ANNOTATING EARLY MODERN ENGLISH TEXTS IN THE SCONE CORPUS¹

Gabriel Amores Carredano
Julia Fernández Cuesta
Luisa García García
Universidad de Sevilla

1. Introduction

The Seville e-corpus of northern English (SCONE) is a part, and to some extent we could say also a corollary of our research for a previous project on the history of northern English [BFF2001-3135]. Since we started working on the evolution of written northern English in 2001 we have collected and analysed a considerable number of texts from the 7th to the 16th centuries. Our original idea was to develop a web interface in order to make this textual material available to the academic community, especially the manuscripts that have never been edited, or not in a way that makes them suitable for serious linguistic analysis. The result of this new project will be an e-corpus which will include both the edition of the manuscripts and information about the language at different linguistic levels, including spelling/phonology, morphosyntax and lexis.²

Since the texts included in SCONE have been the basis of our study of written northern English and have, therefore, been analysed in detail, we later thought that, besides an edition of the texts, it would be useful to tag them partially, highlighting the features that characterise them as northern. We have, therefore, developed some criteria for a ‘selective tagging’, which are explained below (section 4). Our tagging is obviously biased, since what it does is to show the results of our study on Northern English (most of them already in press)³, directly on the texts. But it can be useful in the sense that it will give other scholars the opportunity to see, as it were, the ‘rough material’ of our research and the way in which we have proceeded in our analysis. Besides, the interface will also allow users to perform searches which are not focused on dialectal variation, e.g. studies on the evolution of a given morphological feature, or on the syntax of different text types, etc.

In a previous paper presented at the 6th International Conference on Middle English (July 2008) we described the texts which are to be included in our e-corpus, explained the purpose of SCONE, and the methodology employed: general editorial criteria and the reasons for the ‘selective tagging’ already mentioned: dialect features to be highlighted as ‘distinct’ at the different linguistic levels. The present article takes a step further and incorporates, at the theoretical and methodological levels, the comments and general feedback received at the ICOME conference by various colleagues, and also the results of own discussions on the development of SCONE to date. Some of the ideas expressed here as regards the purpose and

¹ We are grateful to Marcelle Cole and Christopher Langmuir for comments and suggestions on a draft version of this paper.
² This new project [HUM2007-62926/FILO] is funded by the Spanish Ministry of Science and Technology, the National Programme for Scientific Research Development and Technological Innovation and the European Regional Development Fund.
methodology of the e-corpus were already outlined in Cambridge, so there will be some inevitable repetitions. However, we have tried to explain the main developments since last summer and in what ways we have taken the project further.

In this article we focus on Early Modern English, so we have chosen a text from this period to illustrate the criteria followed in editing the texts, the features that are to be tagged and the procedure employed in developing the web interface. As will be explained in section 5., for Early Modern English we have mainly analysed legal documents (wills and testaments and civic records) with the aim to see what northern features are more resilient to the process of standardisation in this type of texts.

2. Corpus description

SCONE consists of northern written material (manuscripts and inscriptions) ranging from the 7th to the 16th centuries. From Early Northumbrian we have included all the extant material: runic and non-runic inscriptions and the northern manuscripts of the three Northumbrian poems Caedmon’s Hymn, Bede’s Death Song and The Leiden Riddle. There are no other extant northern texts until the second half of the 10th century. From this period we have included all the late runic and non-runic inscriptions and excerpts of the glosses to the Lindisfarne Gospels and of the Northumbrian section of the Rushworth Gospels (Ru2), written in a variety known as ‘late Northumbrian’. For this period the e-texts are based on rigorous diplomatic editions by Okasha, Page, Smith Skeat and Smith.

Taking LAEME as our source for Early Middle English we will include all the shorter texts and fragments from longer ones. Those we have analysed so far are listed in the appendix. Since LAEME is now available online we will include at least excerpts from all northern Early Middle English texts for the sake of completeness. For late Middle English and Early Modern English we have used microfilmed copies of original manuscripts (see appendix for a list of the manuscripts). A number of these texts have been edited, but some of the editions are not diplomatic (Testamenta Eboracensia, York Civic Records) and, although useful to some extent, they can never be taken as the basis of serious linguistic study.

In our corpus we have tried to include texts representing different genres (literary, religious, scientific, legal, letters, etc.) with the aim of assessing to what extent the distribution and frequency of dialectal features is influenced by text-type. However, this is not always possible; in the case of eModE we have only analysed legal texts so far (testaments and civic records) and just one private document (Machyn’s diary).

