Helsinki. The compilers of CEECE are Terttu Nevalainen (leader), Helena Raumolin-Brunberg, Samuli Kaislaniemi, Mikko Hakala, Minna Nevala, Arja Nurmi, Minna Palander-Collin, Tanja Säily, and Anni Sairio.
NEW METHODS FOR
HISTORICAL SOCIOLINGUISTICS

CHANGES WITH NO VARIABLE

Nominal suffix -ity: no perfect synonyms; e.g. -ness also attaches to native bases.
• real → reality ≠ reality (Riddle 1985)
• kind → kindness, *kindity

Periphrastic do: should the variable include
• all finite verb forms: I have been thinking about this (Frank 1985)
• all finite verb forms that do not have another auxiliary: I am uncertain
• cases with a bare main verb that could be used with do: I think so (Ellegård 1953)?
If we choose the last option, how can we be sure which verbs to include? How do we count verbs for any of these in an untagged corpus?

Solution: abandon variable and simply count normalised frequencies (Nurmi 1996).
• How to establish statistical significance?

PRODUCTIVITY OF THE SUFFIX -ITY

Productivity can be measured in type frequency, or the number of different words containing the suffix in a corpus. Type frequency does not grow linearly with corpus size → cannot be normalised.

• How to compare figures from subcorpora? Säily & Suomela (2009): Divide corpus into cases with a bare main verb that could be used with do: I think so (Ellegård 1953)?
If we choose the last option, how can we be sure which verbs to include? How do we count verbs for any of these in an untagged corpus?

Solution: abandon variable and simply count normalised frequencies (Nurmi 1996).
• How to establish statistical significance?

As in the 17th century (Säily & Suomela 2009), the productivity of -ity continues to increase throughout the 18th century.
• The first subcorpus (1680–1719) uses significantly fewer -ity types than randomly composed subcorpora of the same size
• The last subcorpus (1760–1800) uses significantly more -ity types than its random counterparts (see Figure 1)

The change seems to be led by the professional class and by men writing to close friends (Säily forthcoming).

FROM PERIPHRASTIC DO TO DO-SUPPORT

During the eighteenth century the process of regulation for periphrastic do was in its final stages. As far as negation and inversion are concerned, the use of do was mostly in the final stage of the s-curve while do in affirmative statements declined steadily in frequency (Figure 2) and appeared more and more with a specific group of verbs (expressing emotions, mental processes and speech acts). This would seem to have been the pre-final stage of development towards polarity emphasis that appears in Present-day English.

1) He knows I do love him, & being certain of that he laughs at every objection that is started (Sarah Lennox, 1781)

Significant decline over time (Figure 3):
• The period 1680–1719 uses do significantly more frequently, while 1760–1800 uses do significantly less frequently, than randomly composed subcorpora of the same size.
• There are linguistic variables (such as the above-mentioned verbs) that go together with do in affirmative statements, but social embedding of the final stage of change seems to be entirely lacking (Nurmi forthcoming).

CONCLUSION

• Statistical significance for changes lacking a variable: Säily & Suomela’s (2009) non-parametric method of permutation testing

Figure 4 shows another example, the progressive (Sairio forthcoming).
LINGUISTIC CHANGE in its social contexts in eighteenth-century English (3/4):

Mikko Laitinen
Department of Languages
Linnaeus University
Minna Palander-Collin
Department of Modern Languages
University of Helsinki
Anni Sairio
Department of Modern Languages
University of Helsinki

INTRODUCTION

We explore three ongoing changes. The paradigm shifts in indefinite pronouns started in the 16th century, whereas its and the progressive were first introduced in the 17th century. What was the role of gender variation in these changes?

INDEFINITE PRON. IN -BODY / -ONE

The variable consists of four paradigmatic variants, two of which are discussed here: the compounds in -body and -one (Ex. 1 and 2):

1) According to promise I send you Mr. Frankland’s further account of what I proposed. You see how it is, and so can

2) and though he may have used some freedom that way formerly, yet I hear of late that his conduct has been such no one need be uneasy on that score. (William Steer, 1711)

The variable excludes partitive structures, and the results only include semantically generic references.

Previous studies by Nevalainen & Raumolin-Brunberg (2003: 124) show that the change towards the indefinites in -body during the 18th century. The change towards the indefinites in -body is linguistically motivated, replacing the old -man forms, and it is led by women. In the spread of -one, men lead in the incipient stages (p < .01), but the differences are leveled in the early/mid-18th century, and women gradually take the lead when the change nears the mid-range.

