A Corpus of Narrative Etymologies
from Proto-Old English to Early Middle English (CoNE)
and accompanying Corpus of Changes (CC)

Introduction: concepts and terminology

All of our reconstructions are hypotheses; we do not *know* that they ever existed, yet we posit them in order to explain certain facts. (Fortson 2010: 40)

The only explanation of a linguistic form is an older linguistic form. (Lehmann 1952: 23n)

Table of Contents

11.2 Identifying loans
11.3 'Scandinavian influence'

See also the additional explanatory documents:

- About CoNE
- References
- Conventions and Abbreviations
- Template Description
- Grammel Commentary
- Change Descriptors
- LAEME Lexical Specifiers
- LAEME Text Keys

1. What is CoNE?

CoNE provides etymologies for the lexemes and grammatical elements of Germanic vocabulary attested in the *Linguistic Atlas of Early Middle English* (LAEME) Corpus of Tagged Texts (CTT). It differs in kind from earlier work in that it etymologises not simply headwords but all variant forms of a given item. This is accomplished by invoking a Corpus of Changes (CC), a subset of which is shown explicitly to operate in each etymology.

1.1 Introduction

CoNE, then, is a corpus of etymologies of a special kind (see §§2–3 below), conjoined with its sub-corpus of linguistic changes (phonological, morphological and orthographical) that bring the etymologies into being. Each change included in CC is identified by a 2–5 character abbreviation in (( )), e.g. ((IU)) for *i*-Umlaut. This serves as a mnemonic for unpacking etymological sequences, and as a link from the etymological histories proper to the CC. The organisation of these abbreviated names in CC is alphabetical rather than chronological, since it would be impossible to produce a relative chronological ordering of 'all the phonological and morphological changes from AD 350 to the end of the period covered by LAEME': see further this Introduction §10.5. There is also a set of Special Codes that act as further exegesis or annotation to etymologised forms, e.g. identifying them as unchanged descendants of Old English forms, scribal errors, etc. These are identified by short abbreviations in ([ ]), e.g. ([FFE]) Figural Formation Error. The material etymologised consists of the Germanic vocabulary in the LAEME CTT. Individual tagged texts are referred to by their filename and text number in LAEME: e.g. the text sample in the LAEME CTT taken from Cambridge, Corpus Christi College 402, *Ancrene Wisse* is referred to as text # 272 corpart.tag. For general instructions on the use of CoNE see the additional explanatory documents. A good deal of the necessary information for use is transparent, and it is possible at least to get started using CoNE on the basis of the website labels and instructions and the material in this Introduction.

The etymologies consist of two parts: a linear etymological sequence (either single or branching) leading to the 'canonical' Old English form(s) that may be presumed to be input to the Middle English material; and etymologies of both the phonology and morphology of all LAEME CTT form types, and where of historical or linguistic interest the orthography. In this way every formal variant in the CTT will be accounted for. Within the Old English etymologies and the CC, as well as the notes to various etymologies, all reconstructed forms are starred and in bold type; all orthographic citations from whatever language (except CTT items: see below) are in italics; the transcription of orthographic forms are in square brackets; and all glosses and names of lexemes are in small caps. Individual texts and descriptions of individual Changes and Special Codes are accessible through hyperlinks.

All CTT forms (unlike Old English forms or Middle English forms from outside the corpus, which are italic) are cited in LAEME internal format. For full details see LAEME, Introduction, Chapter 3, §3.4. Briefly, transcriptions are made using upper case for 'plain text' manuscript letters. Thus manuscript *herte* is transcribed HERTE. We reserve lower case letters in transcriptions for three different functions: the traditional expansion of abbreviations (see LAEME Introduction, Chapter 3, §3.4.5.1); non-alphabetical diacritics (see §3.4.9); and transcription of non-Roman letters: ae = æ (aesc), d = ð (edh), g = ᵹ (insular 'g'), w = ƿ (wynn), y = þ (thorn), z = ȝ (yogh). CTT forms are organised according to their individual 'tags'. A tag consists of a lexical element (lexel) and a grammatical element (grammel). A grammel indicates the form's part-of-speech and, where relevant, its morphological structure. An argument's grammel may also indicate its function in context.[1] Lexels are normally the Modern English equivalent of the Middle English word. Where there is no suitable or unambiguous Modern English equivalent an Old English, Old Scandinavian, Middle English or composite etymon is used (see LAEME, Introduction, Chapter 4, §4.3). The shape of a tag in LAEME is $lexel/grammel_FORM. In lexels the capitalisation conventions for Old English letters are reversed: e.g. $bough/n_BOz, $la:Yian/viK2_LAd+I. For a full description of the structure of LAEME grammels see LAEME Introduction, Chapter 4, §4.4, and for a checklist see LAEME, Auxiliary Data Sets, Keys to LAEME Tags, Key to the Grammels. For details of notations, conventions, abbreviations, the full shape of an etymology and a list of the texts in the LAEME CTT with filenames and numbers, see the additional explanatory documents.

Vowel length is represented in non-orthographic forms with a colon; in orthographic forms it is represented according to the convention for the language in question, e.g. macron for Old English or Latin, acute for Old Norse.[2] Consonant length in non-orthographic forms is not represented, as it would be in phonetic fieldwork, with a colon, but by doubling of the consonant character. This is inconsistent with our convention for vowels, but we use this convention largely because it has been traditional since C19, and is used in all the major handbooks, even the most recent. This tradition is partly due to the early influence of spelling on historical reconstruction: all the old Germanic orthographies use double graphs for long consonants, and this led to the process of consonant lengthening being called *Geminierung*

---

[1] E.g. $man/nOd = MAN/noun object direct. $ always introduces a tag and / separates lexel from grammel.
[2] Strictly speaking, what is represented is 'editorial' vowel length, i.e. the citations given in the literature for lexical or etymological identification. Neither Latin or Old English used macrons in their writing.

DOUBLING,[3] lit. TWINNING or *Verdoppelung* DOUBLING in the handbooks. As a result of this German tradition the normal term in English for consonant lengthening is *gemination*.

Litterae ('letters' as superordinate abstract elements) are cited in ' ' (e.g. 'a', 't': see LAEME, Introduction Chapter 2, §§2.3.1–2 for discussion of the theory of *littera*). Glosses and lexeme identifications are in small capitals except when lexels from the LAEME corpus are directly cited, in which case they will be in plain type with preceding $, as in the tagged texts.

CoNE will list the actual form *types* occurring in the CTT, 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 ((PNRO)) pre-nasal rounding, so from that point of view all occurrences of HOND are interchangeable, and the form type is treated only once in the etymology of $hand/n, regardless of how many tokens actually occur. Frequencies of occurrence of different forms can be obtained from individual text dictionaries in LAEME.

As historians our overriding concern is that information must not be lost; so even apparent nonsense forms are 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 are commented on and we attempt to explain those we can. Even those coded ([PSE]) (for 'probably scribal error') often carry an explanation of how the error might have occurred; others are coded ([NEF]) (for 'no explanation found') and left for posterity. This does not imply of course that we do not 'believe in' scribal mistakes; but we think that scribal errors are less common than is generally believed, 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.

1.2 Terminology and concepts

1.2.1 Introduction
Most of the terminology and concepts used in the etymologies and the CC will be familiar to anyone with some background in linguistics, particularly phonetics and the elements of historical linguistics. Some however is highly specialized, belonging to the domain of Indo-European and Germanic historical linguistics. Since this terminology may be unfamiliar, we

---

[3] Ultimately < L *geminātio* TWINNING, cf. *gemellus* A TWIN, astrological *Gemini*. Some scholars consider gemination and length to be different linguistic properties: so Hogg (1992b: §3.3.1.4) on distributional grounds, but we do not find his argument convincing. He does not make it clear what the phonetic or phonological distinction is, and argues primarily on the basis of variation between -CC and -C spellings in word-final position. But since in our model long vowels and diphthongs are VV, it is difficult to see what difference there can be between say [tt] and [t:] except a notational one. The problem here is that it has generally been a western Indo-European tradition to mark length on consonants by doubling, but not on vowels. In fact generally vowel length is not marked consistently in Indo-European orthographies (unlike in Finnish or Hungarian); though there is (rather rare) vowel doubling in some Old English texts, and it is regular in Dutch in closed syllables, and to some extent in German. These languages however do not have long consonants, but use doubling diacritically, to indicate a preceding short vowel. It could even be said that the colon for length in modern phonetic transcription is really a marker of gemination, since e.g. [e:] means 'the value [e] carried over to another mora', i.e. [e:] = [ee]. We make this assumption throughout.

give a brief guide to some terms and concepts specific to the field. The material most likely to be problematic is in the domain of morphology and word-formation.

1.2.2 Noun classification

One potentially troublesome area is the taxonomy of nouns and adjectives by stem-classes, e.g. *a*-stem, *ō*-stem, etc. These classifications are opaque without some knowledge of phenomena belonging to an earlier stage of Germanic, and indirectly to PIE. We give a brief overview of older Indo-European and Germanic word-structure.

A noun is normally conceived of as a complex structure, consisting of a root (a string carrying the lexical content), a stem, consisting of the root plus (in most cases) an element usually called a 'thematic vowel' (typically not meaningful),[4] and an inflectional ending. So OE *stān* nom sg STONE, an *a*-stem, is reconstructed as having a Proto-Germanic structure like:

*stɑin (root) -ɑ- (thematic vowel) -z (inflection)[5]

The root plus thematic vowel constitute the stem, and the stem with its inflection constitutes the word. Most nouns are thematic, but there are some in which the endings are added directly to the root, so called athematic or 'root-nouns' or 'consonant stems'.[6] Nouns do not show their (a)thematicity directly in attested Germanic, since by the time of the earliest records the tripartite structure shown above has been largely dismantled and is visible mainly in fragments; but examples from a more archaic Indo-European language – Latin – will illustrate. The parts are not labeled but the structures should be apparent:

hort-u-s GARDEN vs re:k-s (spelled *rēx*) KING

So the terminology in general use for Old English is archaistic, referring to a stage of the language before the emergence of the attested Old English forms. Moreover, a good deal of the earlier history is simply invisible, and the Old English forms lose their historical context. The conventions, however, are assumed in the specialist literature, and without some grasp of their meaning much in the handbooks, in CoNE and in CC and its annotations will be opaque. The Germanic nouns and adjectives are classified according to stem, and stem-class largely determines the shape of inflectional paradigms (e.g. an *a*-stem has a genitive sg in -*s*, and an *ō*-stem has one in -*e*). Gender is given as part of the classification, since most stem classes contain nouns of at least two genders, though a few have only one, like the Germanic *ō*-stems, which are only feminine. To illustrate once again from Latin, the *o*-stems (what the textbooks call the 'second declension') contain both masculines (*hort-u-s*) and neuters (*dōn-u-m* GIFT). The name *o*-stem, like Germanic *a*-stem, has an archaic reference

---

[4] In modern Indo-European studies 'thematic' refers only to the ablauting short vowel **e/o**, not any others (see Fortson 2010: 84). We will use the term more widely here, as is often done in Germanic studies and general linguistic discourse. See Trask (1996 s.v. thematic vowel).
[5] Technically, the thematic vowel and inflection had already coalesced in Germanic (cf. Hogg & Fulk 2011: §1.2). We keep them separated here for historical clarity; we are not portraying a strictly synchronic structure.
[6] Verbs as well may be thematic or athematic, but this is not the basis of classification in Germanic linguistics. A clearer idea of this distinction can be obtained from any Latin grammar, by comparing the verb conjugations.

here, as the thematic vowel is IE **-o-** which raised to [u] in Latin.[7] This way of thinking and representing things is natural to etymological discourse.

Adjectives are also described in the handbooks as belonging to stem-classes, but the membership is not quite the same. In particular, most adjectives have multiple membership, which is not usually very clearly explained. Thus in Wright & Wright (1925: §424) the adjective *glæd* GLAD is said to be an *a-*, *ō*-stem. The reason for this is that adjectives do not have independent declension, but agree with the noun they modify. A noun normally has one gender, but an adjective normally has three, since it has to concord with masculine, neuter and feminine nouns. So in the strong masculine and neuter (see following section) *glæd* takes on a modified form of the *a*-stem declension, and in the feminine it behaves like a modified *ō*-stem. Thus masculine and neuter nom sg *glæd*, feminine nom sg *gladu, -o*. What we mean by 'modified' forms of noun-class paradigms will be discussed in the following section. Similarly, in Latin the adjective meaning GOOD is normally cited as *bonus, bona, bonum*,[8] the nom sg forms that concord with nouns of the three genders.[9]

1.2.3 Strong vs weak

The terms 'strong' and 'weak' are used in Germanic linguistics in at least three different ways, which can be a source of confusion. There are strong and weak nouns, adjectives and verbs.

The strong/weak distinction is probably most familiar in the verb. A strong verb is one whose past tense is formed by change of the root vowel (e.g. PDE *ride* vs *rode*). A weak verb is one whose past is formed by a 'dental suffix', e.g. *walk* vs *walk-ed*. It is important to note that 'strong' does not mean 'irregular': PDE *buy* is weak, because its past does in fact retain the dental suffix, if in a disguised form: the historically correct segmentation is *bough-t*. The strong verbs (and a few other small classes we will not deal with here) are an Indo-European inheritance; the vowel-alternations are due to Indo-European ablaut, which is a system of vowel alternations that is pervasive through the lexicon (see Lass 1994: ch. 5 for an introduction). The vowel alternations involve more categories in Old English than they do in Modern English, since the past sg and pl are distinguished. The typical 'principal parts' of an Old English strong verb (*helpan* HELP) would be:

Present	Past sg	Past pl	Past Participle
help-an	healp	hulp-on	holp-en

---

[7] To illustrate the intricacy of the observation and argument that support our reconstructions, consider first that the Latin *o*-stem = Gmc *a*-stem equation is borne out by a general correspondence: IE **o** usually emerges as Gmc *a*: L *octō* EIGHT vs Go *ahtau*: thus the *o*-stem/*a*-stem correspondence. And the entire nom sg ending, which has been lost (or nearly so) appears in early borrowings into Finnish: *kuningas* KING, corresponds to ON *konungr* where -*r* is a remnant of **-ɑ-z**, and to OE *cyning*, where the suffix is completely gone. If the Finnish form did not exist, our reconstruction on other grounds would be **kuning-ɑ-z**, which suggests how good our methods are at their best.

