A LINGUISTIC ATLAS OF EARLY MIDDLE ENGLISH

INTRODUCTION

PART IV: HISTORICAL UTILITIES

CHAPTER 8

THE CORPORA OF ETYMOLOGIES AND CHANGES

Roger Lass

8. Etymology in a historical atlas

8.1 Principles

8.1.1 Introduction

Linguistic atlas databases are not customarily equipped with a full etymological apparatus. *LAEME*, however, is a rather special kind of atlas. It is specifically *historical*: change in time is projected as variation in space (and vice versa). To put it another way, every text-form or cartographical configuration is taken to be a 'still' or snapshot of an episode in a long and continuous narrative. Such a structure is not achievable without the underlying story being present in as fully realised a form as possible (cf. §2.4).

But before we describe exactly what we are doing, we provide some perspective on the term 'etymology' itself, since the evolution of its senses is complex, and even modern usage is often unclear. The earliest uses of the term refer to 'true meanings' or 'original meanings' (as in the debate in Plato's *Cratylus* on whether words have meaning by 'nature' or 'convention'). This line of thinking is developed further in mediaeval etymological theory–prototypically in Isidore's *Etymologiae*.[1]

Another common classical and post-classical usage is not so much diachronic as synchronic: the sections in Latin grammars devoted to 'etymologia' are typically concerned with derivation and word-formation. OED has English citations from the 16th century onwards illustrating their sense 1.a 'the process of tracing out and describing the elements of a word with their modifications of form and sense'.

But it is not difficult to project the notion of 'tracing' from the synchronic to the diachronic arena. By the 18th century this conversion was commonplace: OED cites Watts *Logic* I.iv.sec. 1 'The tracing of a word to its original (which is called etymology)'. This is the sense we intend here, though the notions 'tracing' and 'original' require finer specification. Crucially, our notion of etymology is processual ('tracing'), not static; and its conceptual framework includes much more than the usual postulation of distant protoforms and listing of cognates. Modern dictionaries tend to concentrate on 'origin'; the processual side, the central historical concern, is often entirely neglected.

Even in current technical usage 'etymology' has a number of overlapping senses, partly dependent on who is using the term — the practice of lexicographers is rather different from that of narrative historians, for instance.

8.1.2 'Standard' etymologies

It is important that we make clear here which senses are incorporated in our praxis, which produces objects quite unlike what one would find in a standard etymological dictionary. Consider for instance the contents of the 'etymology' section of the entry for *sit*, *v*. in the latest version of the OED. This is presented as a single block of

---

[1] This often involves the use of punning tropes built out of accidental quasi-homophones like *lucus a non lucendo* ('a grove [*luc-us*] is so called because it is not light [*luc-endo*]'). For an elaborate late mediaeval instance of this search for 'true meaning' praxis see the speculations on St Cecilia's name at the opening of Chaucer's Second Nun's Tale, *CT* VIII. This motif continues well into the 19th century, and still surfaces; for detailed treatment of the pre- and post-neogrammarian debate on etymology see Morpurgo Davies 1998, index s.v. 'etymology'. We will not be concerned with the search for 'semantic originals' here, though in principle the issue could arise: e.g. on the basis of cognate evidence, the *w*-forms of the verb *be* like *was*, etc. can be referred to an older non-copular, non-existential locative sense 'dwell, remain' (cf. Skr *vásati* 'he dwells', OIr *feiss* 'remain'). But this is beyond our time-frame.

unparagraphed text, but the information that the editors see as constituting 'the etymology' falls naturally under three distinct headings:

> I. Common Teut.: OE. *sittan* (*sæt*, *sæ´ton*, *geseten*) = OFris. *sitta* (WFris. *sitte*), MDu. *sitten*, *zitten* (Du. *zitten*), OS. *sittian*, *sittean* (MLG. and LG. *sitten*), OHG. *sizzan* , *sizzen* (G. *sitzen*); ON. and Icel. *sitja* (Norw. *sitja*, *sitta*, *sita*), MSw. *sitia*, *sittia*, Sw *sitta*; Da. *sidde*.
>
> II. Teut. type **sitjan*, for which Goth. had *sitan*.
>
> III. The stem **set*-, pre-Teut. **sed*-, is widely represented in the cognate languages as in Lith. *sedeti*, Lat. *sedēre*, Gr. *'*[illegible] (cf. *'*[illegible] seat)

Such an etymology is essentially static and ahistorical, though it does contain some minor and largely implicit elements of storytelling. What counts as the 'etymology of' an item in this standard dictionary format is mainly a list of cognates at different temporal levels.

List I gives Germanic, presented in terms of increasing filiational distance from English: Ingvaeonic (that is 'Anglo-Frisian', Low German proper, Netherlandic), High German, North Germanic. The bracketed Old English 'principal parts' define the verb implicitly as strong, class 5.

List II gives a putative protoform at the temporal level we would now call (late) Proto-Germanic. Further morphological information is given: the shape of the reconstructed infinitive suffix **-jan* tells us that though *sit* is historically a strong verb, it belongs to a special subclass, called 'weak presents' in the standard grammars. This accounts for its unexpected present tense vocalism and geminate consonant (the 'expected' infinitive would be **setan*).

List III takes us back first to an earlier stage of Proto-Germanic (**set*-, before raising and gemination induced by the following **-j*-); and finally to an Indo-European **sed*- (though it is not labelled as such), represented as usual by the *e*-grade of the root.

An OED etymology then is a list of presumed cognates, with protoforms for salient levels of temporal resolution, and a minimum of explicit historical narrative. One could of course extract from this material at least part of the story of the present stem: **sed*- > **set*- > **sit*-. But accounting in any satisfactory detail for the actual shape of the Old English or Modern English or any other forms requires much more information; none is given, though for certain classes of users it may be extractable from implicit cues. For instance, if you have some training in historical linguistics with an Indo-Europian bias, the pairing of **sed*- and **set*- invokes the Grimm's Law subchange of stop-devoicing in the particular instance **d* > *t*. For historical Germanists the reconstruction **sitjan* provides the environment for West Germanic Gemination (-VC + heterosyllabic **-j*-); other changes, like the loss of **-j*- after the new heavy -VCC syllable created by gemination, have to be known or inferred. For the non-present forms, only OE *sæt* is relevant to modern English (though this is not the case for older states, including the period covered by *LAEME*), since the rest of the paradigm has been lost. If the original ablaut were to be reconstructed so that ModE *sat* < OE *sæt* were as transparent as *sit*, the editors would have to have added something like 'Teut. **sat*' at level II, which would reflect level III 'pre-Teut. **sod*-' (perfect *o*-grade of the root, plus the same Grimm's Law subchange).

8.2 LAEME etymologies

By contrast with the standard OED model, a *LAEME* etymology unpacks and makes explicit the narrative that brings a particular attested form into existence.[2] Each etymology is a step-by-step history of the form it labels. The characterisation of an etymology here follows that sketched in Lass (1972) and developed further in Lass (1997: ch. 5, esp. 135–37). Most important, *an etymology is not a list of cognates* (and vice versa). Indeed cognate material is conceptually subsidiary; its main purpose is to provide the backup for ultimate rather than proximate reconstruction, i.e. the earliest input forms rather than the more recent ones. At the level of early Middle English, Indo-European, and even Proto-Germanic and Proto-West Germanic, are irrelevant. However, cognate dialects of roughly the same temporal locus as Old English, or its immediate predecessors, may be important since some of these contributed to the Germanic population that settled the British Isles in the 5th century. The later forms are reconstructed or noted on the basis of attestation within the language in question, not at the level of (sub)family or any other higher-order taxonomic entity. Cognate lists as such belong to what philosophers of science sometimes call the 'context of justification'; they are not integral to etymological narrative proper.

8.3 Etymologies as narratives

8.3.1 The shape of an etymology

Our definition is a replacement and enrichment of the standard classical and postclassical definitions of what an 'etymology' is. It is not only not original in principle, but something of the sort is assumed, if not always explicitly, by many historical linguists today.[3] An etymology is indeed still a 'set of forms', but they are organised as a sequence of 'episodes' in time, not a set of static correspondences.

Our procedure is to start from a lemma[4] represented by one or more text-forms in the corpus (e.g. $sit/v), and project an ancestor or etymon at a particular temporal level. We then create a sequence in which each successive change is marked and named. The conceptual framework is essentially 'Darwinian': the underlying model (or metaphor) is a population of variants moving through time, with differential selection of particular variants making up the 'event landscape' that defines the narrative. Here is a macro-level example from a different domain. Visualise a linguistic item (word, affix) as a heritable character[5] like any other — blue eyes, the

---

[2] Our definition, as will become apparent, requires answers to types of questions that are not even approachable through the cognate lists. Given the material in the entry and the ModE paradigm, what happened to the past plural vocalism (the long vowel in *sæ 'ton*), and why is the modern past participle *sat* not **setten*?

[3] But not all. See for instance the critical discussion of etymology by 'mass comparison' and 'inspection lists' vs 'processual' or 'properly historical' praxis in Campbell 1990 and Lass 1997: 159-69.

[4] In etymological discussion, the term 'lemma' may denote a lexel, a grammel or a complete tag. In other words, it indicates any object which could be input to an etymological narrative. When citing lexels we will use the form $X, and when citing grammels we will use the form /x.

