A Corpus of Late Eighteenth-Century Prose

Linda van Bergen and David Denison
University of Edinburgh / University of Manchester
l.vanbergen@ed.ac.uk / david.denison@manchester.ac.uk

1. Introduction

A Corpus of Late Eighteenth-Century Prose is made up of unpublished letters transcribed from originals, about 300,000 words in all, and offered in plain-text and HTML versions. The letters, held by the John Rylands University Library, date from the period 1761-90. The project was made possible by funding from the John Rylands Research Institute, which allowed Linda van Bergen and Joana Soliva to work part-time on the transcription of the material.

There is a great deal of hitherto unedited material in the John Rylands University Library of Manchester. From this, we selected the Legh of Lyme Muniments, an archive associated with the Legh family at Lyme Hall in Cheshire.[1] We subsequently homed in on letters written to Richard Orford, a steward of Peter Legh the Younger, as being of suitable date, suitably extensive, of varied and practical content, and having at least *some* connecting thread. An advantage of the material is that it is of interest to historians as much as to linguists (about half of the requests for access have come from historians), so that any future extension of the project might enlist support from either scholarly community. To some extent the corpus is opportunistic in origin, and we are modest in our aspirations for it, other than that it should be a useful resource.

The corpus was mainly intended to help fill the gap between the major diachronic corpora of English (the Helsinki Corpus, the Helsinki Corpus of Older Scots, etc.), most of which stop in the early eighteenth century at the latest, and modern multi-genre corpora, which start in the 1960s.[2] It was also specifically designed to illustrate non-literary English and English relatively uninfluenced by prescriptivist ideas, in the belief that it might help with research into change in (ordinary, spoken) language in the Late Modern English period. It would also be of interest to scholars working on dialectal English of the north-west (north Cheshire and south Lancashire in particular). It has become clear to us since the corpus was compiled that there has been a recent upsurge of interest in the study of non-standard documents within languages that have a standardised form; see for example in this connection Fairman (2000),

Elspaß (2002; 2004) and Vandenbussche (2002), and the conference 'Language History from Below: Linguistic Variation in the Germanic Languages from 1700 to 2000' (Bristol, 2005).

The following short letter will serve to illustrate everyday subject-matter in the corpus, and grammar that would not be standard in Present-day English. First we show the original of the letter.

[insert Figure 1 about here]

We quote the letter from the plain-text version of the corpus, adding line numbers to facilitate a brief linguistic commentary; an explanation of coding symbols is given in the Appendix.

<A GRIMSHAW JAMES>
<O ?1777>

M=r=. Orford
Lyme
by Manchester
Cheshire

<P>

         Haydock Octo=r=. 27=th=.
Sir

      Iohn Hall and Tho=s=. Harrison
came to Haydock on saturday ^night^ about six
oClock and brought with them the two
Mares and sixty sheep, and they themselves
5   two swine for they was both drunk,
the parcel that was sent for your Daughter
they have lost it but where they kn{ow}
not, but they say they lost it at S{*...} [^seal^]
or before they came [^one or two letters rubbed out^] there, I hope you {*...} [^seal^]
10   make them to pay for it, it will be a
warning to them for the future and to others
that is sent [^word crossed out^] on business if they had
any other Message they know no more than
the horse they rid on I have got no Money
15   of Unsworth, I hope your Ears are so
that you have no Occasion for Larding
        J am Your hble Servent
        Ja=s=. Grimshaw

The spelling in this letter happens to be standard, the punctuation less so. Some other forms are odd by present-day standards though not necessarily non-standard at the time of writing.

Common nouns sporadically have initial capitals, something not infrequent in eighteenth-century spelling, and <I> and <J> can still be treated as variants of the same letter, hence *Iohn* (line 1) and *J* (17). Abbreviation was common to save time and postage costs, hence *Tho*[s] for *Thomas* (1) and *Ja*[s] for *James* (18), *hble* for *humble* in the closing formula (17). Past tense *rid* for *rode* is historically correct for the past plural and is actually recorded by *OED* as a variant for the past tense that was still current around 1909, when the fascicle containing *ride* v. was published (s.v., A.2(?)). *Have got* (14) is not yet a mere synonym for *have* 'possess' and still means 'have obtained'. Causative *make* (10) is nowadays usually followed by a plain infinitive except when passive, but here is construed with a *to*-infinitive even though itself active, as was still common at the time. *They was* (5) has remained widespread in dialect, and notice *is* with a plural subject in line 12. The syntax of *the parcel … they have lost it* (6-7) is clumsy: the object noun phrase is topicalised (Denison, 1998, pp. 237-8) and then repeated in a resumptive pronoun that is strictly speaking unnecessary. (*Larding* (16) is probably a reference to the application of grease as ointment: in the next letter Grimshaw writes 'I am sorry to here you are no better of your head'.)

