Mark Faulkner*

Habemus Corpora: Reapproaching Philological Problems in the Age of 'Big' Data

https://doi.org/10.1515/ang-2021-0006

Abstract: This paper demonstrates the potential of new methodologies for using existing corpora of medieval English to better contextualise linguistic variants, a major task of philology and a key underpinning of our ability to answer major literary-historical questions, such as when, where and to what purpose medieval texts and manuscripts were produced. The primary focus of the article is the assistance these methods can offer in dating the composition of texts, which it illustrates with a case study of the "Old" English Life of St Neot, uniquely preserved in the mid-twelfth-century South-Eastern homiliary, London, British Library, Cotton Vespasian D.xiv, fols. 4–169. While the Life has recently been dated around 1100, examining its orthography, lexis, syntax and style alongside that of all other English-language texts surviving from before 1150 using new techniques for searching the *Dictionary of Old English Corpus* suggests it is very unlikely to be this late. The article closes with some reflections on what book-historical research should prioritise as it further evolves into the digital age.

Key terms: philology, text dating, corpus linguistics, Old English, Middle English, London, British Library, Cotton Vespasian D.xiv, fols. 4–169, Old English Life of St Neot

1 Introduction

Book history and philology share, or ought to share, a symbiotic relationship. On the one hand, linguistic analysis of the texts which a manuscript contains often provides the key to unlocking the process by which it was compiled – a telling example here being Celia Sisam (1951)'s philological demonstration, subsequently developed in a more book-historical direction by Ralph Hanna (2009), that when the *Lambeth Homilies* were produced around 1200, their scribe drew on

***Corresponding author: Mark Faulkner,** Trinity College Dublin
E-Mail: mark.faulkner@tcd.ie

at least two different exemplars. On the other hand, recent work, particularly that of Margaret Laing and Roger Lass using the *Linguistic Atlas of Early Middle English* (e.g. Laing and Lass 2009; Lass and Laing 2012; Lass 2015), has demonstrated again and again that the optimal unit for primary philological analysis is not the edited text that purports to reconstruct the authorial original, but the stints individual scribes contributed to particular manuscripts.

I write as a philologist, rather than a book historian, but do so in the consciousness that much of what I do as a philologist is informed by book history and many of the results I generate as a philologist will be of potential utility to book historians, not least with regard to some of the most important questions that any codex can pose, namely when, where and why it was produced. Philology ultimately depends on placing the features of a particular manuscript text in the context of broader linguistic agglomerations such as dialects or phases in the history of a language. Typically, researchers have done this through recourse to dictionaries, grammars and handbooks. But these reference works do not cover all features, are often non-committal on the distribution of particular variants and ultimately rest on datasets that are limited and perhaps unreliable.

This paper is primarily concerned with how new ways to use existing corpora of medieval texts can bring about an epistemological shift in book history and literary studies. This 'corpus philology', as I have called it elsewhere (Faulkner forthcoming a), uses methodologies from corpus linguistics to attempt to expand and, where necessary, correct the picture of the distribution of particular linguistic features in handbooks and dictionaries, thereby enabling the better placing of texts and manuscripts in time and space, a necessity for both philologists and book historians. It exemplifies these methodological possibilities with reference primarily to the dating of the "Old" English Life of St Neot, a work it has recently been suggested was composed around 1100 (Younge 2012), but which detailed linguistic comparison with the surviving corpus of Old English shows is very unlikely to have been written that late and is instead best taken as an eleventh-century work. The methodologies of this new approach are still evolving and how far we can trust their evidence remains an open question; much of this paper is therefore tentative and the claims it advances are therefore couched with deliberate caution, but their potential is clear. For this author, book histories in the digital age will evolve most fruitfully if investment in new resources is accompanied by the efficient use of what we already have.

2 Resources and Approaches

Medievalists are truly fortunate in the range of corpora available to them. For 'Old' English, almost all surviving independent texts are available digitally, as are (albeit sometimes indirectly) samples of the language of most surviving manuscripts. This means one can often make claims about a form vis-à-vis all surviving texts or relative to a large proportion of surviving manuscripts of a particular date. For 'Middle' English, however, though extensive, the coverage is less thorough and, more concerningly, somewhat arbitrary.[1] This is one reason why historical linguists making detailed studies of particular phenomena have often preferred to collect data from resources like the *Helsinki Corpus*, which offers subsets of the surviving Old and Middle English texts balanced for genre, region of production and date, largely disavowing the possibility of using other, more exhaustive text archives. These historical linguists' focus on single linguistic phenomena and smaller datasets have generally allowed rigorous checking of each possible instance; for the methodologies advocated here, where one might want to consider the distribution of a hundred or more linguistic forms across several thousand texts, such exemplary accuracy is necessarily in part sacrificed to economy. But even if there is noise in the data, the overall picture that emerges from using the available corpora with the methods outlined below is surely robust, much as 'big' data approaches in industry can detect broader trends accurately, without necessarily classifying every single datapoint precisely.

The principal corpus used in this paper to contextualise linguistic features of the Life of St Neot and thereby attempt to date it is the *Dictionary of Old English Corpus* (*DOEC*), which offers a machine-readable version of almost every text composed in English before 1150. Unlike, say, the *Linguistic Atlas of Early Middle English*, *DOEC* has no tagging, so the recovery of relevant forms is necessarily reliant on searching for what the user perceives to be the relevant surface orthographical patterns. Thus, someone interested in knowing some texts in which a phrase like *mid mycelre blisse* 'with great joy' is used could employ a simple, fragmentary search for that string. But they would have to run a considerable number

1 I place 'Old' and 'Middle' in inverted commas to indicate that avowedly Old English corpora include some texts likely to have been composed in the twelfth century, after the point some scholars regard Middle English as having begun (e.g. the English translation of Ralph d'Escures' Homily for the Assumption of the Virgin Mary, which has been referred to as the earliest Middle English text (Clark 1970: lii n. 1)), while Middle English ones include some undoubtedly composed in the tenth or eleventh centuries (e.g. the *Linguistic Atlas of Early Middle English*'s text of the *Wintney Rule of St Benedict*, and the Ælfrician pieces printed by Morris (1867–1868) in the course of editing the *Lambeth Homilies*, which are reproduced in the *Corpus of Middle English Prose and Verse*).

of other searches, considering possible variant spellings like *micelre* 'much' before they could be confident they had recovered all the relevant instances. This can be a laborious process. Moreover, since *DOEC* does not directly provide any metadata about the date, origin or transmission of the texts it contains, generalising on the basis of results obtained from it to make statements like 'this form is tenth century' or 'this form is south-eastern' is at present far from easy, since it depends on having at one's fingertips some basic text-critical knowledge of the more than 3,000 texts *DOEC* contains. In the case of high frequency items, it can require summarising at a glance the distribution of several hundred hits.

The best way of avoiding these difficulties is to side-step the array of search options offered in the online version of *DOEC*, take advantage of the compilers' generosity in making the previous (2000) version of the corpus available for download via the *Oxford Text Archive* and interrogate it via a third-party corpus search program like AntConc (Anthony 2019). Such programs usually allow users to search using regular expressions ('regex'), which, when used with care, can economically permit account to be taken of predictable spelling variation. Thus, if one wanted to confirm the grammars' statement that <b> for /v/ is an "extremely rare" spelling in most dialects after the eighth century (Hogg 1992: § 2.54), one could identify a list of lexemes that contain the phoneme, then devise regex searches to recover all predictable spellings. Thus for *dēofol* 'devil', one might run the following regex search:

```
[Dd](io|ia|eo|[aeiouyæ])[fuvb][aeiouyæ]?l
```

Square brackets tell the program to look for any one of the letters they contain, thus on their first occurrence <D> or <d> (a precaution necessary since regex searches are case sensitive), round brackets tell it to search for any of the alternatives demarked by pipes (i.e. <-io->, <-ia-> etc.), while a question mark indicates that segment may or may not be present (and is here used to take into account the possibility the unstressed vowel in *dēofol* has been syncopated); strings are unbounded, so this search will match any compound in which *dēofol* is the second element and any inflected form of the noun. This regular expression thus allows one to search for 560 putative spelling variants at one time. It is important to emphasise, however, that it will not necessarily recover all forms of *dēofol* in the corpus (e.g. an instance spelled *ðiofel* would not be returned), nor will all the forms it returns necessarily be instances of this lexeme (manual inspection of each hit is necessary to confirm that).[2] The results of a regex search are therefore

2 The anonymous reviewer kindly points out the existence of a full list of attested forms of *dēofol* in Feulner (2000: 192–195). For words beginning A–I, like *dēofol*, it is also possible to consult the list of

only as good as the formulation of that search, and some awareness of the range of spellings that might be expected in a text or language variety is necessary to formulate one that is likely to be reliable.[3]

Run across the entirety of *DOEC*, regex searches can, as we shall see, help illustrate the distribution across time and space of particular variants, the currency of words and idioms and highlight hitherto unknown interrelationships between different texts, as this article will go on to show. To interpret the results of such searches appropriate metadata about the different texts *DOEC* contains is however a necessity. While the corpus itself provides this only very minimally, each text it contains is equipped with a number (e.g. B3.4.37), which cross-references it to the "List of Old English Texts" (Cameron 1973; updated in Healey and Venezky 1980; and subsequently as part of *DOEC* and *DOE*), which provides references to a range of relevant handlists (principally Ker 1957; Sawyer 1968; and Okasha 1971) that give an indication of the date and (where available) origin or provenance of the witnesses that preserve the text. That "List of Old English Texts" also indicates which edition the compilers of *DOEC* took their text from and, by consulting that edition, it is usually possible to identify which witness it used as its base manuscript, and to assign the language of that editorial text to a particular date and place. By associating the results of a regex search with this metadata, we can plot the distribution of particular forms over time and space.[4] Thus using the regex string above but stipulating the medial consonant in *dēofol* must be <b> yields 31 instances, largely concentrated in the late Northumbrian of the glosses to *Lindisfarne Gospels* and *Durham Ritual*, which Aldred produced in the second half of the tenth century; in other dialects, it is indeed rare after the eighth century.

Because it is primarily a corpus of editorial attempts at reconstructing authorial works, *DOEC* does have a number of drawbacks, principally that variants present in a textual tradition but not in an editor's base manuscript will not be found there. Despite this, most extant manuscripts are represented at least to some ex-

forms offered by the *Dictionary of Old English*; however, this luxury is not yet available for words beginning with subsequent letters of the alphabet. Even for earlier letters, regex searches occasionally throw up forms in the *Corpus* which are not listed in the relevant entry in the *Dictionary*.

3 It might be natural to suppose that a solution to the risk of missing relevant matches would be to include even more putative spelling variants in the regex search, but while this might increase the recall (the number of relevant forms recovered), it would almost inevitably diminish the precision (the proportion of forms recalled that are relevant), thus necessitating additional post-processing.