ITS VS. OF IT

Its has been available as the 3rd person neuter possessive pronoun from the beginning of the 17th century, and by the 1650s, its has gained the dominant position (Nevalainen & Raumolin-Brunberg 1994: 176). The use of its increased rapidly in the 18th century, reaching nearly 80% towards the end of the century (Fig. 3). Of it is the other main variant during the time period (Ex. 3), and it is still used today:

3) I promisd you an account of Sherborne, before I had seen it, or knew what I undertook. I imagind it to be one of those fine old Seats of which there are Numbers scattered over England. But this is so pecuarial and its Situation of so uncommon a kind, that it merits a more particular description. The House is in the form of an H. The body of it, which was built by Sir Walter Rawleigh, consists of four Stories, with four six-angled Towers at the ends. (Alexander Pope, 1717?)

THE PROGRESSIVE

The being construction has increased throughout the modern period (Kranich 2009), and by the end of the 18th century, it had developed in all tenses (Rissanen 1999). The progressive is associated with “typically more spontaneous, unmonitored, colloquial” language use (Kranich 2009: 102), illustrated in the aspetual usage in Ex. 4:

4) Mr Fox is hurrying me to death to get out of his chair in which I am sitting and writing. So adieu. (Sarah Lennox, 1762)

Fig. 5 shows the diachronic development of the progressive. The increase is particularly prominent in the present tense.

Conclusions: the change is not completed during the 18th century. The change towards the forms in -body is linguistically motivated, replacing the old -man forms, and it is led by women. In the spread of -one, men lead in the progressive more than men, and it increases particularly in the present tense. After mid-18th century, nuclear family correspondence becomes the most common register.
INTRODUCTION

We study the tail end of linguistic change in one verbal and two pronominal processes of change. All three were completed (i.e. passed the 85% mark in their respective S-curves) in the 18th century. How similar were these three in social terms?

VERBAL -S

The generalization of -s as the 3rd-person present indicative suffix is one of lexical diffusion. It started in a few high-frequency verbs. But once it progressed to low-frequency verbs, these soon overtook the more frequent verbs. But once it progressed to low-frequency present indicative suffix is one of the verbs. Chapter 1 shows the diachronic development of the two outgoing indefinites in the CEEC400.

Figure 1. The spread of -s in HAVE.

The last verb to complete the process was HAVE (Fig. 1). As with the other verbs, women were ahead of men. The upper ranks completed the change earlier than the lower (Table 1).

Table 1. The diffusion of has by social status (> 85%, cells highlighted in grey).

<table>
<thead>
<tr>
<th>SOCIAL RANK</th>
<th>1680–1719</th>
<th>1720–1799</th>
<th>1760–1800</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nobility</td>
<td>82%</td>
<td>75%</td>
<td>85%</td>
</tr>
<tr>
<td>Gentry</td>
<td>94%</td>
<td>79%</td>
<td>72%</td>
</tr>
<tr>
<td>Professionals</td>
<td>94%</td>
<td>79%</td>
<td>72%</td>
</tr>
<tr>
<td>Clergy</td>
<td>94%</td>
<td>79%</td>
<td>72%</td>
</tr>
<tr>
<td>Common people</td>
<td>72%</td>
<td>79%</td>
<td>85%</td>
</tr>
</tbody>
</table>

Twelve out of the 37 writers with 6 or more instances of the HAVE variable in 1680–99 used hath throughout, but only one out of the 52 writers in 1780–1800. One of the few who varied their usage was Sir Thomas Browne, writing to his daughter in 1682 (Ex. 1):

1) Our Tommy has had a grievous Cof and feavor... These 3 days hee hath been in better temper and prettie clearly...

The 2nd-person pronoun thou was already marked by the end of the 17th century. By the middle of the 18th century, the distinction between thou and you started to relate to register variation: the use of thou was connected with the language of poetry and religious prose and prayer.

All users of thou in the data mostly used it to their family and friends. Its 18th-century context, thou becomes a status-marker to women and younger family members and other relatives, as well as an intimacy-marker to close friends.

Figure 2. The use of thou in letters to family and friends.

One of the most prolific users of thou is Ignatius Sancho (1729?–1780). Deeply religious, Sancho shows certain linguistic features that could be considered archaic already in the 18th century (Ex. 2).

2) Poor blundering M[eux], I pity thee - thou art a banger in every thing - ask the girls else. - You know nothing of figures - you write a wretched hand - thou hast a nonsensical style - almost as disagreeable as thy heart - thy heart, though better than thy head - and which I wish from my soul (as it now is) was the worst heart in the three kingdoms - thy heart is a silly one - a poor cowardly heart - that would shrink at mere trifles - though there were no danger of fine or imprisonment: (Ignatius Sancho to John Meheux, 1779?)

INDEFINITES

The change in the indefinite pronouns with singular (generic) human reference consists of four paradigmatic variants. Two of them are on the increase in the 18th century, i.e. -body and -one, and the older variants, -man and the independent forms are lost (cf. also Poster 3). Figure 1 shows the diachronic development of the two outgoing indefinites in the CEEC400.

