Dating hitherto undated Old English texts based on text-internal criteria*

Richard Zimmermann
University of Geneva

Abstract

Nineteen undated early English (c. 800-1200 A.D.) texts are grouped into their respective period of origin using a Naïve Bayes classifier trained on fourteen syntactic features from the general early English text corpus. It is assumed that syntactic features are less prone to adaptations during the copying process of manuscripts than phonology or lexis so that syntax can be effectively used for exemplar dating. The classification procedure leads to results that are mostly unambiguous, frequently converge with traditional arguments for specific dates of composition, and can therefore be regarded as plausible. The proposed periodisation of early English texts can be employed in future studies of diachronic English syntax.

1 Introduction

The English language lends itself eminently to diachronic investigations on account of its massive text corpus, which spans more than one thousand years of linguistic history. In order to take advantage of this rich textual material, however, it is necessary to first discern its exact chronological order. Unfortunately, this is not always an easy task - especially for early English texts.

I define as 'early English' a period that comprises Old and the early stages of Middle English, c. 800-1200 A.D. A fair number of texts survives from this period whose date of composition has not been definitively established. In this paper, I attempt to determine the points of origin for nineteen such writings using a probabilistic text classification technique called *Naïve Bayes*. I will proceed as follows: First, I will group early English texts whose date of composition is relatively well known into five distinct periods. These texts will provide the training material for the probabilistic Naïve Bayes classifier model. Next, I present the nineteen undated early English texts that are the actual concern of this study. These texts will form the application material for the classifier; their most probable periods of origin are to be determined. In the subsequent section, I will describe fourteen syntactic features on the basis of which the probabilistic classifier will make its decisions. The workings of a Naïve Bayes classifier are then explained and the accuracy of the specific model used is evaluated. Finally, the probabilistic classifier will be utilised to ascertain the dates of composition for the nineteen undated texts. The results are discussed in some detail. The conclusion follows.

*I would like to thank the audience of SWELL 2013, Zurich for helpful comments.

2 Early English texts

In this section I will give an overview over the textual material employed in this paper. First, I list texts that can be dated with a relatively high degree of certainty, henceforth collectively referred to as the *general early English text corpus*, and categorize them into five periods. Secondly, I will present nineteen texts that are chronologically indeterminate. I will call the set of these texts the *undated early English texts*.

2.1 The general early English text corpus

I divided the general early English text corpus into five periods, 850-900, 900-990, 990-1100, 1100-1200 and 1200-1230. The last period should be regarded as a mere control class - none of the undated early English texts would actually be expected to have been composed quite so late. In what follows, I specify the texts included in each period.

850-900. The earliest seizable number of English texts that have been handed down to us come from the 9th century. They are largely a product of King Alfred's (849-899) educational reform program (Waite 2000), which spawned the translation of several Latin historical, religious and philosophical works into the Old English vernacular. The most important Alfredian texts are *Bede's Ecclesiastical History*, *Gregory's Pastoral Care*, *Boethius' Consolation of Philosophy*, *Orosius' History* and *Augustine's Soliloquies*. Alfred also initiated the recording of annals in Old English, collectively known as *The Anglo-Saxon Chronicle*, which were produced continually for over two and a half centuries (Swanton 1996). Further, *The Old English Martyrology*, a catalogue of martyrs' passions arranged in a calendar in the order of their respective feasts, survives in five manuscripts, two of which come from the late ninth century (Cross 1986, 275). In general, the text accords well with the ninth-century prose context so that it was probably composed between 850-900 (Bately 1988, 118, 135). Finally, another two Old English pieces of prose, travelogues uniquely attested in the famous *Beowulf* manuscript, are likely to originate from the 9th century, *The Letter of Alexander to Aristotle* (Sisam 1953, 85, 88) and *Marvels of the East* (Vleeskruyer 1953, 50, fn.1).

