

A LINGUISTIC ATLAS OF EARLY MIDDLE ENGLISH

INTRODUCTION

PART I: BACKGROUND

CHAPTER 2

INTERPRETING MIDDLE ENGLISH

Roger Lass and Margaret Laing

2. Interpreting written Middle English¹

The letters of the medieval roman alphabet are culturally invested symbols, they have a history, and they have names. Their history informs their use, as also do their phonic implications. In ways that speech is not, writing is subject to design: analysis must take account of the doctrine of *littera*, of the conceptual categories of the designers. The evolved orthographies of the later middle ages, moreover, may have extensive grammars of interchange, the cumulative and partly systematised legacies of sound-change and calligraphic development. Middle English spellings do not exist *in vacuo*: they are products of a generative system (Benskin 1991: 226).

2.1 Introduction

Our first concern in *LAEME*, as it was in *LALME*, is the recording of text languages and the display of their features in profiles and on regional maps.² This has to be done in two stages. The first order taxonomy is orthographic: the primary written evidence of the manuscripts. But the creation of insightful interpretative maps (such as the *LALME* Dot Maps) involves consideration of one major second order property — phonetic substance. In *LAEME* (like *LALME*) we present first order Item Maps and implicitly second order Feature Maps; but for reasons that will become apparent below we also present maps explicitly indicating reconstructed phonetic substance. This reconstruction is particularly necessary for smoothing out the surface ‘nubbliness’ of the early Middle English continuum (Laing and Lass 2003, and see below §2.3.3).

Early Middle English writing is often extremely complex and difficult to interpret. It is above all not uniform. Like late Middle English, it varies from region to region and from scribe to scribe. Much more commonly than in late Middle English, early Middle English usage may also vary from text to text copied by the same scribe and (depending on textual histories) also from portion to portion of the same text (cf. §1.4 esp. n. 18).

The main reason for this complexity is the contingent rupture of the tradition of writing English that occurred after the Norman Conquest (cf. §1.2). A set of ‘text communities’ that had been accustomed to producing a large portion of its official and other documentation in its own vernacular was tipped into a new praxis. For a century English was generally not used as a written medium except for some copying of Old English legal documents into monastic registers and cartularies. In some centres Old English religious texts also continued to be copied. Although these ‘transitional’ texts show some orthographic developments that may well reflect changes in the contemporary language of the copyists, for the most part they are still recognisably Old English. We have no evidence, until the mid to late 12th century, of what could be called spontaneously produced up-to-date written English. Judging from the evidence of the written English that begins to appear post 1150, in the years after the Conquest there must have developed a kind of diglossia. Spoken English would have been used by the majority in the normal way, and continued to vary and change, differently in different regions, like any natural language; but written expression was nearly exclusively in Latin and French (the language of the conquerors), and scribal employment would have depended on mastery of these traditions. This is not a particularly unusual situation: there are many present-day cultures in which written

¹ Some portions of this chapter are amplified and revised versions of material that appears in Laing 2004, Lass and Laing 2005, Laing and Lass 2006.

² And for *LAEME* also specifically diachronic maps.

and spoken discourse are in different languages (Arabic and Tamil are familiar examples). But the immediate post-Conquest English situation is rather different; in Arabic and Tamil the written language is an archaic and ‘classical’ (if not always directly ancestral) form of the current spoken language: Classical Arabic and written Tamil are still Arabic and Tamil.

By the early Middle English period, four generations after the Conquest, the high prestige languages in the written/spoken diglossia (though distantly cognate) were ‘foreign’: normally second and third languages, formally taught and learned. By the late 12th century it is virtually certain — except in the case of scribes from continually bilingual families (and we do not have the information that would allow us to identify them) — that none of the English scribes writing French were native speakers of French (though some may of course have been coordinate bilinguals).³ It is certain that none were native speakers of Latin, since after the genesis of the Romance vernaculars in the early centuries of this era it is most unlikely that there were any.

This means in effect that when the writing of the native vernacular was revived in the 12th century, the hiatus made it necessary for scribes to design new orthographies to represent the results of a century of massive and transformative change at the phonological level, which is the primary (though not the only) input to orthography.

Since English was not ‘classical’ or institutionalised it did not require fixed spelling: so different styles of orthographic design were able to flourish. And since the one-word/one spelling mode of most modern standards was not the norm, natural phonological and morphological variation was not prevented from surfacing in written forms. Early Middle English texts display a wide range of representational strategies. In addition to simple ‘phoneme’-to-‘grapheme’ mapping,⁴ we find logographic and morphographic writing, as well as literal substitution. For these terms see §§2.2 and 2.3 below.

2.2 *What do writers spell?*

2.2.1 Levels of representation

A spelling system is a mapping of some chosen set (or sets) of linguistic units into a set of visual signs.⁵ The standard inventory of linguistic units is the word, the

³ It is likely that writers of Anglo-Norman at this date would have had English as their first language maintaining Anglo-Norman as an artificial written competence, usually alongside Latin. See Prior (1923: 161–185) and cf. Rothwell (1968: 37–46; 1975–76: 445–466; 1978: 1075–1089; 1983: 258–270) and Short (1980: 467–479; 1992: 229–249; 1996: 153–175).

⁴ The scare-quotes suggest that these terms are of questionable status and utility in the analysis of many early Middle English orthographies. This is not to say that some medieval writers did not employ a praxis in which they would be applicable in the modern structuralist sense: the 12th-century Icelandic First Grammarian (see Benediktsson 1972) is a good example of a writer and theorist whose ideology could be said (without much anachronism) to be a kind of classical structuralism. Though it is interesting that the First Grammarian’s theoretical defence of the biunique commutative mode was thought necessary at all, which suggests that medieval norms were otherwise. There are Middle English scribes also whose praxis could be interpreted more or less this way, though they do not engage in metacommentary.

⁵ Just what kind of units is a complex matter. In the earlier part of this exposition we will be considering phonemes, allophones and other traditional kinds; later on we will adopt a more medieval

morphophonemic representation, the syllable⁶ and the ‘phoneme’. In relatively rare cases allophones of certain phonemes may be represented (e.g. the velar nasal in the elder Futhark, Gothic and Greek).

Modern and many ancient alphabetic scripts tend to represent language at what could be loosely called the phonemic level: a linear string of graphs is a rough icon for a string of structuralist phonemes of the sort one could arrive at by commutation — minimal pair tests or similar procedures. But trying to reverse-engineer historically bequeathed alphabets as redundancy-free and univocal systems of representation would generally disappoint us. This is true of early Middle English more than many other languages, for reasons we will discuss below. If spelling systems were designed that way, an ‘ideal’ alphabet would represent biuniquely: one grapheme per phoneme and *vice versa*. Such systems in fact are rare, and highly unlikely in languages with long histories and little spelling reform. Most orthographies carry considerable historical baggage: e.g. English ‘silent’ final <-e>, <gh> in *eight, night, through*, and <kn-, gn> in *know, gnaw*. They also tend to have much non-biunique representation (/ʃ/ in *shoe, vicious, ocean, nation, passion, chic, schist* ...). This may be exacerbated by intensive contact with other orthographies.⁷

Not all alphabetic systems are pure; many also use other types of representation. We find such mixed systems throughout Germanic. The commonest of the not strictly alphabetic, supraphonemic representational strategies are:

(i) *Logography*. Non-biunique phoneme/grapheme relations can enable a form of logographic writing: the spelling indicates not only a phonological string, but one particular member of a set of homophones: E *right, wright, rite*, Afrikaans *ys* ‘ice’, *eis* ‘claim’, both /eis/. These distinctions often reflect historical origin: *right* < OE *riht*, *wright* < OE *wyrhta*, *rite* < OF *rite*; but sometimes they do not, as in E *deer, dear* < OE *dēor, dēore*. We do not use ‘logographic’ in the way that ‘ideographic’ was used in older descriptions of Chinese: 1, 2, @, & are not logographs, but rather icons or pictographs. In our definition, there are two main types of logography: (a) the discrimination of homophones in such a way as not to violate the graphotactic/representational rules of the system, e.g. *dear/deer*; (b) the consistent assignment of particular spellings to particular words where other spellings would, according to the structure of the system, be allowable, e.g. modern English *bright* with medial <gh>, where **brite* would be equally well-formed.⁸

There is a cline, at one end of which are ‘genuine’ spellings of words (including logographs in the sense defined above, like *dear/deer*), and at the other are pure icons. In between is a domain that is of particular interest when dealing with medieval texts: abbreviation. Even abbreviated lexical items constitute a cline, because some retain more phonological clues than others. If *deer/dear* are ‘full’ logographs (since a pronunciation can be inferred from them according to the usual orthographic rules of the language), a further stage toward abstractness on the cline is

and less anachronistic perspective. The mapping need not be one-to-one and in no case does it have to involve one level of unit only.

⁶ We will be concerned here only with segmental styles of representation, as syllabaries are not part of the Germanic tradition. Even the oldest Germanic writing, the runic inscriptions in the Elder Futhark, are in principle alphabetical, if often ‘defectively’ so.

⁷ When discussing modern standard languages we will use the traditional // and < > representations for phonemes and graphemes respectively for convenience and familiarity. Later we will be using other notional devices and a new denotation for < >.