[5] We use the term 'character' in the usual neutral (if originally biological) sense: any definable 'heritable' entity of potential taxonomic or historical significance. For linguistic purposes any item definable in a metalanguage at any level may be a historically interesting character: e.g. aspiration, [p], possession of ejectives, vowel-harmony, gender, nominative/accusative alignment, OV word order,

Hapsburg lip, a reading in a text ... The default process is simple replication: the character is passed down from generation to generation unchanged (like the initial and final consonants of *sit* from Proto-Germanic to the present). Periodically, however, this default sequence is interrupted by 'miscopying' in the medium of replication.[6] When this happens a variant form arises. If such a variant becomes commoner, or in the limiting case goes to stabilisation in the population and replaces its original, there is 'a change': the normally featureless continuum of historical transmission is ruptured by novelty. The terminology here is deliberately domain-neutral, because there is a common formal substrate for all historical change — selective stabilisation (complete or partial) of random variation.[7]

An etymology then is the story of successive copying errors or 'mutations' and their fates in a language conceived as a population of variants replicating in time. At the macro-level the domain for variation is the morph or word. But it is at the micro-level that we generally enter the etymological stream and macro-level change is the end result of cumulative micro-level (i.e. segmental) change. For instance, the macro-level 'replacement' of **sed*- by **set*- is a function of the micro-level (segmental) change **d* > **t*. As we will see, this conception of an etymology is set-theoretic: the history of a form is a set of functions from form-into-form, which are the summations of functions from segment-into-segment.[8]

We assume on uniformitarian grounds that variation is the foundation for change; 'change' is an epiphenomenon of inherent variability. The complex linguistic surfaces we observe are the results of invisible (and for now largely unexplainable) processes of selection. In the simplest formulation, classical exceptionless, 'neogrammarian' change is not a *process* that occurs in languages; it is a piece of *post factum* phenomenology, an *effect* of the contingent stabilisation of variation. So when we see a pairing of temporal states for some character *C* that looks as if there was a direct change *C* > *C*´, it is virtually certain that '>' is a notation for ignorance or lack of evidence. From everything we know about linguistic change in progress, the only way that a sequence *C* > *C*´ can come about is via an intermediate stage where *C* and *C*´ are in variation: *C* > *C*´ is an abbreviation for *C* > *C* ~ *C*´ > *C*´. This holds for 'dialect borrowing' or any other contact phenomenon as well as internal 'evolutive' change.

But things are not always this simple. 'Proper' neogrammarian change takes a long time to complete, but not all change is neogrammarian. The points at which we are permitted to intersect history depend entirely on the contingencies of witness-survival. Therefore we never know in any given case whether we are going to find a completed change of the kind where *C* is associated with a time t_n and *C*´ with t_{n+1}, or whether we will find witnesses in the interstices, and so get a picture of 'irregular'

---

head-initial relative clauses. In linguistics 'heritable' happens to have a cultural rather than genetic sense.

[6] The term 'miscopying' may be open to misinterpretation. It is not evaluative, and does not imply the existence of 'right' or 'wrong' versions of an original. It is a neutral term in stemmatic discourse (though in biology 'mutation' tends to be preferred); it simply reflects the standard assumption that the default in any replication is an identical copy, so we need a term for any replication product that is non-identical to what it replicates, i.e. a source of variation.

[7] For justification of the use of 'ontologically neutral' or 'domain-neutral' language for describing historical linguistic change, see Lass (1997: 109-23, 290-324, and especially 370-83).

[8] This is true at least in the phonology of root morphemes, and to some extent of affixes. There is of course a morphological side to change as well, involving 'replacement' by analogical and other processes which do not involve segment-to-segment mapping proper: see §8.8 below.

transition from t_n to t_{n+1}.[9]

But there is a further complication. Any particular variable state may be the result of one of at least two processes, and it is not always possible to tell which. Say that at t_n our witnesses show only *C*, and at t_{n+1} they show *C* ~ *C´*. On the face of it, there are two conflicting explanations:

(1) Generation of novelty. *C* is 'in process of becoming' *C´*, and we are intersecting history during the variation preceding final selection or stabilisation. If *C*, *C´* are segmental characters that occur in different lexical items of the same etymological class, this is lexical diffusion caught in the act. We must note also that lexical diffusion may and often does abort; the classical S-curve of diffusing change may be asymptotic or broken.

(2) Loss of novelty. *C* has almost completely become *C´* at some (unrepresented) time in the past, but the innovative *C´* is now losing ground to *C*, and the novelty is in process of being deselected. If the same conditions hold that apply in (1), then we have an example of the inverse of lexical diffusion — lexical fading caught in the act. And as above, fading may be aborted too. The usual configuration for loss is the inverse of that for the generation of novelty: the S-curve read backwards.

It is worth observing that such potential undecidability can be resolved only by the consilient testimony of other witnesses, and by inferences from additional knowledge (if it is available) of the history in question. One important source of further knowledge is the matching of minority occurrences in early sources with majority occurrences in late ones. Let an early Middle English textual witness show spellings with <u> for OE *ō* (e.g. <gud> for GOOD) alongside expected spellings in <o(o)>. Our knowledge of the subsequent history of the language tells us that the <u> spellings — assuming they represent something like [u:] — are innovative. We infer this since [u:] is the value stabilised in our first reliable phonetic descriptions from the 16th century; and ModE [ʊ] continues a shortening of older [u:].

Assessing direction on 'universal' grounds is unsafe except in certain quite tightly circumscribed cases. While some directionalities are statistically more likely than others, this is probably not the case for the majority of possible changes.[10] So

---

[9] This is probably not a matter of deep theoretical significance in any case, though some scholars (notably Labov 1981, Kiparsky 1974) have argued that 'neogrammarian' and 'diffusing' change are separate natural kinds. We are satisfied with the refutation of this position in Bybee 2001.

[10] When there is no compelling evidence one way or the other, assimilation is preferred to dissimilation, loss of material is preferred to gain. More generally, reductive or inertial changes are more likely (*ceteris paribus*) than their opposites. Some apparently amplificative changes however are inertial, e.g. segmental insertions deriving from retiming. An example of apparent 'addition' of material that is in fact assimilatory and inertial ('labour-saving') is the insertion of a stop between a nasal and a nonnasal consonant, as in the familiar *thunder* < [θunr-], or the common ModE [t]-epenthesis between [n] and [s] in *prince*, *chance*, etc. Similarly diphthongisation before liquids, which can be taken as retiming of the vocalic gesture of the liquid (e.g. [a] > [au] before dark [ɬ] is not really 'addition' of a segment, but anticipation of the [u]-coloured coarticulatory gesture). Such 'motivated' changes are distinct in principle and mechanism from 'arbitrary' (e.g. phonotactically rather than phonetically motivated) epentheses like [e]-insertion before initial [sC] in Spanish (*escuela* < L *schōla*). Changes that do not involve either inertia or phonotactics, e.g. context-free vowel raisings, lowerings, frontings, roundings, chain-shifts (if they exist), diphthongisations and the like are probably stochastic; they are a function of constant background mutation, and not motivated.

etymological trajectories are in general not *deducible*; the logic of storytelling is uncertain, and heavily dependent on chance survival combined with inferential gap-filling.

We can give a somewhat more formal characterisation. Schematically, the *LAEME* etymologies are sequences of strings of the form *XCY*, where *C* is the character relevant to a particular change and *X*, *Y* are contextual variables. Let representations of the form $((I_i))$, $((I_j))$ be innovations, and *C*, *C′*, *C″* be variants of the character in question. Each form following the sequence '$((I_m))$ >' is the result of the intervention of that particular copying error in the processes of intergenerational descent. **XCY* is the 'ultimate' etymon, the reconstructed form chosen as input to the entire sequence. The output form of the entire sequence is the 'proximate' etymon. So the basic 'form of a *LAEME* etymology' is:

*XCY $((I_1))$ > X C′Y $((I_2))$ > X C″ Y > ...

The *ex hypothesi* necessary states of variation such as *C* ~ *C′*, *C′* ~ *C″* are omitted because they are irrelevant to the exposition, but are assumed to have been present. Indeed, for the later stages, for which we can have concrete evidence, they are often manifest in the texts. The last form in the sequence will be the 'target'. This may be either an 'intermediate etymon' (the Old English ancestor of a Middle English form), or the Middle English form(s) themselves, as attested in the *LAEME* corpus, which serve as inputs to mapmaking and other analytical operations.

This characterisation suggests a set-theoretic definition of an etymology. We begin by defining a 'change'. Let an input string to any change be an ordered set of segments $I = \{s_i, s_j, ...\}$, and an output string be an ordered set $O = \{s_{i+1}, s_{j+1}, ... \}$. Let there be a mapping *M* (including the identity relation) for any $s_m \in I$ such that $M(m_I \rightarrow m_O)$. Non-identical mappings maybe one-to-one, many-to-one, one-to-many, one- or many-to zero (deletion), zero-to-one (epenthesis), or reorderings (metathesis.) Each such mapping is 'a change'; the 'etymology of' a form *F* then is the set of mappings from the first input string to the last — i.e. a set of sets defined by their shared mappings.

Such a mapping would also naturally be represented as a graph, as in the etymology of ModE *sat* below (the bold arrows represent non-identical mappings, or 'sound changes'):

IE input	s	o	d
	□	□	□
Grimm's Law devoicing	s	o	**t**
	□	□	□
IE **o* > Gmc *a*	s	**a**	t
	□	□	□
First Fronting	s	**æ**	t
	□	□	□
Late OE *æ/a* merger	s	**a**	t
	□	□	□
17th c. raising	s	**æ**	t

We can then simply define a 'sound change' as a set of forms connected by a mapping.

8.3.2 What etymological representations mean

The further back we go in time, the less linguistic information survives to us and therefore the more invariant and simpler language states appear to be. No matter how complex or variable the materials whose ancestry we are reconstructing, we tend to speak as if they ultimately to go back to a *single* ancestral item. For instance, it would be quite normal to say that the vowels in ModE *cat*, *face*, *far*, *war* 'go back to ME *a*' (mediated of course by the appropriate sound changes).