900-990. Several Old English texts were composed in the 10th century. Firstly, there are two large collections of sermons, *The Blickling Homilies*, whose manuscript date is fixed at 971 by an explicit reference in homily IX (Morris 1874-1880, viii), and *The Vercelli Homilies*, most items of which are likely to date from the 10th century (Scragg 1992, xxxviii). Secondly, Bishop Æthelwold's (909-984) Benedictine reform movement led to the translation of several monastic texts. The *Rule of St. Benedict* dates to 964-975 with a focus on the years around 970 (Gretsch 2004, 233). The *Enlarged Rule of Chrodegang of Metz* was translated either at the end of the 10th century, based on the fact that the translation includes names of clergymen associated with the Old Minster, Winchester, during the episcopate of Bishop Ælfheah (984-1006) (Förster 1933) or even earlier, in the 940s or 950s, as suggested by its general prose style (Drout 2004, 342-7). Thirdly, the *West-Saxon Gospels*, a complete Old English translations of the four, canonical bible gospels, were produced at the close of the 10th century, possibly at around 990 (Skeat 1874, xii).

990-1100. The turn of the first millennium witnessed the creation of works by two of the most prolific Old English writers. The first, Ælfric of Eynsham (c. 955 - c. 1010), composed all of his works between about 990 and the time of his death (Hill 2009). Ælfric wrote three series of sermons, *The Catholic Homilies I & II* and *Lives of Saints* as well as supplemental homilies, several letters, an Old English rendition of the first books of the Old Testament and a translation of Bede's computus *De temporibus anni*. The second is Wulfstan of York

(died 1023), an influential political and clerical figure in the early 11th century (Whitelock 1942). Wulfstan is the author of a substantial number of homiletic writings and a philosophical work known as *The Institutes of Polity*. In addition, the earliest surviving romance written in English, *Apollonius of Tyre*, was probably translated at the beginning of the 11th century (Goolden 1958, xxvii-xxxii), as were the first sixteen chapters of *Alcuin's De Virtutibus et Vitiis* (Wallach 1955, 192). A science handbook, called *Byrhtferth's manual*, is known to date from 1011 (Henel 1943, 290, fn.4). Furthermore, a substantial number of *Anglo-Saxon Charters* were issued during this and the preceding two time periods (Sawyer 1968).

1100-1200. Unfortunately, the English language is attested only sporadically following the period immediately after Ælfric and Wulfstan, and hardly any English texts survive at all from after the Norman Conquest in 1066. Among the small number of prose texts from the 12th century, there are two short pieces dated to about 1108-1122, *Sermo in festis Sancti Marie* and a translation of a part of *Elucidarius* (Förster 1920, 58-60, 63). Six pieces of the *Lambeth Homilies*, which otherwise contain revisions of earlier Old English material, were probably written in the 12th century (Sisam 1951). Similarly, the *Trinity Homilies* plausibly contain 12th century compositions, but it is difficult to separate those from transliterations of potentially earlier texts. Five homilies are shared between these two homiletic collections (Laing 2000). The religious prose dialogue *Vices and Virtues* is uniquely attested in an early 13th century manuscript, but its language reflects a stage of English somewhat earlier than that (Hall 1920, 443-4). Finally, for this as well as the preceding three periods, there are sections of the *Anglo-Saxon Chronicle* whose dates of composition fall into the respective time spans outlined.

1200-1230. Lastly, texts from the early 13th century (or possibly just slightly earlier) include the *Ancrene Riwle* (Dobson 1972) and the five homilies of the *Katherine Group* (Milett 1996).

Table 1 below summarizes the five reference periods thus established.

2.2 Undated early English texts

The following nineteen early English texts have dates of composition that are either ascertained imprecisely or entirely unknown.

1. *Saint Chad*. The saint's life *The Life of St. Chad* occurs uniquely in a twelfth century manuscript. Vleeskruyer concludes from a number of rare, archaic words that the original was composed early, preferably between 850-875 (1953, 70). On the other hand, Napier believes that the text could have been composed as late as the first half of the 10th century (1888, 139). Thus, while it is sufficiently clear that *St. Chad* must be of an early provenance, the text could still fall into periods 1 or 2.