4 More precisely, what we are primarily plotting is the place of a particular form in scribes' passive repertoires (though of course any given instance of a form might have been the active innovation of the scribe that wrote it). For the understanding of scribal behaviours and notions of repertoire operationalised in this paper, see especially Benskin and Laing (1981) and Laing (2004), as well as Laing and Lass (2013: § 1.4.5, 1.5.6).

tent in the *Corpus*: thus if one were interested in analysing, say, the language of the mid-eleventh-century homiliary, Cambridge, Trinity College, B.15.34, use of the additional metadata described in the previous paragraph reveals that *DOEC* contains texts of six of the 28 sermons in that manuscript, collectively totalling almost 10,000 words. This means that prominent patterns of variation in particular manuscripts are unlikely to go unnoticed.

In sum, with *DOEC* and other corpora not discussed in detail here, including the *Manchester Eleventh-Century Spellings Database*, the *York-Toronto-Helsinki Parsed Corpus of Old English Prose*, the *Linguistic Atlas of Early Middle English* and the *Corpus of Middle English Prose and Verse*, the raw materials already exist for a newly data-driven philology and book history. Developing new methodologies to use them will allow us to place 'Old' English orthographic, lexical and stylistic features in time and place with a precision not available to earlier generations of scholars who had to rely solely on wide reading and powerful memories. With these resources and appropriate approaches in hand, we turn to showing how they can be used to contextualise the language of the "Old" English Life of St Neot, and suggest the unlikeliness of a twelfth-century date for its composition.

3 Corpus Philology and Text Dating

Medieval texts rarely bear clear indications of their date of composition, yet a sense of when a work was written is essential to placing it and the manuscripts that transmit it in their appropriate cultural contexts. Dating an undated text (like dating an undated manuscript) relies on placing it relative to other works of known date; if a text, such as the *Eadwine Psalter* gloss, features spellings like <b> for /v/ and other features not common in works composed after the ninth century, then that text is likely to have been originally composed before 800 (Faulkner 2017a). The resources and approaches described in the previous section allow us to identify similarities in the linguistic behaviour of texts with new precision. In this section, we see how these techniques help show a twelfth-century date for the "Old" English Life of St Neot is, if not impossible, at least very unlikely.

The Life is uniquely preserved in the mid-twelfth-century London, British Library, Cotton Vespasian D.xiv.[5] Its date has been controversial in recent years, with Godden (2010) dating it to the early eleventh century and Younge (2012) to a century later. The Life – really a sermon for Neot's feast day – begins with a general exhortation that saints' lights should not be hidden under bushels (129/4–12).

5 The Life is cited from Warner (1917: 129–134), by page and line number.

It then summarises his display of youthful virtue (129/12–24) and sketches his career in religion (129/24–130/29), before narrating two miracles at some length: the first in which a lost shoe is with God's help miraculously returned to Neot (130/29–131/17), the second in which Neot prophesises concerning King Alfred's battles with the Danes (132/16–133/24). It closes by contrasting these *gode dages* 'good days' with the irreligiousness and criminality of the present and the hope that with Neot and other saints as our intercessors, we may come after death to eternal rest (133/24–134/4).

In what follows, I show how corpus-based methods can be used to determine whether an early-twelfth-century date is plausible for its orthography, lexis, style and syntax and contribute to an understanding of its sources, thus moving from the smallest unit of investigation (individual letters and digraphs) to the largest (intertextual relationships). My primary concern is establishing whether the Life is pre- or post-Conquest, rather than attempting a more finely-grained dating.

3.1 Earlier Views Regarding the Date of the Life of St Neot

The composition of the Life of St Neot has been dated anywhere between the tenth and twelfth centuries. Scragg (1979: 261) listed it among "some pieces [which] may belong to the pre-Ælfrician tradition"; Hardy (1862–1871: I, 539) suggested the apocalypticism of its ending pointed to a date before 1000, perhaps shortly after 986, the year of a bad murrain, a reference to which he detected in the mention of the *orefcwealm mycel* 'extensive disease of cattle' which *nu is* 'is now' (133/26–27). Wülker (1885: 494) attributed it to Ælfric; Richards (1981: 263) dated it to "the middle of the eleventh century, surely before 1066"; Lapidge and Keynes (1983: 198) suggested "late eleventh or early twelfth century". Clemoes (1997: 17) described it as "certainly or probably composed in the twelfth century", while Smyth (1995: 327) had it as "early twelfth century"; Dumville and Lapidge (1985: cxvi) placed it in the first half of that century and Roberts (2000: 447) suggested it was composed "near the time of the compilation of the manuscript in the middle of the twelfth century". The early years of the last decade, however, saw the publication of two detailed studies, one arguing for a date between *c.* 1015 and *c.* 1030 (Godden 2010: 222), the other proposing an origin "at the beginning of the twelfth century, within the orbit of the archiepiscopal see of Canterbury and the sphere of influence of St Anselm" (Younge 2012: 368).[6]

6 Godden's (2010: 209) *terminus ante quem* rests on the supposition that the Life was written while Neot's body was still in Cornwall, prior to its removal to Eynesbury, where it is said to rest in a list of

Most of the argumentation about the date of the Life has hitherto focused on determining which period of literary and cultural history is most likely to have given rise to particular aspects of its presentation of the story. These have included its ambivalent portrait of King Alfred and its chronologically problematic assertion that Neot, who it asserts died during that king's reign in the late ninth century, was consecrated to the priesthood by St Elphege (129/24–26), still alive in 1011, an error which Plummer (1902: 55) regarded as "absolutely conclusive" evidence against a date in the tenth century, even though, as Napier pointed out to Stevenson (1904: 258), the honorific *sancte* might here be a later interpolation, with the reference originally having been to another Ælfheah. Thus, much of Younge's argument for attributing the composition of the Life to early-twelfth-century Canterbury depends on identifying similarities between the intellectual environment he detects there and features of the Life. Similarly, Godden's (2010: 201) suggestion of a pre-Conquest date in part rests on his sense that its "confidence in imitating" the language of Ælfric and Wulfstan is "without parallel in the twelfth century". But our knowledge of the literary and cultural histories of the eleventh and twelfth centuries remains imperfect, and such arguments cannot plausibly be decisive on their own. The other major class of evidence that has informed discussion has come from a consideration of the Life's sources, with Godden proposing his *terminus post quem* of 1014 on the basis of what he and others have judged to be the Life's direct use of Wulfstan's *Sermo Lupi ad Anglos* and Younge suggesting one of 1050 on the basis of its supposed reliance on the so-called *Vita I*, to which Dumville and Lapidge (1985: xcvi) tentatively assigned that date.[7]

Language, however, has not loomed large in these discussions, with most commentators giving it only a token nod and Godden and Younge devoting only a couple of pages each to considering it. Both acknowledge, correctly, that the

saints and where their relics are to be found, the *Secgan be þam Godes sanctum*, usually dated 1013x1031.

7 Despite Younge's arguments, there are many details in the Life not found in *Vita I*, as well as a number of places where the Life seems closer to *Vita II* than *Vita I* (e.g. the statement that Neot was visited by angels who *hine gefrefreden 7 wel geherten* 'powerfully comforted and heartened him' (130/22), which as Richards (1981: 274) notes, is close to *Vita II*'s *confortaretur [...] et sustentaretur* 'he was comforted and sustained' (Sollerio et al. 1868: 334, § 25), suggesting it is more probable the "Old" English Life was ultimately reliant on a third source, a common ancestor of the two *vitae*, which does not survive (and is therefore not readily datable). In any case, there is far from consensus about the date of *Vita I*, with some scholars (e.g. Lapidge and Keynes 1983: 198; Hart 1990: 610; Wright 1995: 1) putting it in the late tenth century.

spelling of the copy of the Life in Vespasian D.xiv is scribal and not authorial and therefore cannot provide any straightforward evidence of the text's origins. Though mentioning in passing that "there is [no]thing in phrasing or syntax which might suggest a post-Conquest date, apart from changes which might equally be due to a scribe", the discussion of Godden focuses primarily on the Life's vocabulary, suggesting "there is not a single word in the text that was not current in tenth-century English or earlier" and some of its vocabulary "would have been distinctly obsolescent by the twelfth century" (2010: 201–202). Younge's discussion of language (2012: 353–354) responds only to the specific lexical items discussed by Godden, suggesting *handbreden* 'palms of the hand' (132/5) is used in an intertextual allusion, noting *herelaf* 'remnant of an army' (133/17) was retained by post-Conquest scribes copying Old English texts and there is therefore "little reason to believe it had become obsolete" and observing *æ* 'law' (129/21) was used in the definitely post-Conquest translation of the *Elucidarium*. He concludes that there is little in the overall lexical profile of the Life of St Neot to distinguish it from this and another assuredly post-Conquest composition preserved in Vespasian D.xiv, Ralph d'Escures' sermon on the Assumption of the Virgin.

In what follows, I expand significantly on Godden's and Younge's discussion of language, showing how novel methods for using existing corpora can be used to contextualise features of the Life's spelling, vocabulary, style, syntax and sources more richly than is possible with existing handbooks and dictionaries. The evidence of these methods makes it very unlikely the vast majority of the Life is of early-twelfth-century composition.

3.2 Orthographic Evidence

The orthography of the bulk of Vespasian D.xiv is homogenous and somewhat idiosyncratic, copied by the main scribe from the exemplar that underlies the collection. I have described this language elsewhere and labelled it Hand 1 Language (H1L) (Faulkner 2017b). Any spellings in the Life of St Neot that deviate from the norms otherwise observed in the manuscript are therefore potentially significant, either as residual forms accidentally left 'untranslated' by the scribe who originally rendered it in H1L or innovations introduced by Hand 1 when he copied the Life in the mid-twelfth century as rare lapses from accurate literatim copying. Any one form might in principle derive from either process, but, given the hints that one of the other scribe's unconstrained usage was more modern than H1L (Faulkner 2017b: 306–307), it is a reasonable assumption that forms that were not otherwise commonly in use in the twelfth century derive from an

earlier stage in the Life's transmission.[8] Such forms suggest that the Life of St Neot existed in a pre-H1L form, with one in particular pointing to its existence in written form in the eleventh century.