There are procedural reasons for talking this way; but we must understand what we are not saying. Except as a technical device, and within the framework of a particular kind of strategy of historical reduction, there is no such single object as either 'ME *a*' or 'the vowels in ModE *cat*, *face*, *far*, *war*'. The items whose mutual mappings we have established in some detail as the basis of etymological praxis (the telling of 'true stories') are to a very large extent instrumental fictions.

It is also important to make clear that the characterisation of etymological categories, whether as italicised representations like 'OE short *a*' or apparent phonetic transcriptions like 'OE [a]' are deliberately and necessarily coarse and imprecise (see §2.4.2 above). In any sophisticated sense, 'OE short *a* (as in *catt*)' means as much or as little as 'ModE short *a* (as in *cat*)'. We speak as if we are mapping from single values, but of course we cannot be. Even within modern southern British English 'short *a*' covers a range from low central [ɐ] to low front [a] up through various [æ]-like and [ɛ]-like vowels. If we were to add South Africa and New Zealand to the set we would get values as high as [e], and if we added parts of Scotland and Northern Ireland we would get low back [ɑ] as well. Still, there is a categorial and lexical sense in which we can talk sensibly of 'short *a* words', and this is what we intend by such representations, or by symbols in phonetic brackets.

This is a crucial point, because in our experience more historians than not probably believe that a statement of the type 'OE *a* was [a]', is a legitimate thing to say. It is not. Each category is a *cluster of variants* whose members are unfortunately not available for inspection; we therefore cannot sensibly talk of 'OE *a*' without the complex stipulations implied by the comments above; and this is the sense in which all representations in etymological narratives are to be construed. Therefore all mappings between segments given particular mnemonic shapes are, like all our transcriptions, to be taken in a typological rather than strictly phonetic sense.

As our discussion of variation throughout has made clear, any symbol like

'ME *a*' is not an individual but *the name of a population of variants*, treated uniformly as a matter of convenience and procedural necessity. There is also of course a deficiency of information at the distal end of the sequence and a surplus at the proximal, which is why all historians perform 'end-matching' simplification.

8.4 The Entry point

We introduced the notions 'ultimate' and 'proximate' etyma above. The first choice we have to make is a *terminus a quo* for the category 'etymon of X'. Early Middle English is made up of native elements, and borrowed elements. The latter, for our purpose, may be conceived as superimposed on a native substrate. This will serve as the reference point for characterising post-borrowing histories of foreign items. For native items, an etymological history can be intersected anywhere from the earliest available protolanguage onward. But in a historical resource of this kind, it seems most appropriate to start at something approximating 'English'. Given our purposes, it would be over-informative to go back to Proto-Germanic or even Proto-West Germanic, let alone Proto-IE (PIE **sed*- would not form part of the etymology of $sit/v in the *LAEME* framework). We have selected as our starting point a very early post-migration language state: that is, the reconstructed 'core' of the heterogeneous cluster of mostly Ingvaeonic or North-Sea Germanic dialects brought from the Continent during the 5th century. This would be recognisably the ancestor (or better the ancestral lect-set) of the collection of text languages we know as 'Old English', and hence, at a further and not always direct remove, of the forms in our corpus. But this language cluster would not yet have undergone those changes that characterise the attested Old English dialects (e.g. breaking), and hence the native input to early Middle English.

This may occasionally be problematic: while no particular purpose would be served by including ancient pan-Germanic changes like Grimm's or Verner's Laws, there are some family-wide developments that appear at least partly convergent in the individual Germanic traditions. These developments are also — in the Old English sequence — later than certain Old English-specific changes: e.g. *i*-umlaut, which while visible to some extent in all post-runic Germanic dialects except Gothic, nonetheless must be later than characteristically Old English changes like breaking and First Fronting, because their products are affected by it. So on pragmatic rather than principled grounds we include these changes.

8.5 Substance and 'structure'

> If a change is both phonetically gradual and lexically gradual — that is, if words change gradually, and each word changes at its own rate, then each word will encompass its own range of variation (Bybee 2001:41).

The etymological starting point then is a reconstructed ca. 5th century pre-Old English. It is the assumed predecessor of varieties such as those represented in the earliest inscriptions (e.g. the Ruthwell Cross, the Franks Casket), but distinctly more archaic than these already characteristically English language forms, and even more so than the 8th-century Northumbrian versions of Caedmon's Hymn or the Corpus and Epinal glosses. Practically, the etyma are given shapes that allow them to be mapped into all the 'standard' phonological and morphological changes that appear in the handbooks (except those that we have reason to reject, which will be noted), and

from there into the forms attested in our corpus.

One important issue, which we have repeatedly encountered in early presentations of atlas materials, is what we mean by segmental representations like **sit*-, etc. Do we intend them to be 'phonemic', 'phonetic' or 'abstract'? This is a difficult question to answer, and is traditionally evaded. More language histories than not (at least until very recent times) give reconstructions in 'safe' italics, which make no particular theoretical commitment (as [ ] or / / would). We deal with this issue elsewhere, in our discussion of what we intend by the mapping of Litteral Substitution Sets (LSSs) into Potestatic Substitution Sets (PSSs) (§2.4). That was a matter of orthographic interpretation. Here, however, except for the 'target' forms in our etymologies, we are dealing with some kind of prehistoric 'substance' that is *not* available in contemporary orthography, and therefore not subject to the special considerations involving LSS > PSS mapping, and phonetic interpretation of orthography.

The appropriate strategy is to admit that by and large we do not know what the 'status' of our representations is. A starred reconstruction is an indication of phonetic substance at some level, but not of status within a 'system' in any structuralist (including generativist) sense. The history of the infinitive of the verb *sit* from our entry point would look like this:

*set-j-an > *sit-j-an > *sitt-j-an > sitt-an

There is no need to make a commitment to any particular theory of 'system structure' in such an etymological narrative. Indeed it may not be possible, or if it is possible there may be so much disagreement in the literature about the 'status' of particular items that it is better to avoid the issue and concentrate on what is etymologically relevant to us: the story at the level of segmental representation we have chosen.

So in the etymological narratives themselves, segmental representations and etyma (both base and later) will be unbracketed, e.g. **xaur-j*-'hear'. Where phonemic or phonetic representation is specifically intended (because there is data for making the distinction, and a reason to make it), we use the standard / / or [ ]. The latter will also be used to single out representations of sound types at any level in the commentaries.

8.6 The scribal dimension in etymological narrative

The *LAEME* corpus is made up of written texts. This is obvious, but it has implications that lead us to expand the notion 'etymology'. Our extended definition encompasses aspects of the attested forms that do not (necessarily) have any phonological dimension. If an etymology is a mapping of forms into forms, then orthography — particularly in a set of varieties none of which are 'focused' standards — is as much a part of the story as reconstructed phonology and morphology. We add to the traditional duties of the etymologist a new one: accounting not only for the putative phonological and morphological forms 'underlying' the written ones, but the written ones themselves. We wish also to account for properties that probably or certainly have nothing to do with either sound substance or morphology. A 'form' in the sense in which we use the term is proximately a visual object — its graphic properties are part of its structure and its history. This is of course nothing new: it is merely a restatement, in a slightly different framework and metalanguage, of a point

made canonically by Angus McIntosh half a century ago (McIntosh 1956), and central to the tradition that led to the creation of *LALME* (McIntosh et al. 1986), and to *LAEME* as well.

Our commentary on graphic form, except where it may be relevant to formal interpretation, will not generally deal with the visual characters of *figurae* themselves. Though we may comment on figural choices that can account for misreadings or confusions, e.g. in minim environments or in ambiguities involving long *s* and *f*, where an interpretation may hinge on the conformation of the crossbar of the letter. But here is a more complicated case where letter shapes do indeed bear on interpretation. On fol. 83v, Scribe A of Cambridge, Trinity College 323 (#246), represents the weak past participle suffix as <-ic>, rather than expected <-it>. He has two distinct *figurae* for <c> and <t>, one clearly 't'-shaped (with a horizontal capping cross-bar) the other 'c'-shaped (with the top rounded and drawn down). He normally keeps them distinct. In other 13th-century hands, however, the *figurae* for <c> and <t> are frequently more or less identical to each other, or are represented by a cline of shapes between those unambiguously identifiable as <t> and <c>. In these cases the bar of <t> may often only protrude to the right of the stem not to the left and may or may not be angled down from left to right, while the top of <c> may be horizontal rather than angled down and may begin to the left of the stem resembling a cross stroke. In such hands, context must be the guide to interpretation; it would be misleading to insist on transcribing <c> in 't'-contexts and vice versa. But where a scribe employs two quite distinct symbols, not variations on a theme, if the manuscript *figura* is <c>, regardless of what phonological expectations we have, we represent it as it is written, and comment accordingly.

Our more typical non-phonological spelling exegeses in the etymological narratives will focus on 'unexpected' litteral choices, and unpacking, where possible, the systemic or historical logic behind these choices. Here are two examples, one relatively simple and the other exceedingly complex:

(a) Many SWML texts use <s> (either exclusively or commonly) for the consonant in OE *-iht* words (so <brist, nist> for BRIGHT, NIGHT. This does not imply a local sound change [xt] > [st]; it is an inverse spelling, natural to scribes who also wrote French, based on an Anglo-Norman change [st] > [xt ~ çt] (see Pope 1934: §§1178, 1216 and Laing & Lass 2003 n. 12).