To return to the general content of the corpus, there are no immediate plans to extend it. However, since we were not able to transcribe all the letters contained in this part of the archive within the time available, ideally we would like to include the remaining material as well, if and when resources can be found. We have transcribed about four fifths of the letters sent to Richard Orford. Other sections of correspondence contained in the archive – especially those written to members of the Legh family, containing material on eighteenth-century northern politics – are used comparatively frequently by historians, so expansion in that direction would certainly be useful for them, as well as of probable interest to linguists. Even as it stands, however, the corpus provides easy access to a substantial body of material that otherwise in all likelihood would rarely be used.

As already indicated, we always intended the corpus to be of use to others besides linguists. A number of requests for access have come from historians, mostly without further specification of their specialism, though one did mention a particular interest in mining and smelting.

2. Material included in the corpus

Given the nature of the material, the corpus is unbalanced and heterogeneous. We could have tried to make the corpus more balanced and/or homogeneous in various ways, and indeed our original intention had been to be more selective with respect to the letters included in the corpus than we have ended up being. However, in addition to the practical difficulties involved in selection, we felt that, on balance, this would have had negative effects on the value of the corpus in view of the range of purposes we had in mind for it.

Selection of material (or more specifically, *exclusion* of letters) on the grounds of linguistic value soon turned out to be far from straightforward. The temptation would be to include letters more readily where there was any evidence of non-standard usage or anything else that caught our attention, which would have given a misleading impression of the language by making it look less standard and/or more unlike present-day usage than it actually was. From a more practical perspective, deciphering the letters could be at least as time-consuming as the actual transcription, especially where there was only a small amount of material available for a particular author so that there was little opportunity to become familiar with the handwriting. In other words, selection would have taken up a significant amount of time (without visible results where letters were excluded). Moreover, given that we hoped that the corpus would prove useful to non-linguists, using linguistically-based selection criteria might detract from its value for them. So in practice, the policy adopted was to include letters unless they were both very brief and highly formulaic and thus unlikely to be of use to anyone.

Our own particular goals would have favoured the inclusion of personal letters, since they are the most likely to exhibit informal language. However, there were few letters which could be described as being of a (purely or even predominantly) personal nature. Most were of a 'business' nature, yet many of these seemed quite informal in tone, and they often did not conform completely to standard language norms. (Generally speaking, we found non-standard features to be more frequent and widespread than we had expected them to be in the late eighteenth century.) Moreover, the dividing line between business and personal letters turned out to be very fuzzy; in practice, it was easily and frequently crossed. Since the material seemed valuable for our purposes regardless of the precise nature of the letter, we made no attempt to select (or even distinguish) on this basis. This again had the added

benefit that it was more likely to cater for the needs of any historians who might be interested in the corpus.

Here is a short letter illustrating the easy transition from business to personal matters and from formal to informal style, and presented in its HTML version (apart from colour coding; see the Appendix for coding conventions):

author DICKENSON JOHN

1780

M^r. Orford

Lyme

    new page

     Taxal Thurs: 2^d Nov^r.

     [different hand:] 1780

D^r %Sir!

   An old Woman dyed Yesterday, who was

a Life in a Lease, and by whose Death an Estate descends

to me. Her Brother, (who is to be sure a very <u>honest</u> man)

took her to live with him a few Years ago, & then sold what

little property she had, wch. he did to prevent my having

an Harriot, whenever she dyed, as he knew J was entitled

to the best Good. Pray am J to be fobb'd off with

an old dirty Red Petty Coat, or am J to have Recourse

to the Tenant who occupys the Premises.? Her Brother

likewise intends to have the whole of this Years Rent, &

only leave me the winter Pasture.

   J shall be much obliged to you for yor. friendly advice,

how J am to act in this Affair, — . J am, Sir

    Your obed^t hble %Servt.