2. *Saint Christopher*. Bately (1988, 99, fn.47) summarizes the dating discussion of this saint's life and finds that its suggested dates range from the 9th to the 10th century, i.e. periods 1 or 2.

3. *Mary of Egypt*. The *Life of St. Mary of Egypt* appears in an early 11th century manuscript, which contains mostly works by Ælfric, but differs from them markedly. The dates suggested for this text range from the 9th century, as indicated by the mediocrity of the Latin translation (Magennis 1986, 333-4; Magennis 2002, 12-3) to the late 10th century (Scheil 2000, 138). The text could thus have been composed in periods 1-3.

4. *Saint Euphrosyne*. The *Life of Saint Euphrosyne* appears in the same manuscript as the last item. It may have been composed at a time when the popularity of this saint's cult increased during the 10th century Benedictine reform movement (Donovan 1999, 80). However, in theory, periods 1-3 are all conceivable points of origin for this text.

Period	Texts	# Words	# Observations
850-900	Bede's Ecclesiastical History	c. 310,000	10
	Gregory's Pastoral Care		
	Boethius' Consolation of Philosophy		
	Orosius' History		
	Augustine's Soliloquies		
	The Old English Martyrology		
	The Letter of Alexander to Aristotle		
	Marvels of the East		
	Sections of the Anglo-Saxon Chronicle		
	9^{th} century Charters		
900-990	Blickling Homilies	c. 200,000	7
	Vercelli Homilies		
	Rule of St. Benedict		
	The Rule of Chrodegang of Metz		
	West-Saxon Gospels		
	Sections of the Anglo-Saxon Chronicle		
	10^{th} century Charters		
990-1100	Ælfric's Catholic Homilies I	c. 530,000	14
	Ælfric's Catholic Homilies II		
	Ælfric's Lives of Saints		
	Ælfric's Supplemental Homilies		
	Ælfric's Letters		
	Ælfric's Old Testament		
	Ælfric's De Temporibus Anni		
	Wulfstan's Homilies		
	Wulfstan's Institutes of Polity		
	Apollonius of Tyre		
	Alcuin's De Virtutibus et Vitiis		
	Byrhtferth's Manual		
	Sections of the Anglo-Saxon Chronicle		
	11^{th} century Charters		
1100-1200	Sermo in festis Sancti Marie	c. 80,000	6
	Elucidarius		
	Lambeth Homilies		
	Trinity Homilies		
	Vices and Virtues		
	Sections of the Anglo-Saxon Chronicle		
1200-1230	Ancrene Riwle	c. 100,000	6
	Hali Meidhad		
	Saint Katherin		
	Saint Juliana		
	Saint Margaret		
	Sawles Warde		

Table 1: Periodisation of the general early English text corpus

5. *Saint Eustace*. The saint's lives *Saint Eustace and his Companions* is yet another text that appears in the same manuscript as the preceding two items but is not attributed to Ælfric (Skeat 1881-1900, 452). Like the last two texts, it could thus have been written in periods 1-3.

6. *Seven Sleepers*. The legend of the *Seven Sleepers* is the last item from the same manuscript as the preceding three texts that is not authored by Ælfric. References to this legend usually appear rather late in the Old English period (Magennis 1991, 44). Still, periods 1-3 are theoretically possible for this text.

7. *Holy Rood-Tree*. The *History of the Holy Rood-Tree* is found in a late 12th century manuscript but its archaic language indicates that it was probably copied from an earlier Old English version, perhaps from the first half of the 11th century (Napier 1894, lviii). At any rate, a fragment of the text survives in a mid 11th century manuscript (Ker 1957, No. 73) so that an early Middle English origin can be ruled out. Potential points of origin for this text are therefore periods 1-3.

8. *James the Greater*. I was unable to find any reliable information on the date of composition for this text. Neither the standard edition (Warner 1917) nor Förster's (1920) survey of the manuscript items provide relevant information. The text is found in a mid 12th century manuscript. Hence, it could have originated in any of periods 1-4.