3. Edition of texts

As previously mentioned, for the edition of texts we have used either reliable diplomatic or semi-diplomatic editions, when available, and/or photographs and microfilmed copies of the original manuscripts.

Our transcriptions aim at reproducing the text at a semi-diplomatic level. As regards specific editorial criteria, the features that are to be included are the following:

---
4 See appendix.
5 By kind permission of Margaret Laing, who let us use her unpublished material in 2005. In 2008 the LAEME copyright holders granted us permission to use their recently published material for the purposes of our corpus.
1. Graphemes are transcribed but not allographs, with the exception of <y> for thorn.
2. Special symbols are used for thorn, ash, eth, wynn, yogh and insular ‘g’.
3. Conventional expansions for most of the symbols of abbreviations in the texts are used for the sake of clarity. Abbreviations will be kept in the following cases:
   3.1. For final -es / -is (plural of nouns, present indicative inflexion) because the vowel of the inflexion may convey historical and dialectal information.
   3.2. The abbreviations for ‘and’ will not be expanded, since they can stand for both ‘and’ and ‘ond’, which is relevant for the study of dialect.
   3.3. For the abbreviation of yt we follow LAEME practice and leave it unexpanded: it stands for both ‘that’ and ‘thet’.

We believe that the retention of these abbreviations does not compromise the legibility of the texts, which is one of our purposes.

No editorial punctuation is added. In the transcription of the manuscripts, the conventional symbols for punctuation signs are used: dot, virgula, punctus elevatus, paragraph markers, etc.

Finally, a literal translation will be provided for Old English and a succinct glossary for Middle English and eModE.

4. Tagging

All texts will be partially tagged at word level. We have marked those words which provide a context for any of the features selected for dialectal discrimination, regardless of whether or not the variant found is the typically northern one. The reason for adopting this strategy is that it allows the statistical analysis of the relative frequency of every feature in each text.

This type of analysis enables us to assess what factors may have conditioned the distribution of the features. As is well known, the distribution of variants is not only determined by place and text-type, but is often lexically conditioned. As regards variation according to text-type, our research for eModE (Fernández Cuesta / Rodríguez Ledesma 2004: 9) has shown, for example, that the inflexions -eth for 3sg. present indicative and -ing for the present participle are more frequent in legal than in literary texts, most probably due to the earlier standardisation of the former. The distribution of plosives vs. palatal consonants in words such as CHURCH, EACH and WHICH is an example of a lexically conditioned variant.

There is also a time dimension in our corpus. Since we have analysed texts over a time-span of nine centuries, it is possible to assess the resilience and recession of northern features from Old and Middle English in later periods. This time dimension also enables us to show innovations in these dialects at different periods, such as the Northern Fronting and the Northern Subject Rule in Middle English.

4.1. Diagnostic features

Diagnostic features have been organised according to the linguistic level they belong to (spelling and phonology, morphosyntax and lexis). So far the work that we have carried on has been mainly on the first two levels.

With regard to phonology, we have considered the features in table 1 for Old English and those in table 2 for Middle English and eModE.
### Table 1. Spelling / Phonology

#### 1.1. Old English

1. Reflex of Germanic */æ:*
2. Reflex of Germanic */a/ when followed by a nasal consonant
3. OE Breaking (+ / -)\(^6\)
4. Retraction (+ / -)
5. Palatal diphthongisation (+ / -)
6. I-mutation (+ / -)
7. Back mutation (+ / -)
8. Smoothing (+ / -)
9. Reflex of Germanic */eu/\(^6\)
10. Reflex of Germanic */au/\(^6\)
11. Rounding of [e(:)] to [œ(:)] when preceded by [w] (+ / -)
12. Loss of -n in final position (+ / -)

---

\(^6\) Plus and minus symbols indicate the presence or absence of the feature.
Since one of the characteristics of Old Northumbrian is the variation in the representation of the vowels of the inflexions (especially in the period known as late Northumbrian), we have decided to tag all surface morphology, so that the degree of variability found in these dialects can be assessed and contrasted with that shown in the grammar of other dialects, such as the semi-standard West Saxon, which is also variable, but not to the same extent.

1.2. Middle English / Early Modern English

1. Reflex of Germanic */a/ when followed by /mb, nd, ng/
2. ME Breaking (+ / -)
3. Reflex of OE /a:/
4. Reflex of ME /o:/
5. <f(f)>/<v> in final position
6. ‘velar’ vs. ‘palatal’ in CHURCH, MUCH, SUCH
7. Spelling <s> in SHALL and SHOULD (+ / -)
8. <i> as a diacritic for vowel length
9. Loss of -n in final position (+ / -)

As regard morphosyntax the following items will be tagged:

Table 2. Morphosyntax


As far as vocabulary is concerned, so far we have marked words of Scandinavian origin, providing their etymology. In this new stage of our project Sara Pons Sanz (University of Nottingham) will be in charge of the study of the lexis and will tag the Scandinavian loans, indicating their regional distribution, i.e. whether they are ‘northern’ or have acquired a more general use in the language at a given period.