    John Dickenson

My Son waites upon You ^with this^ pray

don't give him too' much Lyme Beer

& do you see that he does not bring

Johnson along with him.

(A *heriot* or *harriot* is a feudal payment due to a lord on the death of a tenant.) Again, for interest we show the original.

[Insert Figure 2 about here]

The corpus makes no attempt to achieve a balance between the contributions of different authors. We could have simply excluded authors with little material to their names, but such authors were more likely to fall into the category of relatively inexperienced writers exhibiting a higher degree of non-standard usage. Conversely, we also considered imposing an *upper* limit on the amount of material from any one author where a large set of letters was concerned, to avoid over-representation of the language of particular individuals. However, we felt that such cases were too valuable, and could usefully act as sub-sections of the corpus. They allow researchers interested in linguistic aspects to concentrate on individual usage. In addition, these letters would tend to be linked in terms of content, so including all of them gives historians access to as much material as is available in this part of the archive.

We also decided not to impose any restrictions on the provenance of the letters. Aside from the issue of whether doing so might have led to a misrepresentation of the range of Orford's dealings, there were practical problems involved. Firstly, not all letters give an indication of where they were written. Secondly, and more importantly, the place where letters were written and/or where the author lived is not necessarily the same as that where they grew up. There are a number of cases where it is clear that an author who is not based in the north-west of England nevertheless has family living there and/or has previously lived in that area. Conversely, it is entirely possible that some of the people writing from somewhere in the north-west were not actually from the area originally. Exclusion of letters written outside the area would have meant leaving out some valuable material, and it would not have dealt with the real issue. A coherent selection policy in this respect would have involved background research on individuals, which was impractical.

The result of this is that nearly all material present in this subsection of the archive has been included in the corpus (always excepting the final part, which we were unable to transcribe because of restrictions on funding and time available). Of course, we may have erred on the side of inclusiveness as a result, but material could always be excluded from consideration

(or even from the corpus) at a later stage if appropriate. Given that we have not 'meddled' significantly with the content, it should at least provide a fair view of the type of matters dealt with by Orford as well as the language of his correspondents.

Our desire to create a corpus that would be of value to a wide set of researchers rather than one that is tailor-made for one particular type of research project means that it will be necessary to bear the varied nature of the corpus in mind, and the corpus will need to be used with a certain degree of care. (Compare here the NECTE (Allen et al., this volume) and SCOTS (Corbett et al., 2004–) projects, also compiled with a wide range of users in mind.) Depending on the aims of the scholar using the corpus, it may be necessary to select part(s). The risk, of course, is that in trying to please a wider range of researchers, we may have ended up pleasing no one, thus detracting from the usefulness of the corpus rather than increasing it. It will be interesting to see how successful the attempt to create a resource of use to both linguists and historians will turn out to be. The number of requests for the corpus from both linguists and historians so far is certainly encouraging, as is the fact that the distribution of requests between linguists and historians is roughly equal, although the real test will of course be how useful they will find the corpus once they have obtained access to it.

3. Transcription

We decided to adopt something which can be described as being very close to an exact reproduction of the original documents — a so-called diplomatic edition of the text. For obvious reasons, as little as possible was modified (although we have sometimes opted for the nearest present-day equivalent of punctuation, for example, for practical reasons). Lineation has been preserved, mainly because it was very easy to do and would make it easy to find specific parts if anyone wishes to go back to the original text, and in the same spirit page breaks were marked. However, words hyphenated across line-breaks were silently recombined, and we made no attempt to reproduce any other aspect of layout faithfully, except that the web version goes some way towards representing the disposition of tabular material.

Since we were dealing with unedited material, the transcription process was far from straightforward. The documents could be difficult to read in places, some characters could be difficult to distinguish from one another in the handwriting of particular authors,

capitalisation (or lack of it) could be difficult to determine, and so on. We have dealt with such matters as best we could, including explicit marking of readings that we felt to be tentative where the identity of the letter-form was in doubt. One particular and frequent area of doubt was not so marked, however: where the only uncertainty was between upper and lower case forms of a letter. We decided that marking such readings as conjectural was potentially misleading and would have cluttered the text to little purpose; instead we have simply tried to be as consistent in our transcription practice as we could be. Proof-reading after transcription was essential, of course. Inevitably, there will be some errors, but we have done our best to keep these to a minimum. Some inconsistency is likewise impossible to avoid altogether (especially between the two transcribers involved, but also for individual transcribers), although naturally we have tried to limit it as much as we could.