9. *St. Neot*. The *Life of St. Neot* survives in the same 12th century manuscript as the previous item. The saint's life is investigated in a recent article by Godden (2010), who diverges from the opinion held by several scholars assigning the text to the 12th century and asserts instead that it was composed in the period immediately following Ælfric and Wulfstan in the first half of the 11th century (ibid.: 193, 196-7). He believes that the contents of the text, featuring a less than virtuous King Alfred and Vikings as the instrument of divine punishment, accord well with the time when England was ruled by the Danish kings Cnut the Great and his sons (1016-1040) (ibid: 219-20). It would appear, then, that the controversy over the correct date for this text concerns mainly periods 3 or 4.

10. *Saint Margaret* (Corpus). Saint Margaret was a popular saint in medieval England. The 9th century *Old English Martyrology* contains Margaret's passion, but in this text, she is referred to as 'Marina.' One *Life of St. Margaret* was destroyed during the Cottonian fire in 1731. Another two distinct accounts of her passion are now extant in Old English (Clayton and Magennis 1994, 41). The first of these versions is found in an early 12th century manuscript from Cambridge, Corpus Christ College. It has been argued to have been composed not very long before the manuscript date on the basis of its content and vocabulary (ibid: 70, 104, 106). Once again, the potential dates of composition for this text would thus seem to be periods 3 or 4.

11. *Saint Margaret* (Tiberius). The second version of *The Life of Saint Margaret*, found in a Cotton Tiberius manuscript from the middle of the 11th century (Clayton and Magennis 1994, 70), has not been dated definitively. I will therefore assume that the text might have been composed in any of periods 1-3.

12. *Herbarium*. The *Herbarium* is a medical-botanical translation of three originally separate Latin treatises and survives in four 11th and 12th century manuscripts. Arsdall (2002, 76, 103-4) maintains that there is no proof for the existence of the Old English text from long before the copies of the four surviving manuscripts were made and consequently prefers a relatively late date of composition. Meaney (2000, 228) speculates that the text could have been translated during the 10th century from an imported Italian manuscript while in de Vriend's (1984, xlii-xliii) opinion, it is possible that the *Herbarium* was originally translated during the age of Northumbrian supremacy in the 8th century and subsequently copied repeatedly and

adapted to the West-Saxon standard. All in all, then, the date of the *Herbarium* is not clearly established and I will assume that the text could have been composed in periods 1-3.

13. *Quadrupedibus.* The *Medicina de Quadrupedibus* is a separate, three-part appendix to the *Herbarium* and immediately follows it in all four surviving manuscripts (de Vriend 1984, lxii). Therefore, what was said about the *Herbarium* applies to this text as well; potential dates of composition are periods 1-3.

14. *Leechbook.* The name *Bald's Leechbook* is given to three collections of medical remedies found in a mid 10th manuscript (Cockayne 1865, xxiv). Nokes (2004, 56) summarizes a consensus view among scholars that collections I and II were composed early, dating to the 9th century, perhaps as part of King Alfred's educational reform programme, while collection III is a later addition. However, although parts of collection II in particular clearly seem to be Alfredian (Nokes 2004, 68), a significant number of remedies could have been added in the 10th century. Thus, the bulk of the material could have originated in periods 1 or 2.

15. *Lacnunga.* The fourth major Anglo-Saxon medical text, *Lacnunga*, survives in a manuscript from the turn of the first millennium with later additions in the early 11th century (Ker 1957, 306). Since the compiler of the *Lacnunga* is known to have used the *Herbarium*, the former must be younger than the latter (Arsdall 2002, 103). Possible dates for this text are therefore periods 2 or 3.

16.*Nicodemus.* The Old English *Gospel of Nicodemus* is an apocryphal gospel of the life and works of Jesus of Nazareth. It survives in three manuscripts from the late 11th and mid 12th centuries, neither of which is the original Old English version (Cross 1996*b*, 7-8). Marx (1997, 215) believes that the translation could have been made between as early as the mid 10th to the mid 11th century. However, in principle any date of composition before the end of the Old English period is conceivable, i.e. periods 1-3.