5. Text Analysis: Sample

As previously stated, this article focuses on Early Modern English. This period is marked by the diffusion of the process of standardization, whose onset is traditionally dated around 1430. Thus, texts from the late 15\th and 16\th centuries show fewer regional features than those of earlier periods, with texts of a private nature being more prone to retain variation than public texts. Görllach (1999: 506) makes reference only to literary texts, letters and diaries as possible sources for dialect study in this period, excluding formal texts. However, from our analysis of wills and civic records from Yorkshire, we can conclude that legal texts also display regional features and present more variation than might be expected from the rapid advance of standardization in this period. The results of a preliminary study of the distribution of northern features in Yorkshire wills from the 15\th-16\th c. have already been published by Fernández Cuesta & Rodríguez Ledesma (2004). Our research for this period is work in
progress and so far we have only analysed legal and private texts.\textsuperscript{7}

Here follows an example of one of the texts that has already been edited and tagged. First, we make description of the text, its edition with a comment on some of the highlighted features, and, finally, the linguistic analysis which is the basis for the tagging.

\textbf{5.1. Text Description}

We have selected a civic record of the city of York from 1549-50. The transcription has been made from a microfilmed copy of the manuscript.

\textbf{5.2. Text}

‘Streets to be kept clean by householders. Plague lodges on Hob Moor. Castle Mills accident. Inspector of sea fish’

MS reference: York Civic Records, Book XX, f. 1b, 2, 2b

\begin{verbatim}
Quintus die February Anno iiiijth Majestie Edward vijth

Assemblyd in the Coussayll Chambre of Ousebryg of the sayd City the day and yere above sayd when and where William Man delyverd in to the hands of the seid Lord Maior his leesse of the Chappel callyd Seynt George Chappell to the use of this City./ And it ys agreyd by the seid presens yat if the seid William Man tayke any advantaunce or to gyt any {pencyon} of the Kynges "{......}" at any tyme hereafter then he to have ageyn by way of law the seid leesse to help hym for that {....}

Item it ys agreyd by the seid presens for the clensynge and kepyng cleyn of all the streyttts within this City. that every inhabitaunt within the same fromenowfurth shall cause all suche parte of the seide streyttts as liyth before theire dwellynge howses to be twyse clensyd and swepyd every weyk appon payne of every man yat maybe defaulte at any tyme hereafter to forfet therefore xijd to the common Chambre of this City {nones pene}. whiche penalyte shalbe levying by the Constables of suche persons as shall forfet the sayd payne/ Furthermore it ys agreyd by
\end{verbatim}