⁸ Indeed in some Middle English varieties *brite* would be an acceptable spelling for this word.

something like barred thorn <~> indicating THAT.⁹ How the vocalism of this ideal ‘that’ might be realised phonetically is undecidable. So the abbreviation means THAT lexically, but no particular form phonologically, except with respect to the initial consonant.¹⁰ The crossbar has no phonological import, but the thorn limits the range of representations to those beginning with [θ~ð]:¹¹ <~> is not merely a trigger. The same could be said for <S~> for SAINT. These might be called ‘partial’ or ‘impure’ logographs. There is also the occasional use in religious texts of initial letters only to stand for well known or much repeated word sequences (e.g. in Oxford, Merton College 248, for repeated quotations from the *Pater Noster: ore ylk \ d. b. y. g. vs.*, ‘our each day’s bread thou give us’). At the other end of the cline are objects like ampersand <&> and the Tironian sign <Ϟ> for AND, or <xpc> for CHRIST, and <i□c> for JESUS. Even though these last two examples are made of letters, not abstract symbols,¹² like <&> and <Ϟ> they are not subject to phonological extrapolation. All these objects are mere triggers, i.e. their reference is non-phonologically conventional, outside the surrounding alphabetical schema. These are what we call ‘icons’.

(ii) **Morphophonemic writing (‘morphography’)**. While alphabetic praxis generally represents at phoneme level, some languages represent morphophonemically or ‘abstractly’ as well. A systematic example is the writing of final obstruents in continental WGmc (except Yiddish). German, Dutch, Frisian and Afrikaans have final devoicing, but do not systematically indicate this in spelling, particularly for stops. Thus G *Bund* ‘league’, pl *Bünde* /bʊnt, bʏndə/, but *bunt* ‘colourful’, inflected *bunte* /bunt, buntə/.¹³ So, the final <d> in *Bund* (where /d/ according to surface phonotactics is ‘unpronounceable’) is a signal that /d/ will appear if a vowel-initial suffix follows. The writing <-d> then is **indexical**: it marks both morpheme identity and the existence of an alternation.¹⁴

Old English used similar abstract strategies: for instance, fricatives were voiceless initially and finally but voiced foot-medially between voiced sounds, so that a word like *wulf* WOLF [wulf] would have the nom/acc plural *wulfas* [wulvas] with <ƿ> written in both positions regardless of phonetic value (as opposed to ‘concrete’ ModE *wolf/wolves* where the voicing is indicated). This practice was maintained to some degree in early Middle English, though the preferential spelling mode appears to

⁹ We use small capitals to identify lexemes, as is done in Lyons (1977), and for ‘item names’ as in *LALME* and for glossing citation forms in English.

¹⁰ For a detailed explication of the representational status of abbreviated forms in Middle English see M. Benskin in *LALME* III, §14.12. For a full explanation of our editorial practice for the expansion of abbreviations and diacritics in the corpus see chapter 3, §§3.4.5.1 and 3.4.9.

¹¹ We indicate both voiceless and voiced dentals in general discussion. For THAT and other deictic words the decision as to whether initial voicing has taken place would depend on the time and location of the text language.

¹² They are realisations of originally Greek letters: chi, rho, sigma for CHRIST iota, eta, sigma for JESUS. To what extent these origins were known to particular scribes working in medieval England is not known, but some would certainly have been aware of them. These logographs derive from the old *nomina sacra*, which were not supposed to be uttered in full except in the liturgy. The words CHRIST and JESUS were also frequently written in full, whether in Latin, French or English texts.

¹³ Cf. *Laub/Laube* ‘leaf, foliage’, *Tag/Tage* ‘day(s)’ which follow the same principle of alternation: final <b, g> are read as /p, k/.

¹⁴ Sometimes languages change tack; Old High German tended to access the phonemic rather than morphophonemic level even though it had final devoicing: the ‘day’ paradigm was most commonly written *tac* vs *tage*.

have been concrete or phonemic (as far as this term applies: see §2.3).¹⁵ Here are examples of both types of spelling from one early Middle English writer, the Worcester Tremulous Scribe (first half of the 13th century), illustrated by forms of the reflexive suffix -SELF (OE *seolf/sylf*). The fricative in absolute final position is always written <f>, e.g. *sulf*:

Concrete: final <f> ~ medial <u>¹⁶
 P13XM he-sulf
 P23<prX ham-suluen
 P23X hi-sulue

Abstract: final and medial <f>
 P13XM he-sulf
 P110iX [m]e-sulfum
 P12GX †ines-sulfes

(iii) **Use of diacritics.** Many Germanic scripts are rich in devices of this kind, e.g. letters marked with accents or diereses, and various superscript and subscript symbols. Some languages also use doubling of vowel or consonant letters for diacritic purposes.

In early Middle English, there are three main types of sporadic, and for the most part unsystematic, diacritic usage: (a) doubling of consonants to indicate that the preceding vowel is short; (b) doubling of vowels to indicate length; (c) the use of accents on vowels to indicate their quantity.¹⁷

(a) The doubling of a consonant to indicate a preceding short vowel is widespread¹⁸ though not usually at all regular. A frequent use is to disambiguate words that would otherwise be homographs, e.g. GOD and GOOD (*godd* vs. *god(e)*). It is not a consistently applied practice, however. In most scribal systems short vowels before non-geminates are usually followed by single consonant graphs, especially where the spelling gives rise to no ambiguity. Some writing systems have double letters or digraphs like <ck> even after apparently long vowels, e.g. *bock* for BOOK in both Laðamon A, hand A and Laðamon B. Orm is apparently the only author to have developed an explicit system of diacritic consonant doubling.

¹⁵ The variable shift from abstract to concrete spelling strategies had already begun in some Old English varieties. A frequently cited example is *hliuade* 'it towered' at *Beowulf* 1799 vs *hlifade* 81 (infinitive *hlifian*).

¹⁶ These examples are taken from Worcester Cathedral, Chapter Library F. 174 (Ælfric's grammar and glossary). Two facts are of particular interest here: (a) the issue of abstract vs concrete spelling, though it surfaces in the actual written behaviour, was not considered of systematic importance in spite of the fact that the forms cited are in the context of a discussion of paradigm lists; (b) the text was copied from an Old English original, so the concrete spellings are innovations. For the meaning of the tags see chapter 4. All that counts for this display is that 'X' = reflexive.

¹⁷ It is arguable also that some vowel digraphs, e.g. <ea>, which in earlier times were assumed to be diphthongs may at this date have diacritic functions within certain text languages. Such usage must, however, be interpreted in the context of individual scribal systems because there is no consensus as to whether particular combinations imply continued diphthongs, monophthongs or diacritic modifications of the left hand vowel. See also §2.2.2 and n. 19.

¹⁸ This follows a long tradition in West Germanic languages in which <-CC> is usually a marker of a preceding short vowel, due to the fact that most geminates occurred after short vowels.

(b) Doubling of vowels to indicate length also occurs only sporadically: e.g. *deed* DEATH < OE *dēaƿ* in British Library, Royal 8.F.ii. Surprisingly, however, double vowel graphs may also appear for historical short vowels, even those followed by double consonants or consonant clusters (e.g. *-seelf* -SELF in Oxford, Bodleian Library, Laud Misc. 108, hand A). A further complication arises when a double vowel graph, apparently indicating a long vowel, is nevertheless followed by a double consonant graph, which would normally indicate a preceding short vowel (e.g. *soðfastheedd* SOOTHFASTHOOD in the short verse texts in British Library, Arundel 292).

(c) Accents on vowels to indicate quantity are also sporadic. We do not count as an accent the very frequent oblique stroke (like an acute accent) on <i>. It is a device for disambiguating sequences of minim strokes, helpful for all scripts, whether formal or informal. Although it is rarely employed with complete regularity, most scribes use it at least some of the time. The stroke (or dot) may appear on both historically short and long <i> (as well as on <y> and sometimes also on wynn and thorn). Where either the stroke or the dot appears in these contexts, we therefore take it to be integral to the shape of the letter, the equivalent of the dot on modern printed <i> and <j>, rather than having diacritic significance. Much less common is an acute accent on any other vowel but <i>/<y>, but when it occurs it apparently signifies vowel length. Sometimes the accent seems to have an additional function: to disambiguate a single <a> as a content word — e.g. the indefinite article or number one, or the word for EVER — from <a> representing the preposition IN. Even Orm does not use a consistent accent system. In his text, vowels may be marked by one, two or three acute accents. All three seem normally to indicate vowel length. Occasionally, he uses a breve above a vowel to indicate shortness, normally (but not consistently) to disambiguate pairs of homographs, e.g. *lāte* LATE and *lāte* MANNER.

2.2.2 What readers have to know

*A spelling system is a mnemonic for native speakers.*¹⁹ Readers of any language at any period have to be able to cope with redundancy and historicity. The history of English of all periods shows that it is possible for spellings to be intelligible even when they belong to systems that include considerable merger or redundancy. There appears to be no general rule for how much of a language's system actually has to be represented. Middle English generally did not distinguish between its two heights of long mid vowels, using <e(e)> and <o(o)> indifferently for both high mid and low mid. Though the common alternation in modern standard English between <ee> and <ea> (*greet* vs *great*) and between <oo> and <oa> (*brood* vs *broad*) goes back to the use in some Middle English dialects of <a> as a diacritic for the lower of each pair.²⁰

¹⁹ For further discussion see Lass and Laing 2005.