[8] The -*us*, -*um* forms are IE *o*-stems = Gmc *a*-stems, and the -*a* form is an IE *ā*-stem = Gmc *ō*-stem. So the paradigm structures for the adjectives are historically unchanged.

[9] In Latin and many other Indo-European languages it is customary to cite the genders in the order M, F, Nt; in Germanic the customary order is M, Nt, F. This is probably because the paradigms of the M, Nt are nearly identical in Germanic, and less alike in Latin and other languages. For Old English, the handbooks vary, but in CoNE we follow Campbell (1959) in citing the genders in the order M, F, Nt.

(The present system is represented by the infinitive, as is customary in the handbooks.) This verb has of course become weak, though strong forms remained as late as C17: 'He hath holpen his servant Israel', Luke 1:54 in the 1611 Authorised Version.

A typical weak verb (*dēman* JUDGE) would have the principal parts:

Present	Past	Past Participle
dēm-an	dēm-de	(ge-)dēm-ed

The strong vs weak noun distinction refers to inflectional paradigms, and ultimately class membership. We illustrate with a strong masculine *a*-stem noun, *fisc* FISH and a weak masculine *n*-stem, *hunta* HUNTER:

SG	Nom	fisc	hunt-a
	Acc	fisc	hunt-an
	Gen	fisc-es	hunt-an
	Dat	fisc-e	hunt-an
PL	Nom	fisc-as	hunt-an
	Acc	fisc-as	hunt-an
	Gen	fisc-a	hunt-ena
	Dat	fisc-um	hunt-um

The weak nouns (also called *n*-stems because their stem-forming element contained *-n-)[10] are clearly much less grammatically informative than the strong. The weak declension has virtually disappeared: it was common in Old English and gradually became less common throughout Middle English, though the type was analogically extended to some strong nouns in C16–17, e.g. *shoon* SHOES (OE *scōs*). Now, since the case system has been obliterated, the only part of the weak noun that can remain is the nom/acc pl ending: there is only one general English example left, *ox-en* < OE *ox-an*, and the 'double plural' *child-r-en*. Others, e.g. *een* EYES, *shoon* SHOES, survive in Scots.

The strong/weak distinction in adjectives is related to the nominal one, as we will see below. But the use of this terminology with adjectives can be rather confusing, and some newer handbooks tend to substitute the terms 'definite' and 'indefinite'. But the older terms are more familiar and widely used, and therefore less confusing and more in keeping with most of the literature, so we retain them. Basically, the Germanic adjective developed two types of declensions, which were largely syntactically conditioned: the form of the adjective depended on properties of the phrase it was in. Some trace of this system remains in all Germanic languages except English; it will perhaps be easy to illustrate the essential points from a language with a simpler system than Old English, German. The noun *Bier* BEER is neuter. The adjective *gut* GOOD can take two forms, depending on what comes before it in the

---

[10] Cf. the Latin type *hom-ō* MAN, gen sg *hom-in-is*. In forms of this type the thematic element can be complex, i.e. not just a vowel, but a -VN- sequence.

noun phrase: so *gut-es Bier* GOOD BEER, *ein gut-es Bier* A GOOD BEER vs *das gut-e Bier* THE GOOD BEER. The first two forms are strong, and the third weak. Note that the foundation of this system is degree of informativeness: when there is no determiner, or the determiner is not marked for gender, the adjective takes a gender-marking ending (the neuter *-es*); where the gender is marked in the phrase (by the neuter determiner *das*), then the adjective does not mark it.

The point is that the strong declension is more grammatically informative, the weak nearly without grammatical content. We can see this by comparing the strong and weak endings of an adjective, in this case *blind* BLIND, strong masculine vs weak (which does not vary with gender):

SG	Nom	blind	blind-a
	Acc	blind-*ne*	blind-an
	Gen	blind-es	blind-an
	Dat	blind-*um*	blind-an
PL	Nom	blind-e	blind-an
	Acc	blind-e	blind-an
	Gen	blind-*ra*	blind-ra, -ena
	Dat	blind-um	blind-um

The endings are not identical to the noun endings. Those italicised in the strong adjective are derived from the pronoun declension: compare dat sg *-um* and the masc dat sg determiner *þǣm*, where *-m* is the sign of dative non-feminine: cf. German dat sg *dem*. The same holds for acc sg: determiner *þone* where *-ne* is the case marker for accusative masculine: cf. German acc sg *den*. It is clear from the noun examples given earlier that almost all of the weak adjective forms are derived from the weak *n*-stem declension, except for *-ra* which is generally thought to be an analogical transfer from the strong declension.

2. The Nature of Narrative Etymologies

2.1 Introduction
LAEME, like any historical linguistic atlas, maps change in time 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. LAEME, Introduction, Chapter 2, §2.4). CoNE aims in the first instance to provide an additional layer of historical narrative to the spellings recorded in LAEME. Its methodology can be used to handle any corpus of material subject to etymology. This project is a beginning. It would in principle be possible in the future not only to tackle the non-Germanic vocabulary in LAEME but also historical spellings accessed from any period of English or indeed any other historical linguistic source.

2.2 'Standard' etymologies

Before we describe exactly what we are doing in CoNE, 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*.[11]

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 C16 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 C18 this was commonplace: OED cites Watts *Logic* I.iv. § 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. 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 OED (taken from the 1989 edition, which best exemplifies traditional usage). This is presented as a single block of 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, sǽton, geseten*) = OFris. *sitta* [...] 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*.

---

[11] 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 C19, 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.

III. The stem *set-, pre-Teut. *sed-, is widely represented in the cognate languages, as in Lith. *sedeti*, Lat. *sedēre* [...]

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.[12]

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[13] 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-European 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 -VCC syllable created by gemination, have to be known or inferred. For the non-present forms, only OE *sæt* is relevant to Modern English since the rest of the paradigm has been lost.[14] If the original ablaut were to be reconstructed so that PDE *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).[15]

---

[12] At least in the phonological sense: there is usually semantic narrative, often – as in OED – highly detailed. Our concern here is only with phonological, morphological and orthographic etymology.

[13] The use of italics for reconstructed forms is not CoNE practice. We do it in this section to follow OED usage and keep the discourse unified.

[14] The other forms, however, will also be relevant to older periods, including that covered by LAEME.

[15] For an introduction to ablaut and the sense of terms like '*e*-grade', etc. see Lass (1994: ch. 5).

3. CoNE etymologies

3.1 Etymologies as narratives

By contrast with the standard model, a CoNE etymology unpacks and makes explicit the narrative that brings each particular attested form into existence. 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 temporal level of early Middle English, even Proto-Germanic or Proto-West Germanic, let alone Indo-European, are relevant primarily to support the larger-scale shapes of reconstructions. An example of this support would be deciding whether an Old English final -*er* represents the expansion of an original final consonant cluster, or whether the vowel is original. So in *æcer* FIELD, the Latin root-structure suggests original root-final ***-gr-** (nom sg *ager*, gen sg *agr-i*). On the other hand, *ofer* OVER is cognate to L *super*,[16] where the Latin form suggests that the -*er* is original. 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 began to settle the British Isles in C4. 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. Older and less closely related cognate material may also be useful in supporting reconstructions, or making sense of what seem odd forms or alternations. Cognate lists as such belong to what philosophers of science sometimes call the 'context of justification'; they are not integral to etymological narrative proper.

3.2 The shape of an etymology

Our definition is not original in principle: something of the sort is assumed, at least implicitly, by many historical linguists today.[17] Rather, it is a replacement and enrichment of the standard classical and postclassical definitions of an etymology. Our etymologies are indeed still sets 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 tag represented by one or more text-forms in LAEME CTT (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 and assigned a phonological structure. The conceptual framework is essentially 'Darwinian': the underlying model (or metaphor) is a population of variants moving through time, with innovation and 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

---

[16] This form should really be represented *s-uper*: the Old English form is cognate only to what follows the hyphen. The initial *s-* is an example of '*s*-mobile', the prefixation of an ***s-** to an Indo-European root with no apparent semantic implications. For discussion see Lass (1994: §5.6).
[17] 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).

(word, affix) as a heritable character[18] like any other, e.g. blue eyes, the 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 'error' in the medium of replication.[19] (There are no perfect proof-readers.) 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 innovation stemming from random variation.[20]

An etymology then is the story of successive replication 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 derivative of the micro-level (segmental) change *d > *t. 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.[21]

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 an *effect* of the contingent stabilisation of variation over time. 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 close to 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 $C > C \sim C' > C'$. (In fact there would in any real-world case be

---

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

[19] The term 'error' 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.

[20] For justification of the use of 'domain-neutral' or 'ontologically neutral' language for describing historical linguistic change, see Lass (1997: 109-23, 290-324, and especially 370-83). For the details of the biological metaphor and the concept of replication Lass (1997: 111-13), Croft (2000).

[21] 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, through processes, which do not involve segment-to-segment mapping proper, but the replacement and manipulation of complex objects. This is typically the case with analogy: see §9 below.

many such sequences of variation; this is just exemplary.) 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 more complex transition.[22]

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 $C \sim C'$. 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.[23] 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.

Such undecidability can be resolved only by the consilient testimony of other witnesses, and by inferences from additional knowledge (if available) of the history in question. One potential 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 $\bar{o}$ (e.g. *gud* for GOOD) alongside expected spellings in *o(o)*.[24] Our knowledge of the subsequent history of the language tells us that the u spellings –

---

[22] 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).

[23] An excellent example is the *i*-umlaut of the Germanic back vowels as shown in the C8 Corpus glossary. There are forms of three types: A. those showing the reconstructed unchanged shape, e.g. *unsmoþi* ROUGH; B. those showing a front-rounded vowel graph (the expected output of ((IU))) but with the umlaut environment still intact, e.g. *unsmoeði*; and C. those showing the mutated vowel and destruction of the environment, i.e. phonologised forms: there are none for ROUGH but an example with Gmc **u** is *mygg* GNAT < **muɣ-jɑ.** See further Lass (1992b: 62). The assumption is that these forms do not represent 'mere' spelling variants, but true variants of the type familiar from sociolinguistics: each writing of a word constitutes an 'informant's utterance'.

[24] Specifically an early and non-northern text. In the North, *gud* might imply [gy:d] and be a result of a change called Northern Fronting.

assuming they represent something like [u:] – are innovative. We infer this since [u:] is the value stabilised in our first reliable phonetic descriptions from C16; and PDE [ʊ] continues a shortening of older [u:].

Etymological trajectories are in general not *deducible*; the logic of storytelling is uncertain, and heavily dependent on chance survival combined with inferential gap-filling. It is generally unsafe to predict the direction of change on 'universal' grounds. Some directionalities are statistically more likely than others, but most are undecidable. Though there are some likelihoods that can be used as guides. 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 ('labour-saving'), e.g. segmental insertions deriving from retiming. An example of apparent addition of material that is in fact assimilatory and inertial is the insertion of a stop between a nasal and a nonnasal consonant, as in the familiar *thunder* < *þunr-*, or the common PDE [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 [ɫ] is not really 'addition' of a segment, but anticipation of the [u]-coloured coarticulatory gesture of the dark [ɫ]). Such 'motivated' changes are distinct in principle and mechanism from 'arbitrary' ones (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, diphthongisations and the like are probably stochastic; they are a function of constant background mutation, and not motivated.

3.3 Interlude: what etymological categories mean
The further back we go in time, the less linguistic information survives, 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 go back to a *single* ancestral item. For instance, it would be quite normal to say that the vowels in PDE *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 PDE *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 certain 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 [ɑ]' are deliberately and necessarily coarse and imprecise (see LAEME, Introduction, Chapter 2, §2.4.2, and §6 below). In any sophisticated sense, 'OE short *a* (as in *catt*)' means as much or as little as 'PDE 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 African and New Zealand English varieties 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.[25]

This is a crucial point, because in our experience more historians than not probably believe that a statement of the type 'OE *a* was [ɑ]', is a legitimate thing to say.[26] 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, and against a background of possible or likely variation. It may very well be that all or most dialects of Old English had a back vowel as one of the reflexes of WGmc *ɑ, but we must avoid claims of uniformity that amount to reification of the language as a single homogenous object. Especially as there is variation not only in space but also in time.

There are of course some cases in which unitary phonetic characterisations can be made cross-dialectally, and hence for 'the language', for reasons other than the necessity in some types of historical discourse for the existence of an invariant subject-matter. We can often reconstruct values which are both fairly precise and cross-varietal. OE *a* for instance can be safely assigned the overall value [ɑ], on the basis of its patterning with back consonants like [x] and back vowels like [u] in the triggering of certain changes like breaking and back umlaut.

As our discussion of variation throughout has made clear, any representation like 'ME *a*' is not an individual variant 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 reify and simplify.

This sounds like an undoing of our project; but considered carefully, it is not. It is in fact nothing more than a cautious and epistemologically responsible attitude, based on both the properties of natural languages and the demands of historiography. It is certainly clear to begin with that we cannot talk safely of a single value for a category necessarily characterising 'Middle English as a whole', any more than we can of contemporary English. Each dialect (manifested for our purposes as an individual text language) had its own value, distinct from at least some others certainly at the level of (irrecoverable) fine detail, and not unlikely in more major characteristics. For instance, the short low vowel at any time and in any text language may have been back or front; if, say, in homorganic lengthening environments such as before [nd, ld] it shows up with *o* or *oa* spellings, we assume raising

---

[25] We are speaking here in a historical phonetic sense. The same phenomena led John Wells to the creation of his immensely useful 'lexical sets' like TRAP, etc., which are of great value in modern sociophonetics precisely because of their location of items in system space and avoidance of phonetic specification (Wells 1982).