(b) In the *LAEME* corpus sample from British Library Cotton Caligula A.ix (#278, 'Layamon A', Hand B), the following spellings appear for $hand/n: <hond(e)> 8x, <hande> 1x and <heond(e)> 3x.[11] Both <a> and <o> for OE *a* before nasals are to be expected in the SWML, the latter more frequently; <eo> is apparently 'odd' and requires explanation. Other spelling patterns in this text-language make it clear what is happening. For instance $7/qc has the alternants <seoue, soue>: the source is back-mutation of OE *e*, hence OE short *eo* (here <o> presumably = [o]: see §8.7.3 below). If [o] < *eo* can be spelled both <eo> and <o>, then by an orthographic transposition not uncommon in some of our sources, any [o] of whatever origin can also be spelled <eo>. And indeed we find such spellings in this text for $for/cj, $forth/av, $sorrowful/aj (OE *for*, *forþ*, *sorh*-). The reflex of OE *ēo* is commonly spelled <eo> as well (e.g. in $free/aj < *frēo)*. So, unsurprisingly, this is also a possible spelling for *ō* in $brother and $book for instance (OE *brōþor, bōc*). But interestingly, in these words

[11] In some instances the <n> is written out, in others abbreviated; since this discussion focuses on the nuclear vowel, we allow the full form to stand for both types. Similarly the presence or absence of an inflectional or otiose <-e> is irrelevant for any individual case.

we also find <a>; so it seems that this scribe has an LSS {'eo', 'o', 'a'} that can be used both for [o] and [o:]. The same LSS also occurs for OE *ā*, e.g. in $loth/aj < OE *lāþ*. [12]

These few instances from the much larger number in the text sample suggest that <heond> derives from the existence of the LSS specified above. We reconstruct the enabling reasoning this way: if a word can be spelled with 'o', it can also be spelled with 'eo'; and since 'o, eo' and 'a' can appear in the same LSS, a word like $hand with the 'canonical' or 'traditional' spellings <hond, hand> can therefore be spelled <heond>. An extension of the same reasoning then allows for <a> in the LSSs for OE *ō* words: this is a function of the 'a~o' alternation in $hand. *The logic is LSS-logic, not phonetic, phonemic or etymological.* This scribe's spelling praxis is explicable, but not in familiar modern terms. And this explication is part of the history of the text forms.

So the basic level of accountability is litteral: since we are so often concerned with litteral substitution sets, scribal choice in representation (or even the 'decorative', 'stylistic' and nonrepresentational aspects of forms) are part of the historical record, and therefore also part of the etymology in our wider sense. And to clarify another point in which our etymologies differ from standard lexicographic practice, we are not producing etymologies of words but of *forms*: the 'etymology of HAND' takes OE *hand ~ hond* as its primary input; but the phonological and graphic details are stated for each distinct form, not the word 'as a whole'.

8.7 Special Etymological problems

8.7.1 Preliminaries

Not all etymologies are straightforward. Some are exceedingly complex and multi-stranded, others have gaps due to failures of knowledge or unresolved controversy. One of the causes of multi-strandedness is the gradual simplification and dissolution of the relatively tight and unified structure of some morphological paradigms at earlier stages of Germanic. Another is the existence of multiple possible root-forms for a given lexeme, due to ablaut, differential stem-formation in a paradigm, or the inhomogeneity of the input dialect clusters. ('Pre-Old English' was no more a unitary or monolithic language than 'Old English' or for that matter 'Modern English'.) There are also many indeterminacies in individual narratives. The basic principles will be raised below and details will be treated in the individual etymologies. In many cases there simply is no known etymology, or only a very conjectural one; where relevant this will be indicated in the individual etymological commentary.

8.7.2 The Old English diphthong inputs

There is fair consensus on the broad phonetic values to be associated with most of the graphs used to spell Old English. The exception is the digraph spellings <ea, eo, ie>, which have been subject to a great deal of controversy (see the discussion and bibliography in Lass and Anderson (1975: 75–82)). The view that these spellings represented monophthongs is now largely discredited, but there is still no full consensus on the nature of the diphthongs. The view espoused in Lass and Anderson (1975) and most recently taken up in Hogg (1992: §§16-34) is that the Old English

[12] Strictly speaking, as a subset of the full LSS, which is {'a', 'ai', 'e', 'eo', 'æ', 'o'}. The coexistence of 'a' and 'o' in this LSS allows the use of 'a' for [o:]: see below.

diphthongs were 'height-harmonic': the two elements agreed in height, the second assimilating to the first. Thus the full history of what comes to be written *ea* as in *eald* OLD would be *[a] > *[æ] > *[æu] > [æa], with the original [-u] diphthong (for which there is orthographic evidence) subject to a process that Lass & Anderson called 'Diphthong Height Harmony'. The same kind of diphthongs arose from other Old English processes like back mutation. So the graphic sequences *ea*, *eo* in OE represented [ae(:)a, e(:)o]. The nature of *ie* (largely restricted to early West Saxon, but certainly part of the lineages of some text languages attested in the SW and SWML) is problematic: it has been characterised variously as [ie, iu, iy, iə]. Where there are developments that give evidence about its nature (e.g. its split into [i] and [y]), [iy] would appear to be the representation of choice (Hogg 1992: §§5.74, 5.82). We assume this value for the lineages that could have their origin in early West Saxon, because it accounts most neatly for the subsequent developments: see the change corpus under ((IES)).[13]

8.7.3 <eo> spellings

These present a number of difficulties. First, the number of historical categories they can represent is dauntingly large. In our sample from one text alone, British Library Cotton Otho C.XIII (#280, 'Laʒamon B'), <eo> can represent at least the following categories: OE *y*, *ȳ*, *eo*, *ēa*, *ǣ*[1], *ǣ*[2] (Lass and Laing 2005). But the worst difficulty, paradoxically, appears to arise when there is no doubt that the input to a given form is at some historical stage OE *eo/ēo*. The scribe of Otho for instance spells these categories <eo, e, u, i>, while Cotton Caligula A.ix (# 277, Laʒamon A, hand A), has all of these plus <o>. At least some of this orthographic variation is likely to represent phonic variation at some level. Consideration of the individual scribal spelling systems makes it possible to assign values to <e, u, o, i> without much uncertainty. But <eo>, given what appears to be the history of the Old English categories traditionally represented this way (see ((EOM)) in the Change Corpus) could in any instance represent either [e(:)] or [o(:)], and it is generally impossible to tell which. The spelling <beo-> for the root of $be/v in Laʒamon A, hand A, occurs 46 times in the sample, but it can never serve as a univocal target for an etymology. Such spellings cannot at this date represent anything diphthongal of the type [e:o], we do not think the evidence that they represent front rounded vowels is convincing (Lass and Laing 2005). The only other likely possibilities are [e:] or [o:], but any one instance could represent either.[14] (It is not relevant that the [e:] variant *eventually* became the majority form and stabilised in all modern varieties; the textual attestations are nevertheless ambiguous.)

8.7.4 <ea> spellings

As in the case of <eo> above, this spelling became 'free' when the diphthongs it

[13] It is however also possible that at least some Old English varieties also had diphthongs of the 'classical' Germanic type with first elements of various heights and high second elements, e.g. [æi, æu], etc. If this is the case, the so-called 'new' Middle English diphthongs in [-i, -u] were not new at all in type, but were (except in the case of borrowings like French *oi*, *ui*) simply continuations of nuclei that had been present from the beginning. (See the discussion in the Change Corpus under ((CV)) Coda Vocalisation.)

[14] Potentially, some instances could be resolvable if the <eo> form occurs in rhyme position. But in practice, Laʒamon A has very irregular rhyme usage and even apparently decisive rhyme evidence may be a chimera.

represented in earlier Old English monophthongised; it could then be used for the entire range of categories that constituted its historical inputs, as well as for similar sound types. Earlier OE <ea> indicated something like [æ(:)a]; by the 11th century the long diphthong had merged with *ǣ*. This later probably raised to [□ː] (if in fact what we reconstruct as late OE [æː] was not already [□ː] or [æː~□ː]). The short diphthong merged with *æ*. Since, depending on location and scribal preference, the latter category could develop into eME [a] or [e], this leaves <ea> as a possible spelling for [□ː], and for [a] and [e] of any origins. This difficulty is particularly acute in the AB language texts, which use <ea> extensively to represent the reflexes of a variety of historical categories. Cambridge, Corpus Christi College 402 (#272, *Ancrene Wisse* sample) uses <ea> in forms which appear to go back to OE *æ*, *ǣ*, *ea*, *ēa*, *a*, *ā* and OF *a*, *ē*. Furthermore, it alternates with <e, eo> in some *ǣ*-words, and with <a> in one *ǣ*-word. It will, as in the case of <eo>, not always be clear for texts of this kind precisely how the target of the etymological sequence should be characterised.

8.8 Morphological indeterminacy: 'pathways' and 'replacements'

It would be simple if the etymology of a corpus form were only the etymology of the root: the lemma $heart/n could simply have the base etymon **xert-*, and the changes leading to the various forms of the root like <heort-, hert-, hort-> etc. would constitute the entire entry. But inflectable words have more complex histories, often involving nonphonological processes; the affixes attached to roots are often unexpected, and require their own historical commentary.[15] There can be so much history in an inflected form that it becomes necessary for the etymological entries to separate morphological from phonological development; the commonest case is one in which the same root form in the same grammatical frame, even in the same text language, may have affixes that apparently reflect different historical classes, only one of which (according to the standard grammars) is 'original'.