²⁰ <ea> occurs in many 13th-century texts. Whether it is diacritic in these uses is debatable: it is not clear whether early <ea> is supposed to represent [æ] as the diacritic interpretation suggests, or whether it is an alternative writing ('litteral substitution': see §§2.3.2 ff. below) for [e] or sometimes [a].

But a literate native speaker can routinely understand spellings that may seem strikingly ‘defective’.²¹ All spelling systems have built-in redundancy, and interpretation of even bizarre spellings is possible as long as the reader knows the system and has a good idea in advance of what a word is likely to be, or what the range of choices is. In the present context, no reader who knows English would have any difficulty reconstituting the defective representations <spllng> or <rthgrphc>.

There are, however, complications. If an orthography lasts long enough it will tend to represent ‘ghost contrasts’ due to sound change not indicated by spelling change: e.g. for PDE, except for some Northern Scots, <kn-> vs <n->, and for many dialects <wh-> vs <w->. This kind of purely orthographic pseudo-contrast is generally removable only by deliberate spelling reform. A segment lost in isolation may remain as a diacritic or alternation index: e.g. in English non-rhotic dialects postvocalic <r> is a marker of length and sometimes quality for a preceding vowel; and final <-r> is only a tag for recovering the lost /r/ in external sandhi (*law*, *lore* /lɔ:/). It surfaces phonetically in *lore* when it is followed by a vowel, but not in *law* except in dialects with ‘intrusive *r*’. In whatever mode of representation, there is no *a priori* reason to expect consistency: e.g. English and German are close to ‘proper alphabetic’ systems (representing mostly at classical phonemic level), but with a fair number of logographs, morphophonemic writings and ‘unexpected’ spellings.

If a spelling system is a mnemonic for native speakers, as historians we have no right to expect systems that cohere with our modern European standard-language ideology of ‘good’ spelling practice, or with orthographic models derived from particular formal linguistic theories. Almost all the early Middle English scribes whose work survives were clerics or other institutionally trained writers. We can assume that the spellings they employed were interpretable to their readers. The systems they designed or adopted show differing degrees of internal variation and structural flexibility. Our task is to develop a hermeneutic that provides (as far as possible non-anachronistically) an interpretation for the work of all sane scribes.

2.3 Literally speaking

2.3.1 The doctrine of *littera*

It may ... be questioned whether, if *letter* had been retained in something like its traditional functional sense, the need for a phoneme theory would ever have arisen — though we should, certainly, have subtle theories of the *letter* in its place. (Abercrombie 1965 [1949]: 84)

Up to this point we have been using standard terminology: ‘phoneme’, ‘grapheme’, etc., and representing these theoretical objects with the usual bracketing. We depart now from this framework, for reasons which will become evident. The most important of these is that such concepts do not always characterise what our scribes appear to be doing. They are frequently not ‘structuralists’, and it seems to us better to use a theoretical framework and notation that cohere more closely with what scribes would have experienced in their education — though we will take considerable liberties in exposition.

One work which probably all scribes would have been exposed to in the course of their training as the indispensable foundation of a medieval orthographic

²¹ For instance, the Cypriot syllabary failed to represent half the vowels and two-thirds of the consonantal contrasts (Lass 1997: 51–2); and Latin did not represent vowel length, which means that roughly half of the possible graphic word-forms available were potentially ambiguous.

education, would have been the *Ars maior* of Aelius Donatus (fl. 5th century AD). In book I is a statement that can be taken as canonical:

Littera est pars minima vocis articulatae ... littera est vox, quae scribi potest individua ... accidunt cuique littera tria, nomen figura potestas, quaeritur enim, quid vocatur littera, qua figura sit, qua possit.

Here is slightly exegetical translation.²²

Littera is the smallest unit of articulated sound ... littera is (a) sound which is capable of being written alone ... littera has three properties: name, shape, power [= sound value]. For one must ask what the littera is called, what its shape is, and what its power is.

The *littera* is clearly an abstract object, which under the classical interpretation is a member of a universal phonetic alphabet. Each *littera* in this view has only one possible *potestas*. Thus Quintilian (*Institutiones* I), commenting on some deficiencies of the Latin writing system, notes that certain ‘necessariae litterarum’ present in Greek are lacking in Latin, and that for writing e.g. *seruus* and *uulgus* ‘Aeolicam digammon desideratur’ (‘the Aeolic digamma = [w] is wanted’). This aspect of the theory of *littera* (which is notable in the English tradition in John Hart’s 1569 discussions of the ‘abuse’ of letters, and appears as late as Wallis 1653) will not be treated here; Middle English scribes were not generally concerned with *litterae* in this sense; indeed our interpretive task would be much easier if they had been.

The stream of *litterae* in writing is represented by a sequence of *figurae*; indeed this is the way the *littera* becomes visible. Discourse about the detail of *figurae* is most often the preserve of palaeographical interpretation. For discourse about the interpretation of spellings once established by a palaeographical reading the term *littera* may rather be used. For instance, it does not matter for the interpretation of a spelling whether a scribe uses a short <s>, a long <s>, a sigma-shaped <s> or a bean-shaped <s>. Certain types of script dictate the use of certain types of *figurae* in certain contexts: e.g. in Textura script, the use of 2-shaped <r> after a letter with a rightward facing bow (<o>, , <p>). Except where the concepts *littera* and *figura* overlap, (e.g. when the *litterae* thorn and wynn are realised by identical *figurae*) at the level of system we therefore talk about a *littera* as the superordinate for all the different possible *figurae* that different scribes, or any one scribe, may adopt for it.

For notation we will follow the conventions developed by Michael Benskin (1997: 1 n. 1 and 2001: 194 n. 4) and used by us in Laing and Lass (2003). We put *litterae* in inverted commas, the *figurae* that occur in particular manuscript systems in angled brackets (not to be interpreted as ‘graphemes’), and *potestates* in square brackets: so ‘e’, <e>, [e].²³ Glosses and the names of lexical items will be in small capitals. Standard citation forms, manuscript forms without specific literal reference, etymological categories and reconstructions will be in ‘uninformative’ italics.

2.3.2 Substitution sets

On one parameter, Middle English spelling systems can be loosely classified into two types (Laing 1999, Laing and Lass 2003): ‘economical’ and ‘prodigal’. An

²² Modified somewhat from Laing and Lass 2003: 298; original taken from Benediktsson 1972: §2.2. For detailed discussion of medieval orthographical theory and reform also see Benediktsson.

²³ On the degree of ‘precision’ with which the contents of [] are to be read see §2.4 below.

economical system makes some approach towards the ideal of one *littera* one *potestas*; a prodigal system allows considerable multivocal relationship. These notions of course are relative: all systems are somewhere on a cline between the two, though the two ends of the cline can be easily recognised. And even the most prodigal systems may be economical in some particulars. For instance, we do not know of any system that uses ‘p’ in a strikingly multivocal way; but we know of many that use ‘eo’ for a very large number of *potestates* (Lass and Laing 2005). Similarly we do not know of any early Middle English systems with a plethora of representations for the *potestas* [i],²⁴ but many with large substitution-sets for [x] (see below).

We can illustrate the flavour of the two system types by looking at the spellings of OE *-ht* (as in NIGHT, THOUGHT and NOT < OE *noht*)²⁵ in four hands in the same manuscript: Cambridge, Trinity College (323) (hereafter Trinity) (SW Midlands, second half of the 13th century). Here are the patterns for the four main hands that contribute text in English, with frequencies of different spellings:

Hand A: ‘st’ 89, ‘t’ 55, ‘tt’ 4, ‘cst’ 3, ‘ct’ 2, ‘th’ 2, ‘chit’ 1, ‘cht’ 1, ‘cðth’ 1, ‘s†’ 1, ‘th’ 1, ‘thth’ 1, ‘tth’ 1

Hand B: ‘st’ 90, ‘t’ 21, ‘tt’ 3 ‘d’ 1

Hand C: ‘t’ 19, ‘tt’ 3

Hand D: ‘t’ 11, ‘st’ 10, ‘cht’ 8, ‘ch’ 4, ‘ct’ 3, ‘d’ 2, ‘th’ 2, ‘tht’ 2, ‘ðt’ 2, ‘dt’ 1, ‘tf’ 1
‘tt’ 1

It is clear that Hands B and C are (at least relatively) systematically economical in their approach to words containing this historical sequence; Hands A and D are profligate.