The decrease of the forms in -man is coupled with an increase in the indefinites in -body, and in the chronologically later process, the loss of the independent forms coincides with the increase in the one indefinites. Women are roughly one generation ahead of men in both of these processes.

The loss of the independent forms is nearing completion in the second half of the 17th century and is completed by nearly all social ranks by the early/mid-18th century (Table 2).

Table 2. The loss of independent indefinites by social status (cells highlighted in grey < 15%).

<table>
<thead>
<tr>
<th>SOCIAL RANK</th>
<th>1680–1719</th>
<th>1720–1799</th>
<th>1760–1800</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nobility</td>
<td>98%</td>
<td>95%</td>
<td>92%</td>
</tr>
<tr>
<td>Gentry</td>
<td>98%</td>
<td>95%</td>
<td>92%</td>
</tr>
<tr>
<td>Professionals</td>
<td>98%</td>
<td>95%</td>
<td>92%</td>
</tr>
<tr>
<td>Clergy</td>
<td>98%</td>
<td>95%</td>
<td>92%</td>
</tr>
<tr>
<td>Common people</td>
<td>98%</td>
<td>95%</td>
<td>92%</td>
</tr>
</tbody>
</table>

Conservative individuals (with total variant frequencies >10 and whose share of the outgoing variants is >30%) are male, representing various social groups. None of them are upwardly mobile. See Ex. 3:

3) You may depend on it, no more shall come into any mans hands but your own, (Francis Blomefield, 1736)
LINGUISTIC CHANGE IN ITS SOCIAL CONTEXTS IN EIGHTEENTH-CENTURY ENGLISH

Terttu Nevalainen, Minna Palander-Collin, Tanja Säily, Mikko Hakala, Samuli Kaislaniemi, Mikko Laitinen, Minna Nevala, Arja Nurmi, Anni Sairio
Poster presentation, ICAME 34, Santiago de Compostela, 22–26 May 2013

These posters present new results of ongoing sociolinguistic research by the C18 Research Group. The studies deal with the sociolinguistics of a dozen changes in eighteenth-century English based on data extracted from the Late Modern English part (1680–1800) of the socially stratified Corpus of Early English Correspondence (see CoRD for details). The changes explored include linguistic features whose sociolinguistic trajectories are well known up to 1700 (e.g. Nevalainen & Raumolin-Brunberg 2003), but the later developments of which have not been fully explored. Their spread in the language community coincides with different stages of the S-curve, including changes progressing in mid-range and those nearing completion, but also stable variation and features difficult to conceptualize as sociolinguistic variables.

On the basis of earlier research, the impact of different social variables on the choice of linguistic variants may depend on the stage of the change: the most variation is observed when the process in in mid-range whereas the range of social variation narrows down when the change is nearing completion (e.g. Kurki 2004, Tagliamonte & D’Arcy 2009: 99). The main aim of the posters is thus to explore how social variables such as gender, age and social rank may co-vary with the linguistic features at different stages. In addition, the analysis accounts for individual variation by looking for outliers, who are either conservative or progressive with respect to a particular process of change (cf. Nevalainen, Raumolin-Brunberg & Mannila 2011). When interpreting the results, changes in eighteenth-century society need to be taken into account, including changes in societal structure and social climate, ongoing language standardization processes and the emerging normative influence as well as a newly developed consciousness of letter writing as a social accomplishment. These factors affect the nature of the available data and pose further methodological challenges that are tackled in new ways.

The four posters map the sociolinguistic questions and challenges in the eighteenth century context and present the results thematically as follows: 1) presentation of the 2-million-word corpus, 2) new quantitative methods for historical sociolinguistics, 3) the sociolinguistics of ongoing changes in mid-range, and 4) the sociolinguistics of changes nearing completion. The new methods to be discussed provide statistical solutions for cases that are difficult to model in terms of the S-curve including the productivity of the suffixes -ness and -ity, the development of the auxiliary do in affirmative statements, and the rise of the progressive. The ongoing changes focus on the development of the noun subject and object of the gerund, the variation of its vs. of it, and the indefinite pronouns ending in -body vs. -one. The changes nearing completion are represented by the variation of has/does/says vs. hath/doth/saith, the pronoun thou, and the indefinite pronoun -man. These multiple changes highlight the variety and complexity of sociolinguistic factors in the interpretation of language change.

References (abstract & posters)


Nurmi, Arja. 1996. “Periphrastic do and be + ing: Interconnected developments?” In Terttu Nevalainen & Helena Raumolin-Brunberg (eds), Sociolinguistics and Language History: Studies Based on the Corpus of Early English Correspondence. Amsterdam: Rodopi. 151–165.


Nurmi, Arja. In progress. “From do-perphrasis to do-support: The social embedding of a change nearing completion”. In Nevalainen et al. (eds).


Säily, Tanja. In progress. “Productivity of the suffixes -ness and -ity”. In Nevalainen et al. (eds).