Despite the thoroughness with which the contents of Vespasian D.xiv were in general translated into H1L, there are clear indications that the Life of St Neot once existed in a form of Old English that lacked distinctive H1L features. These include two spellings with <e> rather than H1L's usual <æ> for the i-umlaut of /aN/: *asende* 'sent forth' (132/6), beside typical *sænden* (130/19) and *gesænde* (131/12);[9] and, in the following sentence, *engles* 'angels' (132/6), beside H1L's usual *ængles* (130/22, 25).[10] Its exemplar also seems to have had the dative plural of at least one noun in <-um>, rather than H1L's preferred <-en>, as the Life's *mid [...] wyrtgemangum* 'with [...] herbs' (132/11–12) suggests.[11] Post-Conquest compositions continued to use <-um> among other spellings for the dative plural, so its employment in an earlier stage of the transmission of the Life (like the use of <e> for the i-umlaut of /aN/) cannot be taken as evidence for a pre-Conquest origin for the text, but it, like the <e>-spellings, is a useful indication that it was composed at least two stages prior to its instantiation in Vespasian D.xiv.

More directly suggestive of a pre-Conquest date for the Life's composition is the form *simble* 'always' (131/31), though one would not necessarily know this from the grammars, which note the epenthesised consonant but do not comment on the chronological distribution of spellings showing it (Campbell 1959: § 478.1;

8 As an example of a form that is very likely to be Hand 1's innovation, consider *wæstdæles* 'western parts' (130/7), where the <-æ-> in *wæst-* (beside <e> x 3 elsewhere in H1L (90/25, 31; 104/26)) suggests he, like Hand 2, no longer heeded Old English norms for representing low front vowels (Faulkner 2017b: 306 n. 34).

9 As Faulkner (2017b: 315) reports, H1L has *sænden* (53 instances), *senden* (3 instances). The other two instances of *senden* are in Ælfric's First Series Homily for the Passion of St John (57/19) and the Gospel of Nichodemus (83/20). The Life of St Neot also has a higher than average proportion of <-e->-spellings (<e> 4 instances, <æ> 1 instance) in *wænden*, in which H1L as a whole has <e>:<æ> in a ratio of 1:2.

10 H1L has *æng(e)l* x 135, *engel* x 4 (Faulkner 2017b: 314). The other three <e-> spellings are in Warner 4 (8/7), Warner 28 (69/18) and Warner 37 (111/23), all Ælfrician items, where Vespasian D. xiv's ultimate (if not necessarily proximate) exemplar would have had <e->. Lest anyone contemplate the proximity of *asende* and *engles* in the Life might indicate an interpolation post-dating the text's translation into H1L, it is worth noting that the sense in which *asende* is used here seems to have been obsolescent in the twelfth century (see Section 3.3 below).

11 There are 128 instances of word-final <-um> in Hand 1's portion of the manuscript; manually analysing these (the majority of which are part of Latin proper nouns, the Old English pronoun *sum* or Old English adjectives like *gehyrsum*) yields just two other instances of a dative plural inflected in this way: *þingum* (55/1) and *mæignum* (55/7), both in Warner 19, a copy of Ælfric's sermon for the Passion of St John.

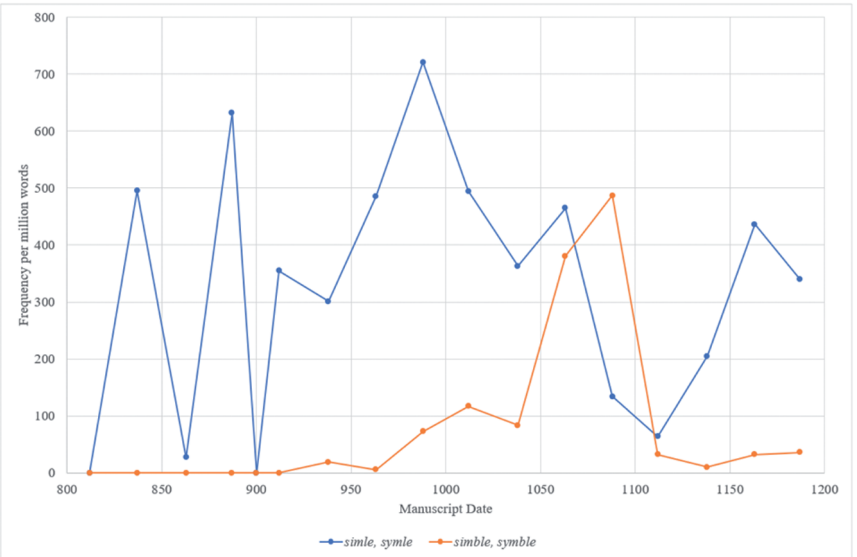

Figure 1: Frequency of spellings *simle*, *symle*, *simble*, *symble* in *DOEC*

Hogg 1992: § 7.87; etc). This is the only instance of a <-b->-spelling in H1L, which otherwise has *symle* or *simle* in roughly equal proportions. Existing resources suggest spellings with <-b-> are sufficiently rare in post-Conquest English to suggest the form is unlikely to be Hand 1's innovation.[12] It is therefore probable it has been carried over from an earlier phase in the Life of St Neot's transmission. Using a technique I have written about in more detail elsewhere (Faulkner forthcoming a), it is possible to use *DOEC* to see how frequently a particular spelling was used in manuscripts of different dates (Figure 1).

One major caveat should be noted before interpreting this graph, which is that the data must include some forms of *symbel* 'feast' and since that noun may have been obsolescent in 'Middle' English, this may lead us to overestimate the proportion of <-b->-spellings for *symle* 'always' in the pre-Conquest period relative to their use thereafter.[13] Nonetheless, the graph clearly shows that <-b->-forms

12 Thus the quotations given in *MED* **simle** adv. include only two such spellings (this one, and one from Warner 48, copied by Hand 2), while *LAEME*'s only <-b->-spelling (as it happens, its only instance of the lexel $simle) is in a copy of an ostensibly pre-Conquest document from the Holme cartulary.

13 Two readers questioned the presentation of this statement that the data contains forms of *symbel* as a probability not a fact, but this is an inherent aspect of the methodology being advanced here. One could of course check manually what proportion of all 1,780 datapoints are forms of

were practically non-existent before the final quarter of the tenth century and very scarce in the twelfth; their heyday was therefore the eleventh century. While it might be tempting to use the spelling as evidence that the Life of St Neot was composed sometime in the second half of that century, it should be remembered that these figures primarily represent scribal practice copying earlier texts, which was often conservative; there may thus be some lag between <-b->-spellings' actual vogue and the peak on the graph. While *simble* does not help establish a *terminus post quem* for the composition of the Life, since in theory a text composed in the ninth century might contain a characteristically eleventh-century spelling as a result of its transcription by a translating scribe in that period, it does however suggest that the Life is not likely to have been first composed in the twelfth century.

3.3 Lexical Evidence

The frequency with which particular words are used changes over time. A text that contains the word *vulgarian* is, on the surface of it, over five times more likely to have been written around 1930 than in the nineteenth century. One that uses the word *dashed* is slightly more likely to have been written in the nineteenth century than in the present, and it is conceivable the skew would be even more telling if there was a way to look only at the distribution of the adjective indicating 'the state of being accursed' and not past tenses of the verb *dash* in its various meanings ('struck', 'ruined', 'excised', 'ran', etc.).[14] One would not necessarily know these things from historical dictionaries, which typically only provide basic indications of the temporal distribution for the majority of words, most notably

symbel 'feast' but this would require considerable time, a time that would have to be multiplied by perhaps a hundred if all the different potential dating indicators the Life yields and which were considered in the research that underlies this paper had to be similarly checked. And here we are trying to date just one text, when one might envisage another study tackling ten or twenty texts at a time.

14 These impressionistic deductions about the likely date of texts using these words rest on the Google Ngrams of their distribution, which map their use by date of publication in at least 8m volumes and 0.5tn words of English (for further discussion, see Faulkner forthcoming a). I invoke Google Ngrams here rather than a corpus, and couch the observations from it deliberately imprecisely, to make a closer parallel to results drawn from *DOEC*, which (despite the name) is not a corpus and thus not necessarily representative, rather just a very large repository of texts.

their earliest attestation.[15] Corpus-based analyses of vocabulary, like the claims about the distribution of *vulgarian* and *dashed* extrapolated from Google Ngrams above, can however illuminate words' currency with more precision, and applying such methods to the Life of St Neot strongly suggests the life was composed before the Conquest and at any rate before the early twelfth century.

As has already been noted, though primarily a corpus of texts, *DOEC* is also indirectly a corpus of the language of particular manuscripts and scribes, whose work can be dated palaeographically. No text copied in the first half of the eleventh century can be a post-Conquest composition; one transcribed in the mid-twelfth century might be. For the purposes of considering the vocabulary of the Life, the threshold for a text to putatively be of post-Conquest composition has been set with hands dating palaeographically later than the middle of the second half of the eleventh century ("s. xi^{2}"), a decision in part principled, in part practical.[16] I have therefore examined the distribution of each word used in the Life of St Neot but not present in the *MED* (or attested there only from twelfth-century copies of Old English texts) in *DOEC*, using the list of spellings attested in the *DOE* for words beginning with letters A through I and regex for those that start with subsequent letters. These words I have classified as follows:[17]

> Type (i): words for which there is no evidence independent of the Life that they were in active use in the post-Conquest period; e.g. *onette* 'hastened' (130/34).

15 A work like the *Middle English Dictionary* with a commitment to accurately representing the period of a word's currency through its selection of quotations (Lewis et al 2007: 18) in principle makes it possible to evaluate a word's use with considerable precision. In practice, however, the *MED*'s reliability on particular words' currency is inconsistent, in part because of its extensive and not always critical incorporation of a very significant body of quotations from twelfth-century copies of Old English texts. The *Dictionary of Old English* only sporadically characterises a word's diachronic, dialectal and diaphasic currency with comments like "mainly in Ælfric" or "in texts of ca. 1200 and later".

16 The practical reason is that *DOEC* contains almost 350,000 words copied s. xi^{2} and xi$^{3/4}$, so including such texts would make interpreting search results considerably more laborious; the principled one that the data primarily represents scribe's passive usage, and there is likely to be a lag between a word's actual frequency in active use and its frequency there. Setting the threshold with s. xi^{2} does however mean it is not strictly speaking possible to disavow the Life of St Neot's lexis originated between 1066 and *c.* 1085.

17 These three types may be seen as sub-categories of Types A (words not found in *MED*) and B (words found in *MED*, but there only attested in late copies of Old English texts) in my earlier typology for categorising why a twelfth-century annotator glossed some words and not others in a copy of Ælfric's First Series of *Catholic Homilies* (Faulkner 2012: 193).

Type (ii): words for which there is no evidence independent of the Life that they were in active use in the post-Conquest period, although the morphemes and/or lexemes that comprised them remained so in 'Middle' English; e.g. *forraðe* 'very quickly' (133/1).

Type (iii): words for which there is no evidence independent of the Life that they were in active use in the post-Conquest period, although closely-related words were in active use in 'Middle' English; e.g. *earfoðe* 'hardship' (132/36), alongside surviving *arveð* 'difficult', *arveðlīch* 'miserable', etc.