A set of *litterae* in variation for the same *potestas* or etymological category we call a *Litteral Substitution Set* (LSS: cf. Laing 1999). Thus Hand C has for OE *-ht* the LSS {‘t’, ‘tt’}, Hand B has the LSS {‘st’, ‘t’, ‘tt’ ‘d’}, Hand D has the LSS {‘t’, ‘st’, ‘cht’, ‘ch’, ‘ct’, ‘d’, ‘th’, ‘tht’, ‘ðt’, ‘dt’, ‘tf’, ‘tt’} and Hand A has the LSS {‘st’, ‘t’, ‘tt’, ‘cst’, ‘ct’, ‘th’, ‘chit’, ‘cht’, ‘cðth’, ‘s†’, ‘th’, ‘thth’, ‘tth’}. The inverse of an LSS we call a *Potestatic Substitution Set*²⁶ (PSS: Laing and Lass 2003: 262–263). So in Trinity Hand D the littera ‘ð’ maps to the PSS {[h], [x], [j], [w], [ɣ]}. For instance [h]: *ðu* HOW < OE *hū*; [x]²⁷ *driðten* LORD < OE *dryhten*; [j]: *ðe* YE < OE *gē*; [w]:

²⁴ Of course there are some variant spellings for [i] in Middle English, including occasionally unexpected ones. In later Middle English there are two common possibilities, with the more frequent adoption of <y> beside <i> to represent [i]. In early Middle English <y> is also found for [i] but is less frequently so used.

²⁵ For justification of the choice of a historical category as identifier see §2.3.4,

²⁶ The term ‘potestatic’ is our coinage. For further discussion of the significance of *potestas* see §2.3.3 below.

²⁷ We use [x] to represent a high tongue-body fricative, without commitment as to palatal or velar articulation; this is a conservative convenience. It is clear that at some point in history (whether in Old English or Middle English we do not know) historical *x was realised as a velar after back vowels and a palatal after front vowels. This distinction (or at least some distinction, and this is the most likely) can be justified on the basis of the fact that *x after front vowels tends to delete, whereas after back vowels it has the alternative option of becoming [f] (*dwarf* < OE *dwearh*, etc.). Also diphthongisations before

roōen TO ROW < OE *rōwan*; [ʏ]: *daōes* DAYS < OE *dagas*.²⁸ A system that is prodigal in one direction is likely to be so in the other — prodigality is a fundamental design style.²⁹

Within the framework of the ‘doctrine of *littera*’ this substitutive praxis is not ‘classical’ but revisionist. According to Donatus’ definition, each *littera* has a *potestas*, as inseparable from it as its name and shape (*accidunt cuique littera tria ...*). Just as the *potestas* is a local property (‘accident’) of the *littera*, so each *littera* would seem to be appropriately connected with just one *potestas*; at least nothing in the text appears to grant a licence for multiple representation. Certainly according to Quintilian — and we would imagine, three centuries later, to Donatus — a foundational principle of the doctrine was the univocal binding of *figura* and *potestas* in a single universal unit. So allowing one *potestas* to be represented by multiple *litterae* would be a violation. This does not vitiate our use of the doctrine of *littera* as a hermeneutic device; but it suggests that the medieval notion is different in major ways from the late antique one.

2.3.3 Why potestatic interpretation is necessary

Since all our Middle English witnesses are written texts, why should we attempt to assign *potestates* to *litterae* at all? It would seem on the face of it that the obvious strategy would be to map literal representations, and construct an orthographic dialect geography and history of early Middle English. This would appear to be a ‘safe’ strategy, as it would require minimal use of inference, and give an accurate picture of the distribution of forms.

But would it? The very possibility of literal substitution raises some serious problems, as does the proposed restriction to spelling. Since we are interested in *a language*, and spellings are spellings of words and morphemes, it would be arbitrary and unnecessarily constraining (and information-destroying) not to consider also the ‘reality’ or phonetic substance that underlies the spellings. Language is phonic before it is orthographical; and without some substance there would be nothing for the LSSs to represent.³⁰ And once the notion of *littera* is accepted, *potestas* (except in the special case discussed below in §2.3.5) is indissolubly bound to it. In fact as we will see even the simplest taxonomic judgements (e.g. the grouping of spelling types) depend ultimately on phonetic judgements.

The statement in *LALME* (vol. 1, 6) that the maps constitute ‘a dialect atlas of written Middle English’, and that texts are ‘treated as examples of a system of written language in its own right’ is often misinterpreted. The emphasis on the independent

suspected palatals and velars are different: e.g. *ei* in *feighten* FIGHT in some varieties vs *ou* in *doughter* DAUGHTER, < OE *feohtan*, *dohtor*.

²⁸ On the face of it we might assume that ‘ō’ simply represents [j] in all these cases. The reason we do not is that on historical grounds it maps to segments that we believe had the sounds in question, and we do not have evidence for sound changes that would for instance map [y] to [j] in back vowel environments as in DAYS. But there is no systematic reason why ‘ō’ should not have been used for all these *potestates*. It is also possible that at least some of them, in this particular idiolect, may have represented variant pronunciations: e.g. what we represent as [y] in *daōes* DAYS < OE *dagas* may very well have been [w], as it eventually became everywhere.

²⁹ For further discussion of Trinity and other prodigal hands see Laing 1999.

³⁰ This was already understood in classical times. Cf. Quintilian, *Institutiones* I: ‘Hic enim est usus litterarum, ut custodiant voces, et velut depositum reddant legentibus’. [This indeed is the use of letters, that they should guard the sounds, and like something given into their care return them to the readers.]

value of written evidence was particularly apposite two decades ago, given the post-Bloomfieldian view that was current then (and to a large extent still is) that writing is of no independent linguistic interest, but merely ‘parasitic on’ speech. But this must not be misunderstood and taken to imply that phonological interpretation is *per se* unnecessary. The *LALME* editors take no such line. They were fully aware of the potential phonological implications of their data. *LALME* is rich in phonological commentary, while the series of Dot Maps (vol. 1) crucially depends on acknowledging the relationship between sound and symbol.

The history of a language cannot be restricted to its orthography. We take spelling with the utmost seriousness, in no sense ‘merely’ as indicative of phonology; but we take phonology equally seriously. *LAEME* is not an atlas of early Middle English orthographic forms, but an atlas (like *LALME*) of both first-order data and the second-order but equally important information deducible, or otherwise arguable, from it. The history it portrays is that of orthography, phonology and their interactions. The inclusion of both is logically necessary: we could not assume that any two orthographic objects represented ‘the same’ item unless we assumed a system-level rather than utterance-level entity lying behind them, and tying them together as *sames* (see also §2.4.2).

But beyond these general linguistic considerations are specifically dialectological ones. In the *LAEME* materials we find frequent surface irregularity and lack of directional or graded variation of spellings across space. There is an apparent mismatch between this and the uniformitarian imperative (except in the face of clear disconfirmatory evidence) that any cluster of adjacent cognate dialects forms a continuum.³¹ One way of extracting the expected underlying continuum from the apparent scatter of divergent forms is by reducing the ‘free’ distribution of many spellings to a smaller number of types, and then tracking these across space.

Consider for instance a sample of the attested spellings for *SUCH* in the *LAEME* corpus: *swylk*, *swilke*, *swilk*, *suylk*, *suilke*, *sulk*, *swiche*, *suich*, *swyche*, *sweche*, *swoche*, *suche*, *soche*, *soch*, *sich*, *seche*. Are these all ‘individuals’, or do they reduce to a smaller set of ‘types’? Entirely by eye, invoking no phonetic intuitions they could be reduced to a certain extent: *-ch-* vs *-lk-* and *sw-* vs *su-* vs *so-* vs *si-* vs *se*. But the maximally parsimonious sorting will also invoke assumptions about what the spellings represent phonetically. E.g. it seems reasonable to take the *sw-*, *su-* forms as beginning with [sw] and all those with initial *s-* + V as not having a following [w]. It is reasonable to assume that those ending in *-k* have a final velar (which typically correlates with initial *sw-*), and those in *-ch* have a final palatal. In effect we can subdivide the attested material into more abstract types: *swVch* vs *swVlk* as against *sVch*. Further, we can group nuclear *i*, *y* together, and *e*, *o*, *u* as constituting different groups. We can take nuclear *u*, *o* as being more closely related by virtue of their being back vowels, though given the presence of an LSS {‘u’, ‘o’} in some text languages they may both represent [u]. By any criteria, *e* would be something quite different. It is these phonetically based representations that serve to identify the continuum, which tends to be occluded if the set of attested spellings is too rich. The richness may be made more tractable by subdivision according to phonetic as well as orthographic criteria.

Potestas then is an ‘accident’ of *littera* only in the technical Latin grammatical sense ‘something that belongs to’ it (cf. Lewis and Short 1879 s.v. *ac-cidō* B.3); it is an essential and inseparable part of the concept and its interpretation.

³¹ For further discussion see Laing and Lass 2003.