We have included in the corpus nearly all the information contained in the letters. The only things normally omitted were calculations and other scribbles that seemed unrelated to the content of the letter, and anything (e.g. drawn plans) that could not be transcribed using text, although even then we generally note such omissions. Where possible, we marked in-text corrections and deletions and noted illegible material. The reasoning was that we could not be certain of what information individual researchers were going to need, and that it was easier and less time-consuming to include it now than have to go back to the originals at a later stage.

The lack of normalisation of spelling and the inclusion of detailed information, especially the use of in-text comments, means that users of the corpus have a version of the text that allows them to reconstruct most properties of the original letters. While this has obvious advantages, the disadvantage is that it hampers text searches, since words may be spelled in various ways and phrases may be broken up by in-text comments. The latter problem could be solved by the user of the corpus if required: it should be relatively easy to produce a version of the corpus without comments, although it would need to be used side by side with the original version of the corpus to make sure that any comments set aside are in fact irrelevant to the data found. There is, however, no easy solution to the other problem: lemmatisation or producing a parallel normalised version are the obvious possibilities, but either would be time-consuming. We have gone a small way towards facilitating normalisation by transcribing what we call 'deviant word division', such as *a fore* for *afore*, as <a_fore>, and 'deviant word joining', such as in *Iam* for *I am*, as <I %am>.

4. Coding

The coding of the corpus (especially for the plain text version) is largely based on that used in the Helsinki Corpus, mainly for practical reasons. Almost anyone using historical corpora of English will be familiar with that particular corpus and its coding conventions (Kytö, 1993). It makes an ascii version of the corpus possible (still required for certain types of text retrieval software, and compatible with virtually all types), while allowing additional information to be included. In a few cases we have adapted the codes or added new ones, such as for example a code for tentative readings. We also allowed the use of extended ascii (the upper 128 symbols) where this was of use, for example for <½ ¼ £>, which the Helsinki compilers, working with DOS in the 1980s and 1990s, avoided.[3]

The extent of text-level coding was radically reduced from what the Helsinki Corpus contains. While it would be desirable to add more information at this level, it would have taken up a significant portion of the time available to do the necessary background research on the individual authors. In some cases information was obvious from the letters themselves, but where it was not, it would often have been difficult if not impossible to find out.

We decided to provide the corpus initially in two versions. One was as a single text file (c. 1.6 Mb), suitable for use in concordancing programs. The other is designed for ease of reading in a web browser and uses the standard rendering of HTML to convey superscript, interlineations, underlining, strike-through and so on, with a larger font size to signal the author at the start of a letter. It also employs different colours as a quick visual clue to editorial decisions, following the general convention that red signifies conjectural, crossed-out or illegible text, while blue indicates editorial material. For some of the tabular material, HTML is an improvement over plain text. Here is part of a letter from J. Hancock in 1788 which throws interesting light on the expenses attendant on the serving of a writ by a 'Sherriffs offecer':

To A %Pair of Superfind 4 {*}d Hose / forgot the price given but think	3/6
To Attending Wm. Frith and Serveing with a writen notice to Quit his Livelihood	6/8
To Iames Miller and PPott to Bear witness that the Said notice was Properly Served Tow Glasssd Gim	0.0.3
To Wm. Frith to heal his wounds Ginn	0 ½
& to writeing this Acctd	0 0
	£:10:5½

Our attempt to preserve the tabular form may be compared with the original, displayed below.

[Insert Figure 3 about here]

Another advantage of the HTML version is line length, since various in-line comments are obviated. (Remember that we are committed to preserving original manuscript lineation.) Compare what appears in the text version, all on one line, with its HTML equivalent. Here is the opening of a letter by Shaw Allanson in 1789, first in the original.