Within each subclass, words are counted according to whether they occur in no other texts which may be post-Conquest compositions or in a number of texts which could be (primarily because they survive only in post-Conquest manuscripts and their date is un(der)considered or remains controversial). Thus *arfæste* 'benevolent' (131/11) is categorised as Type (ii) only with caveats, since it is also found in the poem *Durham*, the date of which remains uncertain (see most recently Appleton 2016; Bailey and Cambridge 2016). To avoid the dangers of prejudging the dates of particular texts, I have tended to use this caveat category fairly liberally, applying it to texts often thought post-Conquest (e.g. the Life of St Margaret in Cambridge, Corpus Christi College, 303) and those usually assumed to be pre-Conquest (e.g. some of the sermons uniquely preserved in Cambridge, Corpus Christi College, 302). I would hope it will in due course prove possible to assign each of these texts a secure date; in the meantime, we must try to date the Life of St Neot (and in due course them) with regard primarily to securely-dated texts, for instance the post-Conquest annals of the Peterborough Chronicle and works whose post-Conquest date is guaranteed by the late composition of their sources, e.g. the 'Old' English Honorius and the sermon printed from Cotton Vespasian A.xxii as *An Bispel* (on the date of which, see e.g. Millett 2010: 229).

Analysing the vocabulary of the Life of St Neot according to this typology yields the following results:[18]

18 In the rightmost column of this and subsequent tables, 'token' refers to the total number of words in a text, 'type' to the number of distinct words it uses. The lexemes belonging to each category in the Life of St Neot and the comparanda Honorius 1 and Honorius 2 are itemised in the Appendix.

Table 1: Lexical profile of the Life of St Neot (1999 tokens, *c.* 550 types)

Type	Frequency (caveats)	Frequency (no caveats)	Frequency no caveats / 1000 tokens	Frequency no caveats / 1000 types
(i): no independent evidence lexeme in active use after 1066	8	12	6.0	21.8
(ii): no independent evidence lexeme in active use after 1066 but all component lexemes survived into 'Middle' English	7	12	6.0	21.8
(iii): no independent evidence lexeme in active use after 1066, but related lexeme(s) survived into 'Middle' English	3	9	4.5	16.4
Total	18	33	16.5	60.0

Words not independently evidenced as in active use after the Conquest thus loom large in the Life of St Neot, with 6 per cent of its lexicon not found in any other certain post-Conquest composition. But evaluating the significance of this figure is tricky. The lexicon of English underwent significant change in the course of the twelfth century and the quantity of writing in English diminished significantly, at least after 1140 (Faulkner forthcoming b, esp. Ch. 3), and it may be that a text composed in the early twelfth century might provide the latest attestation for quite a number of pre-Conquest English words. The Life of St Neot's lexical profile can therefore only have meaning in comparison to similar figures from known twelfth-century compositions. Ideally we would compare its profile with multiple such texts; below, as a preliminary, are those for the two excerpts from Honorius's *Elucidarium* (Warner 45, 46) translated into English sometime in the first half of the twelfth century, not necessarily by the same author.

Comparison with the lexical profiles for Honorius 1 and Honorius 2 suggests that if the Life of St Neot is a post-Conquest composition, it contains an unusually high proportion of 'Old' English vocabulary not otherwise known to have been in use after 1066. Here are the figures for Honorius 1:

Table 2: Lexical profile of Honorius 1 (1511 tokens, *c.* 345 types)

Type	Frequency (caveats)	Frequency (no caveats)	Frequency no caveats / 1000 tokens	Frequency no caveats / 1000 types
(i): no independent evidence lexeme in active use after 1066	2	3	2.0	8.7
(ii): no independent evidence lexeme in active use after 1066 but all component lexemes survived into 'Middle' English	1	8	5.3	23.2
(iii): no independent evidence lexeme in active use after 1066, but related lexeme(s) survived into 'Middle' English	0	0	0	0
Total	3	11	7.3	31.9

Honorius 1 thus employs just over half the number of words that the Life of St Neot does which were not demonstrably in use in the post-Conquest period. But this figure perhaps understates the difference between the two texts, since a proportion of Honorius 1's Type (ii) words are hapax legomena or exceptionally rare in Old English and probably the author's own coinages from elements still available in Middle English.[19] The figures for Honorius 2 are considerably lower again, with just 1 per cent of its lexicon not otherwise attested in use after 1066:

19 These include *hindernysse* 'wickedness' (142/24), which *DOE* **hindernes** gives as a hapax legomenon; and *teonlease* 'free from suffering' (143/5, 12), given by *B-T Enlarged Addenda* **tēonleas** only from this text and which does not appear to occur otherwise in the *DOEC*. This apparent fondness for coining words also gives one to wonder whether some of the other words he uses which are attested in Old English but only relatively rarely might be his own recoinages rather than part of a lexicon he inherited, e.g. *gewittscipe* 'knowledge' (140/31, as part of the prepositional phrase *beo his gewittscipe* 'with his knowledge', translating *scienter* 'knowingly'), apparently only found otherwise once in the *Old English Bede*, and there used in the sense 'condition of being a witness in a transaction'.

Table 3: Lexical profile of Honorius 2 (583 tokens, c. 165 types)

Type	Frequency (caveats)	Frequency (no caveats)	Frequency no caveats / 1000 tokens	Frequency no caveats / 1000 types
(i): no independent evidence lexeme in active use after 1066	1	1	1.7	6.0
(ii): no independent evidence lexeme in active use after 1066 but all component lexemes survived into 'Middle' English	0	1	1.7	6.0
(iii): no independent evidence lexeme in active use after 1066, but related lexeme(s) survived into 'Middle' English	0	0	0	0
Total	1	2	3.4	12.0

Comparison with the lexical profiles of Honorius 1 and Honorius 2 does therefore suggest that if the Life of St Neot is a post-Conquest composition, it uses an unusually high number of words that seem to have retired from active use by that period. The Life also uses fewer neologisms than Honorius 1 and 2,[20] as well as a greater number of words which while surviving into 'Middle' English were no longer used there in their 'Old' English senses,[21] though here we are reliant on the dictionaries and the argument is to some extent qualitative.

20 Pace Godden (2010: 202), the surviving text of the Life does however contain a handful of words that do not appear to have been current in tenth-century English, including *seinte* (130/3, 134/1 and in the rubric which introduces the text in Vespasian D.xiv); *ætfaren* 'depart' (131/10), which is a hapax legomenon in Old English and otherwise only attested in Layamon's *Brut*; and *aweigward* 'transitory' (133/37), attested only here in Old English, but used much more commonly in Middle English, albeit primarily as an adverb meaning 'in a different direction'. However, Honorius 1 has three words derived from likely Norse loans (*unscellice* 'unwisely' (141/2), *litigere* 'painter' (141/9) and *lit* 'colour' (141/10)), as well as *emcristene* 'fellow Christians' (141/30), *unfere* 'incapacitated' (141/38; 142/18) and *unfernysse* 'incapacity' (142/16); and *gefostrede* 'fostered' (142/37). The shorter Honorius 2 has Norse-derived *scele* 'discretion' (145/4), as well as *sicerlice* 'assuredly' (145/12).

21 Words the Life uses in senses apparently no longer found in Middle English texts include *to þycgene* 'to drink' (130/10); *adreah* 'lived' (130/14); *wise* 'occurrence' (131/7); *asende* 'sent forth [his soul]' (132/6); *aleigd* 'reburied' (132/15); and *gebege* 'will convert' (133/13). Honorius 2 does use a

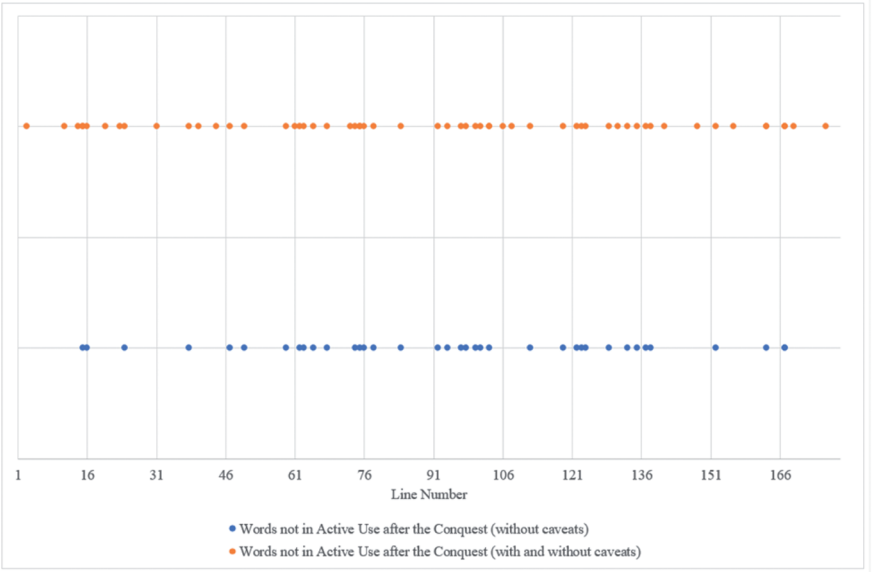

Figure 2: Distribution of obsolescent words in the Life of St Neot

The differences between the lexical profile of the Life of St Neot and that of the two translations from the Old English Honorius preserved alongside it in Vespasian D.xiv thus suggest the Life of St Neot is likely to have been written at a time when a greater portion of Old English vocabulary was current and very probably before the final quarter of the eleventh century, by when some 6 per cent of its vocabulary is not otherwise attested in active use. It is also notable that the obsolescent words the Life uses are evenly distributed throughout the text.

This graph plots the occurrence of obsolete words through the textual space of the Life, with the x-axis corresponding to the cumulative line number in Warner's edition (e.g. *gelustfullunge* 'pleasure' at Warner 129/18 occurs in line 15 of the text since the Life begins at Warner 129/4) and the blue dots representing instances of words that can be shown not to have been in active use after the Conquest without caveat, and the orange dots such words and those for which some caveats exist. While the graph does show there are some passages as much as fifteen lines long that contain no words that would become obsolete soon after

number of words in obsolescent senses, including *ȝinȝran* 'followers' (144/12) and *abiten* 'eat' (144/22), but Honorius 1 balances this by employing some in novel senses, including *beteon* (142/7) in the sense 'bestow'.

the Conquest, these never correspond to what could plausibly be seen as discrete textual units and therefore additions or interpolations subsequent to the composition of the bulk of the Life. Thus the gap between *geðeah* in line 24 (Warner 129/27) and *wæterseaðes* 'springs' in line 38 (Warner 130/9) encompasses the second half of a long sentence about Neot's ascetic practices, his visits to Rome and the opening part of the description of his way of life at *Neotes stoca*. The lexical evidence therefore suggests that the vast majority of the Life as it is preserved in Vespasian D.xiv was composed before the Conquest, or at least before the final quarter of the eleventh century.