2.3.4 Litteral or potestatic substitution?

Since spelling ultimately has some relation, however indirect, with sound substance, considerations of litteral substitution often raise potestatic questions. Some of the choices in one of the Trinity hands illustrated in §2.3.2 above will make this clear, and suggest the kinds of knowledge that have to be brought to bear on matters of litteral interpretation. First we should explain ‘st’ as a representation for historical *-ht* (expected synchronic [xt]). This is a choice found, beside other representations, in three of the hands, and it might seem phonologically odd. The compound littera³² ‘st’ is also used in this text for historical *-st* (e.g. in *beste* BEST); its use for *-ht* however does not reflect a change [xt] > [st] in English, but is an inverse spelling based on a French change [st] > [xt ~ çt ~ ht] (Pope 1934: §§1178(ii), 1216).³³

There is also nothing problematic about ‘ct’, ‘cht’, or ‘ōt’: they are to be read in the first instance as some obstruent before [t], and by various historical arguments (see §2.4.1) as more precisely something in the palatal or velar region. The litteral sequences ‘ct’ and ‘cht’ have a history going back to early Old English, and form part of a series that later resolved primarily to ‘ht’, which indeed remained in many early Middle English text languages. But other substitutions are not that straightforward, and may have potestatic implications as well. Difficulties arise here with ‘t’, ‘tt’, ‘th’ and ‘tht’. The following are the reflex-sets of four Old English items in Trinity Hand D (tagging is omitted as only the root phonology is at issue; capitalisation is as in the manuscript):

ēht PROPERTY: Acte 1, hachte 1, achte 2, haite 1
dryhten LORD: drichen 1, dristin 2, driōten1, Drittin 1
cniht KNIGHT: cniches 1, cnith 1, cnit 1, Cnites 1
liht LIGHT: litht 1

We group ‘t’ and ‘tt’ together, since they appear to show absence of the reflex of OE *h*: at least we assume they do because this fricative was lost eventually, and these spellings could easily be interpreted as early variants showing loss. That is, while ‘ct’ and ‘cht’ may be interpreted as [xt],³⁴ the general usage of this writer (and others), and the history, suggest that ‘t’ or ‘tt’ are better taken as lacking the preceding

³² *Litterae* may be either simple or compound. Simple *litterae* represent one segment: ‘t’, ‘h’, ‘c’, etc. Compound *litterae* use two or more simple *litterae* for a single segment — almost like a separate *littera* in its own right: ‘th’, ‘ch’, ‘ij’ where <j> is a *figura* of ‘i’). A simple *littera* may also be used for a chain of *potestates*, as in Oxford Merton 248 where <yng> = THING and presumably ‘y’ therefore stands for the sequence [θi].

³³ This justification of ‘st’ also allows us a possible way of dealing with the apparently extraordinary ‘ft’ (only in *nitf* NIGHT, once). This can be seen as a complex error, an inversion of expected ‘ft’, where ‘f’ is either a ‘default fricative’ writing (on the principle that if the reader knows that some fricative is meant he will be able to figure out which one), or inversion plus a miswriting of ‘f’ for ‘s’. This error is easy to account for as in most early Middle English scripts ‘f’ and long ‘s’ are identical except for the cross stroke of ‘f’. The error is more often evidenced in the accidental omission of a cross stroke giving ‘s’ for required ‘f’.

³⁴ Although, note also <kt> for this segment in some early Middle English spelling systems, which suggests the possibility that in some reflexes of OE *-ht* the fricative may have become a stop. All spellings must be interpreted in relation to others within the individual spelling system. From system to system, the same spellings may represent different sounds, just as different spellings may represent the same sounds.

fricative.³⁵ The other problematic cases are ‘th’ and ‘tht’. On the face of it, they might appear to represent [θ(t)]; is there evidence for this? There is no doubt that *-ht* can become [θ]; this is attested in modern Scots at least.³⁶ So the development itself is not outlandish, though its current regional restriction makes it somewhat problematic to attribute it to an early Middle English SW Midlands text. But an examination of Hand D’s system of writing makes the interpretation of ‘th’ as [θ] more doubtful. There are six examples of ‘th’ in this text, all representing historic *-t-* (*sitthest* SITTEST < *sittest*), or *-ht* (*cnith* KNIGHT < *cniht*, *mitht* MIGHT < *miht*, *liht* LIGHT < *liht*, *noth* NOT = NOUGHT < *nāwiht* > *nāht*, *sothede* SOT-HOOD, FOLLY < OF *sot*). There are no initial ‘th’ for OE *þ* (hand D’s system has ‘þ’ throughout); and postvocally in stressed syllables there are 20 examples of ‘þ’. So from a systematic point of view it does not look as if ‘th’ is a likely writing for [θ] in this text. This emphasises the important point that spellings cannot be looked at in isolation: one must consider the system employed in a particular text language before making interpretive judgements.³⁷

2.3.5 Zero

It should be clear now that our interpretation of the litteral praxis of our scribes is not intended to be purely synchronic. What we take as the content or import of *litterae* is often etymological. This is because it is often not possible to produce the kind of synchronic analyses of our text languages that would allow for an implementation of the Saussurean dichotomy — even if we considered it desirable (see the discussions in Lass 1997: chapter 1, Williamson 2004). The most convenient descriptions of the import of particular *litterae* are often historical: we do not know precisely what ‘cht’ in Trinity D means synchronically, but ‘OE *-ht*’ is a useful and accurate label for it, and one that is relevant to the entire subject matter of *LAEME*.

One case where this essentially historical analysis becomes particularly relevant is in dealing with ‘zero elements’; it is not really possible to characterise these intelligibly except on historical grounds. Zero is not an element in classical theory of *littera*; we introduce it here because both zero *litterae* and zero *potestates* are well attested in our materials, and their use was familiar and apparently unproblematic for our scribes.

A simple case of zero *potestas* is ‘excrecent *h*’, the writing of ‘h’ in places where history tells us that it should not have had potestatic import.³⁸ Two examples from Trinity Hand D again shown as reflexes of OE forms:

³⁵ Given the general history of English, it is unlikely that ‘tt’ stands for a geminate, since this strategy of compensatory lengthening after preceding fricative loss is typically Scandinavian, not West Germanic: cf. Sw *natt* vs E *night*.

³⁶ *LAS* III records [θ] for *-ht* in DAUGHTER, MIGHT in Moray, Banff, Aberdeenshire, Kincardine and Angus, and in MIGHT only in one location in Fife. We are grateful to Keith Williamson for this reference.

³⁷ The significance of the usage of individual texts is emphasised by the form *biþoutthe* BETHOUGHT in Trinity Hand A; here the <ou> suggests diphthongisation, which would presuppose the presence (at some historical point) of [xt]. It is not decidable whether this means that <tth> = [t], or whether the spelling represents an earlier stage in the development of this lect (or that of the exemplar) before deletion of [x].

³⁸ Except possibly as a spoken hypercorrection: cf. *My Fair Lady*’s ‘urricanes arldly hever appen’. We have, however, no direct means of knowing where or whether written hypercorrections in early Middle English had their counterparts in speech.

ǣht PROPERTY: Acte 1, hachte 1, achte 2, haite 1
ic I: ich 7, ic 5, Hi 1 Hic 1, I, 1, hich 1, y 1

For the category OE word-initial vowel (*#V-*) the LSS is then {0, 'h'}, but the PSS may very probably be {0}. The reason we can infer this is the existence of its inverse, e.g.

hæbbe HAVE: abbe 2, haue 1
hit IT: hit 8, it 7, hid 1

In these cases the 'h' is etymological, so that forms without it represent what we could call a zero *littera*. The LSS for OE *h-* then is {0, 'h'}, and the PSS is {0, [h]}. Of course in one sense the use of a *littera* with zero *potestas* as above is simply what is known as 'inverse spelling'; the reason we want to rename it for the moment is to make it coherent with the rest of what appears to be a reasonable model for scribal praxis.

2.3.6 Substance and structure: why we choose *littera* as our basic unit

The theory of *littera* enables us to bypass a question that for our purposes is vexatious rather than informative: whether we intend our representations to be 'phonemic', 'phonetic', 'abstract' or anything else within contemporary metalanguage. This issue is also avoided by the handbook tradition with its 'uninformative' italics.

Of course, we do not know in precise phonetic detail what any historical spellings represent; but in terms of the parameters of *LAEME*, this question is largely irrelevant. A representation of a Middle English form in [] is an indication of phonetic substance at some level, but not of status within a system in any structuralist (including generativist) sense. As indicated above, systemic status rather is relevant within the confines of specific text languages. We disavow any attempt to set up an overall pattern or system for 'Middle English in general'. In the first place we do not think it can be done, and in the second our aims are comparative and variationist, not reductive and generalist. As dialectologists we are interested in description and comparison and as linguistic historians we have to work within the constraints of the data that survive. That does not mean that we eschew interpretation, extrapolation or theoretical argumentation. Simply, there is no need to make a commitment to any particular theory of system structure in an endeavour such as *LAEME*. *LAEME* provides the tools to test and to create such theories. The option for producing a structural analysis of any kind is always open to the user. Indeed, if such analyses turn out to be possible or useful, the corpus provides precisely the kind of material one could use to make them.³⁹

In *LAEME*, we are concerned with word- or affix-histories and distributions.⁴⁰ Whether a given symbol is to be taken as representing something 'emic' or 'etic' is a

³⁹ For a detailed argument to the effect that 'structuralist' systems in any sense may not be proper objects of linguistic enquiry, and do not correspond to the way speakers store and learn language, see Bybee 2001. The core of her argument is that the proper phonological prime is the stored pronunciation of an individual word; her model accounts more elegantly for the kind of variation we actually find in written texts and modern spoken languages than models that abstract away from variation to an invariance putatively 'underlying' it.