\textsuperscript{7} We are currently working on literary texts from the Modern English period, especially Cumberland. At a later stage of the project some of these texts will be included in the corpus with the view to assessing the resilience of some dialect features in written vernacular varieties of the north of England.
the sayd presents yat all the inhabitaunts within this City shall cause ye common Streyttts and frunts before theire doers to be sufficiently pavyd as often tymes as neyd shalbe appon a resonable warnyng appon payne of forfettyng of iij shillings iiij pences {nones pene} to the seid commun Chambre to be levyed as is aforesayd 30
Item it ys agreyd by the seid presents yat of suche money as nowe is and hereaft shalbe gatheryd of devocyon within this City there shalbe (fo. 2) two lodgs mayde of hobmore for the inhabitacyon of suche folks as ar infectyd with the playg of pestylency for the saifgard of this City provyded always always yat if yt chaune hereafter no persons to be visyted with with the sayd playg then can conveniently inhabet and dwell within the lodgs of the seid mower./ that then so persons so infectyd to inhabet and dwell withoute laythorpe posteron there to be ordered as the seyd presents shall farther devyse as the cas shall requyre for the preservacon of the seid City etc 40
Moreover it ys agreyd by the said presents for the mayntenaunce of the auncyent libertyes of this City yat Myles Newton and John Emondson nowe beyng the Kynges Majestie Coroners within the Cownty of the City of york shall byforce of theire offyce of Coronershyp go and syt supervisum Corporis of William Watson otherwyse callyd William Clerk son of John Watson alias clerk bocher who was drownyd the iiijth day of this instaunct moneth a littyll frome the Castell mylnes in the Cownty of the sayd City with the greyt owtragous fluyd of water that then came owte of the water of fosse with suche vyolens yat the sayd water dyd then and there overthrow and whelme over the bote yat the sayd William Watson Robert Welshe maryner Rychard dyconson shipwright and one boye of Fraunces Trotter was then in for to sayf twoo sheyps of ye sayd John Watson yat then had been drownyd within seynt George if yat they had not beyn takyn owte of the seyd close by the sayd William Watson and other abovenamyd yat was then in the sayd bote with hym. Any Instruccyon or commandment gevyn unto the seyde coroners heretofore in writyng under the lord Admyrall seyll of England of his offyce of the Admyraltie in the `last` tyme of the Mairaltie of the Right Worshipfull `master`George Gayll in any thynge to the contrary notwithstondyng. 60
Item it ys fully concluyd and agreyd by the seyd presents for a common (fo. 2b) that William Man one of the common servaunts of this City shalbe surveyour of all suche freshe sea fishe as at any tyme hereafter shalbe brought to fossebrige within this City to thentent yat all the inhabitaunts within the seid City and Suburbes of the same shalbe well servyd therwith at a resonable pryce for their money./ And fromehensfurth to suffer no parte of the seyd fyshe to be conveyed away in bots nayther to the towne of Duncaster nor to none other place or places contrary the auncyent statuts and ordynaunces of the sayd City./ And also yat he shall suffer no suche kynde of fyshe to be conveyed owte of this City at any tyme hereafter of horslods but suche as the Mayor of the said City for the tyme beyng shall suffer to passe by his
discresson by speciall commaundment by hym to be gevyn to
the said William Man or to any other commom {.}. usyng and
excersisyng the sayd offyce for a common well./ And the
seid William Man to have a reward for his paynes takyng
in this Cas at Duncaster next to come accordyng to his
deservyng to be payd fyth of the common chambre of this
City. Withoute takyng of any {l..ber} appon payne of forfeettyng
of his of offyce of "common" serauntshyp etc./

Item it ys agreyd by the seid presens that the letter of
attourney of Course concernyng the offyce of the said lord Maior
of {---tourshyp} and also the offyce of the Sheriffs of this city
shalbe sealyd with the common seell of this city and to be
sent furthwith unto Mr. Cristofer Smyth and to Mr. {;}ebold
beyng attourneyes fo this city. the kynges court of {........}

Glossary:
Bocher: ‘butcher’
Mairaltie: the office of a mayor
Thentent: the intent, ‘in order that’
Castell mylnes: Castle Mills (Richmond)

As has been mentioned before, we have marked all the words where there is context for a
particular feature, regardless of whether the ‘northern’ variant appears or not. Since this text is
from the Early Modern English period, many of the features that are characteristic of northern
Middle English have already been standardised. The spelling <a>, for instance, as a reflex of
OE /a:/, is only found in one instance: overthraw (l. 50). Elsewhere standard influenced
spellings are found, which is the general trend in legal texts of the period (Fernández Cuesta /
Rodríguez Ledesma 2004). The same can be said about typical northern spellings suggesting
velar /g/, /k/ instead of palatal /dʒ, /tʃ/. They are only found in two instances: Fossebrige (l.
64) and Ousebryg (l. 2).

On the other hand, late orthographic innovations, such as the use of <i> as a diacritic to
indicate vowel length, are widely illustrated in the text. This feature, which is characteristic of
Scottish texts from earlier periods, is common in northern English from the end of the
fifteenth century, especially in semi-standardised texts (Benskin 1989:16). It is found in
words that had /a:/ in Middle English, such as: made, safe and take, but it also appears in
words that had ME /e:/: cleyn (l. 15), streytts (ll. 15, 18, 27), weyk, (l. 20), neyd (l. 28), etc.

The standardisation of written English, which starts in England in the 15th c., is especially
visible in legal texts from this period. This text illustrates how two of the most characteristic
morphological features of northern English (verbal -s for 3sg. present indicative and the
ending -and for the present participle) have already been replaced by standard ones: -eth
(liyth (l. 18)), and -ing (clensynge (l. 15), kepyng (l.15), dwellynge (l. 19), etc).
Standardisation did not proceed at the same pace in all text-types; as our own research has
shown, the early appearance of these standard features in 15th c. legal texts seems to point to
an earlier standardisation of this text-type, as opposed to, for example, the literary genre
(Fernández Cuesta / Rodríguez Ledesma 2004).
5.4. Analysis