[Insert Figure 4 about here]

For our edited illustration, hard line-breaks in both files are represented by <¶>, and colours are removed from the HTML version:

Text version:
Yours I Rec'd yesterday [^"{day}" added above the line, and subsequently crossed out^] and as such have¶
been with M=r= Rigby this morn=g= and has¶
given him 34=£=..3=s=..6=d=

HTML version:
Yours I Rec'd yesterday ^~~day~~^ and as such have¶

been with Mr Rigby this morn[g] and has¶

given him $34^{\pounds}..3^{s}..6^{d}$

To avoid excessively long downloads, the HTML version comes as a sequence of three files, (909, 883 and 804 Kb), plus a synopsis of the coding conventions (6 Kb) which is given here as an Appendix. The four files are notionally numbered as 1, 2, 3 and 0 (coding conventions), with links at head and foot of each web page to allow a reader to move easily back and forth through the set. Conversion was an extremely time-consuming business. Some of the decisions may not have been optimal. For example, long dash is coded as the string <—> rather than simply as <—> or <–>, so that the HTML version would look unnecessarily opaque if loaded direct into a concordancer. On the other hand, the frequent changes of font colour, often within a word, make the HTML version unsuitable for searching anyway, unless the search engine can ignore HTML tags. (The simple search functions within commonly-used browsers do indeed ignore attributes like font colour and superscript.)

Lou Burnard of the Oxford Text Archive has raised with us the possibility of a conversion to XML ('extensible mark-up language'), which would allow a more structured and standardised coding of textual features. Indeed, at the Workshop on which these volumes are based there was some discussion of the merits of TEI-conformant coding ('Text Encoding Initiative'), of which XML is an exemplar. We were not sufficiently convinced of the utility or practicality of XML to attempt it ourselves for the original release of the corpus, since XML-aware readers and editors are not yet sufficiently widely available: without such a reader, XML tags would intrude on the displayed text and interfere with searches. However, a conversion may follow eventually (see also note 3 on Unicode).

What is lacking altogether is grammatical coding, whether word-class tagging or sentence parsing. As linguists we would certainly have welcomed a corpus equipped with such tools, however problematic a particular analysis might have been, but our resources simply did not run to this.

5. Conclusion

We have not yet had much information on the purposes to which others are putting the corpus, apart from the pleasing discovery in it of the earliest known modern uses of the

elliptical adverb or interjection *please* without following *to*-infinitive (Tieken-Boon van Ostade and Faya, 2004):

> You'l please return the apointment that I may destroy it (I. Hodson, 23 June 1775)
>
> (Please see over) (James Hammond, 27 June 1778)
>
> Please Remembar mee to my Ant and All my Cosins and My Uncall Iohn (John Mercer, 12 April 1789)

Not only do these examples antedate *OED* by over a century (s.v. *please* v. 6c), they may support a different syntactic origin from that proposed by *OED*. Since we ourselves have not yet used the corpus for linguistic investigation, our review of its utility and its limitations is necessarily provisional. A recurrent theme in this paper, as in the team's discussions while we were working on the corpus, was the difficult choice between size of corpus and richness of annotation. Some of the desiderata of annotation or coding were not really consistent with each other, while others would have required a flow-chart of stages, each a prerequisite for the next. For example, a version of the text with normalised spelling would certainly have been a helpful – though not perhaps an essential – preliminary to tagging, and certainly to production of parallel aligned texts. On balance, though, it seemed best to concentrate our efforts on the original text.

References

Allen, W., J. Beal, K. Corrigan, H. Moisl and C. Rowe. (this volume). A linguistic time-capsule: the Newcastle Electronic Corpus of Tyneside English.

Corbett, J., W. Anderson, D. Beavan, F. Edmonds, L. Edmonds, C. O'Hare, J. Anderson, C. Kay and J. Stuart-Smith. (2004-). Scottish Corpus of Texts and Speech. http://www.scottishcorpus.ac.uk/

De Smet, H. (forthcoming, 2005). A corpus of late Modern English texts. *ICAME*.

Denison, D. (1998). Syntax. In Romaine, S. (ed.) *The Cambridge history of the English language*, vol. 4, *1776-1997*, pp. 92-329. Cambridge: Cambridge University Press.

Elspaß, S. (2002). Standard German in the 19th century? (Counter-)evidence from the private correspondence of 'ordinary people'. In Linn, A. R. and McLelland, N. (eds.). *Standardization: studies from the Germanic languages*, pp. 43-65. (= Current Issues in Linguistic Theory, 235.) Amsterdam and Philadelphia: John Benjamins.

Elspaß, S. (2004). A view 'from below': historical sociolinguistics and the study of present-day language change. Paper due to be presented at Sociolinguistics Symposium 15, Newcastle upon Tyne.