⁴⁰ This is a function of the fact that our primary evidence for geographical placement is the *assemblage*, not the system. See chapter 1, §§1.1 and 1.5.

consideration that belongs to a different kind of discourse. The previous discussion of substitution has already indicated why formal structural analysis based on distinctiveness is not appropriate for application to most of the text languages in the *LAEME* corpus. Many of these languages, as the previous discussion of substitution suggests, have writing systems that display two distinctly non-structuralist properties: (a) their creators are not particularly concerned with biunique grapheme/phoneme mapping; and (b) the orthographies (and frequently the languages they represent) are highly variable. The evidence for structuralist ‘eme’ systems is too weak to make it sensible for us to use this kind of modelling. Distinctiveness in the usual sense does not seem to be a high priority, which makes techniques like commutation nearly useless.

A case in point is Trinity Hand A. For taxonomic convenience we set up a typological inventory (based on the handbook accounts of ‘the Middle English sound system at large’) from which the *potestates* represented by any early Middle English orthography would be likely to have been drawn.⁴¹

| Short | | Long | | Diphthongal |
|-------|---|------|----|----------------------|
| i | u | i: | u: | iu au eu ou ai oi ui |
| e | o | e: | o: | |
| a | | ɛ: | ɔ: | |
| | | a: | | |

Here we differentiate between ‘inventory’ as the set of available sound types and ‘system’ as the deployment of a subset of these types in the phonology of a language.

The total inventory of *litterae* that occur in Hand A’s output in what we assume are accented syllables is:

‘a’, ‘e’, ‘i’, ‘o’, ‘u’
‘ai’, ‘au’, ‘ay’, ‘ea’, ‘ei’, ‘eo’, ‘eu’, ‘ey’, ‘ie’, ‘oe’, ‘oi’, ‘oo’, ‘ou’, ‘ui’

These nineteen *litterae* happen to be the same numerically as the number of vowel types one would expect in a maximal system; but the arrangement seems odd. We might expect 12 monophthongs and 7 diphthongs. However, if (as is usual in Middle English vowel orthographies) length is not represented and ‘open’ and ‘close’ long mid vowels are not distinguished, the 12 monophthongs would be represented by just the first five *litterae* listed in our inventory.⁴² We then have an expectation of 7 possible diphthongs with 15 digraphs actually occurring in Hand A’s output. The problem then is matching all 19 vowel-type representations to the likely potestatic

⁴¹ This inventory is modified from Lass (1992: 47–50). We assume, following the arguments in Lass and Laing 2005, that there would not have been any front rounded vowels in early Middle English, e.g. the reflex of *ēo* would have been [e:] or [o:]. For more theoretical discussion of the question of specifying *potestates* see §2.4 below.

⁴² Note that this commonplace expectation is already a violation of any neat phoneme/grapheme correspondence.

inventory. Some cases are relatively simple. E.g. ‘i’ appears in many forms where we would expect it, like (*h*)*ic* I, *min(e)* MY/MINE. On the basis of standard protocols (see §2.4.1 below) we can assign the potestates [i], [i:] respectively.⁴³ And in general throughout this text wherever we find reflexes of OE *i* or *ī* they will be represented by ‘i’.

But now consider the representation of stressed nuclei in the following items from this text language:

| <i>Item</i> | <i>LSS</i> | <i>Source</i> | <i>Potestas</i> |
|-----------------|----------------------------------|----------------|----------------------------|
| FOURTEEN | {‘oi’} | OE <i>ēo</i> | [e:] or [o:] ⁴⁴ |
| CROSS | {‘oi’} | OF <i>oi/o</i> | [oi] or [o] |
| <i>cū</i> KNOWN | {‘oi’} | OE <i>ū</i> | [u:] |
| FORSOOTH | {‘oi’} | OE <i>ō</i> | [o:] |
| GHOST | {‘oi’, ‘o’, ‘a’} | OE <i>ā</i> | [ɔ:]/[a:] |
| GOD | {‘o’, ‘oi’} | OE <i>o</i> | [o] |
| GOLD | {‘o’, ‘oi’} | OE <i>o</i> | [o(:)] ⁴⁵ |
| GOOD | {‘o’, ‘oi’, ‘ohi’} ⁴⁶ | OE <i>ō</i> | [o:] |
| OATH | {‘oi’} | OE <i>ā</i> | [ɔ:] |
| PUT | {‘oi’, ‘u’} | OE <i>u</i> | [u] |
| SOUND (adj) | {‘oi’} | OE <i>u</i> | [u(:)] |
| WRING (pt sg) | {‘oi’} | OE <i>aNC</i> | [o] |
| WROTH | {‘oi’} | OE <i>ā</i> | [ɔ:] |

LSSs containing both simplex litterae and digraphs are frequent in this text: e.g. the 3rd person plural genitive pronoun THEIR is represented with nuclear {‘e’, ‘ei’, ‘eo’, ‘uo’}. If we take one member of this set — ‘ei’, historical evidence suggests that there should be at least two or perhaps three *potestates* associated with this *littera*: [ai] as in *awei* AWAY, [e:] as in *feit* FEET, and whatever ‘ei’ might be presumed to stand for in unstressed syllables, as in *afteir* AFTER (see note 53 below). It is therefore clear that digraphs need not represent diphthongs (though they may); and that there is no way that sense can be made of a system like this from the inside except (a) historically, and (b) in terms of literal substitution. The scribe’s eschewing of biuniqueness makes a standard synchronic phonemic analysis impossible: there are too many overlaps and failures to distinguish what we know independently to be etymologically distinct and unmerged categories for such an analysis. Hand A’s orthographic system has a number of LSSs for the nuclear vowels of particular lexemes, and only that; assignment of potestatic representations is at least partially an etymological act.

⁴³ Short [i] might arguably better be represented as [ɪ], but this is controversial (see Lass 1999a: §§2.1.2.1, 2.4.1.3), and we think this fineness of resolution is unattainable for early Middle English. Let [i] stand for ‘the highest short front vowel’, regardless of precise quality.

⁴⁴ It is probably more likely that [o:] is represented, since ‘oi’ seems otherwise to be used exclusively for back vowels; but both are possibilities: see note 40 above and Lass and Laing 2005.

⁴⁵ The vowel would be long if homorganic lengthening had occurred; but this is not decidable on internal evidence. The same applies to SOUND below.

⁴⁶ This scribe writes excrescent ‘h’ so frequently that it is likely he did not pronounce [h] at all; the spelling *gohid* then for GOOD might very well have a purely ‘decorative’ ‘h’, so that ‘ohi’ is potestatically equivalent to ‘oi’. Cf. also his *fehīd*, *fehīt* (~ *feid*, *feit*) for FEET.

Such orthographies are not based on ‘structural’ analyses of a lect (quite unlike the First Grammarian’s earlier analysis of Icelandic, which is essentially based on minimal pairs). Rather they are complex systems whose inventors apparently lack certain interests that they ‘ought to have’. Hand A writes a language that, with some knowledge of both prior and subsequent linguistic history, we can interpret perfectly well. But it is not clear on orthographic evidence, which is our primary data, what the internal structure of its sound system is.⁴⁷ We cannot know from the evidence of the LSSs themselves whether any two or more etymological categories have merged or have remained separate. It should be clear at this point why we are interested in the histories of forms and of individual systems, not in an ‘overall system’.

Before we produce a more detailed model of litteral praxis, it is important to scrutinise the theoretical framework that enables us to assign potestatic values to written materials from the past.

2.4 Potestatic interpretation

2.4.1 Protocols for potestatic interpretation

How do we assign a credible phonetic value to an orthographic string from our period? We have so far taken the ‘canonical’ handbook representations shown in the inventories above as unproblematical. But the principles involved are not without difficulties, and these deserve to be made explicit. We will take as an example a token spelled *niht* and clearly meaning NIGHT in an early Middle English text. What are our sources of information as to its likely phonetic representation?

(i) Comparative evidence from ancient Indo-European. Forms like L *noct-*, Gr *núkt-*, etc. imply (given our background knowledge of Indo-European comparative linguistics) a distant reconstruction **nokt-*. This already suggests that the shape of the root in our form ought to be CVCC. We can also use descriptive phonetic evidence from those ancient languages that have such a tradition, e.g. Greek and Latin (cf. Allen 1965: 14–16). There is comparative evidence from modern Gmc as well: G *Nacht*, Du *nacht* cohere with the suggested ancient Indo-European forms having undergone spirantisation of their penultimate consonant. Sw *natt* [nat:], ModE *night* [nait] suggest further major changes: here deletion of the segment represented as [k] in Latin and [x] in German, and compensatory lengthening of the final consonant in Swedish, and lengthening and diphthongisation of the root vowel in English.

(ii) The history of forms within Old English. We find among others *naeht*, *neaht* and *niht*; further exploration of Old English sound changes suggests that the original Germanic vowel was **a*, with later changes involving fronting to **æ*, ‘breaking’ to *ea*,

⁴⁷ Within the system there may indeed be more than one set of protocols for mapping symbol to sound. Extensive variability in a lect can result from different degrees of ‘archaism’ and innovativeness in unstable coexistence resulting in a variable system. This does not, however, imply a mixed system. Such internal variation in a scribe’s usage may genuinely represent his active repertoire. In practice, with copied texts, this variation is very often impossible to separate from the results of constrained selection (see §1.5.6 and references there cited), but such objects can still be considered single systems. Genuinely mixed systems, those of *Mischsprachen*, are statically layered rather than variable (though there may of course be variation within the static layers).

and raising eventually to *i*, giving a high vowel in English whereas all other Germanic dialects have a low vowel.