1. Phonology and spelling

<table>
<thead>
<tr>
<th>Feature</th>
<th>SCONETag</th>
<th>Instances in the text</th>
</tr>
</thead>
</table>
| Reflex of OE /a:/ | oe-aa | <a>: overthrow (OVERTHROW) 50  
|  |  | <o>: furthermore 25, one 52, 62, none 69, mo (MORE) 35, so 37 (2x), moreover 41, go 44, also 70, 84, two 33  
|  |  | <oo> twoo 52 |
| <V+i/y> to show vowel length | i | Present: agreyd (AGREED) 7, 14, 25, 31, 41, 61, 82, cley (CLEAN) 15, streytts (STREETS) 15, 18, 27, weyk (WEEK) 20, neyd (NEED) 28, belyn (BEEN) 54, mayde (MADE) 21, 33, playg (PLAGUE) 34, 36, greyt (GREAT) 48, sheyps (SHEEP) 52, sayf (SAFE) 52, saigard (SAFEGUARD) 34, seylle (SEAL) 58, tayke (TAKE) 8, liyth (LIE) 18  
|  |  | Absent: leesse (LEASE) 5, 11, cas (CASE) 39, sealyd (SEAL) 85, seell (SEAL) 85, takyng (TAKE) 80 |
| Reflex of OE <a/o> when followed by /mb, nd, ng/ng, nd, mb> | g-and | <o>: notwithstondiyyng (NOTWITHSTANDING) 60 |
| Reflex of OE /o:/ | me-oo | <u>: fluyd (FLOOD) 48  
|  |  | <o>: doers (DOOR) 27, hobmore (Hob MOOR) 33  
|  |  | <ow> mower (MOOR) 37 |
| Velar vs. palatal | k | Velar: Fossebrige (BRIDGE) 64, Ousebryg (BRIDGE) 2  
|  |  | Palatal: suche 18, 24, 31, 33, 49, 63, 71, 72  
|  |  | whiche 23 |
| sal/suld vs. shall/should | sh | shall (SHALL) 17, 24, 26, 39, 43, 70, 73  
|  |  | shalbe 23, 28, 32 (x2), 62, 64, 65, 85 |
| <f, ff> vs. <v> in final position | ff | sayf (SAVE) 52  
|  |  | have 10, gevn 56, 74 |

2. Morphology

<table>
<thead>
<tr>
<th>3pl. pronoun subject</th>
<th>they-subj</th>
<th>they 54</th>
</tr>
</thead>
<tbody>
<tr>
<td>3pl. pronoun possessive</td>
<td>they-poss</td>
<td>theire 19, 27, 44, their 66</td>
</tr>
<tr>
<td>3sg.pres.ind. &lt;-es / -is, -ys / -(e)th&gt;</td>
<td>sg3</td>
<td>-th: liyth (LIE) 18</td>
</tr>
<tr>
<td>Pres.ind.pl. of be</td>
<td>be</td>
<td>ar (BE) 34</td>
</tr>
<tr>
<td>Pres.part.</td>
<td>ing</td>
<td>-ynge: clensynge 15, kepynge 15, dwellynge (DWELLING) 19, belyn (BEING) 42, 73, 87, warnynge 29, wrytyng 57, notwithstondynge 60, leyynge 23, forfettynge 29, 80, usynge 75, excersisyng (EXERCISE) 76, taykyng 77, 80, accordyng 78, deservyng 79, concernyng 83.</td>
</tr>
</tbody>
</table>
3. Lexis

| Scandinavian borrowings | scan | forth | FORTH | fromenowfurth | 17, fromehensfurth | 66, furthwith | 86, laythorpe | 38 |

There is an example of split genitive with zero morpheme: *under the Lord Admyrall seyll of England* ‘under the Lord Admiral of England’s seal’. The verb *was* in *other abovenamyd that was then in the sayd bote with hym* follows the Northern Subject Rule.

6. The SCONE electronic corpus

The next sections of the paper describe the methodology followed in creating the electronic version of the corpus described in sections 1 to 5, and the current status of the web interface.

The SCONE corpus is TEI-compatible. The Text Encoding Initiative (TEI) stands out as the best documented and organised standard for the representation of texts in digital form. Its first guidelines appeared in 1994, as the result of a project sponsored by the Association for Computers and the Humanities, the Association for Computational Linguistics and the Association for Literary and Linguistic Computing.

In addition to dictating or ‘suggesting’ how digital texts should be annotated, the TEI consortium offers a number of publicly available modules, examples and software, so that, a very small portion of the analytical description, if anything, should be redesigned.