3.4 Stylistic Evidence

Regex searches can also be used to investigate collocation, and thereby linguistic patterns of two or more words' length, like idioms. When contextualising the language of the Life of St Neot, we might, for instance, be interested in what other texts use the idiom *cēpan flēames* 'take flight', which it uses to describe the cowardly retreat of King Alfred's troops and then Alfred himself from combat with the Danes (132/19–20, 22). We can accomplish this with a regular expression search:

```
[Ff]l(ea|ie|y|i)m.?.?\W+(\w+\W+){0,5}[Cc](ie|y|i|e)p|[Cc][æe]p.?.?.?.?\W+(\w+\W+){0,5}[Ff]l(ea|æ)m
```

This string searches for an inflected or uninflected form of *flēam* (spelled <fleam> or <flæm>) adjacent to or within five words of an inflected form of *cēpan* (with the stem spelled <cep-> or <cæp->). The search returns nine hits, all of which, with the exception of that here, are from the works of Ælfric.[22]

In interpreting the significance of this overlap between the phrasing of the Life of St Neot and the works of Ælfric, we should first note that it is not likely to be simply a phrase that occurred spontaneously to both authors. As *DOE* shows, while *cēpan* + genitive was used in two other Old English idioms meaning 'to be intent on something', namely *þances cēpan* 'to be grateful' and *hearmes cēpan* 'to intend harm against', it seems unlikely that the author of the Life and Ælfric would have hit on the same expression simply by chance, particularly as other idioms were available in Old English texts to express the concept 'take flight', including *weorþan*/*bēon on flēame*, *dælan flēame*, *ætwindan mid flēame*, *ge-*

22 The dominance of Ælfric's works in attestations of the idiom is noted in *DOE* **flēam**, 2.b.ii *cepan fleames*, which is described as occurring "mainly in Ælfric".

windan on flēam, *sceacan*/*gewitan on flēam* and *(ge)wyrcan flēam(es)*.[23] The question remains what significance should be attached to this shared usage.

In answering this, it is first worth noting that the idiom is not attested in 'Middle' English, nor (as the search of *DOEC* shows) are there any examples of its active use after 1066.[24] Its use in the Life of St Neot therefore provides at least some further indication that the Life is unlikely to have been composed after the Conquest. Second, its use might suggest that the author of the Life was familiar with or trained in the modes of expression habitual to Old English literate culture or even knew the works of Ælfric. However, it is rarely easy to judge whether a particular correspondence simply reflects two authors' shared knowledge of an idiom they both independently acquired when they, presumably as children, learned to speak English, a shared intellectual training, or deliberate imitation by one of a written work by the other. But it seems *a priori* more likely that the shared use of an idiom that has some specialised or metaphorical meaning indicates a common intellectual training rather than shared use of the English language. Depending on the distribution of the idiom in surviving texts, that training might have been the kind of training all persons of literary interests in pre-Conquest England could access (presumably in monasteries), or something more specific, such as the training given in the monasteries responsible for the phenomenon of Winchester words (Hofstetter 1987). If the correspondence to another written text is particularly close and distinctive, it may, on the other hand, reflect the use of that text as a source by the author of the other (see Section 3.6 below). Thus, since, as the *DOE* shows, other Old English texts used different synonyms for *flēames cēpan* and our corpus work confirms that *flēames cēpan* was used by no one other than Ælfric, its employment here almost certainly reflects (chances of survival notwithstanding) at the very least a shared training more specific than that available in all monasteries, while the lack of any closer correspondence between the two uses in the Life of St Neot and any of the seven instances in Ælfric's works suggests it is not simply a borrowing from one of these texts.

Some of the correspondences between the phrasing of the Life of St Neot and other Old English texts that can be found through collocation searches in *DOEC* necessarily indicate little more than the likely pre-Conquest date of the Life. As

23 These are just the idioms involving *flēam*; for others, see *HTOE* 01.14.05.13|20 (vi.) **run away/flee**. By contrast, though the Life shares the idiom *mid ealre heorte* 'wholeheartedly' (131/31) with a number of other Old English texts, the general availability of *eall* in the sense 'whole, entire, all' and the frequency of *heort* in Old English discourse, makes this unlikely to be an overlap of any significance.

24 *MED* **flēm** n.(1) gives the verbs that collocate with it in 'Middle' English as **astellen**, **maken**, **nimen**, **taken ~** and **setten to**. *LAEME* $fle:am collocates only with *make*.

with the words from the Life's vocabulary discussed in Section 3.3, these are often idioms that can be shown not to have been current in what we now call 'Middle' English; indeed some of them can be shown not to have been in active use after the Norman Conquest. Thus, as far as I am aware, no 'Middle' English sermon addresses its audience as *menn þa leofeste* 'most beloved people' (129/4), which is ubiquitous in 'Old' English homilies, though how far its presence here tells against a twelfth-century date is limited by at least ten of these being extant only in twelfth-century manuscripts, and therefore at least possibly of post-Conquest composition.[25] The idiom *līf alætan* 'to give up one's life', used in the description of Neot's death (131/13), is found a total of fourteen times in 'Old' English but not in any manuscript copied after the Conquest and in 'Middle' English the verb it features meant only 'drain off' or 'dissolve', making the idiom's use in the Life a further indication that it is unlikely to have been composed after 1066.[26] Other idioms found in the Life that are either rare or unattested after the Conquest and in 'Middle' English and that therefore point in a similar direction include *līfes weg* '[correct] way of life' (129/7),[27] and *Drihtenes þrōwunge* 'the Gospel narrative of the Lord's passion' (131/34–35).[28] These show a debt to the phraseology of pre-Conquest sermons, presumably as a consequence of an active training in the medium.

Some of the correspondences, however, suggest a more specific debt. The Life uses several idioms which are markedly Ælfrician, including *on ǣr* 'beforehand' (132/9, 132/36), the formula *X wide sprang* 'X diffused widely' (133/18), not found otherwise outside poetry, Ælfric's works and the Life, *þurh Godes gast* 'by means of God's spirit' (discussed below, Section 3.6) as well as *flēames cēpan* 'take flight'

25 For instance, *men þa leofestan* opens an excommunication formula uniquely preserved in Cambridge, Corpus Christi College, 303 (Treharne 1995: 210).

26 *DOE* **alætan** 4.c.; *MED* **aleten**. The idiom was however in use around the mid-eleventh century, as the annals for 1045 in the 'D' Chronicle (Cubbin 1996: 67) and for 1053 in the 'C' Chronicle (O'Brien O'Keeffe 2001: 115) show.

27 The idiom occurs 24 other times in *DOEC*, including in the gloss to the *Benedictine Rule* where it translates *viam vite* (Logeman 1888: 3/10), but not other than in the Life of St Neot in a manuscript dated later than s. xi^2; *MED* **wei** 7b (a) **lifes (lifli) ~** (with this apparently the only instance). In *LAEME*, *lifes weg* occurs only in the *Wintney Rule of St Benedict*, but it is possible that God showing *þe wit ant te wei of lif* translation to Adam and Eve in the Katherine Group *Life of St Katherine* (d'Ardenne and Dobson 1981: 48/328) is a post-Conquest instance of the same idiom.

28 There appear to be only three other instances of *Drihtenes þrowung* referring to a written text, two of which are in the later version of Vercelli 1 in Oxford, Bodleian Library, Bodley 340 + 342 and Cambridge, Corpus Christi College, 162 (Scragg 1992: 19/142–143, 35/320), the other in the fragmentary translation of the *Regularis Concordia* in Cambridge, Corpus Christi College, 201, p. 1–178 (Zupitza 1890: 16/219). *MED* **throuinge** ger.(2), 1. (d) has as its sole citation this passage from the Life of St Neot.

discussed above.[29] But some point more broadly than just to Ælfric. These would include the phrase *nan mænnisc tunge hit eall asecgan ne mæig* 'no human tongue might narrate it all' (131/22), which bears some resemblance to a phrase from the Life of St Machutus, a work generally dated to the early eleventh century and assigned to a Winchester milieu,[30] as well as the 'Winchester words' *gelaðung* 'church' (129/8) and *bereowsede* 'repented' (130/2) (Hofstetter 1987: 244).[31] All of these idioms are unattested in active use after *c.* 1075 and therefore provide further evidence that the Life is unlikely to have been composed after the Conquest.

3.5 Syntactic Evidence

Pending the development of an automated part of speech tagger and parser for Old and Middle English (or the investment of significant research funding in performing this tagging manually), syntax is (alongside semantics) the trickiest linguistic feature on which to collect reliable data. The handbooks (e.g. Mustanoja 1960; Visser 1963–1973; Kisbye 1971–1972; Mitchell 1985) therefore remain essential reference. But while orthographic and lexical fashions often change rapidly, so that many words used widely in 'Old' English writing disappeared almost immediately after the Conquest, syntactic features often obsolesce over much longer time periods, necessitating robust data about the proportions of particular constructions in use at a given time. This makes a handbook statement that there is an instance of a particular construction in one thirteenth-century text a fact of only questionable utility. Consequently, they must be supplemented with corpus-based approaches. Unfortunately, however, *YCOE* and its kinsman *PPCME2* are not entirely straightforward to use, limited in their coverage and not always

29 *MED* **on** prep. 18. (f) **~ er**, with the only citation from Belfour 4, now attributed to Ælfric. In the *DOEC*, the idiom is found 43 other times, with the only instances in manuscripts copied later than s. xi^2 from late manuscripts of Ælfrician works and Æthelwold's account of the refoundation of the monasteries. 33 of the 43 instances are from texts by Ælfric. For *X wide sprang*, see Godden (2010: 200). Compare also the lexeme *madmfaten* 'treasure chests' (132/23), which outside of *Beowulf* 2405a is used only here and five times by Ælfric.

30 In the Life of St Machutus, *ne mæg nan mænnisc tunge areccean* 'no human tongue might tell' how assiduous he was in prayer (Yerkes 1984: 69/2).

31 *Behreosian* is amply attested in Middle English, but there is little evidence this usage of *gelaðung* survived: see *MED* **ilathung(e** ger. (in various twelfth-century copies of Old English texts and the Tremulous Hand's copy of the Nicene Creed, the original date of composition of which remains uncertain); and **lāthinge** n. (a) **godes (holi, ilefful) ~**, where all the quotations with this meaning are from twelfth-century copies of Old English texts.

tagged in sufficiently fine-grained detail to permit locating the constructions that might interest us.[32] Some interesting work has however been done on their potential utility for dating texts, chiefly in an unpublished paper by Zimmermann, developed by Ecay and Pintzuk (2016).