Let us take $heart and $soul as typical examples. According to the handbooks, $heart 'is' a weak feminine (*n*-stem) noun, whose etymon should be **xert-o-n-*; that is, it is roughly the historical equivalent of the Latin '3rd declension' *n*-stems like *homo* (gen sg *hom-in-is*, etc.).[16] Other Germanic evidence bears this out (e.g. German gen sg, nom/acc pl *Herze-n*). The same sources classify $soul as a heavy-base strong feminine (*ō*-stem), with zero nom sg and the rest of the case/number endings (except dat pl) vocalic. This means is that $heart and $soul ought to have been associated 'in Old English' with these paradigms:[17]

---

[15] This is why the traditional handbooks so often (correctly) divide their subject matter into *Laut- und Formenlehre*. The *Formenlehre* is not comprehensible without the *Lautlehre*, but the latter does not exhaust it: much morphological change is non-phonologically driven morph-substitution, not segment-to-segment mapping.

[16] For the problems involved in claiming that an Old English or other Old Germanic word 'belongs to' a particular inflectional class see Lass (1986, 1991). The actual vowel that occurs in the *-Vn-* element varies according to dialect; the vowel was subject to ablaut, so that the input to any given Germanic language could have been 'normal' grade **-en-*, **-on-*, zero-grade **-n-*, or lengthened *e*- and *o*-grades (cf. Campbell 1959: § 616).

[17] The scare quotes round 'In Old English' are just a reminder that the term does not denote a uniform language, but like Middle English a set of text-languages, only more fragmentarily attested and less well-provenanced.

sg		pl	sg		pl
nom	heort-e	heort-an	nom	sāwol	sāwl-a, -e
gen	heort-an	heort-ena	gen	sāwl-e	sāwl-a
dat	heort-an	heort-um	dat	sāwl-e	sāwl-um
acc	heort-an	heort-an	acc	sāwl-e	sāwl-e

The gen/dat/acc sg and nom/acc pl of $heart ended in *-an*, which by regular developments should give Middle English forms for those categories ending in *-en* or (later, reduced) *-e*. The same should also apply to the dat pl, since as early as the 10th century there is widespread (if not categorical) merger of final *-m* and *-n* in most varieties of Old English. As for $soul, *-en* might be expected in earlier texts in dat pl, but the only other ending throughout the paradigm ought to be *-e*. These two paradigms are very different from that of the strong masculine *a*-stem, which can be represented by the ancestor of $stone (m):

sg		pl
nom	stān	stān-as
gen	stān-es	stān-a
dat	stān-e	stān-um
acc	stān	stān-as

Now consider these genitive and plural forms for $heart and $soul from a number of corpus samples of different dates and provenances: [18]

London British Library Arundel 57, *Ayenbite of Inwyt* (# 291)
$heart/npl hert-en 7x
$soul/npl zaul-en 3x zaul-es 5x

Cambridge Corpus Christi College 444, *Genesis and Exodus* (#155)
$heart/nG hert-e 1x hert-es 1x
$heart/npl hert-es 1x
$soul/nG soul-es/sopl-es 3x

London British Library Cotton Cleopatra C.vi, hand A (main hand) *Ancrene Riwle* (#273)
$heart/nG heort-e 4x
$heart/npl heort-en 2x
$soul/nG saul-e 7x saul-es 1x
$soul/npl saul-e 2x

Cambridge Corpus Christi College 402, *Ancrene Wisse* (#272)
$heart/nG heort-e 6x heort-es 1x

[18] For expository purposes all plural grammels except genitive (i.e. /npl, /nplOd, /nplOi, npl<pr) have been amalgamated. Differential plural case endings are for the most part lost in early Middle English and here we are interested rather in which historical declension a form appears to 'belong' to. Note that tokens in genitive plural function only occur in the present set of examples in $soul in Trinity Homilies: see below.

$soul/nG sapl-e 7x sapl-es 1x
$soul/npl sapl-en 1x sapl-es 1x

Cambridge, Trinity College B.14.54 (335), Trinity Homilies, Hand B (#1300)
$heart/nG he(o)rt-e 5x
$heart/npl he(o)rt-e 7x he(o)rt-es 3x
$soul/nG saul-e/soul-e 11x saul-es 2x
$soul/npl saul-e/soul-e 3x soul-es 1x
$soul/nplG saul-e 1x saul-ene 1x

There are two interesting features in this collection of forms, both relevant to the task of etymologising. One is the ease with which these nouns appear to shift declensions, not only to the more innovative *a*-stem *s*-genitive and plural, but also in other directions. Trinity Homilies *saul-ene* not only shows a shift from strong to weak, but perhaps also an unexpected full form of the oldest type of genitive plural. The other is the indeterminacy of a number of the endings that do occur: what for instance is the source of the gen sg *-e* in *heort-e* in Trinity Homilies or in *Genesis and Exodus*? If we take the canonical paradigm as our source, then these are examples of the 'standard' development *-an* > *-en* > *-e*. But can we actually claim this with any certainty, given the other changes? How do we know that the genitive *-e* is not in fact (in those particular forms) a transfer from strong feminine, i.e. the original type associated with $soul? The answer is that we do not and cannot.

So a given affix, say an *-e*, could, in a particular instance, be the result of a 'natural phonetic pathway' like the attrition discussed above; it could as easily be one of a set of 'chunks' that exist in a morphological search space. Whether this kind of equivocal history is involved is a decision that must be made for each individual case. For instance, given the original morphology for $heart, plural or genitive affixal *-es* cannot represent the output of a stepwise pathway, because there is no known pathway of that shape: [n] does not map into [s]. Therefore the inflection must be a substituted whole: *-es* has its own independent history, and can only be interpreted as an alternative to, not a development of, *-en* or *-e*. Thus whether or not an affix is taken to be the output of a sequence of segmental mappings or a substitution-as-a-whole is contingent on our knowledge of the inventory of likely pathways.

So for many affixes, assigning a source involves decisions as to what 'space' the development has been operating in. Is a text form *-e* the result of (a) a phonological mapping like *-an* > *-en* > *-e*, or (b) a substitution of some other affix with a determinable source, or (c) the choice of a 'generalised *-e*' marking some function like 'oblique', with no decidable source? The chain of custody is broken in all cases where a number of narratives could converge on the same result. Therefore, in a system whose affixal morphology works on different principles from the canonical Old English one, we may not be able to choose among (a) – (c). The only really clear cases are those in which there is sufficient phonological substance to the affix for us to assign it a definitive source, e.g. if it ends in *-n* or *-s*.

8.9 Loanwords

8.9.1 Types of borrowings

Every language has an inventory of formatives: lexical items and grammatical items

(inflexions and derivational markers).[19] Under normal conditions, the bulk of these have been there 'since the beginning', either of the language itself as a member of a distinct lineage or even of the macrofamily it belongs to. *Father* and *me* have been 'in English' from the beginning in the second sense (their roots can be reconstructed for Proto-Indo-European). The agentive ending *-er* has been in the English lineage at least since the formation of a Northwest Indo-European subgroup (if indeed it is cognate to L *-ārius*). Other forms in the inventory represent incursions via contact with other languages: *paternal* and the derivational suffix *-ity* are of this type, as are the pronouns *they*, *their*, *them*. The Indian grammatical tradition makes a useful distinction between two kinds of Sanskrit loans in the vernacular languages: *tatsama* ('same as') and *tadbhabva* ('originating in').[20] Tatsamas can usually be recognised by their shape (e.g. retention of archaic morphology, not having undergone certain Middle Indic sound changes); tadbhavas have been modified morphologically and/or phonologically so that they *look* 'native', but there is evidence of one kind of another that they are reborrowings rather than direct inheritances through the Prakrits.

For our purposes we define as tatsamas loans into English that are formally identifiable as such, on the grounds of orthographic or morphological foreignness. Old and early Middle English however are not a rich source of such forms: with few exceptions, the extensive loanword component of the *LAEME* corpus consists of tadbhavas, many so deeply integrated and transformed that identifying them can be exceedingly difficult, especially if the source is genetically close. For instance it tends to be harder to identify Dutch or Scandinavian loans than Latin or Romance ones (see §8.9.2 below).

Although tatsamas are in principle the easiest type of loan to identify, in our materials there are very few 'complete' ones. That is, the 'foreignness' of even the least integrated loans from the genetically most distant sources tends to be only partial. With few exceptions the form of a word or affix does not in itself give a clear indication of its source. We require external evidence and particular strategies of inference for identifying loans. Since virtually all orthographic systems in our data allow for some degree of variation, the same item can appear as both tatsama and tadbhava or a 'mixture' in the same text language. E.g. in the *LAEME* sample from Oxford, Bodleian Library, Laud Misc 108, Hand A, $pharisee/n is written <Phariseu>1x, <pharisee> 2x, and in the plural <phariseus> 2x and <fariseus> 1x. If we take <ph> as a marker of foreignness, then all these forms are tatsamas except <fariseus> which is a tadbhava. Many loan-forms have elements of both tatsama and tadbhava. The spelling of the only example of $physician/npl <Fisicieins> (from the same text language) lacks the <ph> foreignness marker, and might be reasonably classified as a pure tadbhava, except that the <-ieins> ending looks rather non-English. Firm identification as a loan will depend on our external knowledge of Germanic and Romance and Greek. Other writers too may show mixed strategies. Our sample of the *Ayenbite of Inwyt* shows the following spellings: $physic/n <fisike> 3x, <fizike> 1x; $philosopher/n <filosofe>1x and <filizophe> 1x. These are mostly tadbhavas: in this text language <z> is not a 'foreign letter', as it might be in some others, since Dan Michel quite uniformly uses it for native initial **s-*. The form

[19] Under one interpretation inflexional processes like Ablaut could be considered members of inventories too; though we will not do this here, but rather treat them in terms of their segmental reflexes.

[20] These are abbreviations respectively of *saṃskr̥tatsama* 'the same as in Sanskrit' and *saṃskr̥tadbhava* 'originating in Sanskrit'. For discussion and context see Masica (1991: §4.2).