The TEI standard is organised as a core module which is common to all TEI documents, and a series of specialised modules for the representation of specific types of text, such as Corpus texts, Printed Dictionaries, Manuscript Description, Transcribed Speech, Verse Structures, and some others. These modules do not work as closed entities. Instead, they can be augmented and/or reused as part of the definition of other types of document until the combination which best suits the desired description apparatus has been obtained.

The TEI Guidelines use XML (Extensible Mark-up Language) as a metalanguage for defining markup schemes. XML has become the standardised lingua franca in which virtually all electronic corpora are being created nowadays, both inside and outside the TEI guidelines. In addition to XML, the Unicode standard for character specification has certainly brought about a significant improvement in the way corpora can nowadays be created, manipulated and visualised. This is of special relevance when dealing with texts which contain ‘strange’ characters, such as Old and Middle English.

6.1. TEI Customization

An XML file must be well-formed according to a pre-defined structure in the form of a DTD or xml schema. Creating such a structure is not a trivial task. The TEI consortium makes available a Web interface which automatically generates the desired DTD or schema based on the options selected by the user.8

This process starts by selecting the TEI modules which are required in the current customization, and deleting and/or modifying the pre-defined elements and attributes in each module.

---

8 http://tei.oucs.ox.ac.uk/Roma/
Once the different modules, elements and attributes have been chosen, the personal customization can be saved, and the corresponding DTD and/or schema can be generated automatically.

6.2. Oxygen XML editing tool

An additional advantage of making use of the TEI modules is that one of the most popular XML editing software tools, Oxygen (see figure below), is equipped with built-in TEI schemas and stylesheets. Oxygen has also proven an essential tool in this project, since it also equipped with a version control system (SVN) which facilitates the collaborative work of researchers contributing from different sites.

7. The road to XML annotation

This section describes the particular TEI customization which has been necessary to encode the analytic apparatus described in section 4 above. Elements from 9 TEI modules have been necessary in our customization: header, core, textstructure, analysis, linking, msdescription, namesdates, transcr, and verse. Only the most relevant elements will be described below.

7.1. Manuscript Description

The first part of the tagged text is devoted to describing the manuscript under analysis. As pointed out in section 5.1. and 5.2., the following information is provided for each text:

- Title of the text
- Manuscript Reference (repository and identification number)
• LALME Code (where applicable)
• Text-Type, according to the predefined taxonomy of text classes
• Author (Scribe)
• Period
• Dialect Area
• Source of the manuscript

What follows is the actual tagging of those parameters in the sample text:

```
<msDesc>
  <msIdentifier>
    <settlement>York</settlement>
    <repository>City Archives</repository>
    <altIdentifier>
      <collection>York Civic Records</collection>
      <idno>Book XX, ff. 1b,2,2b</idno>
    </altIdentifier>
  </msIdentifier>
  <msContents class="#Legal">
    <msItem>
      <author>Anonymous</author>
      <textLang mainLang="en">York city</textLang>
      <title>Streets to be kept clean by householders. Plague lodges on
      <placeName>Hob Moor</placeName>.
      <placeName>Castle Mills</placeName> accident. Inspector of sea fish</title>
    </msItem>
  </msContents>
  <history>
    <origin notBefore="1549" notAfter="1550"
      period="Early Modern English"/>
  </history>
  <additional>
    <adminInfo>
      <recordHist>
        <source>Microfilmed copy from original manuscript</source>
      </recordHist>
    </adminInfo>
  </additional>
</msDesc>
```

7.2. Text XML tagging

Once the manuscript has been identified, the actual text is tagged according to the diagnostic features presented in section 4.1.
As described above, a total of 36 diagnostic features have been identified. They are declared only once and shared by all XML texts in an external file which is included via an include instruction:

```xml
<back>
  <xi:include href="DiagFeats.xml"></xi:include>
</back>
```

Each diagnostic feature is given a short identifier (xml:id) and a description, which will be later displayed in the html version. The most appropriate TEI element to group the diagnostic features is the interpGrp tag: “(interpretation) summarizes a specific interpretative annotation which can be linked to a span of text.”

```xml
<interpGrp xmlns="http://www.tei-c.org/ns/1.0" type="DiagFeats">
  <interp xml:id="g-ae">Reflex of Germanic */ae:/</interp>
  ...
  <interp xml:id="scan">Scandinavian loan</interp>
</interpGrp>
```