In his study, Zimmermann (n. d.: 7–14) selected fourteen syntactic variables from the period 850–1230, calculating the frequency with which the major variants were used in a series of datable texts, trained a naïve Bayes classifier on these works, then asked it to date a number of undated texts, including the Life of St Neot, which the classifier assigned to either 900–990 (in its view, a 68.9 % chance) or 990–1100 (a 28.9 % chance).[33] It judged there was a small probability (2.2 %) it was from 850–900, but, tellingly, no chance at all it was from later than 1100. The syntactic criteria used are largely features of language users' internal grammar that would have operated below writers' thresholds of consciousness. But some are more superficial, for instance his "GEN criterion", which concerns the position of a genitive relative to its head noun. Both the 'conservative' postnominal position and the (in the longue durée) 'innovative' prenominal one are used in the Life of St Neot. Thus both appear in the sentence describing the location of Neot's cell (130/7–8):

> Seo is wæstdæles **þysses landes**, ten milen fram Petrocesstowe þa me hatt **Neotes** stoca.
> 'This [deserted place] is in the western parts **of this land**, ten miles from Padstow, [a place] which is called **Neot's** dwelling'.

In total, there are 43 noun phrases containing within them a genitive in the Life; in eight that genitive appears after the head noun of the noun phrase; in 37 (82.2 %) it appears before it, in what is the 'innovative' order.[34] In his training set,

32 One might wonder, for instance, how late the kind of long-distance case (and gender) agreement of anaphors on show in this sentence from the Life survived the Conquest: *se Godes þeowe Sanctus Neotus his lif adreah […] þeowwigende* ***þan*** [DAT. SG. MASC.], *þe hine to þeowe geceas, þæt wæs* ***Gode sylfen*** [DAT. SG. MASC.] (130/10–13). How late authors observed the need for such agreement cannot, as far as I am aware, be readily discovered from any corpus, since *YCOE* and *PPCME2* do not indicate any kind of syntactic relationship between *þan* and *Gode sylfen*.

33 The same set of variables is used by Ecay and Pintzuk (2016), where further discussion may be found.

34 These figures include two instances where the genitive noun phrase is wrapped around the head noun: *on* ***sanctes Ælfeges*** *dagen* ***þæs halgen biscopes*** 'in the days of St Elphege, the holy bishop' and *for* ***Cristenes folcas*** *gearnunge and* ***rihtwisra heafodmanna*** 'in return the merits of the people and righteous rulers'. I have counted the genitives in these as appearing both before and after the head noun. My counts for the Life of St Neot derive from searches conducted using *YCOE*'s proprietary software, CorpusSearch, and subsequent checking, which revealed several examples which might not necessarily be genitives, e.g. *he* [scil. Neot] *geneosode Romeburh […] 7 þære his*

Zimmermann (2014: Table 3) found the rate of the use of the innovative order to be 80.8 % in texts composed between 910 and 990, 84.9 % in those composed between 990 and 1100 and 96.3 % in texts composed between 1100 and 1200, suggesting that the Life's positioning of the genitive may be one reason the classifier assigned it a date prior to 1100. More work is clearly necessary, as is increasing the number of texts available for investigation, but it is worth emphasising that the results from Zimmermann's classifier accord well with the findings concerning other aspects of the Life's language advanced in this paper.

3.6 Evidence from Longer Phrasal Correspondences with Other Texts

The techniques for proximity searching used in Section 3.4 to identify modes of expression shared between different bodies of texts sometimes throw up closer parallels that suggest a direct relationship between one text and another, a type of correspondence that may add to scholarship's understanding of that text's sources, deriving from more traditional philological methods. Thus, searching for instances of the phrase *þurh Godes gast* in *DOEC* reveals it was a favourite turn of phrase for Ælfric, who is responsible for seventeen of the twenty other instances of it in the corpus. But closer examination of those hits also shows that the Life of St Neot's reference to King Alfred having *manega bec þurh Godes gast gedyhte* 'composed many books by means of God's spirit' (133/21–22) is particularly close to a phrase from Ælfric's Second Series Homily for the Assumption of St Mary, which refers to *ealle ða halgan bec ðe þurh godes muð oððe ðurh Godes gast gedihte wæron* 'all the holy books which were composed through God's mouth or by means of God's spirit' (Godden 1979: 259/129–130). The similarity here is not close or protracted enough to suggest that the author of the Life of St Neot had this homily in front of him as he worked, but it further confirms the Life's author's familiarity with the Ælfrician oeuvre already flagged by Godden when he, for instance, points out the similarity of the opening of the Life to two passages, one the opening lines of his Life of the Forty Soldiers from his *Lives of Saints*, the other the Preface to the First Series of *Catholic Homilies* (Godden 2010: 198–199):

synnen forgyfenysse underfeng 'Neot visited the City of Rome [...] and there received forgiveness for his sins' (130/3–4), where *synnen* is tagged N^G (for a noun in the genitive), but might better be seen as dative.

> We wylleð eow cyðen beo sumen dæle emb þyssen halgen, þe we todæig wurðigeð, þæt eower geleafe þe trumre seo for þan mancynn behofeð godcundre lare (129/4–5).
> 'Since mankind requires divine instruction, we wish to make known to you something concerning this holy one, who we honour today, so that your faith will be the stronger'.

> We wyllað eow gereccan þæra feowertigra cempena ðrowunge, þæt eower geleafa þe trumre sy þonne ge gehyrað hu þegenlice hi þrowodon for Criste (Skeat 1881–1900: 238/1–2).
> 'We wish to narrate to you the passion of the forty soldiers, so that your faith will be the stronger when you hear how loyally they suffered for Christ'.

> ic gedyrstlæhte […] þæt ic ðas gesetnysse undergann […] for ðam ðe menn behofiað godre lare swiðost on þisum timan þe is geendung þyssere worulde (Clemoes 1997: 174/57–58).
> 'I presumed to undertake this composition because men most need good teaching in this period, which is the end of this world'.

The lexical parallel between the Life of the Forty Soldiers and that of Neot and the fact that both passages come from the very beginning of their respective texts strongly suggests that the author of the latter had access to a text of the former.

Unless Ælfric's turn of phrase in all these passages was based on a series of now lost works by another author or authors, or was happened upon independently by the author of the Life of St Neot, then that Life must postdate Ælfric's composition of the First and Second Series of *Catholic Homilies* and *Lives of Saints*, expressions from all three collections being there paralleled. It is also possible it postdates Wulfstan's publication of the *Sermo Lupi ad Anglos* in 1014.[35] But while such source study can provide a *terminus post quem*, it cannot, as Godden (2010: 201) has tentatively suggested, furnish a *terminus ante quem*. This is shown by a parallel between the Life of St Neot and the opening of the putatively post-Conquest version of the *Trinubium Annae* appended to the end of Ralph d'Escures' Homily for the Assumption of the Virgin (139/3–4):

> We wylleð eow nu \bi/ sum dæl gereccen emben […]
> 'We wish to narrate to you something concerning […]'

Whether it was the translator of the *Trinubium Annae* or the person who appended it to Ralph's sermon who lifted this opening, and whether he took it directly from Ælfric's Life of the Forty Soldiers or from the Life of St Neot itself, it shows that post-Conquest authors or compilers were just as able to borrow from pre-Conquest texts as those writing before 1066. While it is therefore difficult to think of

35 Whitelock (1952: 18) noted parallels between the closing peroration of the Life and the *Sermo Lupi* as well as to Wulfstan's works more generally, which are supplemented by Godden (2010: 197–198). But none of the parallels is particularly close, so, as Godden (2010: 209 n. 74) observes, the debts may be to a lost sermon in Wulfstan's style, not necessarily the *Sermo Lupi*.

a circumstance in which source study could give robust evidence for the Life's *terminus ante quem*, it does corroborate that the Life cannot have been written before the final decade of the tenth century, when Ælfric published his first two series of *Catholic Homilies*.

3.7 Summary

The evidence amassed above cannot prove that the Life of St Neot was not written in the early twelfth century. But, by using innovative methodologies for searching *DOEC* texts to place the language of the Life in the context of almost all the texts that survive from before 1150, some of which are of known date, as we have done, does suggest it is very unlikely to have been written that late. Rather, traces of antecedent spellings, its lexis, its syntax and its style all suggest a date before *c.* 1075. These methods therefore help us adjudicate, as decisively as the surviving evidence allows, between the competing hypotheses of Godden and Younge. The Life is thus in practice best thought of as a pre-Conquest work, with all the consequences that entails for how we write the literary history of the eleventh and twelfth centuries and the story of Vespasian D.xiv's compilation.

4 Looking Forward

An unfortunate consequence of the contemporary division of the various intellectual concerns and methodologies of nineteenth- and early-twentieth-century philology between departments of literature and linguistics has been that techniques developed in linguistics for uncovering the distribution of variants in particular corpora of texts have not been transferred into literary studies and daughter disciplines like book history, in spite of their power to provide data with which to answer questions like when and how works were composed and how certain collections of texts came together in particular manuscripts. The present paper has attempted to demonstrate some ways in which we can make much more extensive use of existing resources like *DOEC* and other corpora.

It has also tried to point out where the existing corpora are limited and cannot at present substantiate the kind of conclusions one might like to be able to draw. Chief among desiderata would be work to link the metadata of existing corpora of Old English like *DOEC* and the *Manchester Eleventh-Century Spellings Database*, so that it is possible to interrogate both simultaneously and thus capture forms used by scribes in manuscripts not used in the editions used by *DOEC*. Another would be a corpus of early Middle English that is more extensive than the *Linguis-*

tic Atlas of Early Middle English, ideally one that is as exhaustive as *DOEC*. Plain text files with editorial interventions flagged would be perfectly adequate, and it should be possible to source a majority from existing online resources like *CMEPV* and the optical character recognition of published editions, using Laing (1993) as a preliminary guide to the texts that should be included. The development of a part-of-speech tagger and syntactic parser for Old and Middle English would also be a valuable innovation, facilitating a much wider sense of the currency of particular structures than is at present available from the limited and narrowly canonical set of texts available in *YCOE* and *PPCME2*. Better metadata that allowed the user to distinguish between usages that occur in different portions of composite texts would also be a valuable addition to both the *DOE* and the *MED*: the appearance of a word in the original Ælfrician part of *De initio creaturae* from Cotton Vespasian A.xxii is much less significant than its appearance in the new opening added in the twelfth century as evidence of its currency after the Conquest, for instance.