<filosophe> however is a tadbhava at the left but a tatsama at the right.

8.9.2 Identifying loans

Since one of the tasks of etymology is identifying beginnings, loans must be distinguished from native items. Depending on perspective, we will use the term 'native' in either of two senses: (a) 'Germanic' in general; (b) 'not from any Germanic language other than English or its precursors'. In the *LAEME* context, therefore, while *they* is 'native' in the first sense (not coming from any Indo-European subfamily outside Germanic), it is a loan in the second, since it is from North Germanic. Under this interpretation it will therefore have the same status as a form with a Greek or Latin or French original. On the other hand *sea* is native in the second sense, since it is a Proto-Germanic inheritance, and has been 'in Old English' from the beginning. We would also classify as native in the second sense forms that do not have a from-the-beginning provenance, but belong to lower-level dialect-groupings: e.g. *brook*, which is attested only in Ingvaeonic. Identification of loans, except for forms with tatsama-like properties, is then to a large extent a matter of knowledge of external history.

However, our judgements at least of intra-Germanic 'foreignness' may be skewed by the contingencies of survival. We must always bear in mind, given the fragmentariness of the early record, that absence of evidence is not evidence of absence. Here is a familiar and exemplary case, which we will use to illustrate the kinds of argumentative strategies problematic in the difficult matter of identification of loans from other Germanic dialect-clusters. In a discussion of putative Scandinavian elements in modern Norfolk varieties, Poussa (1993: 37ff) ascribes the form (*a*)*thwart* to Scandinavian borrowing.[21] The tenor of her argument is that since Scandinavian expressions such as OIc *um þvert* 'diagonally' do occur, but there is no form of this shape in Old English, the word must be Scandinavian. Now there are two things wrong with this argument. First, there is at least a cognate attested in OE: *þweorh* 'crooked, perverse' (not to mention derivates like *þweorian* 'oppose', *þweorlic* 'contrary, adverse', *þweornes* 'peversity'). And second, cognate forms occur not only throughout Germanic with similar meanings (Go *þwairhs* 'angry, cross', OHG *dwerh*, *twerh*), but also in other IE groups (L *torquēre* 'twist'). This suggests an IE **tVr-k-*, with both *e*- and *o*-grades surviving in Germanic.

The existence of the root in Gothic, and of related forms in West Germanic dialects genetically closer to English than Scandinavian (even in Ingvaeonic: Du *dwars* with historical suffixal -*s*) makes something near a *prima facie* case for *thwart* not being a loan, but a native formation which happens to be unattested. If there were an OE **þweort*, it would of course come down as *thwart*, by the same historical sequence that gives *dark*, *heart* < *deorc*, *heorte*.[22] But Poussa's logic is interesting, and a similar logic confuses a good deal of the ascription of Old English and Middle English lexis to Scandinavian. Her argument can be unpacked as follows:

1. Form F is attested in place P.

---

[21] The discussion here is partly based on Lass (1997: 203ff).

[22] If we take it that the original root ends in *-*r*, then there might just be a Scandinavian element in the form *thwart*: Jespersen (1905: 83) suggests that the final -*t* is a NGmc neuter marker. On the other hand this neuter marker (<IE *-*d* as in L *quo-d*) does occur in Old English as a suffixal element as well, e.g. in *þæ-t*, *hwæ-t*.

2. F does not occur (in this precise shape) in attested Old English.
3. There is a well-documented Scandinavian settlement history in P.
4. F or something very like F does occur in attested Scandinavian.
5. Therefore F-in-P is from Scandinavian F.

But in cases of suspected borrowing, there is an alternative, and in the end safer kind of argument:

1*. Form F is attested in P.
2*. F is not *attested* in OE.
3*. There is a well-documented Scandinavian settlement history in P.
4*. But extant West Germanic (and/or Gothic) cognates, along with the Scandinavian ones, indicate that F is in fact a common Germanic or even Indo-European inheritance.
5*. Though there is no 'proof' that F-in-P could not be a Scandinavian loan, it could just as well be that its non-attestation in Old English is contingent (only some 25,000 lexical items survive).
5**. Therefore Scandinavian provenance is at best not proven, but of course possible.

The general principle is: regardless of history and attestation, looking for Scandinavian sources, where the phonology or morphology do not specifically demand them, is non-parsimonious, and does not serve as evidence for borrowing.
How then ought one to identify a Scandinavian loan? The ideal case is one in which phonological developments are displayed that cannot belong to Old English. If in addition the sense of the form coheres with a Scandinavian origin, one can be as close to certain as possible. An example of this kind is $window/n. The normal Old English word for this object is *ēagþyrel* 'eyehole', and the normal Sc word is *vindauga* 'wind-eye'. The forms occurring in the *LAEME* corpus are satisfactory from both points of view:

$window/n windoun, wyndow, ƿindoge, ƿindohe, ƿindoƿ-

The sense is Scandinavian, and the development of the second element coheres with Scandinavian rather than Old English (Gmc **aug-ō-* > Sc *auga*, OE *ēage* and none of the second elements in the corpus could be from OE *ēa*.) Therefore on both counts this is a loan.

8.9.3 'Scandinavian influence'
Despite the caveats in the preceding section, there are many undoubted Scandinavian loans in English. Nobody doubts the Norse provenance of *sky*, *egg*, *fellow*, *take*, *wing*, *husband*, *they*, *their*, *them*.[23] Given what we know of patterns of settlement, and the likelihood that there was extensive English/Norse bilingualism,[24] it would indeed be surprising if this was not the case.

[23] For a dated but still useful survey see Serjeantson 1935: ch. IV; for a modern discussion with emphasis on the sociolinguistic milieu see Townend 2002.
[24] According to some there was even mutual comprehensibility. This has been well (if still controversially) argued by Townend 2002.

But there is another category of English/Scandinavian interaction that is conceptually problematic, and still not 'solved' to everybody's satisfaction. This is the problem of forms, which appear not to be direct borrowings, but show what in the literature is often called 'Scandinavian influence' (so for instance OED s.v. *give*, v.). These forms are not obviously loans, but they deviate in certain ways from what their Old English etyma would lead us to expect, and this deviation is in a 'Scandinavian direction'.

The most salient example is the presence of unpalatalised velars in words where an Old English original is attested, and would be expected to have a palatal. Typical cases are *give*, *kirk* and *skirt* (OE *g*(*i*)*efan*, *cyr*(*i*)*ce*, *scyrt* with initial [j, tʃ, ʃ] respectively). These forms have Scandinavian cognates with velars, and it would seem at first as if they could be taken as simple loans. But their status on reflection is difficult: are they core-lexis loans from North Germanic, or could they represent something else? The nature of the forms seems at first to make direct borrowing unlikely: would a language with a perfectly good word for 'give' borrow a cognate that differs primarily in its initial consonant, giving rise to a phonotactically inadmissible sequence?[25] There are two solutions to this problem: (a) despite the apparent unlikeliness of this kind of phonologically selective borrowing, even in a multilingual setting, this is precisely what happened; or (b) the dialects of Old English spoken in the relevant areas lacked palatalisation (perhaps under 'Scandinavian influence', to come full circle). This would be the methodologically preferable solution, but unfortunately the evidence is against it: all the Old English dialects, even Old Northumbrian where we might expect this, show palatal diphthongisation in precisely these positions, i.e. after historical initial **k*, **g*, **sk*. Therefore velar palatalisation and palatalisation of **sk* were apparently universal in Old English, and option (b) is not available.

Serjeantson (1935: 75-6) gives a somewhat simplistic account which may have some plausibility. She suggests that on the basis of existing cognate pairs like OE *sciftan* SHIFT, O Scan *skifta*, bilingual speakers of Old English might have been tempted to 'reconstruct' a nonpalatalised form even where one did not exist. This would presume some kind of prestige, a desire for what she calls a 'Scandinavian flavour'. Given this tendency,

> a Scandinavian initial *g* might be substituted for the corresponding English *ȝ* (= *y*- [j]), since the one is often equivalent to the other in cognates. This, it seems, was what happened in *give* ... which eventually took the place of M.E. *ȝeve*, *ȝive*, *yive*.

This assumes in the usual way that the 'agent' of the borrowing was an Old English speaker influenced by Scandinavian. But there is another kind of borrowing, called 'imposition' by Townend (2006: 71-2): a Scandinavian speaker speaking English serves as the source, by using his native pronunciation rather than the English one (preserving his 'articulatory habits', following the model in van Coetsem 1988). Then for whatever reason this new pronunciation is picked up by the native English speakers and nativised.

Despite the mild attractiveness of these accounts we are agnostic. We would

[25] Or *possibly* inadmissible: it depends on when the form came into being. If it was created during the Old English period the sequences [ge-] or [gi-] would have been disallowed; in very late Old English or early Middle English the situation would have been different. We simply do not know when and where velar palatalisation ceased to be active.

prefer to leave the problem unsolved; in the etymologies, forms like GIVE with an initial velar will simply be marked as having traversed an alternative pathway, where velar palatalisation was lost.

8.10 The structure and use of the Corpus of Etymologies (CE)

8.10.1 Structure of entries
Each entry has three parts:

(a) Lemma and presumed etymon (or etyma if there is not a unique original), with commentary if relevant.

(b) The sequence of changes (if any) leading to the attested Old English form(s) presumed to be ancestral to the attested corpus forms. These link to descriptions of the changes in the Corpus of Changes (CC).