Fairman, T. (2000). English pauper letters 1800-34, and the English language. In Barton, D. and Hall, N. (eds.). *Letter writing as a social practice*, pp. 63-82. Amsterdam and Philadelphia: John Benjamins.

Fitzmaurice, S. (this volume). Questions of standardization and representativeness in the development of social networks based corpora: the story of the Network of Eighteenth-Century English Texts.

Kytö, M. (1993). *Manual to the diachronic part of the Helsinki Corpus of English Texts: coding conventions and lists of source texts*, 2nd edn. Helsinki: Helsinki University Press for Department of English, University of Helsinki.

Kytö, M. and M. Rissanen. (1999). English historical corpora: report on developments in 1998. *ICAME*, **23**, 175-188.

OED. OED Online. http://dictionary.oed.com/

Tieken-Boon van Ostade, I. and F. Faya. (2004). Saying "Please!" in eighteenth and nineteenth-century English. Paper presented at Second International Conference on the English Language in the Late Modern Period 1700–1900, Vigo, November 2004.

Vandenbussche, W. (2002). Dutch orthography in lower, middle and upper class documents in 19th-century Flanders. In Linn, A. R. and McLelland, N. (eds.). *Standardization: studies from the Germanic languages*, pp. 27-42. (= Current Issues in Linguistic Theory, 235.) Amsterdam and Philadelphia: John Benjamins.

Appendix

The following material forms the 'zeroth' of the four pages which make up the web version of the corpus. The files are not publicly available on the WWW, as we ask interested scholars to submit an access request form before providing the information needed to access or download them. However, the file of coding conventions reproduced here in monochrome will gladly be emailed without formality to any interested scholar.

The English language of the north-west in the late Modern English period:

A Corpus of late 18c Prose, coding conventions

For conditions of use see readme.txt. If unavailable, please get and return access request form.

0 | 1 | 2 | 3

The corpus is available as a single text file, **orford.txt** [1.6 Mb] or as three linked HTML files, **orford1.htm**, **orford2.htm** and **orford3.htm** [approx. 800-910 Kb each]. (An XML version may be on the way.) The text version is coded in a similar way to the Helsinki Corpus, whereas the HTML version aims for greater readability, following the general convention that red signifies conjectural, crossed-out or illegible text , while blue indicates editorial material.[4]

Coding conventions		
meaning	**text version**	**HTML version**
writer's name	<A XXXX>	**author XXXX**
date (year)	<O nnnn>	**nnnn**
new page	<P>	*new page*
new page with indication to turn from foot of previous page	<P Turn over/turn/Please to turn>	*new page* Turn over
new page with word repeated at foot of previous page	<P xxxx>	*new page* xxxx
underline	(_xxxx_)	xxxx

superscript	yy=xxxx=yy	yy^{xxxx}yy
subscript	yy=xxxx=yy (+ note indicating **sub**script)	yy$_{xxxx}$yy
interlineation	yy^xxxx^yy	yy^^{xxxx}^yy
deviant word division e.g. "a fore" for "afore"	a_fore	a_fore
deviant word joining e.g. "Iam" for "I am"	I %am	I %am
abbreviation indicated by author	~	~
conjectural reading	{xxxx}	xxxx
crossing out or rubbing out	[^ "XXXX" crossed/rubbed out^]	~~XXXX~~
crossing out or rubbing out with some part uncertain	[^ "YYXXXXYY" crossed/rubbed out?^] or [^ "YY{XXXX}YY" crossed/rubbed out^]	~~YY{XXXX}YY~~ or ~~YYXXXXYY~~
illegible (number of asterisks indicates approx number of letters)	{**}	{**}
illegible (see below)	{*...}	{*...}
our comment	[^xxxx^]	[xxxx]

For illegible text the number of asterisks estimates number of letters where possible; otherwise {*...} is used. If there is a specific cause for the illegibility, this is specified in a comment.

The HTML version does not distinguish crossing out from rubbing out, and for readability it ignores any uncertainty in the words crossed or rubbed out. In such cases the text version preserves more detailed information. The comment [corrected] indicates a correction made in the original letter by the author. It is often difficult to determine whether a letter-form is upper or lower case though the letter itself is not in doubt: we have not marked such readings as tentative and have merely tried to be reasonably consistent. You will therefore need to consult the original documents if capitalisation is of particular importance to you.