(iii) Later developments in English. Spelling, rhyme practice and phonetic description from the 16th–17th centuries suggest coexistence of types with medial [x], [ç] or [h] (generally preceded by short vowels), and forms with a long vowel but no consonant before the final [t] (see Lass 1999a: 116–18 for details). We have here a common type of trajectory, in which a consonant weakens from stop (Indo-European) to oral fricative, to [h], and then to zero.

(iv) Modern dialect evidence. Forms of ‘night’ with a nonanterior fricative exist in Scots, e.g. [nixt]; unless we want to propose something as odd as late [x]-epenthesis, which the historical evidence in any case does not support, we must interpret this as a survival.

Historical reasoning is typically not ‘linear’ but ‘reticulate’. We make claims on the basis of convergence or ‘consilience’ of many different arguments from different temporal strata and theoretical positions. In the case above we use evidence from ancient forms, subsequent forms, and the period under investigation. This argument has been a particularly simple one. Characterisation of vowels, for instance, may be more complex, involving detailed comparative reconstruction, judgements on the range of likely *potestates* of Latin and Greek *litterae*, the behaviour of Latin and Greek loans in Germanic and studies of 16th–17th-century orthoepic testimony. We will not go into any further examples: our judgements are based on general consideration of the available evidence and the long tradition of scholarship that has led to consensus.

2.4.2. Level of resolution

Assuming that we need some kind of phonetic representation in order to do our work, the Hard Question remains: how fine should our phonetic resolution be? In our representation of historical categories and *potestates* we have used conventional IPA symbols broadly, as signs for ‘ranges’ of phonetic quality. This is necessary in historical reconstruction, where we do not have live informants. There are two polar approaches to this problem in the literature. One is the ‘hard structuralist’ assumption that all that counts are systems of oppositions, and that phonetic value means little. The other is a strongly realist view that would have the outputs of reconstruction be something close to ‘pronunciations’. The second may be desirable but is unattainable. The first is untenable, not only for the reasons already demonstrated above, but also for the reason argued below.

What do we actually mean by the phonetic symbols we use? We could be phonetically agnostic and say simply that they stand for points of opposition in a contrast space. This way we assign them nothing but ‘mnemonic’ phonetic value.⁴⁸ So for mnemonic convenience we might say that Old English had two vowel sets:

| | |
|---|---|
| i | u |
| e | o |
| æ | a |

⁴⁸ Cf. Jakobson and Halle, 1965:11: ‘all phonemes denote nothing but mere otherness’.

and call them ‘front’ vs. ‘back’, but not mean anything more (despite the mnemonic transparency) than if we wrote:

| | |
|------|------|
| Fred | Sue |
| Joe | May |
| Tom | Jean |

We could call these two sets ‘men’ vs ‘women’, and still capture the insight that there are two groups of entities, and each set has some major feature in common which distinguishes it from the other. And yet as a matter of historical fact there is a well-known major ‘sound change’ (or ‘sex change’) in which Women vowels become Men vowels if followed by Fred in the next syllable. (This is an austere ‘classical’ structuralist version of *i*-umlaut.) If we contrast this statement with the phonetically specific alternative ‘Back vowels become Front if followed by [i] in the next syllable’, this, along with the claim that every distinct phonological string represents a real word, may be sufficient to show that *some* phonetic specification is necessary. Without it, history becomes arbitrary and no longer ‘naturalistic’; e.g. assimilation is no more transparent than any arbitrary (even cross-linguistically unattested) change.

Somewhere between [Fred] and [i] (= ‘precisely the value of Cardinal Vowel 1’) is the domain in which we operate in historical reconstruction. In practice we limit both the number of phonetic parameters and the number of points on each parameter to a ‘safe’ minimum. This allows for naturalistic specification at a fineness of detail adequate for our purposes and historically insightful, which is in keeping with what we know of both historical input and present development — as well as with what we know of language history in general. Our experience over decades of work on the history of English suggests that for our period and its predecessors, we can retrieve distinctions of only a limited degree of fineness, but with a certain confidence.

The level of phonetic representation we choose might best be called ‘**poorly resolved broad transcription**’ (cf. Laing and Lass 2003). This is, we think, the right way to represent most historical sound substance. We operate on the assumption that our reconstructions are well enough supported so that if a responsible phonetician equipped with a time machine were able to hear the items represented, the symbol in question would be a reasonable transcriptional response. This is partly standard wishful thinking, and partly our positive assessment of the results of work in comparative and historical linguistics over the past two centuries. For instance, we would be very much surprised if what we choose to represent as [me:] (the oblique form of the 1 pers sg pronoun), would have anything other than a labial nasal in first position and a mid front vowel (probably high mid) in second position.⁴⁹ So each segmental representation is fuzzy; it is equivalent to a penumbral range around the denotational centre of the equivalent IPA symbol. Therefore the use of [] in the maps,

⁴⁹ This is a simple case. In more complex situations, our symbols may have greater latitude: [æ] represents a ‘low front vowel’, but how low (especially given the range of phone-types in modern English typically transcribed as [æ]) is not resolvable. We use the IPA alphabet throughout, with one exception: [a] stands not for a low front vowel but simply ‘a low unrounded vowel that is not [æ]’. The arguments in Lass 1976: chapter 4 to the effect that ME *a* was a front vowel have not worn well; the comparative base for that reconstruction was too constrained to yield as clear and univocal a result as it seemed to. In any case nothing particular hangs on the quality of the short low vowel — unlike the long one, which judging from its history in our period and later must almost surely have been back. We use [a] to avoid the overfine distinctions that would be implied by distinguishing [a, ɐ, A]). It is not possible to specify whether [a] is back or central — though we do not believe that ‘central’ is a distinctive vowel category at any point in Old or Middle English (Lass 1999a: 87–91, 133–35).

etymologies and statements of changes is essentially ‘typological’ rather than ‘phonetic’ in any more highly resolved sense. It is a representation of ‘sound substance’ at an unspecified but undoubtedly coarse level.

2.5 Modelling *litteral* praxis

What implications does the theory of *littera* have for modelling the writing process in early Middle English scribes? It is clear from the discussion above that any adequate model would look rather different from one based on standard phoneme/grapheme mappings. We also suspect that its basic construal would be different. Writing in early Middle English is a procedure that requires more intervention from the agent than the (apparently) unproblematic writing behaviour of modern standard-language literates. We do not (except rarely) make orthographic *choices* when we write; the majority of our scribes, judging from their substitutional habits, appear to do precisely that. We cannot of course reconstruct their individual motivations for particular choices, but we can model their behaviour based on the phenomenology of their texts.⁵⁰ This notion of choice holds even when (as may be the majority case) the scribe is copying an exemplar not his own. As the well motivated distinction between literatim and translating scribes makes clear, every act of writing represents a choice point. The choices, however, are more constrained for the literatim scribe though they may well not have been for the writer of his exemplar.

There are two aspects to writing: it is an action, but it is also a system or set of protocols. In other words, before he actually writes, the scribe asks:⁵¹ How shall I represent this linguistic object? What are my options? Which strategy type (economical or profligate) shall I use? How constrained am I by pre-existing design? What is the difference in the range of choices I have if I am copying as opposed to composing? We can assume that in most cases an author would not produce a fair copy in the first instance. His rough copy might be the exemplar for himself or for some other scribe; we rarely know which. Orm, however, gives us considerable insight into authorial scribal choices and how they may change through time. We would presume that Orm was less constrained (except by his own orthographic theory) than copyists of other people’s work, because everything was coming out of his head. He was not working with an exemplar in the strict sense. If he had a rough copy at any stage in his work, his editing processes on it are lost, and the transitions between it and his fair copy are now invisible. But his habit of continuing to work on the copy of the *Ormulum* that survives, and the layered corrections it evidences, illustrate how even self imposed constraints can evolve.⁵²

We suggest that a reasonable model for general scribal practice is defined by a ‘scribal lexicon’; this consists of a set of PSS and LSS listings, and a set of word and affix templates. The general listings would look like this:⁵³

1. For each *potestas* there is a canonical LSS, e.g. [o:] <=> {‘o’, ‘oi’, ‘ohi’}

⁵⁰ For detailed discussion of particular cases see Laing 2004.

⁵¹ What follows is not to be taken literally. It is an attempt at the imaginative reconstruction of a conscious process.

⁵² See further Nils-Lennart Johannesson: <http://www.english.su.se/nlj/ormproj/ormulum.htm>.

⁵³ Note that <=> below indicates bidirectional mapping: neither *littera* nor *potestas* is to be taken as having priority.

2. For each *littera* there is a canonical PSS, e.g. ‘oi’ \Leftrightarrow {[o], [o:], [ɔ:], [u:], [u]}

The inventory of these constitutes the superstructure of the lexicon, the material out of which word and affix shapes may be constructed.

Subsets of the canonical LSSs and PSSs are then associated *with particular forms*. This is an important point: for most of our sources there is at least some degree of lexical specificity, such that for a given *potestas* in a given form only a subset (most often but not always a proper subset) of a given LSS is available. We would then conceive typical lexical representations (here from Trinity Hand A) as being of this type:

GOOD n.