(c) The actual forms attested in the *LAEME* corpus, with later changes leading to them.[26]

The forms under (b) and (c) may be sub-grouped according to which categories of changes apply. In the case of items with multiple histories, these are grouped as separate 'pathways'. So for $night/n, there are two etyma with different Old English and hence Middle English outcomes: **naxt* representing the bare root, and **naxt-i* representing the stem (the second undergoes *i*-umlaut among other changes not relevant for the first). Etymologically incomprehensible or otherwise odd outcomes are commented on as fully as the 'normal' ones. Some of these would be dismissed by editors as 'scribal errors' (e.g. <NITF> for $night: but see the CE s.v. $night/n and Manual Chapter 2, §2.3.4 n.33) for an analysis that probably saves it).[27]

All changes in the sequence are bracketed ((...)), and the names inside the brackets are linked to descriptions and commentary in the Change Corpus. The etymological portion of *LAEME* then consists of two subcorpora: (a) the CE proper: etymologies for all lemmata (lexels and grammels) in the central tagdic, and (b) the CC, a linked sub-corpus of changes (phonological and morphological) that are used in representing the histories of the forms.

8.10.2 A Sample Etymological Entry with links
The overall structure and use (and what we described rather abstractly above as 'narrative') can best be explained by a worked example: here the etymology for the lemma $hill/n. We present it below with the subparts of the entry given expository labels, and a printed version of the results that would be obtained by following the hyperlinks to the various changes listed. This will give an overall picture of what an etymology looks like as an accessible utility within the interactive portion of *LAEME*.

---

[26] More accurately, the actual form *types* occurring in the corpus, not all tokens. If a form of type T exists, whatever the textual source, it will (in general) have an etymology of the form E, and this will stand for all tokens of that type. For instance wherever $hand/n appears spelled <HOND>, it will have undergone the change called ((PNR)) 'Pre-Nasal Rounding', so from that point of view all occurrences of <HOND> are interchangeable, and the form type is listed only once. Frequencies of occurrence of different forms can be obtained from individual textdics or the tagdic.

[27] As historians our overriding concern is that information must not be lost; so even apparent nonsense is recorded, partly under the assumption that nonsense can and often does turn into sense when somebody with a different imagination looks at it from a different direction. These 'odd' forms will be commented on and we will attempt to explain those we can; others will simply be left for posterity. This does not imply of course that we do not 'believe in' scribal mistakes; but it is not always *a priori* determinable whether a peculiar looking or uninterpretable form is in fact a 'mistake', and in any case even patent errors are part of the documentary record.

Conventions:

(a) Material in 12pt is the etymology itself, with reminders in italics (which will not appear in the CE itself) of what each section contains. Forms in the narrative etymology are in plain text, except for Old English outcomes which are in italics.

(b) Material after in 10pt is what would be the result of following the changes in (( )) as links to the CC.

(c) In the change entries, the following flags are used: %a = analogical; %l = lexically sensitive; %m = morphologically conditioned; %n= no traditional name (change name given in this corpus); %p = pretextual, i.e. before earliest attestation of OE; %r = regionally restricted; { } enclose regional abbreviations; %v = variable, sporadic or irregular change.

1. *Lemma and ultimate root etymology*

$hill/n (= noun)
SM/F (= strong masculine or feminine)
*xull-i; *i*-stem; geminate original, cf. L *coll-i-s*.

2. *Pre-Middle English history*
*xull-i ((IU)) > *xyll-i ((HVD)) > xyll *hyll*

 ((IU))%p I-Umlaut [OE]
Also *i*-mutation. Back vowels front without change of rounding, and low front [æ] > [e] before *i, *j. IU is one of the major sources of OE paradigmatic alternations, both inflexional and derivational, many of which survive into ME and beyond. E.g.. sg/pl pairs like *man/men* < **mann/*mann-iz*; the /ps ~/pt alternations in verbs like tell < **tal-j-an* vs *told* < **tal-d-æ*; alternations between positive and cpv/sup adjectives like *old/elder* < **ald/*ald-i-ra*; derivational relations like those between *lāc* STATE and *-lǣcan* MOVE INTO A STATE < **laik/*-laik-j-an*. The basic pattern is: *[a(:]) > [æ(:)], with later raising before nasals as in *sendan* SEND < **sænd-j-an* < **sand-j-an*; *[o(:)] > [oe(:)] (later [e(:)]), e.g. *dehter* DAUGHTERS < *doehter* < **doxtr-iz*; *[u(:)] > [y(:)] as in *mȳs* MICE < **mūs-iz*, pl of mūs MOUSE.

IU of the diphthongs is complex (for details see Campbell 1959: §190), but in outline WS *ēa/ea*, *ēo/eo* collapse in WS *īe/ ie*, (*ē* in other dialects): *bieldan* MAKE BOLD cf. *beald* BOLD, *līehtan* SHINE, cf. *lēoht* LIGHT, both **-jan* verbs. Not all diphthongs mutate in all dialects: for the exceptions and general patterning see Campbell §§ 200ff.

IU produces the first front-rounded vowels in English, as well as increasing the lexical incidence of *æ*, *e*. [y, oe] are phonologised when the umlaut triggers are deleted, e.g.by ((HVD)), or otherwise neutralised.

Because this change is later than velar palatalisation ((VP)), the new front vowels produced by it do not cause palatalisation: thus a velar in *cynn* KIN < **kunj-* as opposed to a palatal in *cinn* CHIN < **kinni*.

 ((HVD)): High Vowel Deletion [OE]
{=*i,* u] delete after stressed final heavy syllable (VV, VCC); thus *sunu* SON but *hand* HAND, both with original thematic **-u (<*sun-u-z, *hand-u-z)*.

3. *Form-types occurring in corpus + post-Old English history and comment; post-Old English changes in left margin in (( )).*

((EU)) > HIL(L)-E-, HYL[28]

[28] Entries are typically given in conflated form; HIL(L)-E- stands for the set HIL HILLE HILL+EN. That is, L(L) indicates that forms with both single and double L occur in the corpus; the first hyphen

 ((EU)) %n, r{E}: Eastern U nrounding [ME]
[y (:)] > [i (:)]

((WYR)) > HUL(L)E-

 ((WYR))%n, v, r{W}: Western y-Retraction [?OE/ME]
[y(:)] > [u(:)].The handbook consens us is that in the West OE [y(:)] remains as ME [y(:)], spelled <u>'after the French style'; we adopt the position of Lass & Laing 2005 that this is at best not proven, and that [y] when spelled <u> has retracted and merged with [u]. This is borne out by lack of distinctive reflexes for putative ME [y(:)], and the fact that OE and French [y] fall in with ME [u] (e.g. *cudgel*, *judge* respectively).

((KC)) > HELL-

 ((KC)) %r{Kt}: Kentish Collapse [OE]
[æ, e(:), y(:)] > [e(:)], This effectively destroys the contrast [a]: [æ] in some varieties of Old Kentish, and collapses [y(:)] with [e(:)]. There are other later changes that have the same result, but these examples are from *Ayenbite,* hence known to be Kentish.

4. *Morphology: remarks on forms on line below; unchanged developments from Old English uncommented; all morphological changes marked in left margin, conventions as under phonology. For tagging conventions see chapter 4.*

/n hil(l)(e), hul
final E otiose
/n<pr helle, hil(l)(e), hyl, hul(l)(e)
Final E otiose, or relic of old dat sg
/nOd hil, hul
/npl hull+es
((IA)) /npl<pr +hull+es, hell+es
/nplOd hell+es
((DS)) /npl<pr hull+en, hill+en
{=Weak instead of expected strong form=}

 ((IA))%m, v Intraparadigmatic Analogy [OE, ME]
Cover term for analogical changes within the paradigm, e.g. replacement of old dat or gen pl by nom/acc pl.

 ((DS))%m,l,v: Declension Shift [?OE, ME] This characterises apparent shifts from a historically expected declension to some other, e.g. from weak to strong noun signalled by a genitive sg for $heart like <hert+es> (masculine/neuter a-stem type) rather than expected <hert+e(n)>. In effect, the form so marked represents a type not associated with the historically 'original' declension class. These shifts are calculated from 'canonical' OE class assignments, defined by postulating (counterfactually but conveniently) a 'variation-free' ancestral state. But it is not possible to exclude the same kind of variation in OE, even quite early, as examination of the 'early, late and dialectal forms' in Campbell 1959 makes clear.

8.10.3 Lemmas and links
Each lexical category lemma in the tagged corpus is linked only once to a base-form in the corpus of etymologies (CE): e.g. all verb forms are accessed via a link from the present system, with the minimal grammatical tag '/v', nouns and pronouns via a link from the nom sg, adjectives from the positive. Nonlexical categories (e.g. affixes)

---

indicates that there are some forms with a final E, which is not part of the phonological analysis, for reasons detailed in the notes; the second hyphen indicates that there may be some (inflectional) material following the root, which is treated under the morphology.

have their own tags and etymologies. So for instance all forms of each individual strong verb are linked to a complex entry that gives the histories of all the tense/number forms: $bear/vpt13 would link to $bear/v (with SK4 on the line below), which would be the headword for the whole system, and show the developments of the various ablaut grades, analogical weak forms, class transfers, etc., with commentary (see below).

Nominal compounds (e.g. $fish-net/n) are lemmatised not under their heads but under the first lexical element; if they do not contain unique forms of the root they are not further distinguished within the etymological entry. There are cross-references to the second elements as well, so $fish-net/n, while lemmatised as a subcase of $fish/n, would cross-refer to $net/n. This apparently redundant listing allows for the identification of compounds as lexical items in their own right as well as their component simplex elements. Though this may look like excessive scupulousness, it has a solid linguistic motivation. Most compounds (especially *tatpurusas* like this) are not strictly compositional (the relation between 'fish' and 'net' is conventional rather than derivable from linguistic structure *per se*); therefore the cross-listing gives a better picture of the vocabulary than would be obtained if the redundancy were bypassed and there were entries only for $fish and $net, but not for $fish-net as well.