Hyphenation at the end of a line has not been preserved and the whole of the word has been put at the end of the first line, unless there is a reasonable chance that the hyphen belongs to the word form. Otherwise, lineation has been preserved, except in some cases where the text runs parallel. The remainder of the layout has largely been ignored.

Accounts or calculations — the terms are used interchangeably — are sometimes omitted, but if so, this is noted in the text. London postmarks are not generally noted.

0 | *1* | *2* | *3*
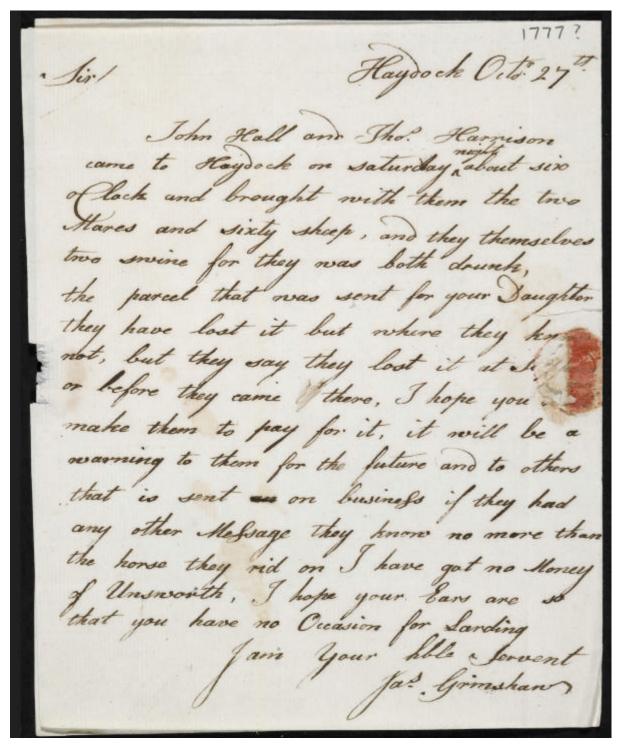
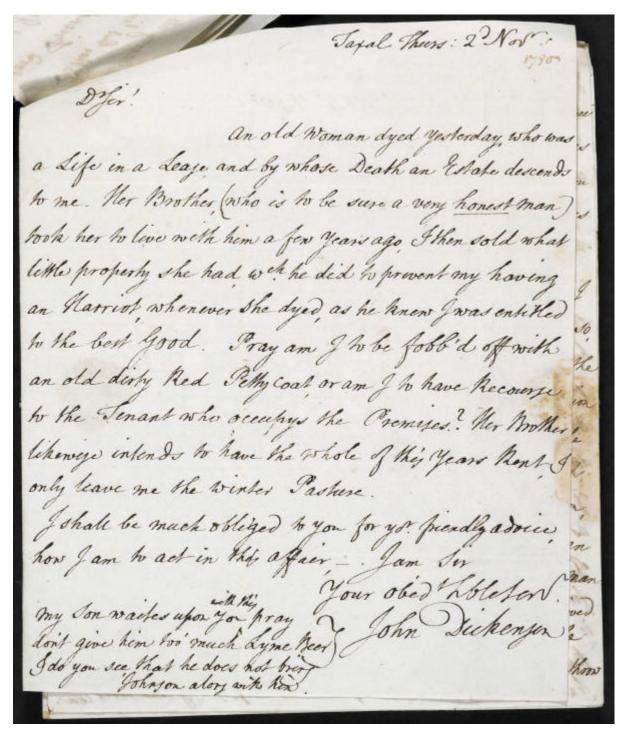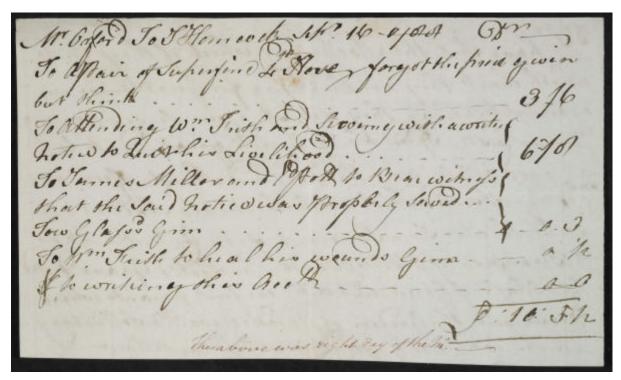This document last updated 29 April 2003.