[g] \Leftrightarrow {‘g’}
[o:] \Leftrightarrow {‘o’, ‘oi’, ‘ohi’}
[d] \Leftrightarrow {‘d’}
#

We include # ‘word terminus’ because certain substitutions are limited to positions adjacent to # (see below); in complex words we include + ‘formative boundary’, which may control processes such as syncope in verbal endings. An example illustrating the role of # would be:

FEET

[f] \Leftrightarrow {‘f’}
[e:] \Leftrightarrow {‘ei’, ‘ehi’}
[t] \Leftrightarrow {‘t’, ‘d’}
#

<

Another important phenomenon has to be accounted for: this is zero (see the discussion under §2.3.5 above). Aside from mere absence, zero may be manifested as either excrescent ‘h’ or ‘otiose’ final ‘e’. Zero for etymological ‘h’ can be represented in the model so far, by simply having zero as a member of the LSS for initial [h]. Many forms have a final ‘e’ that is not etymologically justified, but is presumably ‘decorative’: e.g. in Trinity Hand A (which has fewer than many other texts) in nominative *tuelue* TWELVE, *breste* BREAST < OE *twelf*, *brēost*. Below are models for words with excrescent ‘h’ and otiose final ‘e’, where empty [] indicates a *potestas*-free but *littera*-accepting slot:

OATH

[]⁵⁴ \Leftrightarrow {‘h’}
[ɔ:] \Leftrightarrow {‘oi’}
[t] \Leftrightarrow {‘t’}⁵⁵

⁵⁴ We leave the first potestatic slot null under the default assumption that this scribe was a total non-pronouncer of ‘h’. This is of course not necessarily the case: [h] and zero could be variant members of a PSS. Either situation could yield the same orthographic results.

#

BREAST

#

[b] <=> {'b'}

[r] <=> {'r'}

[e(:)] <=> {'e'}⁵⁶

[s] <=> {'s'}

[t] <=> {'t'}

[] <=> {'e'}

#

In both these examples the forms represented are the only ones in the text; to illustrate just how complex the substitutions in both zero and filled positions can be in a single word, consider the spellings of AFTER (adverb, conjunction and preposition) also from Trinity Hand A:

hefter, Afteir, afeir, aftir, aster, efter, -hefteir, -hefter, -hester

Here we for the first time we encounter multiple *potestates* in a single slot:

#

{0} <=> {0, 'h'}

{{a}, [e]} <=> {'a', 'e'}⁵⁷

{{f]} <=> {'f', 's'}

{{t]} <=> {'t'}

{{e]} <=> {'e', 'i', 'ei'}⁵⁸

{{r]} <=> {'r'}

#

So the model for litteral praxis consists of the canonical *littera/potestas* mappings plus a set of individual word templates. It is not intended to be a fully developed formal model in the usual sense; rather it is a ‘diagram’ — a procedural layout driven by the actual data presented to us by our text witnesses.

This chapter has provided the background for interpreting the orthography of Middle English forms and demonstrates the necessity of phonetic reconstruction within the *LAEME* framework. In the next chapter we describe the corpus of forms itself.

⁵⁵ It might seem curious that {[t], [θ]} is not given as the canonical PSS for the final segment. Hand A uses the littera ‘†’ for OE *ʰ* initially and medially but never finally. The presumption is that in final position it has become a stop.

⁵⁶ It is uncertain whether pre-cluster shortening has taken place in this form.

⁵⁷ In a SWML text (like this one) the presence of ‘e’ for OE *æ* could well be a reflex of second fronting. The presence also of ‘a’ might indicate phonetic variability. This is our assumption here. It is also possible, however, that both ‘e’ and ‘a’ spellings represent different orthographic variants for unchanged [æ], but we take the traditional position that OE *æ* had merged phonetically with OE *a*, and that the quality [æ] did not arise again until the early Modern English period.

⁵⁸ The traditional assumption is that ‘e’ in unstressed syllables in Middle English represents [ɔ]; for arguments against this view, assigning the development of [ɔ] to a much later period, see Lass 1999a: 133–5.

References

- Abercrombie, D. 1965 [1949]. What is a letter? In D. Abercrombie, *Studies in phonetics and linguistics*, 76–85. London: OUP. Original version in *Lingua* 2.
- Allen, W.S. 1965. *Vox Latina*. Cambridge University Press.
- Benediktsson, H. 1972. *The first grammatical treatise*. University of Iceland Publications in Linguistics 1. Reykjavík: Institute of Nordic Linguistics.
- Benskin, M. 1991. In reply to Dr Burton. *Leeds Studies in English* n.s.22.209–62.
- Benskin, M. 1997. Texts from a township in late medieval Ireland. *Collegium Medievale* 10.91–173.
- Benskin, M. 2001. The language of the English texts. Chapter 4 of T. Hunt (Ed). *Three receptaria from medieval England* (pp 193–230). Medium Ævum Monographs, NS 21. Oxford: The Society for the Study of Medieval Languages and Literature.
- Benskin, M. and Laing, M. 1981. Translations and *Mischsprachen* in Middle English manuscripts. In: Benskin and Samuels 1981: 55–106.
- Benskin, M. and Samuels, M.L. eds. 1981. *So meny people longages and tonges: philological essays in Scots and mediæval English presented to Angus McIntosh*. Edinburgh: The Editors.
- Blake, N.F. ed. 1992. *The Cambridge history of the English language. II, 1066–1476*. Cambridge University Press.
- Bybee, J. 2001. *Phonology and language use*. Cambridge University Press.
- Dossena, M. and Lass R. eds. 2004. *Methods and Data in English Historical Dialectology*. Bern: Peter Lang.
- Hart, J. 1569. *An orthographie*. London: no publisher given.
- Jakobson, R. and Halle, M. 1965. *Fundamentals of language*. The Hague: Mouton.
- Johannesson, N-L. <http://www.english.su.se/nlj/ormproj/ormulum.htm>.
- Laing, M. 1999. Confusion *wrs* confounded: litteral substitution sets in early Middle English writing systems. *Neuphilologische Mitteilungen* 100.251–70.
- Laing, M 2004. Multidimensionality: Time, Space and Stratigraphy in Historical Dialectology. In: Dossena and Lass 2004: 49–96.
- Laing, M. and Lass, R. 2003. Tales of the 1001 nists: the phonological implications of litteral substitution sets in some thirteenth-century South-West Midland texts. *English Language and Linguistics* 7.257–278.

Laing, M. and Lass, R. 2006. Early Middle English dialectology: problems and prospects. In: A van Kemenade and B.L. Loos 2006: 417–51.

LALME = McIntosh, A., Samuels, M.L., Benskin, M. 1986. *A linguistic atlas of late mediæval English*. 4 vols. Aberdeen University Press.

LAS = Mather, J.Y., Speitel, H.H., Leslie, G.W. 1986. *The linguistic atlas of Scotland*, vol 3 Phonology. London: Croom Helm.

Lass, R. 1992. Phonology and morphology. In: Blake 1992: 23–206.

Lass, R. 1997. *Historical linguistics and language change*. Cambridge University Press.

Lass, R. 1999a. Phonology and morphology. In: R. Lass 1999b: 56–186.

Lass, R. ed. 1999b. *The Cambridge history of the English language. III, 1476–1776*. Cambridge University Press.

Lass, R. and Laing, M. 2005. Are front rounded vowels retained in West Midland Middle English? In: Ritt and Schendl 2005: 280–90.

Lewis, C.T. and Short, C. 1879. *A Latin dictionary*. Oxford: Clarendon Press.

Lyons, J. 1977. *Semantics*. 2 vols. Cambridge: Cambridge University Press.

Pope, M.K. 1934. *From Latin to Modern French with especial consideration of Anglo-Norman*. University of Manchester Press.

Prior, O.H. 1923. Remarques sur l'Anglo-Normand. *Romania* 49.161–85.

Ritt, N. and Schendl, H. eds. 2005. *Rethinking Middle English. Linguistic and literary approaches*. Frankfurt: Peter Lang Verlag.

Rothwell, W. 1968. The teaching of French in medieval England. *Modern Language Review* 63.37–46.

Rothwell, W. 1975–76. The rôle of French in thirteenth-century England. *Bulletin of the John Rylands Library* 58. 445–66.

Rothwell, W. 1978. A quelle époque a-t-on cessé de parler français en Angleterre? In *Mélanges de philologie romaine offerts à Charles Camproux*, 1075–89. Montpellier: CEO.

Rothwell, W. 1983. Language and government in medieval England. *Zeitschrift für französische Sprache und Literatur* 93.258–70.

Short, I. 1980. On bilingualism in Anglo-Norman England. *Romance Philology* 33.467–79.

Short, I. 1992. Patrons and polyglots: French literature in twelfth-century England. *Anglo-Norman Studies* 14.229–49.

Short, I. 1996. *Tam Angli quam Franci*: self-definition in Anglo-Norman England. *Anglo-Norman Studies* 18.153–75.

van Kemenade, A. and Loos B.L. (eds.) 2006. *Handbook of the History of English*. Oxford: Blackwell.

Wallis, J. 1653. *Grammatica linguae Anglicanae*. London: A. Millar.

Williamson, K. 2004. On chronicity and space(s) in historical dialectology. In: Dossena and Lass 2004: 97–136.