In perhaps the majority of cases, the relation between lemma and etymological entry is straightforward: $hill/n links as above to the etymology for that word. But in cases where paradigms are complex (as in the strong verbs), or where suppletion is involved (as in verbs like $go, $be, or adjectives like $good), the linkage may be rather less direct. As a matter of convenience we have lemmatised these 'irregular' forms in a somewhat inconsistent way, in fact a way we are not always entirely happy with on theoretical grounds. These choices are based on the traditional lemmatisation in dictionaries like Clark Hall & Meritt (1960), the OED, the MED and the like, as well as general traditions that for instance consider 'be-', 'was/were' and 'am/is/are' to be 'forms of the verb $be'. The following links may be worth noting:

1. $be/v in the corpus covers all forms usually associated with that verb-complex (see examples above). Thus 'was' and 'were' are not separately lemmatised, but are accessed from $be/v. Any lemma with that character sequence will link to the headword $be in the CE, and the various suppletive stems 'constituting' the verb will be listed and etymologised under that heading.

2. Contrariwise, $worse and $worst are lemmatised separately, rather than (as would be consistent but perhaps less convenient and intuitive) accessed directly from $good/aj. But the links from those forms will in fact take the user to the etymological entry for $good in any case, and they will be etymologised and commented on under that heading. The 'suppletive past tense of $go' also has a separate lemma, $e:ode/v; this is rather easier to justify, as there are other forms that could have the same function, like <WENT(-E)>; this however is not an 'orphan' like $e:ode, but also the past of $wendan/v, to which it is linked. It will be up to the user to decide from context which instances of <WENT(-E)> are in fact pasts of $go, and which really 'belong to' $wendan.

3. Verbs with complex paradigms, either standard strong verbs belonging to particular 'ablaut classes' in the handbooks, the preterite presents, or 'anomalous' verbs like $do, $will, are listed in the CE somewhat differently from the way they are listed in individual textdics and the tagdic. There the listing of form types is alphabetical: e.g. the past participle (pp) comes before the present tense (ps) which comes before the past tense (pt). In the etymological listings the following conventions are adopted:

(a) Each strong verbis listed with its traditional class-number: e.g. $find/v (with SK3 on the line below) = 'FIND, strong verb, class 3'. This is to facilitate comparison with the standard handbook listings. Within the entry, the listings will be nonalphabetical, but will follow the standard mnemonic 'principal parts', as in the paradigm *findan/fand/fundon/funden*. The order then will be present system (infinitive,

imperative, present-tense forms), past singular, past plural, and past participle, with indicative and subjunctive in that order under each heading; the present participle will follow, and then verbal nouns and other root-derivatives, e.g. *-ung* nouns (as it happens there are none for $find) and derived verbs (e.g. weak K2 $fandian 'tempt, try, test' which is a formation off the past singular stem of $find). These derivatives will be listed separately in the etymological corpus, but cross-referred to the root form, which we will assume to be the strong verb, and flagged with '□'. (In this case the procedure is at least historically justified, as weak K2 is by and large a derivational class, historically akin to the Latin 'first conjugation' in -~re: Lass 1993a, b) In general we will try to keep all derivations under the heading of a single root, with cross-referencing, as one of the things we would like the etymologies to do is give a picture of the internal morphological 'coherence' (Lass 1994: ch. 8) of much of the vocabulary.

(b) Following from this, all derived forms will where possible be etymologised under a putative root, unless this simplex root is not attested in the corpus.

4. Special tags. To facilitate etymological comparison and allow retrieval of major morphological reorganisations, some headwords in the CE will carry additional tags to mark their historical class assignments. Thus strong verbs will be marked according to the traditional 'classes' or 'ablaut series' as in the handbooks, and preterite present and 'anomalous' verbs will also be tagged. Examples:

$find/v 'verb'
SK3 'strong class 3'

$sit/v 'verb'
SK5WP 'strong class 5, weak present'

$can/v 'verb'
PP 'preterite present'

$will/v 'verb'
A 'anomalous'

'K' will be used, as elsewhere in *LAEME*, for 'class'. Weak verbs will also be marked according to class, except for the commonest, class 1, which will be unmarked; absence of the marker S 'strong' means that a verb is historically weak:

$think/v

$love/v
K2

$have/v
K3

(Weak class 3 is marginal in OE, and does not show the characteristic markers with any clarity: cf. OE *habban* 'to have' vs OHG *habēn* with its characteristic *-ē-*; but its morphology is idiosyncratic enough to justify special labelling.)

Nouns of Germanic origin will be marked according to the commonest attested Old English stem-class and gender: e.g. 'heart' will be tagged as $heart/n with WF on the line below = 'weak feminine', and $soul as /n with SF on the line below = 'strong feminine'. Which particular subclass of strong or weak nouns a given one belongs to (e.g. *a*-stem, *u*-stem, consonant-stem) will be indicated not in the headword but in commentary. This is again a mnemonic convenience, which also serves to identify

changes in class-membership and morphology: e.g. WF (weak feminine) should predict a gen sg and nom/acc pl in <-E(N)>, but $heart often shows the <-ES> characteristic of the masculine/neuter *a*-stem declension, and $soul shows both this and weak <-(E)N>.

References

Baldi, P. (ed.) 1990. *Linguistic change and reconstruction methodology.* Berlin: Mouton de Gruyter.

Bybee, J. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

Campbell, A. 1959. *Old English grammar*. Oxford: Clarendon Press.

Campbell, L. 1990. Mayan languages and linguistic change. In: Baldi 1990: 115–32

Clark Hall J.R. & Merritt, H. 1960. *A concise Anglo Saxon dictionary*. 4th edn. with aa supplement by Herbert D. Merritt. Cambridge: Cambridge University Press.

Hogg, R.M. 1992. *A Grammar of Old English volume 1 Phonology*. Oxford: Blackwell.

Hymes, D. (ed.) 1974. *Studies in the history of linguistics: traditions and paradigms*. Bloomington: Idiana University Press.

Kastovsky, D. & Szwedek, A. (eds.) 1986. *Linguistics across historical and geographical boundaries: in honour of Jacek Fisiak on the occasion of his fiftieth birthday*. 2 vols. Berlin: Mouton de Gruyter.

Kiparsky, P. 1974. From paleogrammarians to neogrammarians. In: Hymes (ed.), 1974: 331–45.

Labov, W. 1981. Resolving the neogrammarian controversy. *Language* 57: 267–309.

LAEME = Laing, M. & Lass, R. 2008–

Laing, M. & Lass, R. 2003. Tales of the 1001 Nists. The Phonological Implications of Litteral Substitution Sets in 13th-century South-West-Midland texts, *English Language and Linguistics* 7.2:1–22.

Laing, M. & Lass, R. 2008–. *A Linguistic Atlas of Early Middle English 1150–1325* — electronic text corpus with accompanying software (Keith Williamson) index of sources and theoretical introduction. http://www.lel.ed.ac.uk/ihd/laeme1/laeme1.html. Compiled by Margaret Laing and Roger Lass. Edinburgh: ©The University of Edinburgh 2008–

LALME = McIntosh, A., Samuels, M.L. & Benskin, M. (eds.) 1986.

Lass, R. 1972. Review of P.H. Reaney, The origins of English surnames. *Foundations of Language* 9.393-402.

Lass, R, 1986. Words without etyma: Germanic 'tooth'. In: Kastovsky & Szwedek (eds.) 1986: 473–82.

Lass R. 1991. Of data and 'datives': Ruthwell Cross *rodi* again. *Neuphilologische Mitteilungen* 92: 395–403.

Lass, R. 1993a. Old English *-ian*: inflectional or derivational? *Vienna English Working Papers* 2.1, 26-34.

Lass, R. 1993b. Old English class II: More VIEWS. *Vienna English Working Papers* 2.2, 104-111.

Lass, R. 1994. *Old English: a historical linguistic companion*. Cambridge: Cambridge University Press.

Lass, R. 1997. *Historical Linguistics and Language Change*. Cambridge University Press.

Lass, R. & Anderson, J.M. 1975. *Old English phonology*. Cambridge: Cambridge University Press.

Lass, R. and Laing, M. 2005. Are front rounded vowels retained in West Midland Middle English? In Ritt, N. and Schendl, H. (eds.), 280-90.

McIntosh, A. 1956. The analysis of written Middle English. *Transactions of the Philological Society*, 26–55.

McIntosh, A., Samuels, M.L. & Benskin, M. (eds.) 1986. *A Linguistic Atlas of Mediaeval English.* 4 vols. Aberdeen: Aberdeen University Press and Edinburgh: Mercat Press.

Masica, C.P. 1991. *The Indo-Aryan Languages*. Cambridge: Cambridge University Press.

Morpurgo Davies, A. 1998. *Nineteenth-century linguistics.* History of linguistics 4. London: Longman.

Mugglestone, L. (Ed.) 2006. *The Oxford history of English*. Oxford University Press.

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

Poussa, P. 1993. Relitivization and settlement history in North Norfolk. PhD diss, University of Sheffield, unpubl.

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

Serjeantson, M.S. 1935. *A history of foreign words in English*. London: Routledge & Kegan Paul.

Townend, M. 2002. *Language and history in Viking Age England: linguistic relations between speakers of Old Norse and Old English*. Turnhout: Brepols.

Townend, M. 2006. Contacts and conflicts: Latin, Norse, and French. In Mugglestone 2006: 61-85.

Van Coetsem, F. 1988. *Loan phonology and the two transfer types in language contact*. Dordrecht: Foris.