Distinct words: Words exhibiting one of the diagnostic features defined above will be tagged as <distinct>. According to the TEI Guidelines, <distinct> identifies any word or phrase which is regarded as linguistically distinct, for example as archaic, technical, dialectal, non-preferred, etc., or as forming part of a sublanguage. The <distinct> element contains an ana attribute whose value must coincide with the analysis tag previously defined in the diagnostic features under the xml:id attribute. The ana attribute, according to the TEI guidelines “indicates one or more elements containing interpretations of the element on which the ana attribute appears”. The following sample shows the actual tagging of some lines in the sample text. It should be noted that, although not found in this portion of the text, more than one analysis feature may be assigned to each word. The sample also includes some other tags which will be explained below:

```xml
<p n="2">Assemblyd in the Counsayll Chambre of <distinct ana="#k">Ousebryg</distinct></p>
<p n="3">of the sayd City the day &amp; yere abovevesayd when &amp;\p>
<n p="4">where <expan>Willi<ex rend="it"/>m</expan> Man <expan>delyv<ex rend="it"/>er</expan> in to the <distinct ana="#g-and">hands</distinct> of </n>
<p n="5">the seid lord Maior his <distinct ana="#i">leesse</distinct> of {...} the</p>
```

Other annotations: Place names (<placeName>), Person names (<persName>), measure expressions (<measureGrp>), dates (<date>) and expansions (<expan>) have also been tagged. Finally, special characters such as yogh and other typographical conventions have also been included for its correct rendering.
7.3. Html conversion

Once the text has been tagged, it is converted into an html file for its correct visualization in web browsers. This is accomplished by means of an xsl stylesheet especially designed for this project.\footnote{We are especially grateful to Fernando Gómez for designing the stylesheets, and taking care of most of the technical details related to the project.} The figure above shows a short sample of the html version obtained for the same text.
7.4. Web interface

At this point, a first version of the web page of the project has been built (http://ingles3.us.es). The web page contains an introduction to the project, members, publications, and a link to the Three versions of each texts are available for each text: xml, html and plain. The plain version is intended for those researchers willing to create concordances and other type of analyses of the texts.

In addition to making the texts available in those versions, we intend to offer a querying interface through the web page which will allow users to search for specific analysis features across texts, and impose restrictions on the search (i.e. show me all instances of words containing X features in legal texts between such and such period, etc.).

8. References


Fernández Cuesta, Julia and Mª Nieves Rodríguez Ledesma (2004): “Northern Features in 15th-16th-Century Legal Documents from Yorkshire”, in Marina Dossena and Roger Lass (eds.) Methods and Data in English Historical Dialectology, Bern, Peter Lang, 287-308.


9. Appendix: Textual evidence

9.1. Early Northumbrian

*Caedmon’s Hymn:* MS Cambridge University Library Kk, 5.16
*Bede’s Death Song:* MS St Gall 254, fol. 127a.
*Leiden Riddle:* MS Leiden, Voss 106, fol. 25b.

Early runic inscriptions (Ruthwell Cross Inscription, Franks Casket Inscription, etc.)
Early non-runic inscriptions
9.2. Late Northumbrian

Gloss to the *Lindisfarne Gospels*  
Gloss to the *Rushworth Gospels* (Northumbrian section: Rushworth2)  
Late inscriptions

9.3. Early Middle English

Carlisle, Cumbria RO, D/Lons/L Medieval Deeds C1: Gospatric's Writ  
Edinburgh, Royal College of Physicians, MS of Cursor Mundi, hand A, fols. 1r-15v  
Edinburgh, Royal College of Physicians, MS of Cursor Mundi, hand C, fols. 37r-50v  
London, British Library, Cotton Cleopatra B vi, fol. 204v: Pater Noster and Creed  
Oxford Bodleian Library, Bodley 26, fols. 107r-108r: Macaronic sermon  
London, British Library, Cotton Julius A v, fols. 180r-181v: Prophecy of Scottish Wars  
Oxford, Bodleian Library, Digby 2, fol. 6v, hand A lang 2

9.4. Late Middle English

London, British Library, Cotton Vespasian A iii. Hand A. *Cursor Mundi*  
London, British Library, Cotton Vespasian A iii. Hand B. *Cursor Mundi*  
North Yorkshire County Record Office, Northallerton: ZRL 1/20. Indenture  
New York, Pierpont Morgan Library, Glazier collection 39. Charm  
Beverley, Humberside County Record Office: DDHE 19, Section B, f. 133v. Marian verses  
London, British Library, Cotton Nero D iii, f. 176r: Cartulary of the Hospital of St Leonard, York  
Kingston-Upon-Hull, Corporation Archives: Bench Book 2, f. 212 and f. 243. Legal texts  