[26] This and the next two paragraphs are more or less borrowed from Lass (2006: 22–3).

and rounding to [ɔ:], and hence that the basic value was back. We also assume that probably in all pre-Middle English dialects the value was back, because the evidence from sound change (e.g. harmonic and other effects, natural-class groupings) suggests this. On the other hand, we know that judging from what happened in all dialects of English, standard or nonstandard, during the Great Vowel Shift, this vowel must have been front at the time of Open Syllable Lengthening, since it emerges as a mid front vowel or as a diphthongal reflex of one (e.g. in *same, name*). From this kind of evidence, we project to an 'overall value', as a convenient stipulation in those instances where clearer evidence is not available. We therefore characterise the short low vowel as back in Proto-Germanic and assume that it remained so from until some ill-defined period during early Middle English when its lengthened reflexes still appear to be back. We then assume that at some later point, probably around mid-C13, it fronted.

3.4 A CoNE etymology

As we have suggested, CoNE's etymologies are constructed in a very different way from standard dictionary ones. The etymology section of the entry for *hill*, n. in OED is presented as a single text-block; but as with *sit* the information falls under three distinct headings:

> I. Common Teut.: OE. *hyll* str. masc. and fem. = LG. *hull*, Fris. *hel*, MDu. *hille*, *hil*, *hul*.
> II. OTeut. **hulni-z*
> III. Pre-Teut: **kulní-s*; cf. Lith. *kilnus* high, *kalnas* hill, L. *collis* hill, *celsus* lofty, *culmen* top, ...

Such an etymology is static and ahistorical: a list of presumed cognates, with protoforms for salient levels of temporal resolution, and a minimum of explicit historical narrative.

In CoNE we start, for any item, from its initial reconstructed phonetic shape in Proto-Old English and trace its subsequent shapes through Proto-Old English (for this term see §4.2) and Old English, which are rich in changes, and through the sequence of changes within early Middle English, to our end-point: the set of shapes the item displays in the LAEME texts. Thus for *hill* (cognate L *coll-i-s* shows that it was an *i*-stem),[27] we take as input ***xull-i**[28] (nom sg **-z** < **-s** was lost in West Germanic), and explicate its development, showing a single Old English lineage and three forking developments in Middle English:

---

[27] More accurately, it shows the existence of an *i*-stem form that can be used as evidence. Other cognates show that other stem-types existed for this root. The OED gives a different etymon from the handbooks (e.g. Wright & Wright 1925, Campbell 1959) who consider it a *ja*-stem. There seems however, as far as we can tell, no independent evidence for this, and we follow the OED model. If it were a *ja*-stem the etymology would be longer and more complex; but the output would be the same. Any given root could in principle be assigned to a different declension in any language: this one for instance shows forms without umlaut (i.e. with a back root vowel) in LG *hull* and the related OFris *holla* HEAD.

[28] Recall that in CoNE, pre-orthographic material is represented in bold.

Old English etymology:

*xull-i (*i*-umlaut) > ***xyll-i** (high-vowel deletion) > ***xyll** (x-weakening) > [hyll] > OE *hyll*[29]

Note that the OED etymology given above is that of the nom sg, which 'stands for' the item as a whole, or is its *headword*. In CoNE, following conventional Germanist practice, the headword for a noun or adjective is the nominative singular, and that for verbs is the infinitive (unlike the Latin convention where the headword is the present active 1 sg). For nouns and verbs that have root alternations via umlaut or ablaut, etymologies of a set of 'principal parts' will also be provided, i.e. for the consonant-stem nouns like *fōt* FOOT, dat sg/nom/acc pl *fēt*, or for the strong verbs, where the etymologies of the infinitive, past sg, past pl and past participle will be provided (as in *helpan / healp / hulpon /holpen* HELP).

Middle English etymology:

hyll (western y-respelling) > ME *hull*
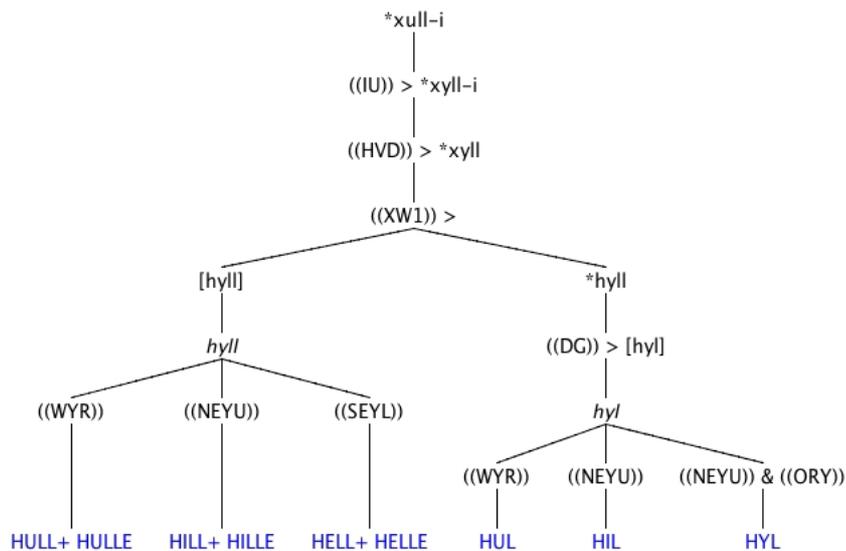*hyll* (south-eastern y-lowering) > ME *hell*
hyll (northern and eastern y-unrounding) > ME *hill*[30]

The three Middle English reflexes of OE *hyll* emerge from separate forks in the pathway, defined by the changes that produced them: *hull*-types are generated by an assumed respelling of OE [y], most commonly and consistently observed in texts from the West Midlands; *hell*-types by south-eastern lowering of OE [y] accompanied by unrounding; *hill*-types by the unrounding that occurred regularly in the North and North-East Midlands. The narrative above (Old English trunk and Middle English branches) constitutes an outline version of the etymology of the item $hill/n in CoNE. Morphological endings have their own separate etymologies, and many items also have derived forms, e.g. denominal verbs, deadjectival or deverbal nouns, compounds and grammaticalisations.

The above etymology, novel as it may seem, is simply a complex branching genealogy. It is a *Stammbaum* or 'family tree' of the type that is used in any stemmatic discipline, like text criticism or evolutionary biology. Each branch represents a lineal split, and each mode an innovation that produces a new lineage member.

---

[29] The change-names in parentheses are for expository purposes. In CoNE itself they are hyperlinks.
[30] The full Middle English etymology will also have a morphological component in CoNE, but we use only the phonological one here for simplicity of illustration. The typographical representation will also be different (all Middle English forms are given in CoNE in LAEME format and all changes are given as hyperlinks); but we are concerned here only with the basic principles, not the presentational details.

```
                              *xull-i
                                │
                       ((IU)) > *xyll-i
                                │
                       ((HVD)) > *xyll
                                │
                          ((XW1)) >
                    ┌───────────┴───────────┐
                 [hyll]                    *hyll
                    │                        │
                  hyll              ((DG)) > [hyl]
          ┌─────────┼─────────┐              │
      ((WYR))   ((NEYU))   ((SEYL))         hyl
          │         │         │      ┌───────┼───────────┐
          │         │         │   ((WYR))  ((NEYU))  ((NEYU)) & ((ORY))
          │         │         │      │         │            │
      HULL+ HULLE  HILL+ HILLE  HELL+ HELLE   HUL       HIL          HYL
```

4 The Entry point

4.1 Preliminaries

In writing an etymology we naturally have to specify an input form; and this ideally involves establishing a temporal location. At the period we are working in, the reconstructed 'named' languages (Proto-Germanic, Ingvaeonic, etc.) cannot be given an accurate temporal location and extension. For instance, Proto-Germanic can probably be located only in terms of 'at least as old as'. The placename *Sylva Baconis* appears in Caesar, *de Bello Gallico*. If *Baconis* (as is generally assumed) represents the IE root **bhVg-** BEECH, OAK (cf. L *fāgus*) then the name is Germanic. Therefore Proto-Germanic must be at least as old as the date of composition of this text, which somewhat antedates 46 BC, when it is first mentioned. Of course Proto-Germanic must in fact be much older than that to have fully developed its characteristics as a definable branch of Indo-European, but we have no way of telling by how much. What are apparently Germanic peoples, according to their ethnonyms, are mentioned as early as the 2nd century BC (Green 1998: 13), but we have no traces of their languages, so the names have to stand as weak evidence that the protolanguage also existed.

We introduced the notions 'ultimate' and 'proximate' etyma in §3.1 above. A CoNE narrative begins with the choice of 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, which for this phase of CoNE will be confined to Germanic loans (except of course for Latin and other loans already nativised in Old English). For native items, an etymological history could be intersected anywhere from the earliest available protolanguage onward. For CoNE the appropriate starting point is the earliest language state that is sufficiently distinct from other Ingvaeonic dialects[31] to be called 'English'. We have selected a very early post-migration

---

[31] See §4.2 for this term.

language state: the linguistic material brought from the Continent starting about C4.[32] This would be recognisably the ancestor (or better the ancestral lect-set) of the collection of text languages we know as 'Old English'. But this language cluster would not yet have undergone those changes that characterise the attested Old English dialects (e.g. breaking), that became the native input to early Middle English, and hence of the forms in the LAEME CTT.

This choice may occasionally be problematic: while no particular purpose would be served by including ancient pan-Germanic changes like Grimm's Law, there are some family-wide developments that appear at least partly convergent in the individual Germanic traditions. Some of these developments are also later than certain Old English-specific changes: e.g. *i*-umlaut, which while visible to some extent in all North and West Germanic dialects of the manuscript tradition, nonetheless must be later than characteristically Old English changes like breaking and first fronting, because their products are affected by it.

We also include changes considerably older than our input stage, if they are necessary for understanding salient properties of Old English (and hence Middle English) forms. These constitute what we call the 'back story'. A typical example is the cluster consisting of Verner's Law and rhotacism. We include this in etymologies when it is the source of otherwise inexplicable morphophonemic alternations, e.g. in the consonantism of infinitive *cēosan* CHOOSE vs past plural *curon*; without these two changes the alternation *s ~ r* would be inexplicable. And without Verner's Law there would be no account for the alternation in the infinitive *weorþan* BECOME and past pl *wurdon*. So however early this change may have been, it is still plays a crucial role in the exposition of Old English morphophonemics.

The most serious problem we face however is deciding the overall content of this input stage. Because of the absence of documentation, we often lack principled grounds for localising particular developments in time. Consider for instance the loss of the nom sg ending in masculine *a*-stem nouns like *stān* STONE. On very good grounds we reconstruct this ending as ***-ɑ-z**, with the final consonant lost in West Germanic.[33] So the input to Old English at some point must have been **staɪn-ɑ*, i.e. stem plus thematic vowel acting as a nom sg affix. But this class of nouns lost its nom sg ending before the attestation of Old English. What is problematic is the time of the loss: had it occurred before Proto-Old English or during? There is no principled way to tell; but we have to decide one way or the other, because in setting up Old English etymologies we use the nom sg as headword according to customary practice. We have a choice between ***staɪn** with the deletion already

[32] At times, for clarity of exposition, we go back further, to the Ingvaeonic cluster from which Proto-Old English emerged, or to the subset of that cluster that consists of English and Frisian. It is a matter of controversy whether there was in fact a genuine 'Anglo-Frisian' subgroup of Ingvaeonic, but we accept it provisionally, since something of the sort appears to figure in many of our etymologies, and there are developments that are shared by English and Frisian but do not occur elsewhere in West Germanic. Typical examples are the reflexes of PGmc ***u(:)** under ((IU)) as *ē/e* shared with Kentish, fronting of WGmc ***ɑ** and velar palatalisation. See Bremmer (2009: §§36–71). The relationship between English and Frisian may however be considerably more complex: the two groups may have been in contact both on the continent and after the invasion of England, and some of the common features may be archaisms in Frisian. On balance however, the claim for a genealogical relationship as the primary cause of the similarities seems necessary, if not an exclusive explanation. For discussion of this complex issue see Bremmer (2009: §§220-25).
[33] It is retained in Gothic, cf. *stains*, and an assimilated version apparently in North Germanic, cf. ON *steinn*).

accomplished,[34] or **stain-ɑ**. On the grounds of a preference for having Old English-typical developments occur as late as possible (we are 'late-preference historians': cf. Lass 1997: 289), we assume loss of the suffix in Proto-Old English rather than take the truncated form as input. This choice factor is again visible in the fact that Hogg & Fulk reconstruct a more archaic dative singular for this noun class than we do.

As we approach the period of textual attestation, location may become somewhat easier: we can increasingly assign changes to particular periods. But there are problems and indeterminacies, many of which, for any given scholar, can have solutions that could only be called 'aesthetic'. Because of serious data-gaps our argumentation cannot be strictly empirical, but rests partly on preference or intuition. These aesthetic preferences can powerfully affect the shapes of the stories historical linguists tell, when – as so often – the outcome is underdetermined by the data. History is after all in one (relatively benign) sense a form of myth-making, and we often construct, when we are not controlled by empirical *force majeure*, the myths that most closely cohere with our intellectual preferences. There are of course different kinds of mythical objects: Zeus vs the loss of the *a*-stem nom sg ending. We are, like any palaeontologists, constructing data-based and largely rationally organised stories (see Lass 1997: ch.1). But we can rarely claim our reconstructions or temporal assignments to be 'true': just that they are better than others, often for instrumental rather than empirical reasons. Note this exemplary statement from Hogg & Fulk (2011: §2.14, n. 1):

> The term Pre-Old English (Pre-OE) is used here to refer to a time when all the PGmc and WGmc changes discussed in Hogg (1992b: chs 3–4) had occurred. It is a variable term, sometimes referring to so early a stage as that in which we find a system of unstressed vowels of the type outlined in Hogg (1992b: §6.1), prior to first fronting and associated changes, both in stressed and unstressed syllables. At other times it may refer to later, though still prehistoric, stages in the development of OE.

This statement shows a rare awareness of this aspect of historians' praxis. It indicates that the kind of approach we have been suggesting is not simply an eccentricity, but is practised by distinguished and authoritative scholars.