Most of the methods demonstrated in this paper have been developed with my own preoccupation with the extent of English composition after the Norman Conquest in mind, with the consequence that I have prioritised strategies that help determine whether a text was written before 1066 or after. But as our methods develop, they should become more fine-grained and better able to distinguish texts of the eleventh century from those of the tenth, and perhaps even those of the late tenth from those of the mid tenth. The precision of our datings will also improve as the methods succeed in dating more texts; the approach outlined in this paper is essentially iterative, so that as more features are established as otherwise unattested in post-Conquest texts, so our grounds for securely dating texts before that period grow more secure. We have the corpora; the question now is how much we can do with them.

Appendix: Lexemes Contributing to the Lexical Profiles in Tables 1, 2, and 3

1 The Life of St Neot

Type (i): no independent evidence lexeme in active use after 1066

Caveats

forhæfednysse 'abstinence' (129/17; 130/11–12) [caveats: *Peri Didaxeon*; *Life of St Giles*; *Instructions to Christians*]; *sacerdhad* 'priesthood' (129/26) [*Life of St Giles*]; *geneosode* 'visited' (130/2–3) [*Life of St Giles*; Writ of the Monk Edwin (Sawyer 1968: no. 1428)]; *glæingde* 'adorned' (130/15) [*Life of St Giles*; *Life of St Nicholas*]; *færinge* 'suddenly' (130/32; 131/9; 132/30) [*Peri Didaxeon*; CCCC 303 *Life of St Margaret*; note also the *MED* **feringes** adj. 'sudden' used in the early fifteenth century, but only in the phrase *feringes deth* 'sudden death']; *þearle* 'severely, vigorously' (132/2; 133/11) [recipe for eye salve added to Cotton Faustina A.x, fols. 102–151]; *handbreden* 'palms of hands' (132/5) [Glossary in Bodley 730]; and *gefean* 'joy' (132/7) [Assmann 14; Bazire and Cross 5; an inedited sermon for the fifth Sunday after Lent (Cameron B3.2.7); Dream Lunary added to Hatton 115 in the twelfth century (Chardonnens 9.2.4/4)].

No Caveats

manðwære 'gentle' (129/19); *onette* 'hastened' (130/34); *færeld* 'journey' (131/1, 4); *emhydiglice* 'carefully' (131/3); *geðwoh* 'washed' (131/10); *tostæncte* 'scattered' (131/34, 36); *angsumest* 'most distressful' (131/37); *forhtode* 'became afraid' (132/2); *mycelæte* 'greedy' (132/32); *siccetunge* 'sighing' (132/34); *herelafe* 'remnant of army' (133/15); and *heof* 'lamentation' (133/26, as part of what may be a borrowing from Blickling 10).

Type (ii): no independent evidence lexeme in active use after 1066 but all component lexemes survived into 'Middle' English

Caveats

forhæfde 'restrained' (129/18) [caveats: *Peri Didaxeon*; *Life of St Giles*; *Instructions for Christians*]; *alefd* 'given' (129/23) [Bazire and Cross 5; inedited sermon for the Fifth Sunday after Epiphany (B3.2.7); *Life of St James*]; *arfæste* 'benevolent' (131/11) [*Durham*]; *oferðeah* 'prospered beyond others' (133/19) [*Life of St Giles*] ; *æighwanen* 'on all sides' (133/26) [*Life of St Giles*; the occurrence in the Life is part of what may be a borrowing from Blickling 10]; *anymen* 'remove' (133/32) [*Life of St Margaret* from CCCC 303]; and *þingeres* 'intercessors' (134/1) [Bazire and Cross 5].

No Caveats

gelustfullunge 'pleasure' (129/18); *lufigendlic* 'amiable' (129/27); *wæterseaðes* 'springs' (130/9, 30); *gebedhuse* 'oratory' (131/11; see also *MED* **bēd(e** n. 2b. **~ hous**, though this is probably a reformation; *forewitegunge* 'foreknowing' (131/20, 28); *wyrtgemangum* 'herbs' (132/11–12); *wælreowen* 'bloodthirsty' (132/18); *forfyhrt* 'very frightened' (132/22); *madmfaten* 'treasure chests' (132/23); *forraðe* 'very quickly' (133/1); *tolysnesse* 'destruction' (133/30, as part of what may be a borrowing from Blickling 10); and *tælnysse* 'reproach' (133/30).

Type (iii): no independent evidence lexeme in active use after 1066, but related lexeme(s) survived into 'Middle' English

Caveats

trumre 'stronger' (129/6) [caveat: *Life of St Quentin*]; *andgitfull* 'endowed with understanding' (129/14) [*Peri Didaxeon*]; *uneðnyssen* 'discomforts' (133/4) [*Life of St Nicholas*].

No Caveats

andigen 'envy' (130/18); *gescyldnysse* 'protection' (130/21); *ridenda* 'riders' (130/33); *geunrotsod* 'made sorrowful' (131/12); *geteignde* 'hastened' (131/14); *bedigligen* 'conceal' (131/30); *stiðlic* 'powerful' (132/21); *hleowwinde* 'warming' (132/28); *earfoðe* 'hardship' (132/36).

2.1 Honorius 1

Type (i): no independent evidence lexeme in active use after 1066

Caveats

drohtnunga 'service' (141/19; 142/2) [caveats: *Life of St Giles*; *Life of St Neot*]; and *unrote* 'sad' (142/16, also found in Honorius 2) [*Adrian and Ritheus*; the homily for the Third Sunday after Epiphany preserved in CCCC 302 and Cotton Faustina A.ix (Cameron B3.2.5)].

No Caveats

ehteres 'persecutors' (141/29; discounting *MED* ?**eghten v.** as evidence the word should be Type (ii) since it is based only on a single doubtful example); *onhagige* 'be within their power' (142/14); and *forelcodan* 'delayed' (142/34).

Type (ii): no independent evidence lexeme in active use after 1066 but all component lexemes survived into 'Middle' English

Caveats

bescufð 'casts down' (141/13) [caveats: *Life of St Nicholas*; Assmann 14]

No Caveats

antimbre 'substance' (140/27 (x2), 28); *tyhtere* 'instigator' (140/29); *gewittscipe* 'knowledge' (140/31); *manslege* 'homicide' (140/34; 141/1, 2); *byrstige* 'afflicted by loss' (141/38; 142/18); *hindernysse* 'wickedness' (142/24); *teonlease* 'free from suffering' (143/5, 12) and *acwelmode* 'killed' (143/23).

Type (iii): no independent evidence lexeme in active use after 1066, but related lexeme(s) survived into 'Middle' English

Caveats

none.

No Caveats

none.

2.2 Honorius 2

Type (i): no independent evidence lexeme in active use after 1066

Caveats

unrote 'sad' (144/25, also found in Honorius 1) [caveats: *Adrian and Ritheus*; the homily for the Third Sunday after Epiphany preserved in CCCC 302 and Cotton Faustina A.ix (Cameron B3.2.5)].

No Caveats

unȝeleafsumnesse 'lack of faith' (145/2).

Type (ii): no independent evidence lexeme in active use after 1066 but all component lexemes survived into 'Middle' English

Caveats

none.

No Caveats

forsacunge 'rejection' (144/25; cf. *MED* **forsāking** ger., apparently a late-fourteenth-century reformation).

Type (iii): no independent evidence lexeme in active use after 1066, but related lexeme(s) survived into 'Middle' English

Caveats

none.

No Caveats

none.

Works Cited

Appleton, Helen. 2016. "The Old English *Durham* and the Cult of St Cuthbert". *Journal of English and Germanic Philology* 115: 346–369.

Anthony, Laurence. 2019. *AntConc*. Version 3.5.8. Tokyo: Waseda University. <https://www.laurenceanthony.net/software> [accessed 27 May 2020].

Bailey, Richard N. and Eric Cambridge. 2016. "Dating the Old English Poem *Durham*". *Medium Ævum* 85: 1–14.

Benskin, Michael and Margaret Laing. 1981. "Translations and Mischsprachen in Middle English Manuscripts". In: Michael Benskin and M. L. Samuels (eds.). *So Meny People Longages and Tonges: Philological Essays in Scots and Mediaeval English Presented to Angus McIntosh*. Edinburgh: Middle English Dialect Project. 55–106.

B-T Enlarged Addenda = Campbell, Alistair. 1972. *Enlarged Addenda and Corrigenda to the Supplement* (1921) of Joseph Bosworth and T. Northcote Toller. 1898. *An Anglo-Saxon Dictionary*. Oxford: Clarendon.

Cameron, Angus. 1973. "A List of Old English Texts". In: Roberta Frank and Angus Cameron (eds.). *A Plan for the Dictionary of Old English*. Toronto: University of Toronto Press. 25–306.

Campbell, Alistair. 1959. *Old English Grammar*. Oxford: Clarendon.

Clark, Cecily. 1970. *The Peterborough Chronicle: 1070–1154*. 2nd ed. Oxford: Clarendon.

Clemoes, Peter (ed.). 1997. *Ælfric's Catholic Homilies: The First Series*. EETS SS 17. Oxford: Oxford University Press.

CMEPV = *Corpus of Middle English Prose and Verse*. 2000. Michigan: University of Michigan. <https://quod.lib.umich.edu/c/cme/> [accessed 27 May 2020].

Cubbin, G. P. (ed.). 1996. *The Anglo-Saxon Chronicle: A Collaborative Edition*. Volume 6: *MS D*. Cambridge: Brewer.

D'Ardenne, S. R. T. O. and E. J. Dobson (eds.). 1981. *Seinte Katerine*. EETS SS 7. London: Oxford University Press.

DOE = *Dictionary of Old English, A to I*. 2018. Ed. Angus Cameron, Ashley Crandell Amos, Antonette diPaolo Healey et al. Toronto: Dictionary of Old English Project. <http://www.doe.utoronto.ca>.

DOEC = *Dictionary of Old English Web Corpus*. 2007. Ed. Antonette diPaolo Healey et al. Toronto: University of Toronto. <https://www.doe.utoronto.ca/pages/pub/web-corpus.html>.

Dumville, David and Michael Lapidge (eds.). 1985. *The Anglo-Saxon Chronicle: A Collaborative Edition*. Volume 17: *The Annals of St Neots with Vita Prima Sancti Neoti*. Cambridge: Brewer.

Ecay, Aaron and Susan Pintzuk. 2016. "The Syntax of Old English Poetry and the Dating of *Beowulf*". In: Leonard Neidorf, Rafael J. Pascual and Tom Shippey (eds.). *Old English Philology: Studies in Honour of R. D. Fulk.* Cambridge: Boydell & Brewer. 219–258.

Faulkner, Mark. 2012. "Archaism, Belatedness and Modernisation: 'Old' English in the Twelfth Century". *Review of English Studies* 63: 179–203.

Faulkner, Mark. 2017a. "The Eadwine Psalter and Twelfth-Century English Vernacular Literary Culture". In: Tamara Atkin and Francis Leneghan (eds.). *The Psalms and Medieval English Literature*. Cambridge: Brewer. 72–107.