Figure 1

Reproduced by courtesy of the University Librarian and Director, The John Rylands University Library, The University of Manchester.

Figure 2

Reproduced by courtesy of the University Librarian and Director, The John Rylands University Library, The University of Manchester.

Figure 3

Reproduced by courtesy of the University Librarian and Director, The John Rylands University Library, The University of Manchester.

Figure 4

Reproduced by courtesy of the University Librarian and Director, The John Rylands University Library, The University of Manchester.

[1] The John Rylands University Library gives the following information on the Legh of Lyme Muniments on its Special Collections website (http://rylibweb.man.ac.uk/data2/spcoll/legh/):
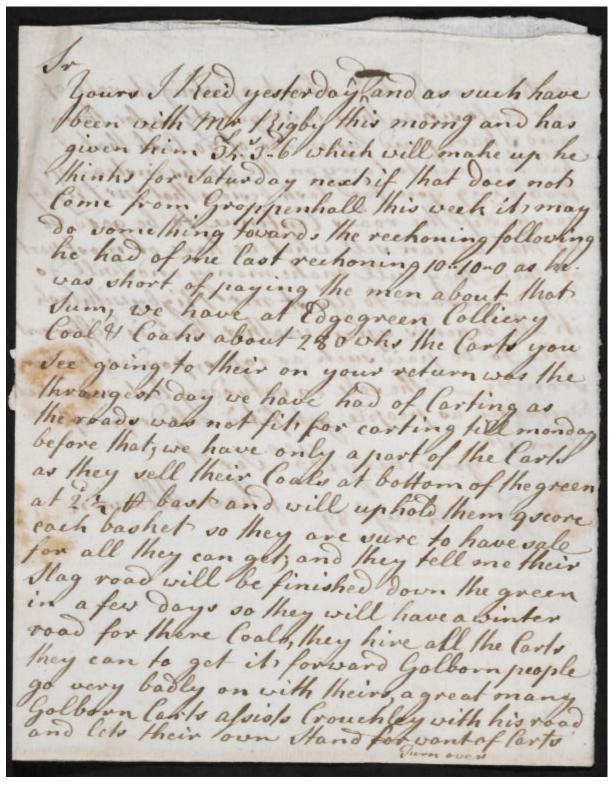
> Extensive family papers of the Leghs of Lyme Park, Cheshire. These comprise muniments of title, including large numbers of medieval deeds and charters, 17th- and 18th-century manorial court records, original architect's plans of Lyme Hall, surveys, wills, abstracts of title, estate correspondence, accounts and other papers. The muniments relate to the Lancashire estates (the manors of Newton and Golborne, and property in Newton-le-Willows, Golborne, Lowton, Haydock, Ashton-in-Makerfield, Ince-in-Makerfield, Warrington, Burtonwood, Poulton and Fearnhead, Bold, Pemberton and Dalton), and the Cheshire estates (with property in Lyme Handley, Disley, Pott Shrigley, Macclesfield, Grappenhall, Norbury, Marple, and Broomedge and Heatley in Lymm).

> There is an extensive and important selection of personal correspondence, dating from the 16th century onwards, including correspondence with members of the Gerard, Egerton and Chicheley families, with much material on 18th-century Northern politics.

Lyme Hall is a fine, originally Tudor house, home to the Legh family for 600 years, now run by the National Trust (http://www.nationaltrust.org.uk/).

[2] Another filler of this gap is CONCE, The Corpus of Nineteenth-Century English (Kytö and Rissanen, 1999, p. 181). Since we conceived our corpus, other corpora of Late Modern English have been announced, including *The Corpus of Late Modern English Texts* (De Smet, forthcoming, 2005) and Fitzmaurice (this volume).

[3] At the workshop at Sociolinguistics Symposium 15 on which the present volumes are based, Wolfgang Schmidle suggested the use of Unicode character encoding, although Ylva Berglund of the Oxford Text Archive felt that the time was not yet ripe. In principle, the universal, device-independent nature of Unicode and the wealth of characters available would both be of obvious advantage. On the other hand, plain text is more compact and — for the time being — more widely usable, and we chose the more conservative option in producing and releasing the corpus. The decision is not irrevocable, however.

[4] Colours have of course been lost in this print-out.