4.2 The ancestral inventories
The Germanic subfamily of Indo-European is now usually divided into three major subgroups: East Germanic (Gothic); West Germanic (English, German, Yiddish, Dutch, Afrikaans, Frisian); and North Germanic (Icelandic, Faroese, Norwegian, Swedish, Danish). This division includes of course the ancestors of these modern languages. The consensus is that the original mother language, Proto-Germanic, a Northwest Indo-European dialect, first split into East and Northwest Germanic, and North and West Germanic split from Northwest at some uncertain early time, probably around C3. Northwest Germanic may actually be attested rather than reconstructed; there is an argument that it survived the split as a kind of 'classical' language in which many of the early runic inscriptions that used to be called 'Scandinavian' are written (Antonsen 1975). This would be like the use of Latin as a classical language at the period of the early Romance vernaculars.

---

[34] This is what the rest of the attested West Germanic languages show, and appears to be the basis of its choice in the handbooks, e.g. Hogg & Fulk 2011: §2.14). Nom sg endings occur in this class only in North Germanic.

West Germanic is divided into two major subgroups: Inland Germanic (modern High German and Yiddish) and a group usually called Ingvaeonic. This contains not only the somewhat hybrid modern Dutch and Low German, but also Old English, Old Frisian, Old Low Franconian, Old Saxon and others, which in contact varieties led to the creation of English, Frisian, Standard Dutch, Rhenish Franconian and many other dialects of what would now be called Dutch and Low German.[35] More accurately, Old English is an extraterritorial offshoot of one Ingvaeonic cluster, the result of the insular developments of a complex dialect mixture. Since the dialects that were brought to England in C3–5 have not survived, and since we know that there was a mixture rather than a single 'Proto-Old English', it is difficult to decide how to characterise the input to the precursor of attested Old English. But one thing we can be sure of: it was heterogeneous and its members were and remained in contact for a considerable time. Borrowing took place both from the Continent to Britain and the other way: cf. OS *steorra* STAR, which is an Old English loan. (For an excellent study of Old English/Old Saxon relations see Adamczyk 2010.)

But the dialects of the settlement period seem to have fallen together sufficiently so that eventually there arose a relatively homogeneous language. At least there did in the informal sense that there existed a spectrum[36] of varieties, which could all sensibly be called 'dialects of Proto-Old English' and not of some other reified language. This covers roughly C4–C8 and has various names in the philological tradition. Hogg (1992a) uses Prehistoric or Proto-Old English. Hogg & Fulk (2011) use Pre-Old English. Older works (e.g. Campbell 1959) use the term Primitive Old English. We will use the term Proto-Old English (abbreviated as PrOE) because the designated language state fits the standard definition of protolanguage in historical linguistic usage. This is the period during which most of the well known sound changes like breaking, *i*-umlaut, etc. took place. The rest of the Old English period has its own nomenclature. Again, we follow Hogg: Early Old English 700–900, Classical Old English from the end of the Alfredian period to about 1000, and Transitional Old English, the period when the Classical tradition is beginning to break down (evident from changes in the inflectional system, merger of unstressed vowels), from early C11. In the CC, however, we conflate the later periods in assigning times to particular changes, and simply use the designation Old English, since (despite the divisions) we are nevertheless dealing with (varieties of) 'the same language'.

To etymologise forms, we need an input; the problem is that there is, as the story above suggests, no 'single language' that is the input to Old English. Even the sparest marginally credible 'creation myth', Bede's Angles, Saxons and Jutes, is over-simple and inaccurate. There is no doubt that the input was heterogeneous, but equally little doubt that one needs a homogeneous input to construct an intelligible etymology. So what do we do? In essence we take an overview of the settlement-period spectrum and develop it into a generalised dialect, dividing it into 'regional' varieties when appropriate (e.g. West Saxon,

---

[35] For an excellent survey of the history of Ingvaeonic and a collection of texts and a grammar, see Markey (1976).

[36] This term echoes a standard usage in psychiatry, where collections of symptom-clusters are gathered under the heading of a 'spectrum': so autism spectrum disorders, depression spectrum disorders. The term is just loose enough to be valuable in talking about dialect clusters too.

Mercian).[37] Paradoxically, there is no such thing as a dialect-free protolanguage, but one cannot do etymology without assuming one.

We note first the main characteristics of Ingvaeonic:

(i)   Late monophthongization of Germanic ***ai**, cf. OE *bān,* OFris *bēn* BONE vs OHG *bein* (this may fall within the Proto-Old English period).

(ii)   Loss of final ***r** < ***z** with lowering in monosyllabic pronouns, cf. OE dat/acc 2 sg *mē*, OFris *mi*/*mē*[38] ME vs OHG *mir*.

(iii)   Loss of nasals before voiceless fricatives, with compensatory lengthening of the preceding vowel: OE *fīf* FIVE, OHG *finf*. As a subcase, ***an** + voiceless fricative > [o:] (according to the handbooks via nasalisation, but this is not certain): OE *gōs* GOOSE, OHG *gans*.

(iv)   a -Vs ending in the *a*-stem nom/acc pl (OE -*as*, OS -*os*).

(v)   Loss of the three-way person distinction in the verb plural, and levelling into one ending. The original paradigm remains in some modern Germanic languages, e.g. German and Icelandic.

Within Ingvaeonic, the Anglo-Frisian subgroup shows additionally fronting (and raising in Old Frisian) of PGmc ***a** in certain contexts: OE *dæg*, OFris *deg* DAY vs. OS *dag*. These features plus what we know of West Germanic give us an outline of the construction of a sound system for the input to what was later to become Old English in the wider sense.

4.3 Proto-Old English and Classical Old English sound inventories
The consensus seems to be that Proto-Old English had a sound inventory of this type:[39]

i	u	i:	u:		iu		p	t		k
e	o	e:	o:		eu		β	ð		ɣ
ɑ	æ:			ɑi	ɑu		f	θ	s	x
							m	n		ŋ

Note the lack of a low back long vowel. (Germanic ***a:** had already fronted leaving a gap subsequently filled by monophthongisation of ***ai**.) The presence of voiced fricatives as counterparts to the voiceless stops may be an oversimplification: see CC s.v. ((VFH)). We

---

[37] The word 'regional' is in inverted commas because regionality may not always be a secure parameter. For a sophisticated discussion see Hogg (1992a: ch. 1).
[38] This change may belong to the more restricted English/Frisian cluster only. We treat it as such in the CC.
[39] Note that we use the term 'sound inventory' rather than 'phonological system'. This is to avoid committing ourselves to a 'status' for the elements involved. See the following note and the discussion in §5.

have listed ***w** as a labial; it could just as well be a velar since it was almost certainly both, like PDE /w/. All the consonants except ***w** and ***j** could also occur geminate, either deriving from original geminates (usually from clusters) as in ***full** FULL (ultimately < IE ***p̥ln-**), or from West Germanic gemination ((WGG)), as in ***kunn-jɑ** KIN.

      After several centuries of sound change this system was radically transformed; it developed among other things a short/long diphthong contrast, a palatal series and front rounded vowels.

      Classical Old English (just post-Alfredian) had the following inventory:[40]

i	y	u	i:	y:	u:	ĭo	io		p	t		k
e	ø	o	e:	ø:	o:	ĕo	eo		b	d		g
æ	ɑ	æ:	ɑ:	æ̆ɑ	æɑ						ʧ	
											ʤ	
									f	θ s	ʃ	x
									v	ð z		ɣ
									m	n		ŋ
										r l		
											j	w

The short diphthongs consist of two 'half-morae', in combination yielding the length of a short vowel; the long diphthongs (unmarked) consist of two normal morae, so they have the same length as a long vowel.[41] All the consonants except [ʃ, j, w] occur geminate as well. Note also that there is now a series of palatal obstruents where before the only palatal was ***j**.

      Our etymological starting point then is a reconstructed Proto-Old English ca. C4. This is the assumed predecessor of the earliest attested varieties such as those represented in the earliest inscriptions (e.g. the Franks Casket, C7 and the Ruthwell Cross, early C8). These are

---

[40] Most readers will be surprised at the inclusion of voiced fricatives and the velar nasal in this inventory. That is because it is not intended to represent a phonemic system. See discussion in the next section. For the meaning of breves over diphthongs see §7.

[41] This (standard) interpretation has been the subject of controversy, and still is. For an overview see Hogg (1992a: §§2.22-2.40). See further §7.2 below.

already characteristically English language forms, and our reconstructed etyma are distinctly more archaic than these, and even more so than the C8 Northumbrian versions of Caedmon's Hymn or the Corpus and Epinal glosses. Our etyma are given shapes that allow them to be progressively mapped into the phonological and morphological changes that create the attested Old English forms. These include the standard changes that appear in the handbooks (except those that we have reason to reject, which will be noted). They also include newly designated changes that are not always standardly treated. It is the various types of later Old English that are the direct inputs to Middle English. We then itemise the changes that allow the Old English forms to be mapped into the Middle English forms attested in our corpus.

5 Level of representation

In setting up vowel and consonant inventories for various stages of the language in the previous section, we partially follow the praxis of Hogg (1992a: §§ 2.41, 2.79). This is based on the view that 'it might be possible to postulate an overall vowel and diphthong system for Old English, but this would be to ignore many differences which occur both dialectally and diatopically' (§2.41). But as is clear from our representations, we draw a different moral.

Hogg's system displays 'attempt only to present the major phonological contrasts potentially present in OE' (ibid.). We however do not represent just 'phonological' (= phonemic) contasts, since there are no phonemic systems underlying our transcriptions. We represent surface phonetic segments only, and all distinctions are at the phonetic level. We take them as equally 'part of the system'. The reason for this is our particular historical remit. For instance, it is true that within a phonemic discourse framework OE *f* represented /f/, with the voiced allophone [v] intervocalically: hence *lufu* LOVE would be phonemically /lufu/ but phonetically [luvu]. But this will not do for an etymological undertaking that attempts to capture both the phonetic surface and the reasons for particular orthographic choices in Middle English. Even if in one kind of discourse [v] is 'really /f/', we need a source for the Middle English spellings *u*, *v* and what they presumably represent in this word. Its phonemic status is not relevant to our aims.

In constructing etymologies we are writing the histories of surface word-forms, not of anything abstract or 'underlying'. This is especially true because we consider etymologies to be visual as well as phonic objects, and Middle English spelling to be based to a large degree on non-phonemic considerations, either phonetic or purely orthographic (see §7 below). This of course runs counter to the classic dictum that scribes 'write phonemes': but that is simply not a true claim. To take one example, the velar nasal, though not phonemic, has a separate representation in Greek, Gothic and the elder Futhark.

There is also considerable evidence, which will be apparent in many of the spellings cited in CoNE, that Middle English scribes (more so than Old English ones) often lacked a particular interest in distinctiveness as the basis for constructing spelling systems. Thus we see scribes with enormous numbers of spellings for the same category, and/or complex non-biunique symbol/category relations. On the other hand many do not indicate contrasts that we know on historical grounds must have existed, e.g. [eː] vs [ɛː] and [oː] vs [ɔː]. In addition, perhaps the majority of scribal systems are variable, often to a high degree, and there may be frequent overlap, as well as the variable instantiation of changes in progress. Lexical diffusion is often captured in spelling, which overrides any possible attempt, for the category

in process of change, to anchor spelling to distinctiveness. This is of course especially the case when either splits or mergers are in progress. And of course during any diffusional or variational process, phoneme-membership will in any case be fuzzy.

So we try to avoid invoking pan-varietal 'overall patterns', and do not presuppose particular phonemic analyses for 'the language as a whole'. There are likely to have been multiple, if overlapping, systems, and there is no universally accepted 'canonical' analysis. In any case, even if one should want to treat this material phonemically, there are many varieties of phonemic analysis, and not all will yield the same results. Over the years scholars have come to forget 'the non-uniqueness of phonemic solutions of phonetic systems' so elegantly established by Chao (1934).[42] The general view of Old English phonology we adopt here is that given in Lass (1994), which in most details is in agreement with that in Hogg (1992a). Our view of Middle English is primarily informed by our data-source, the LAEME CTT, supported by knowledge also of *LALME* and MED. Where we depart from any handbook consensus, we will indicate it, and if necessary cite work that supports our position.

6. The phonetic primitives used in CoNE

A number of typefaces are used in CoNE, and some confusion about their meanings may arise. This section deals with the historical phonetic categories used in CoNE whether reconstructed or attested, and their representations. That is, it specifies the mode of phonetic representation for any pre-modern category. Bold face signals the supposed value of a reconstructed form and square brackets signal the supposed value of an attested form. The values of phonetic symbols in boldface in an etymology or commentary or in the CC, and those used in square brackets in a phonetic representation of an attested Old or Middle English item or category, are the same. These representations are to be taken, as is normally the case in historical work, as at least partly typological: in Laing & Lass (2003: 276) historical representations of this kind are referred to as 'poorly resolved broad phonetic transcriptions'. But they do use the symbols of the IPA alphabet, if occasionally in somewhat loose or restricted senses, which will where necessary be specified below. More precisely, each symbol, given a latitude not uncommon in professional usage, 'covers a set of event-types that phoneticians could agree to hear as 'the same'' (Lass 1997: 271). That is, we cannot specify fine detail like precise degree of vowel height or voice onset time, but we can characterise a segment in terms of broad phonetic parameters, like height or backness or continuancy or voice. So we can specify a segment type as distinct from any other, and as being capable of being perceived by scribes and therefore represented in their orthographies. Or we can specify it as distinct on historical grounds, e.g. because it has particular phonetic effects (e.g. induces a certain vowel harmony); or on the basis of later reflex types that some other segment-type does not have. Given the necessary coarseness of historical representation, we can still set up a usable and easily interpretable inventory of the phonetic categories we will be using. (Of course we may use much finer transcription of modern examples in commentaries.) The inventory we allow for the construction of historical representations in CoNE and CC is as follows:

---

[42] And see Lass (1981) for examples of intractably non-unique phoneme assignments in a Modern English dialect.