Faulkner, Mark. 2017b. "Linguistic Evidence for the Compilation of Twelfth-Century Manuscripts Containing Old English: The Case of Cotton Vespasian D. xiv". *Neuphilologische Mitteilungen* 118: 279–316.

Faulkner, Mark. Forthcoming a. "Corpus Philology, Big Dating and Bottom-Up Periodization". In: Stephen Pink and Anthony Lappin (eds.). *Dark Archives: The Medieval Unread and Unreadable.* Oxford: Society for the Study of Medieval Languages and Literature.

Faulkner, Mark. Forthcoming b. *A New Literary History of the Long Twelfth Century: Language and Literature between Old and Middle English*. Cambridge: Cambridge University Press

Feulner, Anna Helene. 2000. *Die griechischen Lehnwörter im Altenglischen*. Frankfurt a.M.: Lang.

Godden, M. R. (ed.). 1979. *Ælfric's Catholic Homilies: The Second Series*. EETS SS 5. London: Oxford University Press.

Godden, M. R. 2010. "The Old English Life of St Neot and the Legends of King Alfred". *Anglo-Saxon England* 39: 193–225.

Hanna, Ralph. 2009. "Lambeth Palace Library, MS. 487: Some Problems of Early Thirteenth-Century Textual Transmission". In: Cate Gunn and Catherine Innes-Parker (eds.). *Texts and Traditions of Medieval Pastoral Care: Essays in Honour of Bella Millett*. Woodbridge: York Medieval Press. 78–88.

Hardy, Thomas Duffus. 1862–1871. *Descriptive Catalogue of Materials Relating to the History of Great Britain and Ireland*. RS 26. 3 vols. London: Longman, Green, Longman and Roberts.

Hart, Cyril. 1990. *The Danelaw*. London: Hambledon.

Healey, Antonette diPaolo and Richard L. Venezky. 1980. *A Microfiche Concordance to Old English: The List of Texts and Index of Editions*. Toronto: Dictionary of Old English Project.

Helsinki Corpus = Rissanen, Matti et al. (eds.). 1991. *The Helsinki Corpus of English Texts*. <http://www.helsinki.fi/varieng/CoRD/corpora/HelsinkiCorpus/> [accessed 27 May 2020].

Hofstetter, Walter. 1987. *Winchester und der spätaltenglische Sprachgebrauch*. Munich: Fink.

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

HTOE = The Historical Thesaurus of English. Version 4.21. 2020. Glasgow: University of Glasgow. <https://ht.ac.uk> [accessed 27 May 2020].

Ker, N. R. 1957. *Catalogue of Manuscripts Containing Anglo-Saxon*. Oxford: Clarendon.

Kisbye, Torben. 1971–1972. *An Historical Outline of English Syntax*. 2 vols. Aarhus: Akademisk Boghandel.

LAEME = A Linguistic Atlas of Early Middle English 1150–1325. 2013–. Ed. Margaret Laing. Version 3.2. <http://www.lel.ed.ac.uk/ihd/laeme2/laeme2.html> [accessed 27 May 2020].

Laing, Margaret. 1993. *Catalogue of Sources for a Linguistic Atlas of Early Medieval English*. Cambridge: Brewer.

Laing, Margaret. 2004. "Multidimensionality: Time, Space and Stratigraphy in Historical Dialectology". In: Marina Dossena and Roger Lass (eds.). *Methods and Data in Historical Dialectology*. Bern: Lang. 49–96.

Laing, Margaret and Roger Lass. 2009. "Shape-Shifting, Sound-Change and the Genesis of Prodigal Writing Systems". *English Language & Linguistics* 13: 1–31.
Laing, Margaret and Roger Lass. 2013. "Introduction". In: Margaret Laing (ed.). *A Linguistic Atlas of Early Middle English, 1150–1325*. Version 3.2. Edinburgh: The University of Edinburgh. <http://www.lel.ed.ac.uk/ihd/laeme2/laeme2.html> [accessed 27 May 2020].
Lapidge, Michael and Simon Keynes (ed. and trans.). 1983. *Alfred the Great: Asser's Life of King Alfred and Other Contemporary Sources*. Harmondsworth: Penguin.
Lass, Roger. 2015. "Interpreting Alphabetic Orthographies: Early Middle English Spelling". In: Patrick Honeybone and Joseph Salmons (eds.). *The Oxford Handbook of Historical Phonology*. Oxford: Oxford University Press. 100–120.
Lass, Roger and Margaret Laing. 2012. "'Ea' in Early Middle English: From Diphthong to Digraph". In: David Denison, Ricardo Bermúdez-Otero, Chris McCully and Emma Moore (eds.). *Analysing Older English*. Cambridge: Cambridge University Press. 75–117.
Lewis, Robert E., Mary Jane Williams and Marilyn Sue Miller. 2007. *Middle English Dictionary: Plan and Bibliography*. 2nd ed. Ann Arbor, MI: University of Michigan Press.
Logeman, Henri (ed.). 1888. *The Rule of St Benet: Latin and Anglo-Saxon Interlinear Version*. EETS OS 90. London: Trübner.
MED = *Middle English Dictionary*. 1952–2001. Ed. Hans Kurath, Sherman M. Kuhn, John Reidy and Robert E. Lewis. Ann Arbor, MI: University of Michigan Press. <http://quod.lib.umich.edu/m/med> [accessed 27 May 2020].
Millett, Bella. 2010. "Change and Continuity: The English Sermon before 1250". In: Elaine Treharne and Greg Walker (eds.). *The Oxford Handbook of Medieval Literature in English*. Oxford: Clarendon. 221–239.
Mitchell, Bruce. 1985. *Old English Syntax*. 2 vols. Oxford: Clarendon.
Morris, Richard (ed.). 1867–1868. *Old English Homilies of the Twelfth Century: First Series*. EETS OS 29, 34. London: Trübner.
Mustanoja, Tauno F. 1960. *A Middle English Syntax*. Part 1: *Parts of Speech*. Helsinki: Société Néophilologique.
O'Brien O'Keeffe, Katherine (ed.). 2001. *The Anglo-Saxon Chronicle: A Collaborative Edition*. Volume 5: *MS C*. Cambridge: Brewer.
OED = *The Oxford English Dictionary*. 2000–. 3rd ed. online. Oxford: Oxford University Press. <https://www.oed.com/> [accessed 27 May 2020].
Okasha, Elisabeth. 1971. *Hand-List of Anglo-Saxon Non-Runic Inscriptions*. Cambridge: Cambridge University Press.
Plummer, Charles. 1902. *The Life and Times of Alfred the Great*. Oxford: Clarendon.
PPCME2 = *The Penn-Helsinki Parsed Corpus of Middle English*. 2000–. Ed. Anthony Kroch, Ann Taylor and Beatrice Santorini. 2nd ed., release 4. Philadelphia, PA: University of Pennsylvania.
Richards, Mary. 1981. "The Medieval Hagiography of St Neot". *Analecta Bollandiana* 99: 259–278.
Roberts, Jane. 2000. "The English Saints Remembered in Old English Anonymous Homilies". In: Paul E. Szarmach (ed.). *Old English Prose: Basic Readings*. New York: Garland. 433–461.
Sawyer, P. H. 1968. *Anglo-Saxon Charters: An Annotated List and Bibliography*. London: Offices of the Royal Historical Society.
Scragg, Donald G. 1979. "The Corpus of Vernacular Homilies and Prose Saints' Lives before Ælfric". *Anglo-Saxon England* 8: 223–277.
Scragg, Donald G. (ed.). 1992. *The Vercelli Homilies*. EETS OS 300. Oxford: Oxford University Press.

Scragg, Donald G., Alexander Rumble and Kathryn Powell (eds.). 2004. *The Manchester Eleventh Century Spellings Database*. Manchester: Manchester Centre for Anglo-Saxon Studies.
Sisam, Celia. 1951. "The Scribal Tradition of the Lambeth Homilies". *Review of English Studies* 2: 105–113.
Skeat, Walter W. (ed.). 1881–1900. *Ælfric's Lives of Saints*. EETS OS 76, 82, 94, 114. London: Oxford University Press.
Smyth, Alfred P. 1995. *King Alfred the Great*. Oxford: Oxford University Press.
Sollerio, A. Johanne Baptista et al. 1868. *Acta Sanctorum quotquot toto orbe coluntur*. Julii vol. 7. Paris: Palmé.
Stevenson, William Henry. 1904. *Asser's Life of King Alfred*. Oxford: Clarendon.
Treharne, Elaine. 1995. "A Unique Old English Formula for Excommunication from Cambridge Corpus Christ College 303". *Anglo-Saxon England* 24: 185–211.
Visser, Frederik Theodor. 1963–1973. *An Historical Syntax of the English Language*. 3 vols. Leiden: Brill.
Warner, Rubie D-N. (ed.). 1917. *Early English Homilies from the Twelfth-Century MS. Vesp. D.xiv*. EETS OS 152. Oxford: Oxford University Press.
Whitelock, Dorothy. 1952. *Sermo Lupi ad Anglos*. 2nd ed. London: Methuen.
Wright, Neil. 1995. "Alfred Burns the Cakes: the Vita Prima Sancti Neoti, Telesinus and Juvenal". In: Neil Wright (ed.). *History and Literature in Late Antiquity and the Early Medieval West*. Aldershot: Ashgate. XV.
Wülker, Richard. 1885. *Grundriss zur Geschichte der angelsächsischen Literatur*. Leipzig: Veit.
YCOE = The York-Toronto-Helsinki Parsed Corpus of Old English Prose. 2003. Ed. Ann Taylor, Anthony Warner, Susan Pintzuk and Frank Beths. Department of Language and Linguistic Science. York: University of York. <http://www-users.york.ac.uk/~lang22/YCOE/YcoeHome.htm>.
Yerkes, David. 1984. *The Old English Life of Machutus*. Toronto: University of Toronto Press.
Younge, George. 2012. "'Those were Good Days': Representations of the Anglo-Saxon Past in the Old English Homily on Saint Neot". *Review of English Studies* 63: 349–369.
Zimmermann, Richard. 2014. "Dating Hitherto Undated Old English Texts Based on Text-Internal Critera". Unpublished MS. <http://www.old-engli.sh/myres%20documents/Zimmermann_R.(2014)Dating_OE_Texts.pdf> [accessed 27 May 2020].
Zimmermann, Richard (ed.). n. d. *Parsed Corpus of Middle English Poetry*. <https://www.pcmep.net/index.php> [accessed 27 May 2020].
Zupitza, Julius. 1890. "Ein weiteres Bruchstück der Regularis Concordia in altenglischer Sprache". *Archiv für das Studium der neueren Sprachen und Literatur* 84: 1–24.