9.5. Early Modern English

9.5.1. Testaments:

TE1464: Testamenta Eboracensia 34. The will of John Smyth of Rolleston. 1464.  
TE1470-1: Testamenta Eboracensia 43. The will of John Ferriby, Esquire. 1470.  
TE1470-2: Testamenta Eboracensia 45. The will of Thomas Thurland, burgess of Nottingham. 1470.  
TE1471: Testamenta Eboracensia 54. The will of Thomas Beverley of York, merchant. 1471.  
TE1472: Testamenta Eboracensia 60. The will of Ralph Snaith. 1472.  
TE1476: Testamenta Eboracensia 76. The will of Henry Eure, of Old Malton, Esquire. 1476.  
TE1478: Testamenta Eboracensia 91. The will of Alice, widow of Sir Thomas Neville of Liversedge, Knight. 1478.
TE1479-1: Testamenta Eboracensia 92. The will of Joan Candell. 1479.
TE1479-2: Testamenta Eboracensia 95. The will of Isabel Grimston. 1479.
TE1480: Testamenta Eboracensia 98. The will of William Scargill of Lede, Esquire. 1480.
TE1482: Testamenta Eboracensia 109. The will of Sir Hugh Hastings, Knight. 1482.
TE1490: Testamenta Eboracensia 29. The will of John Holme of Huntington, Gentleman. 1490.
TE1496: Testamenta Eboracensia 61. The Will of Sir James Danby of Thorp and Farneley, Knight. 1496.
YW1520: York Clergy Wills 1. The will of William Burton, vicar choral. 1520.
SW1522: Swaledale Wills 1. The will of Henry Snawdon of the Parish of Grinton. 1522.
YW1522: York Clergy Wills 3. The will of Thomas Seton, Chantry Priest in York Minster. 1522.
SW1533: Swaledale Wills 3. The will of William Huchynson of Marske, Yeoman. 1533.
YW1534: York Clergy Wills 13. The will of John Sheffield, prebendary in the chapel of St Mary and the Holy Angels in York Minster. 1534.
SW1539: Swaledale Wills 6. The Will of Brian Clarkson of Satron in Grinton. 1539.
YW1540-1: York Clergy Wills 17. The will of William Crosby, vicar choral. 1540.
YW1540-2: York Clergy Wills 18. The Will of John Rayner, chantry priest in York Minster. 1540
SW1542: Swaledale Wills 10. The will of Sir Edmond Charder of Grinton, priest. 1542.
YW1549-1: York Clergy Wills 32. The will of Thomas Baoke, vicar Choral. 1549.
SW1552: Swaledale Wills 32. The will of Thomas Cooites of Dellwaite in Marske. 1552.
SW1553: Swaledale Wills 37. The will of William Conyers of Marske, Esq. 1553.
YW1564: York Clergy Wills 41. The will of George Williamson, prebendary of Hustwaite. 1564.
SW1569: Swaledale Wills 108. The will of Janet Metcalf of Low Whitaside in Grinton. 1569.
SW1575: Swaledale Wills 106. The will of John Close of Healaugh in Grinton. 1575.
YW1578: York Clergy Wills 53. The will of Anthony Foord, succentor of York. 1578.
YW1585: York Clergy Wills 57. The will of Anthony Iveson, vicar Choral. 1585.
YW1590: York Clergy Wills 58. The will of Henry Moore, prebendary of Bole. 1590.
SW1595: Swaledale Wills 186. The will of Anthony Spenseley of Whitaside in Grinton. 1595.
SW1600: Swaledale Wills 208. The Will of Margaret Metcallf of Muker in Grinton. 1600.

9.5.2. Civic Records:

Award as to a Gutter. Book I, fol. 21b
Honour for the Mayor returning from Parliament. Unseemly language. Books II, III, IV (bound together in one volume), fol. 112
Arrangements for collecting Royal aids. Book IX, fol. 20a

10 The titles of the manuscripts after Angelo Raine.
Bridgemasters to make an inventory of moveables and to repair their property. Selling of corn. Book IX, fol. 24b
A Curfew for poor Labourers. Book XI, fol. 101b
Streets to be kept clean by householders. Plague lodges on Hob Moor. Castle Mills accident. Inspector of sea fish. Book XX, fol. 1b, 2, 2b
Difficulty in getting Bridgemasters in plague time. No Corpus Christi Play this year. Book XX, fol. 16b, 17
Punishment of evil doers. Book XXIII, fol. 13b
Brick stall to be removed. Assessments for poor rate. Fishing day. A Ducking stool. Book XXVII, fol. 243b, 244
Report on Tang Hall lands. Book XXIX, fol. 185b, 186