Vowels. The basic categories are front vs back,[43] high, mid and low, and round vs unround. We assume that most of these could occur long or short.

Front Unrounded. High [i], high mid [e], low mid [ɛ], raised low [æ], low [a].[44]

Front rounded. High [y], high mid [ø].[45]

Back rounded. High [u], high mid [o], low mid [ɔ].

Back unrounded. Low [ɑ].

Two omissions will be obvious to the reader. There is no provision for schwa [ə], or for the centralised high mid vowels [ɪ] and [ʊ]. This is because we do not think they existed in Old or Middle English. (This is a controversial position: references to the argumentation are given below.) That is, we take the position that the variable collapse of unstressed vowels in *e* in later Old English and throughout Middle English does not represent 'reduction to schwa' but simple neutralisation in the quality [e] and hence to merger with the already existing phonemic short /e/ (for detailed arguments see Lass 2009; the same point was in fact made by Luick 1914/40: §440). We also claim, again counter to the received wisdom, that long/short pairs were qualitatively identical throughout Old and Middle English, and that there is no compelling evidence for pairings like [i:]/[ɪ] (*beet/bit*) in any variety of English until after C16 (see Lass 1989, 1999a). All of the vowel qualities in the above inventory occur both short and long throughout our time-span, though certain ones of course may only occur in certain periods (e.g. front rounded vowels are a late development, and to a large extent vanish). Certain of the vowel types could cluster and form diphthongs, which behaved as long vowels in Germanic, either as long or short vowels in Old English and Proto-Old English, and fell in with monophthongs at the end of the OE period. In Middle English new diphthongs developed. Even though the diphthongs were 'unit phonemes', they are more appropriately discussed under other headings, since they are basically combinatorial products of the inventory above, regardless of their function. See §§7.2, 10.4.

6.1 Transcription of low vowels

As we have noted, we adhere strictly to IPA transcriptional conventions. In this framework [a] is solely the symbol for a low front vowel, lower than [æ][46], and [ɑ] is the symbol for a

---

[43] Note that we do not posit the category central for OE or ME vowels, as we find no convincing evidence for it (see references below). Some scholars use 'central' in a very odd way, which we will never do. Thus Hogg (1992b: 101) on early Proto-Old English: 'there was only one short low vowel, which may be best analysed as central, since it had no front or back contrasts at this stage'. This makes 'central' a kind of dustbin category, hierarchically inferior to and different in type from back and front. We think, as most phoneticians do, that it is as 'real' as any other degree of backness, and cannot be used this way.

[44] Although [æ] and [a] are qualitatively distinct, they never contrasted within any lect, though transition from one to the other is a feature of the history of at least southern English as a whole.

[45] It is not determinable whether the mid front rounded vowel in any variety of OE or ME was high or low mid. Our choice of a general symbol is to that extent arbitrary, as there is as far as we know no evidence either way.

[46] The symbol [æ] seems to have an odd status in the IPA: it is defined as being between [ɛ] and [a], and is described (IPA 1999:20) as being in a descending series English *head* [ɛ], English *had* [æ], French *patte* [a]. But in the 1966 full chart of the IPA (unnumbered page preceding 3) while the symbols [i], [e], [ɛ], [a] are marked by boldface dots, [æ] appears to sit on a line in the appropriate region of the vowel quadrilateral, but is

low back vowel. Many writers use [a] for whatever low vowel is present, and this can be extremely misleading or at very least make it unclear what a given category is supposed to be phonetically. In some cases, the symbol is used both for a back and a front vowel: Burrow & Turville Petrie (1992: 10-11), in directions on how to pronounce Middle English, speak of a long vowel /a:/ which is to be pronounced like the vowel in English *father*, and a short vowel /a/ which is to be pronounced like the vowel of French *patte*, i.e. they use the same symbol for [a] and [ɑ] and relegate phonetic characterisation to accompanying prose. Aside from putting what they specifically characterise as 'pronunciations' into phonemic brackets, which is common but improper, they give to all of 'Middle English', regardless of place or time, a single value for the low vowels, which is not historically possible (see below).

Such unspecific use of [a] can obscure the description of modern languages too. Many people use [a] for French and Italian and German low short vowels: it would be appropriate to use [a] for the first two, but not for most varieties of German, where the short low vowel is central. The appropriate symbol is IPA [ɐ]. Further, the use of [a] for any low vowel that happens to be the only one in a system (or if there are long and short low vowels, simply using [a] with a diacritic – the colon in the citation above, or the macron as in Wright & Wright 1925), with phonetic characterisation in accompanying prose, is simply confusing. It also apparently makes (or can be read as making) the erroneous claim that the lowness of the vowels somehow takes hierarchical priority over their front/back locations; i.e. that there is only one underlying low vowel which comes in front and back varieties; this implicit claim would obscure important stages in the development of early English vowel system. These have major systemic effects: they add a new vowel type to the non-Northern dialects, and destroy the back/front symmetry that had been the case from Old English to the early C13, leaving both Northern and Southern systems without a back unrounded vowel either long or short.

This restructuring is effected by three changes in C11 and C12: [ɑ:] either raises and rounds in the south to [ɔ:] via ((ARR)) or fronts in the North to [a:] via ((LAF)), and [ɑ] fronts to [a] via ((AF)).These developments were preceded by a merger of OE [æ] and [ɑ] in [ɑ] via ((ÆM)). (See Lass 1992a: §§2.3.1-2.3.2 for an overview of these complex developments, and see CoNE, CC s.v. ((ARR)), ((LAF)), ((AF)), ((ÆM))

In summary, the use of [a] for any low vowel that happens to be the only one in a system (or if there are long and short low vowels, simply using [a] with a diacritic (the colon in the citation above, or the macron as in Wright & Wright (1925)), with phonetic

---

not given a strict position. This may be a function of non-Cardinal vowels being uncertain, because they do not appear in the original recordings by Daniel Jones (these attempt 'auditory equidistance' between vowels, and [æ] becomes a slightly moveable feast because of its apparent 'non-canonical' status, and because it interrupts the visual equidistance of the vowel quadrilateral.) Incidentally, it appears that Cardinal Vowel 4 [a] has drifted rather backer over time – if that is not an oxymoron, since Cardinal Vowels are supposed to be abstract reference points for describing real-language data. (For the theory of Cardinal Vowels and how they are used in relation to actual utterances, see Jones (1967: §§131-133).) This is rather amusing in its own way, as Jones was a mentor and consultant for a project of Peter Ladefoged's, and Ladefoged worked at Edinburgh with David Abercrombie, who was a pupil of Daniel Jones: (See Wikipedia s.v. Peter Ladefoged). For Jones' original recording see https://www.youtube.com/watch?v=6UAe4p2174. Interestingly, in Ladefoged's later recording, [a] is not only lower than in Daniel Jones' canonical one, but retracted (following the shape of the notional vowel quadrilateral). See http://www.phonetics.ucla.edu/course/chapter9/cardinal/cardinal.html.

characterisation in accompanying prose, is no longer permissible in our view: there are unambiguous standard symbols available for front, central and back low vowels. So we are following a standard practise in using [a] solely for a front low vowel. In doing this we reiterate the canonical position of the IPA, which we consider the gold standard for transcription.

Stops.

　　Labial. [p], [b]

　　Dental. [t], [d]

　　Velar. [k], [g].

The categories 'labial' and 'dental' are used broadly, as is typical in historical discourse, where the bilabial/labiodental and dental/alveolar distinctions are often not discoverable or useful. This is true at least of stops and nasals; the situation with fricatives is different, as will be noted below. As far as the 'dental' (or coronal) stops are concerned, there is no reason why different Old English and Middle English varieties should not have varied in the location of their /t/ and /d/ (and for that matter /n/ and /l/). This is certainly the case in Modern English, and there is no reason why it should not have been so earlier.[47] The question of aspiration of the voiceless stops in certain positions is undecidable: aspiration is widespread in Germanic (most varieties of English, standard northern German, Icelandic, Norwegian, Swedish, Danish), but so is lack of it (Dutch, Afrikaans, many varieties of Yiddish, south German). In general aspiration can be reconstructed only where at least some members of a family indicated it in writing (e.g. Greek, Indic), or where it is noted in an early phonetic tradition (e.g. Latin with respect to Greek aspirated stops in loans, C16 London English from the unambiguous descriptions of Hart 1569).

Affricates.

　　Palatal. [tʃ], [dʒ].

On this underdifferentiated use of the term 'palatal' see the commentary on fricatives below.

Fricatives.

　　Bilabial. [β]

　　Labiodental. [f], [v]

　　Dental. [θ], [ð]

　　Alveolar. [s], [z]

　　Palatal. [ʃ], [ç]

---

[47] The first clear description of /t, d/ in a variety of English (educated London, Hart 1569) is unambiguously dental; in the next century (Holder 1669) alveolars are described and the term 'gingival' is used. But this still only gives us information for two lects, and phoneticians' implicit projections to the rest of the language.

Velar. [x], [ɣ]

Glottal. [h]

The term 'palatal' is used, as typically in historical discourse, for both palatals and palatoalveolars. Their phonological behaviour is sufficiently similar to warrant this: e.g. both [ʃ] and [ç] cause diphthongisation of preceding front mid vowels by insertion of epenthetic [i] in Middle English, and both segment types, at different historical periods, result from the palatalisation of velars. The bilabial/labiodental distinction is made because the two categories are presumed to have occurred at different historical periods, and have different historical origins: the bilabial ***β** was the reflex of PGmc ***β** usually from IE ***bh** in certain environments, and later merged with labiodental [v] (see ((LD)) in the CC.)

Nasals.

Labial. [m]

Dental. [n]

Velar. [ŋ]

As far as we can tell [ŋ] occurred only before velars, and [g] remained after it in all positions: thus *singan* SING [siŋɑn], *þing* THING [θiŋg]. The loss of [g] in these positions took place around C17 in the South and South Midlands, and has not occurred in many varieties in the Northwest Midlands.

Liquids.

Non-lateral [r]

Lateral [l], [ɬ]

The symbol [r] in pre-modern citations represents 'a non-lateral liquid' or 'rhotic' as it is usually called, not the precise IPA value of alveolar trill. The nature of the rhotic (there was only ever one in any Germanic language, unlike the case in some languages like Spanish and Tamil) may have varied from variety to variety all through the history of early Germanic, and in Old and Middle English, as it still does. Taps and approximants are likely at certain times and places, trills uncertain, though present and well described in post-Middle English varieties. It is clear however that the category typically represented in the writing of Early English with 'r' was a complex segment, with at least velar and labial and probably pharyngeal components, judging from its effects on neighbouring vowels (see Lass 1983).

Similarly we assume a single lateral that we write [l], without further specification. The Old English lateral was at one point a dark [ɬ] at least in syllable codas, judging from its patterning with ***r**, ***x** and ***w** in breaking (see ((BRK))). Whether any dialects of Old or Middle English had the common modern southern British and South African and southern US

alternation of clear lateral in onsets and dark in codas is not decidable. There is no evidence for or against dialects with dark [ɫ] initially and medially as well (like present-day vernacular Yorks and Lancs, most of the non-southern US, Australasia). The Middle English evidence does not suggest the existence of any varieties with clear [l] in coda position, as in modern Hiberno-English and Tyneside, though post-breaking they may of course have existed.

Approximants.

> Labial/velar [w]

> Palatal [j]

These segments used to be more commonly called 'semivowels', 'frictionless continuants' or 'glides'. The term we use here seems to be taking over, and is used in the newer historical handbooks as well as phonetic discourse (e.g. Hogg 1992a, Hogg & Fulk 2011; see also Trask 1996: s.v. approximant). This is a problematic category phonetically; quite often for instance English /j, w/ are phonetically nonsyllabic versions of [ɪ, ʊ] (see Trager & Smith 1957). But they do behave like consonants in terms of distribution (occurring only in onsets and codas), and historically it is clear that at least in older Germanic and Old English (the story is less clear for Middle English) they constitute a class of segments distinct from nonsyllabic high vowels like the final segment of the diphthong [au]. It is not certain precisely in what respect they were distinct; most likely greater stricture, just short of producing friction, but there are differences for instance in the Middle English outputs of OE *ēo* and *ēow*, e.g. *bēo* BEE vs *hrēowan* RUE. In addition, if *w* in these forms represented a vowel the result would be a triphthong, which as far as we know is a cluster type generally ruled out in the periods we are concerned with.[48]

7 Prosody

7.1 Stress

In our etymologies we invoke changes like weak final vowel deletion ((WFVD)), in which the trigger is lack of stress. This suggests that we have a model of Old English and Middle English prosody that tells us which syllables are weak and which are strong. We do, but it is not very finely resolved, and often somewhat indirect and gapped. There is a striking lack of hard prosodic evidence in the early Middle English materials; we have much solider evidence

---

[48] The scholarly tradition however posits early 'trimoric' vowels in certain environments, a claim based largely on the differential developments of certain sequences involving laryngeals (for details see Fortson 2010: §15.16, Hogg & Fulk 2011: §§2.6, 2.59, 2.71, 6.71, notes). There are phonological problems with these supposedly extra-heavy nuclei, since they typically seem to have occurred in weak syllables. Ringe (2006a: 73-5) argues that they were not in fact trimoric in the literal sense of being VVV, but differed from other vowels in the same positions in some other property, perhaps tonal accent or glottalisation. Though he still uses the term unqualified in Ringe (2006b). It is not clear how late these nuclei survived, but there are certainly phenomena in OE that can be attributed partly to them, e.g. the vowel of the nominal genitive plural (see ((TVR))). The only historic rather than prehistoric and reconstructed triphthongs in English seem to be those ending in [ə] which have been reported in some English dialects, notably RP, e.g. in perceptually monosyllabic tokens of words in /aɪ/plus final orthographic *r* like *fire*.

for Old English and for later Middle English. This arises, among other things, from the fact that Old English and one tradition of later Middle English verse are alliterative; this allows us to locate the main stress of an alliterating word. There is other evidence also, for Old English, that indicates the position both of secondary stress and of lack of stress (see e.g. Hogg 1992a: §§2.84–2.91).

In early Middle English however, metrical evidence is much scarcer; the alliterative tradition had not yet been revived, and the new Latinate syllabic rather than Germanic stress-based verse type was often highly variable, so that it is not easy to discover syllable prominence (or in many cases even syllable-number) from verse evidence.[49] This Latinate tradition is based essentially on stress-defined metrical feet, but in early Middle English the regularity of foot structure in verse is much looser than in later times, and the counting is rarely tightly constrained.

Our practice, which we think is reasonably safe, is to project forward from the rather better-known Old English scansion system, (using, with some discomfort, verse structure as a proxy for linguistic structure), and backward from later stages of the language. We also take into account orthographic evidence like deleted syllables that we know 'should be there' etymologically, under the assumption that deletion of weak syllables is more likely than deletion of strong syllables. And where we feel confident we use as evidence those few writers we have reason to believe were really accurate versifiers, such as Orm.

The dominant prosodic type in early Middle English, as in Old English, was a left-strong foot, which could be described as trochaic or dactylic (the first commoner, as the dominant Germanic word type is disyllabic; trisyllabic feet probably occurred only in inflected forms). The main word stress therefore fell on the first syllable of the lexical root, in accordance with the historical Germanic Stress Shift ((GSS)). This means that in general affixes were unstressed, except in certain well-defined cases, such as particular classes of prefixes (or better prefixoids or prefix-like elements). As in Modern English and other Germanic languages, we operate with three degrees of stress: primary, secondary, and weak (no stress in effect).

Simplex words consisted of one foot; complex words could consist either of one or two feet, depending on the nature of the elements involved. Thus, using Old English examples to illustrate the main pattern and its origin: in *mann* MAN, HUMAN the one syllable, since this is a lexical word, is assigned primary stress by default, and constitutes what is sometimes called a 'degenerate' foot. (We assume that grammatical words like prepositions and pronouns and conjunctions do not normally bear primary stress, but are weaker than lexical words; affixes are unstressed.) In the dative plural *mánn-um* the first syllable is stressed and the second weak by virtue of its being an inflection, and the word is a trochaic foot; in *món-cÿnn-es* MAN-KIND'S the first syllable bears primary stress, the second bears secondary stress by virtue of being the second lexical element, and the third is unstressed. The second and third syllables together constitute a trochee. (Note that we are using terms

---

[49] Even in later verse in 'iambic pentameter' and other metres of this kind, there are permitted variations, e.g. substitution of a trochee for an iamb in line-initial position or at caesura, occasional substitution of three-syllable feet (e.g. dactyls) for two-syllable ones.

like *trochee, dactyl* in the usual English sense, where they refer to stress configurations, rather than the original Classical sense in which they refer to quantity.)

Prefixes are variably stressed. The basic generalisation is that nominal and adjectival prefixes are stressed, and verbal prefixes unstressed: e.g. *ánd-sàca* APOSTATE vs *on-sácan* DENY, where *on-* is a reduced version (the historical origin of both is ***ɑnd-**). There are however prefix-like elements, normally derived from adverbs, which can be stressed when attached to verbs, as in *ínn-gàngan* ENTER. These are the kind of element that appears in German as 'separable' prefixes, e.g. *éin-gèhen* COME IN, *ich gehe ein* I COME IN. Such prefixes usually also exist as independent adverbs, making these items more like compounds than prefixed verbs. There are also some nouns with unstressed prefixes, e.g. those formed from the usual type of verb, e.g. *for-gífness* FORGIVENESS < *for-gífan*. (For further discussion see Lass 1994: §4.2.3.)

In general this description of Old English stress will hold for the Germanic element in early Middle English as well. It is only later, with increased entry into the language of Latin and French loans, that a new type of stressing, essentially right-strong and based on Romance models, comes into the language.

7.2 Weight

Handbooks and textbooks often talk about 'long' and 'short' syllables, which can be slightly puzzling because a syllable containing no long vowel or geminate consonant can be referred to as long. This terminology is fortunately becoming obsolete (though anyone working in the field will meet it). Nowadays the parameter invoked is not 'length' but 'weight', which can be visualised over-simply but usefully in terms of the number of elements the rhyme of a syllable (the nucleus and any following material) consists of.[50]

We adopt the definition of weight used in the standard Old English handbooks (e.g. Wright & Wright 1925, Campbell 1959, Hogg 1992a, 1992b). This is now somewhat old-fashioned and skirts a number of complex and much discussed theoretical issues. But it is the usual conceptualisation of these properties in historical Germanic studies, and we do not consider it desirable here to lay out the modern (controversial and constantly changing) argumentation, which would require introductions to Optimality Theory, Strict CV Phonology and other newer theoretical frameworks. Certain issues appear to us to belong in technical theoretical discourse rather than in an introduction like this. (For an insightful overview of more modern conceptions of the syllable and length and weight than we use, see Zec 2007 and references, especially §8.4.)

Using the word 'mora' in a relatively traditional sense, we say that a mora is a unit of weight (or length, which is close to the same thing), so that a short vowel or short diphthong

---

[50] 'Rhyme', and the hierarchical model of syllable structure it is associated with, has been extensively argued against, and in modern phonological theory (as opposed to linguistically based philology, which we presume is the appropriate name for the genre of CoNE) is not extensively used. We will not go into the extremely complex and healthily unsettled contemporary debates on syllable structure (see e.g. Zec 2007: 173–244). The model we use here is essentially linear, and unlike many modern ones counts consonants as weight-bearing, in the traditional way. See the following discussion.

is monomoric and a long vowel or diphthong bimoric,[51] i.e. the contrast at the structural (not phonetic) level of weight is V vs VV. We take it that consonants are units of weight as well, so that a short consonant is C and a geminate consonant or cluster is CC. (In Germanic, an affricate is normally the prosodic equivalent of a short consonant, so is C.) We then define syllable weight in the traditional way: omitting the onset (if any), which does not count in the measurement of weight, -V, -VC are light syllables, and -VV, -VVC, -VCC are heavy syllables.

Users acquainted with Latin metrics (where VC counts as heavy), or with various forms of modern theory, will find this assignment surprising. We assume that the status of VC is not a universal, but it may count as light or heavy depending on the language. For instance, in modern North Germanic (except Danish), there is a general constraint that stressed syllables may be only of the types VV, VCC; VC counts with V alone as illegal in stressed syllables, and hence V, VC constitute a natural class in these languages. Since this class contains V alone, we call it light. We maintain that the same classification holds for Old English. For a more technical justification of the model adopted here, see Lass (1994: 34–8) and Trask (1996: s.v. heavy syllable). That discussion is of course old-fashioned, but it accounts better for the facts of West Germanic and Old English than a model where VC is heavy. This framework is still used in the most recent handbooks, e.g. Hogg & Fulk (2011). It is also used in the most recent grammar of Old Frisian (Bremmer 2009: §27, fn.4). For counterarguments see McCully (1992).

Vowels in Old English and Middle English may be bimoric or monomoric, i.e. long or short; but in Old English there are two kinds of complex vowels:

(a) The so-called 'long' but really simply bimoric diphthongs, which count quantitatively as long vowels;

(b) We accept for Old English also the existence of short diphthongs.

We mark the short diphthongs in phonological representations with a breve over the first element, e.g. [ĕo] vs [eo], etc., In orthographical representations we mark the diphthongs in the standard handbook way with a macron over the long and no marking of the short, e.g. *ēo* vs *eo*. Phonetically the short diphthongs were (presumably like such structures in Modern English, Scots and Icelandic) two qualities packed into the timing/weight space normally allotted to one, i.e. di-qualitative but monomoric. The two primary pieces of evidence for short diphthongs are: (a) they pattern together with short vowels historically, e.g. typically merging with them; and (b) a short diphthong followed by a consonant does not make a heavy syllable, but a light one (see above). For a Modern English dialect with long and short diphthongs see Lass (1981).

8 The scribal dimension in etymological narrative

---

[51] The adjective 'moraic' is more usual in contemporary theoretical discourse, but 'moric' has been used in historical Germanic studies at least since the 1960s, and we will retain the older usage since it is that of the modern handbooks (e.g. Hogg 1992a, Lass 1994, Ringe 2006a, Hogg & Fulk 2011).

The LAEME CTT 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 forms 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 over 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.[52] We may, however, 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 presence or absence of a crossbar. But here is a more complicated case where letter shapes do indeed bear on interpretation. On fol. 83v, the scribe of text # 246 tr323at.tag, 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 C13 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. In both these examples we invoke the concept of

---

[52] We use the terminology of the medieval theory of *littera*. The conventions (established by Michael Benskin (1997: 1 n. 1 and 2001: 194 note 4) are as follows. *Littera* is the abstract or superordinate notion of the letter, and (when referred to independently of manuscript citation), *litterae* are enclosed in single inverted commas. *Figura* is the shape of a *littera*. This can refer both to the *figura* that is associated with the *littera* in a particular script or type of script (Textura, Anglicana, charter hand, etc.); or it can indicate a particular realisation within any one script. Manuscript *figurae* are enclosed in angle brackets or are italicised when combined as whole words or longer. *Potestates* are sound values and are represented by IPA symbols in phonetic brackets. As an additional convention, glosses and names of lexical categories are in small capitals. For further explication see LAEME, Introduction, Chapter 2, §2.3.1.

'Litteral Substitution Set' (LSS): a set of *litterae* that can be substituted for each other in the representation of a particular etymological category or set of categories. (For an introduction, see Laing 1999, Laing & Lass 2003, Laing & Lass 2009 and LAEME Introduction, §2.3.) Here are two examples, one simple and the other exceedingly complex:

(a) Many SWML texts use S (either exclusively or commonly) for the medial consonant graph 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). Most of these texts also show HT spellings; a text of this sort would be said to have, for the category 'OE *-ht*', the LSS {'st', 'ht'}.

(b) In the LAEME corpus sample text # 278 layamonAbt.tag ('Layamon A', Hand B), the following spellings appear for $hand/n: HOND(E) 8x, HANDE 1x and HEOND(E) 3x.[53] 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 OE *seofon* SEVEN, with back-umlaut of *e*, hence short *eo*.[54] If the reflex of OE *eo* can be spelled both EO and O, then by an orthographic transposition not uncommon in some of our sources, any category with the spelling *o* as one of its realisations 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/n and $book/n for instance (OE *brōþor, bōc*). But interestingly, in these words we also find A; so it seems that this scribe has a Litteral Substitution Set (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*.[55]

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/n with the 'canonical' or traditional spellings HOND(E), HANDE can therefore be spelled HEOND(E). 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/n. 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.

---

[53] 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.

[54] OE *eo* is traditionally thought to have become a mid front rounded vowel in western ME; there is also an argument that the *o* spellings of this category may represent [o] (Lass & Laing 2005). All that matters for this exposition is that the category represented by *eo* can mean the same thing as orthographic *o* in this text; the phonetic value is not crucial. See the CC s.v. ((EOM)) for further discussion.

[55] 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:], as in BRAyER.

So the basic level of accountability is litteral: since we are so often concerned with litteral substitution sets, scribal choice in representation (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 in LAEME CTT, not for the word 'as a whole'.

9 Fineness of etymology: the size of changes

9.1 Parsimony and realism

Consider the etymology of OE *cynn* KIN, RACE. The nuclear *y* indicates *i*-umlaut of Germanic ***u**, and the geminate *-nn-* indicates West Germanic gemination. This plus comparative evidence like OHG *kunni* tells us that there was a ***j** following the stem, enabling the reconstruction at least of an input ***kun-j-**. The grammatical behaviour of this noun in Old English is that of an *a*-stem neuter, allowing us to assign it to the subclass of *a*-stems with a post-root ***j**, called *ja*-stems (on stem-classes see §1.2). I.e. its West Germanic form would be ***kun-jɑ-z**, becoming PrOE ***kunn-jɑ** by later changes. To reach the attested Old English form one further operation is required, deletion of the suffixal ***-jɑ**.

Now there are in principle two ways of getting rid of a suffix consisting of more than one segment. One is phonological: the deletion of the suffix occurs through a chain of reductive processes, e.g. in this case first deletion of ***j** at the beginning of a weak syllable, and then deletion of the remaining vowel, giving [kynn], naturally written in Old English as *cynn*.[56] These processes are independently motivated: for instance the deletion of ***j** in the onset of a weak syllable (see CC s.v. ((JD))) is needed for among other things the correct form of the weak verb class 1 infinitive suffix *-an* < ***-jɑn**; deletion of final nasals in weak syllables and deletion of final weak vowels are pan-Germanic processes attested at least as early as C5, and are needed elsewhere in our etymologies.

The other way of getting rid of the suffix would be via morphological rather than phonological procedure: specifying the suffix as *ja*-stem nom sg and simply chopping off the whole thing in a single morphological operation that does not refer to the phonology. In CoNE we will where possible take the phonological route. There are a number of reasons for this. First, in general, phonological reductive processes appear to be commoner in languages than morphological ones; second, there is a methodological constraint, at least in some scholars' opinion, that in the absence of compelling evidence the *currently* commonest type of process should be invoked in reconstruction.[57] But more, as noted above, a number of the phonological changes that we would use in getting rid of ***jɑ** and other complex suffixes are independently needed because they operate elsewhere; and in those cases they leave residue behind in such a way that they cannot really be morphological. So why not use these

---

[56] There is variable final degemination in some Old English traditions, so *cyn* also occurs.
[57] See Lass (1997: §1.5). This is referred to there as 'the Uniform Probabilities Principle'. It is part of an epistemological approach to history called 'uniformitarianism', which essentially says that in historiography, the present constrains the past. See the citation above for detailed discussion and references.

operations wherever their structural descriptions are met, and think of what happens in suffixal decay as something like the operation of fast-speech rules? I.e. we have a subset of changes that belong to weak syllables, and they constitute a linguistically identifiable part of the phonology. So we think then of morphological reductions, where possible, as essentially phonological, yielding morphological change as a side-effect. This kind of relation works elsewhere in the morphology: virtually all morphophonemic alternations are the result of purely phonological rules happening to apply in a salient morphological environment. For instance, the stressed vowel alternations in the Old English paradigm of FOOT, say nom sg *fōt* ~ nom pl *fēt*, while morphological in Old English, have nothing morphological in their origins. They are simply due to the presence of historical ***i** in the plural suffix, which umlauted the root vowel and later deleted, morphologising the alternation. In other words, such changes are not 'aimed at' morphological outcomes, but produce them by accident.

We put the deletion of ***-jɑ** this way, as a 'task' that can be accomplished in different ways, because we want to make the point that this is a procedural, not purely linguistic or historical matter. It cannot be emphasised too strongly that procedural and aesthetic judgements have a major role to play in historical reconstruction. There *is* in fact very often no 'history' in the usual sense. That is, it must of course have been there, but we can only see it through the acts we perform to bring it into being. All we have in the case we began with, the *ja*-stem nom sg, is a reconstructed form ending in ***-jɑ** and an attested zero. Since the actual process of deletion is unattested, we have to invent it. And in this case there are two choices, both of which involve types of change that are well attested and unexceptionable, it would seem. Therefore we have to make a choice, on the grounds of general principles and preferences, for one type of historical procedure over another. The guiding principle we choose in this case is our view that the commonest changes are phonological, and that grammatical changes, except through analogy or borrowing (see below and §11), are less preferable in cases where there is also a phonological route available.

Similar judgements have to be made about the size of segmental changes. How big can an internal phonological change be? If we consider modern changes in progress in the sociophonetic literature (e.g. in Chambers 1995: 198, or Labov 1994: ch. 6), it would seem that immediate vowel changes the size of one inter-symbol space in the IPA chart are most unlikely, except through a period of much smaller incremental changes (and cf. also the argument in Lass 1997: ch. 5, especially 228, 236). But history does not permit that kind of delicacy, and we are generally constrained by the size of the intervals in the standard grids, e.g. the IPA alphabet. But, as we did above, we can constrain movement within those grids, in a way that appears similar to what we find in modern materials. And the uniformitarian principles, that the present constrains the past, also constrain our methodology. We will follow the general preference for minimality in phonological change, at least as far as we can implement it.

One way of dong this would be to interpolate segment-types that could be seen as 'transitions' between input and output. Thus we could treat the accepted *i*-umlaut of ***ɑ** not as the change ***ɑ** > [æ] as it is described in all the handbooks and histories, but as involving at very least the intermediate stage ***ɐ**. The claim would be that (a) we know on uniformitarian grounds that the change was gradual, and (b) that at least by introducing a central-vowel

transition state we make the change longer and occurring in smaller stages than the received model would have it. Even though we do indeed believe that direct changes from back to front do not occur in real time except in analogy or borrowing,[58] there is nevertheless no independent evidence for a category of central vowels in older Germanic. We do not, for this reason, adopt general implementation of the transitional segment solution. We thus remain mutually comprehensible with the tradition of Old English studies, and make it clear once again that history does not yield phonetic finesse. But still, within the grid made available by the tradition, we will try to make changes as small as possible, and introduce certain constraints. One important one, which will partly control the way we formulate changes, is that *ceteris paribus* no change involves an alteration of more than one parameter at a time. If the overall historical picture for instance shows that a vowel has shortened and lowered, we would label these as two separate changes.

Similarly we would generally allow deletion of obstruents only through pathways of lenition. Thus what is often referred to in the handbooks as deletion of intervocalic [x] in Old English (as in *sēon* SEE < **sexɑn**) we would represent as having to have occurred via an intermediate stage [h]. This is the standard lenition hierarchy, and we should invoke it unless there are reasons for not doing so. Thus we accomplish this deletion in two stages, weakening of [x] > [h] ((XW1)) and deletion of [h] ((MHDCL)).

The procedures we have been advocating might seem from one point of view unparsimonious. Many scholars recommend making historical derivations as short as possible, rather than lengthening them. We take a uniformitarianist view and model history as far as possible on the properties of change in the present day, which often makes the derivations longer. For a detailed discussion of this problem see Lass (1997: §5.2).[59]

9.2. Macro-changes: analogy and indeterminacy

It would be simple if the etymology of a corpus form were only the etymology of the root: the CoNE 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

---

[58] Of course such changes *appear* in languages in morphophonemic alternations, but these are the result of stabilization after completion of gradual changes. Thus OE *mann ~ men* does not imply a direct change of [ɑ] to [e] at any point in time, but a gradual sequence that ends up in that shape. But a standard internal reconstruction would in fact construe the change as non-gradual, because of the way the procedure generally works. They can also occur in borrowing, as in the case of modern Northern English standardising speakers with [ɑ:] instead of original native [a] in words of the BATH class. This is the result of direct dialect borrowing and there are no intermediate stages. Virtually anything can be borrowed and replace anything else, without regard to the size of the differences between them: there is no phonetic pathway involved here, but more the equivalent of borrowing a word. See the following section.

[59] We must be careful however not to rule out 'macromutations'. Our treatment of change is the same as that typically employed in historical biology: most mutations are very small and most change is very gradual, but there are occasional huge transformations caused by a single gene change. An example is the commonest type of dwarfism in humans, which has enormous effects on different parts of the body, but involves only one gene. It is a 'saltation' or leap, which is generally not favoured. In historical biology (and we are closer to that than anything else), there is a maxim (usually attributed to Linnaeus) which is rarely violated: 'natura non facit saltum' NATURE DOES NOT MAKE LEAPS. In the following sections we will see some ways that such saltations occur.

nonphonological processes; the affixes attached to roots are frequently unexpected, and require their own historical commentary.[60] 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 category, 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 *homō* MAN (gen sg *hom-in-is*, etc.).[61] Other Germanic evidence bears this out (e.g. German nom/acc pl *Herz-en*). The same sources classify $soul as a heavy-root strong feminine (*ō*-stem), with zero nom sg and the rest of the case/number endings (except dat pl) vocalic, mostly *-e* but occasionally *-a* in nom/acc pl. This means that $heart and $soul ought to have been associated 'in Old English' with these paradigms:[62]

Nom sg	heort-e	sāwol	Nom pl	heort-an	sāwl-a, -e
Acc sg	heort-an	sāwl-a, -e	Acc pl	heort-an	sāwl-e
Gen sg	heort-an	sāwl-e	Gen pl	heort-an	sāwl-a
Dat sg	heort-an	sāwl-e	Dat pl	heort-um	sāwl-um

The acc/gen/dat 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, with final nasal deletion) *-e*. The same should also apply to the dat pl, originally *-um*, since as early as C10 there is widespread (if not categorical) merger of final *-m* and *-n* in weak final syllables in most varieties of Old English, and merger of many weak vowels to *a* and then *e*. As for $soul, *-en* might be expected in earlier Middle English 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:

| Nom sg | stān | Nom pl | stān-as |

Acc sg	stān	Acc pl	stān-as
Gen sg	stān-es	Gen pl	stān-a
Dat sg	stān-e	Dat Pl	stān-um

Now consider these genitive and plural forms for $heart and $soul from a number of LAEME CTT samples of different dates and provenances:[63]

ayenbitet.tag (# 291)
$heart/npl HERT+EN 7x
$soul/npl ZAUL+EN 3x ZAUL+ES 5x

genexodt.tag (# 155)
$heart/nG HERT+E 1x HERT+ES 1x
$heart/npl HERT+ES 1x
$soul/nG SOUL+ES 1x SOUL+E^S 1x SOwLES 1x

cleoarat.tag (# 273)
$heart/nG HEORT+E 4x
$heart/npl HEORT+EN 2x
$soul/nG SAUL+E 7x SAUL+ES 1x
$soul/npl SAUL+E 2x

corpart.tag (# 272)
$heart/nG HEORT+E 6x HEORT+ES 1x
$soul/nG SAwL+E 7x SAwL+ES 1x
$soul/npl SAwL+EN 1x SAwL+ES 1x

trhomBt.tag (# 1300)
$heart/nG HE(O)RT+E 5x
$heart/npl HE(O)RT+E 8x HE(O)RT+ES 2x HERT+E^S 1x
$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 high prevalence of analogical extension, i.e. the ease with which these nouns appear to take on forms from other declensions, not only 'in the modern direction' by adopting the *a*-stem *s*-genitive and plural, but also in other directions.

[63] 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 text # 1300 trhomBt.tag (Trinity Homilies): see below.

trhomBt.tag SAUL+ENE not only shows a shift from strong to weak, but perhaps also an unexpected full reflex 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 text # 1300 trhomBt.tag or in text # 155 genexodt.tag? 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? 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 result from one of a set of 'chunks' that exist in a morphological search space, i.e. analogical extension. 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*. The process involved here does not resemble phonological change; it is formally precisely the same as borrowing. In this case the donor paradigm plays the same role as the donor speaker in a language contact situation. Thus whether or not an affix is taken to be the output of a sequence of segmental mappings or a substitution-as-a-whole (analogy) 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, and allowing for macromutation. Is a text form +E the result of (a) a phonological mapping like *-an > -en > -e*, or (b) a whole-item 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 canonical Old English, we may not be able to choose among (a)–(c). In such cases we will instead note the competing possibilities. 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.

But we can come to something of a solution by stipulation. On the basis of our acquaintance with language histories, we claim that paradigm-internal phonological change is to be preferred to paradigm-external analogical change, if the evidence does not make the decision for us.

10 Some problems

10.1 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. ('Proto-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; the same applies to the order and temporal placement of changes. Where relevant indeterminacy will be indicated in the individual etymological commentaries.

10.2 The Old English dialects

The handbooks typically say that certain phenomena occur e.g. 'in Anglian' or 'in West Saxon'. It is important to realise that these 'dialect' names do not generally designate the language of particular geographical regions in the strict sense; they rather denote clusters of text languages showing more in common with each other than with text languages outside the cluster. The 'dialect boundaries' on the traditional maps are assumptions based largely on geopolitical demarcations, and do not – except in the coarsest terms and not for all clusters – designate 'linguistic areas'. For a sophisticated discussion see Hogg (1992a: ch. 1).

The historical situation is made even more difficult by our inability to establish – except very broadly, and not in all cases – any solid lineal relationships between Old English and early Middle English text clusters. As far as we know there is no surviving early Middle English text that can be shown to stand in a direct lineal relationship to any surviving Old English text. (This of course excludes Middle English copies of Old English texts, but that is not the kind of lineality we mean, which is development and transmission of a system over time.) But connections can still be made, though it takes considerable care and discrimination to avoid making them too literally. For instance, -*s* for present 3sg indicative in Old English 'belongs to Northumbrian'; it also belongs to northern and occasionally West Midland texts in the LAEME CTT. So in some sense the endings +ES, +IS for this category in the Edinburgh, Royal College of Physicians MS of *Cursor Mundi* (text # 297 edincmat.tag, text # 298 edincmbt.tag, text # 296 edincmct.tag) are 'continuations' of an Old Northumbrian property; what we cannot establish is a continuity between these particular 's' forms and those in any particular Old Northumbrian texts, and therefore we cannot connect pairs or n-tuples of text-languages across the notional Old English/Middle English 'border'.

Similarly, early West Saxon is the only Old English cluster in which the *i*-umlaut of ***au** is *īe* ([iy]), which in later periods shows up as *ī* (mainly in palatal environments) and *ȳ*. Thus in West Saxon the infinitive HEAR arises via the sequence:

*xaurjan ((AUF)) > ***xæurjan** ((IU)) > ***xiyrjan** ((XW1)) > ***hiyrjan** ((JD)) > [hiyrɑn] > *hīeran*, later *hȳran*

whereas in Anglian the story is:

*xaurjan ((AUF)) > ***xæurjan** ((IU)) > ***xeyrjan** ((EYM)) > ***xe:rjan** ((XW1)) > ***he:rjan** ((JD)) > [he:rɑn] > *hēran*

((EYM)) of course turns ***ey** into ***ee**, the latter equivalent to ***e:**, cf. fn 2 and §7.2 above. The denotations of the changes in (( )) can be found in the CC.

 Therefore any Middle English text with nuclear *u* (one of the reflexes of OE *ȳ*) in HEAR is to that extent 'of West Saxon provenance'. For the two cases discussed, then, there will be conjectural pre-Middle English provenances of 'WS type' and 'non-WS type' respectively. This is the primary sense in which Old English dialects mentioned in the CC entries and the etymological histories of particular forms are to be read.

 So we use Old English dialect names as guides to the likely form shapes that existed in the ancestors of particular groups of text languages. For a rough indication (rich in caveats) to what the geographical loci of some of the named Old English dialect types might have been, see Hogg (1992a: §§1.5–1.12). For some more specific indications, particularly with respect to West Mercian, see Kitson (1990).

10.3 The West Saxon bias

The Old English word-forms and paradigms given in any etymology follow the standard ones given in the handbooks. These are normally West Saxon, which is problematic, as most of our surviving Middle English texts are not of West Saxon provenance, and neither are the modern standard varieties. We follow this practice because these are the most familiar forms, and are taken at large to define 'Old English phonology and morphology'. They are conventional rather than purely genetic items (which given the amount of variation in Old English would not strictly speaking be obtainable anyway); but they serve the purpose of outlining the forms most commonly referred to in the literature. This does not mean of course that perhaps the majority of West Saxon forms are not perfectly good etyma for modern or Middle English ones, since a very large part of the structures of all Old English dialects was the same. In cases where there is disagreement in the literature about morphological paradigms, we adopt a compromise position and generally follow the versions given in Campbell (1959).

 One must be aware of this bias, which even extends itself in a subtle way into sources like the OED. The OED etymology of *hear* gives EWS *híeran* as the first form (the unrevised portions of the OED still use an acute for Old English length), followed by *hýran*, and then at the end of a list gives Anglian *héran*. OED does not make it clear that the West Saxon form cannot be the source of the Modern English verb, which if it came from this would have the vowel */aɪ/ and thus be a homophone of *hire*. The OED simply does not say which lineage the modern forms come from: the assumption is that the user will know that PDE /i:/ presupposes [e:] at least in Middle English, and that would presuppose an Anglian ancestor, i.e. OE *hēran*. This etymology makes sense only to the user who knows that the bulk of modern Standard English phonology is SEML, hence Anglian.

 Anybody working with the literature on Old English is going to have to get used to this, and watch out for certain results of an extreme form of this bias in particular sources.

For instance, Holthausen (1963), an invaluable etymological dictionary, uses early West Saxon lemmata – even if his headwords are unattested. The reader has to know this in some cases in order simply to locate words. A typical case: Holthausen uses *ǣ* for the *i*-umlaut of ***oː**, which in most Old English texts would be *ē*. Therefore if you want to look up what you would most likely know as *grēne* GREEN, you would find nothing listed in the expected place, and no cross-reference, but would have to know the etymological source, paradoxically before you would be able to look up the etymologically relevant material. The place to look would be of course *grǣne*: but you would have to know this, or at least some of the sparse Old English evidence that points to it, such as *groeni* in the Corpus glossary, which shows the umlaut environment still intact.

10.4 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 & 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 & Anderson (1975) and taken up in Hogg (1992a: §§16–34) is that the Old English diphthongs were 'height-harmonic': the two elements agreed in height, the second assimilating to the first. Thus the history of what comes to be written *ea* as in *eald* OLD would be ***ɑ** > ***æ** > [æu] > [æɑ], with the original [-u] diphthong (for which there is orthographic evidence) subject to a process that Lass & Anderson called 'Diphthong Height Harmony' ((DHH)). The same kind of diphthongs arose from other Old English processes such as back umlaut. So the graphic sequences *ea, eo* in Old English represented [æɑ, eo], long and short. The nature of *ie*[64] is problematic: see §10.4.3.

The other problem associated with the diphthongs is the existence of a length contrast, which is taken for granted in the standard literature, but has been argued against primarily on typological grounds. It is true that a classical phonemic contrast of length in diphthongs is not common, but that is not a problem here. First, as we have said, we are not operating on a phonemic level; and second, length contrasts (even phonemic ones) are commoner than has been generally thought, even in English. (For a detailed description of this kind of contrast in a dialect of English see Lass 1981). We assume that the two sets of diphthongs were phonetically identical, but that the two quality-segments of the short one occupied only one temporal slot. See the discussion in §7.2 above.

It is possible, even likely, that there were, in some varieties of Old English, diphthongs of a different type, non-harmonic with a high second mora. In such cases -*eg* as in *weg* WAY would have represented [ei] rather than [ej], i.e. it would be a VV rather than a VC sequence. There also may have been in later Old English closing diphthongs from French like [oi, ui]. See further CC s.v. ((CV)). Thus at some point the 'classical' harmonic diphthongs were lost and replaced by the closing type that were structurally the same as the early Germanic inputs.

---

[64] Largely restricted to early West Saxon, but certainly part of the lineages of some text languages attested in the SW and SWML.

10.4.1 'eo' spellings

The diphthongs spelled 'eo' in Old English seem undoubtedly to have been of the type [eo], suggested by their origins in Gmc **eu** (long) and breaking or back umlaut of **e** (long or short). In early Middle English, 'eo' spellings present a number of difficulties. First, the number of historical categories they can represent is dauntingly large. In our sample from one text alone, text # 280 layamonBOt.tag (Laȝamon B), 'eo' can represent at least the following categories: OE *y, ȳ, eo, ēa, ǣ¹, ǣ²* (Lass & 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 text # 280 layamonBOt.tag for instance spells his reflexes of these categories EO, E, U, I, while text # 277 layamonAat.tag (Laȝamon A, hand A), employs all of these plus O for OE *eo, ēo*. At least some of this orthographic variation is likely to represent phonic non-identity at some level. Consideration of the individual scribal spelling systems makes it possible to assign values to E, U, O, I without much uncertainty. However, given what appears to be the history of the Old English categories traditionally represented as 'eo', an EO spelling in texts in the LAEME CTT could in any instance arguably represent [e(:)], [o(:)] or [ø(:)], and it is generally impossible to tell which.[65] 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.

10.4.2 'ea' spellings

As in the case of 'eo' above, this spelling became 'free' when the diphthongs it 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 being coopted for similar sound types. Earlier OE *ea* indicated something like [æɑ] (long or short); by C11 the long diphthong had merged with *ǣ*. This later probably raised to [ɛ:] (if in fact what we reconstruct as late OE [æ:] was not already [ɛ:], variably or categorically). The short diphthong merged with *æ*. Since, depending on location and scribal preference, the latter category could emerge as eME *a* or *e*, this leaves 'ea' as a possible spelling for [ɛ:], and for [ɑ] 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. corpart.tag, # 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 (for detailed discussion see Lass & Laing 2011).

---

[65] Such spellings cannot at this date represent anything diphthongal of the type [eo]. Potentially, some instances could be resolvable if the EO form occurs in rhyme position. But in practice, text # 277 layamonAat.tag has very irregular rhyme usage and even apparently decisive rhyme evidence may be a chimera. Lass & Laing (2005) argued that the evidence for front rounded vowels was unconvincing, and the *eo* diphthongs had simply merged with the categories represented by either of the components. This claim has been shown by Stenbrenden (2010: ch. 3) to be overstated.

10.4.3 'ie' spellings

This is probably the most controversial and difficult of the digraphs. It occurs only in Early West Saxon and some rare Anglian pronominal forms, but has become an integral part of the history of English in general as displayed in the handbooks and textbooks because of the West Saxon bias referred to above. It has been characterised variously as representing [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 1992a: §§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 Corpus of Changes (CC) under ((IES)).

10.5 Placement in time

The sequences of changes that constitute an etymology are presumed ultimately to be ordered in time: the initiation of a change at least represents a historical event. We specify initiation, because it is possible for changes to take hundreds of years to complete, e.g. the loss of postvocalic /r/ in non-rhotic dialects of English (Lass 1997: ch. 5). Working in prehistory as we do, and with fairly sparse data even in the historical period, it is impossible to establish absolute chronologies for any of our changes: we cannot say in what decade *i*-umlaut began. All historians who deal with pre-attestation language states can usually establish is relative chronologies. In phonology at least, this is made possible by the fact that there are two primary ordering relations a pair of phonological rules can have: feeding (in which one provides the input for the other), and bleeding (in which one destroys a configuration that would be the input for the other). Thus we can tell for instance that weakening of ***x** to ***h** in certain positions would bleed breaking, which requires [x] as its environment; therefore weakening of intervocalic ***x** and deletion of the resultant [h] must come after breaking (see CC s.v. ((BRK)), ((XW1)), ((MHDCL))).[66]

There is however a third possible relation between two changes: mutually unaffecting. In this case the rules do not share any environments, and it does not matter which precedes which, or if they are simultaneous. This can produce serious problems for the historian dealing with pre-orthographic periods: there must have been a real order (including simultaneity as an order) but we can only make one up, and with luck find some principles that would make one sequence preferable to another. But since historiography abhors a vacuum as much as nature does, we may simply have to act arbitrarily, what might be called the 'constructivist' element in historiography. There is a formal problem too with simultaneity; given the format of our etymology and the concept of linear narrative, it is not easy to represent it clearly even if we had definite evidence for it. If we suspect it in particular cases we will say so in commentary on the relevant etymologies.

Another thing we have to consider is that there are two kinds of changes: those that occur once and occupy a delimited region of time and then cease, and those that persist or

---

[66] Changes can also 'compete' for the same environment. There are two late Middle English and Early Modern English changes affecting [x] in word-final position. It can delete and leave a long vowel behind (*through* < *þruh* < *þurh*; or it can become [f]: *dwarf* < *dwearh*. There is no way of predicting which change will 'win' in any given case.

recycle, and keep happening. The latter sort are often variable and lexically or morphologically constrained, so that they may or may not affect items that meet their specifications. A typical example of the first type is palatalization of velars ((VP)) before front vowels in Proto-Old English. We can show that it must have preceded *i*-umlaut, because only original (Germanic or Proto-Old English) front vowels are subject to it, not those which result from umlaut of back vowels. So *cinn* CHIN with an original Germanic front vowel and palatalization vs *cynn* KIN with a secondary front vowel produced by umlaut of original ***u** and no palatalisation. ((VP)) is thus bounded: it occurred and ceased to be active before umlaut. On the other hand the deletion of unstressed final vowels is attested in some items as early in the history of Germanic as Gothic, and occurs in various forms all through Proto-Old English, Old English and Middle English, and never affects all items that could be affected by it, but continually recurs in subsets of its environment. Another example is ((FND)) final nasal deletion, which not only occurs scattered throughout late Old English and Middle English, but is attested as early as the runic inscription on the Gallehus gold horn (Denmark, ca 400): the acc sg *horna* HORN shows loss of original ***-n** < ***-m**, cf. L *cornu-m*. (Note that we are dealing with changes here that not only have a wide distribution in time, but also in space.)

Such phenomena account for a major aspect of the structure of the CC, and the way it differs from the presentations of changes found in the standard handbooks and histories. We present changes like the two discussed above as single items, rather than giving the histories of forms and presenting a change anew for each of its operational periods. In a standard kind of history of English, for instance, deletion of the final nasal in weak syllables in Northumbrian Old English (e.g. *wesa* TO BE vs WS *wesan*) would be treated separately from the later losses in Middle English (e.g. infinitives in -*e* < -*an*).[67] We take the view that the two occurrences are not 'different changes', but the same change repeated at different points in time. This change, ((FND)) final nasal deletion, is in fact the same change as the one in the Gallehus horn form in the previous paragraph, and in the -*a* infinitives of the North Germanic languages (e.g. ON *vera* BE). In other words, it is a typological feature of the history of Germanic that final nasals in weak syllables tend to delete, and the emergence of the change ((FND)) is simply a property of these languages, which can recur at any time.

Those changes that occur from the time of earliest textual attestation to ca 1100 are labelled OE, those datable after 1100 are labelled ME. Some changes 'repeat' in both periods, or can be considered 'habitual' for English, e.g. merger or deletion of vowels in prosodically weak positions. These changes are marked as such, e.g. OE, ME. When changes can be firmly assigned to early or late subperiods this is indicated as well, e.g. lOE, eME. Changes that predate attested Old English are labelled PrOE 'Proto-Old English', PGmc 'Proto-Germanic', WGmc 'West Germanic' or Ing 'Ingvaeonic'.

We must also distinguish changes which start out located at a particular point in time, but then change status, and rather than being 'events' with a temporal boundary, i.e. 'changes' *sensu stricto*, become constraints. A characteristic example is Diphthong Height Harmony ((DHH)), which in Proto-Old English and Old English makes the second element

---

[67] The standard reconstruction of the Germanic infinitive ending is ***-ɑn-ɑ-n**, where the last segment is a neuter marker < IE ***-m**: cf. the *horna* example above.

of a diphthong the same height as the first: so ***æu** > [æɑ], spelled *ēa*. This remains throughout the Old English period as a well-formedness condition on diphthongs, so that when new diphthongisations arise, ((DHH)) is simply a condition determining the height of their second elements. But at a certain point in late Old English, difficult to specify, it becomes inactive, and diphthongs in [-i, -u] appear and gradually become the norm. There are some spellings as early as C8 that suggest that the constraint may have been less than totally firm in some dialects (see discussion under ((CV)) in CC), but these are marginal considering the activity of the constraint during the bulk of the Old English period.

11 Loanwords

11.1 Types of borrowings

Every language has an inventory of formatives: lexical items and grammatical items (inflexions and derivational markers).[68] 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 (IN SANSKRIT) and *tadbhabva* ORIGINATING (IN SANSKRIT): see Masica (1991: §4.2). 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 that they are reborrowings rather than direct inheritances through the Prakrits (vernacular descendants of Sanskrit).

These terms are frequently extended beyond Indic studies into general linguistic discourse. 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 CTT 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 §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

---

[68] 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.

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 text # 1600 laud108at.tag, $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.

11.2 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 CoNE 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.[69] 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.[70] 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 Old English: *þweorh* CROOKED, PERVERSE (not to mention derivates like *þweorian* OPPOSE, *þweorlic* CONTRARY, ADVERSE, *þweornes* PERVERSITY). And second, cognate forms occur not only throughout Germanic with similar meanings (Go *þwairhs* ANGRY, CROSS, OHG *dwerh*, *twerh*), but also in other Indo-European 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

---

[69] Writing narrative etymologies for non-Germanic loans will be left to a later phase of the CoNE project.
[70] The discussion here is partly based on Lass (1997: 203ff).

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*.[71] 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.
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 Old English.
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* EYE-HOLE, and the normal Scandinavian word is *vindauga* WIND-EYE. The forms occurring in the LAEME CTT are satisfactory from both points of view:

$window/n WINDOUN, WYNDOW, wINDOGE, wINDOHE, wINDOw-

---

[71] If we take it that the original root ends in ***-r**, then at first sight there might just be a Scandinavian element in *thwart*: Jespersen (1938: 83) suggests that the final -*t* is a North Germanic neuter marker. On the other hand this same 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*.

The sense is Scandinavian, and the development of the second element coheres with Scandinavian rather than Old English: Gmc ***αυγ-o:** > 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.

11.3 '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.*[72] Given what we know of patterns of settlement, there must have been extensive English/Norse bilingualism of some sort.[73]

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?[74] 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, 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,

---

[72] 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). There is a sophisticated study of Norse loans in early Middle English in Dance (2011).

[73] Some people have claimed that there was even mutual comprehensibility, e.g. Townend (2002).

[74] 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. The initial [g] in <give> is one of the few cases where there may be unambiguous evidence for a Scandinavian source. There is no doubt that the original Germanic vowel is ***e**, and already in Old English there are spellings in *i*, which probably derive from OE *giefan* types, where the *ie* is the result of palatal diphthongisation ((PD)). But Old Danish alone among the Germanic languages appears to have nuclear *i*, so *give* could be a direct borrowing with initial [g].

OSc *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 (76),

> 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 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. For further discussion see Durkin (2009: ch. 5).

These accounts are completely conjectural and we remain agnostic. In the etymologies, forms of GIVE with an initial velar will simply be marked as having traversed an alternative pathway, where velar palatalisation was lost.