17. *Vindicta.* The Old English *Vindicta Salvatoris* is another apocryphal text. It appears immediately after *Nicodemus* in two of its manuscripts and in addition, there is a third copy of the text in a manuscript from the second half of the 11th century that does not include *Nicodemus.* As for the previous item, none of the manuscripts represents the original Old English version (**?**, 76-7). Potential dates of composition are thus periods 1-3.

18. *De Ascensione.* There are various additions in the margins of an early 11th century manuscript primarily known for the Old English *Bede.* One of them is a homily without a standardized name, but referred to in Wanley's catalogue (cited in Hulme 1904) as *Homilia de Ascensione D. N. Jesu Christi.* Its date of composition is unknown. Hulme (1904, 590) speculates that it originated in the late 11th century. In principle, however, periods 1-3 must remain an option.

19. *Distichs of Cato.* The translation of the *Distichs of Cato*, a collection of proverbs often used for Latin teaching, survives in three manuscripts from the late 11th and 12th centuries. The text has received a fair amount of scholarly attention. For instance, Treharne (2003, 468) does not see any certain evidence for the assumption that the text was translated before the 11th century. Hollis & Wright (1992, 18) believe that the composition of the *Distichs of Cato* was begun between the mid 10th to the mid 11th century but that a substantial number of proverbs were added only at the end of the 11th century. It is therefore possible that the origin of this text lies in periods 2 to 3.

Table 2 below summarizes the main points of this synopsis. The table lists the undated Old English texts in the order as discussed above, starting with saint's lives, followed by medical texts, apocrypha and, lastly, others. The next column gives the Dictionary of Old English (DOE) short titles. Indicated after this are information from Ker's (1957) catalogue -

manuscript number, item number and manuscript date for the base manuscript used for the respective DOE / YCOE (2003) editions. The last two columns state for each text the number of words in the YCOE text file - altogether 132,922 words - and possible periods for the date of composition.

Text	DOE	Ker	Ms. date	# Words	Periods
1. Saint Chad	LS 3	333.1	s. XII[1]	2,659	1-2
2. Saint Christopher	LS 4	216.1 [1]	s. X/XI	1,426	1-2
3. Mary of Egypt	LS 23	162.31	s. XI in.	8,181	1-3
4. Saint Euphrosyne	LS 7	162.44	s. XI in.	3,658	1-3
5. Saint Eustace	LS 8	162.41	s. XI in.	5,271	1-3
6. Seven Sleepers	LS 34	162.30	s. XI in	9,143	1-3
7. Holy Rood-Tree	LS 5	310.12	s. XII[2]	6,920	1-3
8. James the Greater	LS 11	209.11	s. XII med.	1,659	1-4
9. Saint Neot	LS 28	209.43	s. XII med.	2,003	3-4
10. Saint Margaret	LS 14	57.23	s. XII[1]	4,196	3-4
11. Saint Margaret	LS 16	186.15	s. XI med.	3,661	1-3
12. Herbarium	Lch I	219.1	s. XI[1]	22,213	1-3
13. Quadrupedibus	Med 1.1	219.1	s. XI[1]	4,276	1-3
14. Leechbook	Lch II	264.1-3	s. X med.	34,727	1-2
15. Lacnunga	Med 3	231.2	s. X/XI, XI[1]	7,099	2-3
16. Nicodemus	Nic(A)	20.2	s. XI[3/4]	8,197	1-3
17. Vindicta	VSal(1)	20.3	s. XI[3/4]	3,655	1-3
18. De Ascensione	Nic(D) [2]	32.13	s. XI[1]	1,798	1-3
19. Distichs of Cato	Prov 1	89.2	s. XI/XII	2,180	2-3

Table 2: Undated early English texts

3 Criteria used for the dating process

I will now turn to the description of the explanatory variables measured for each text as a proxy for the text's relative age. All text-internal criteria express the frequency of the innovative variant as a percentage of all relevant examples. In other words, the variables are all continuous such that a value close to 0% indicates a more conservative and a value close to 100% a more innovative manifestation of the feature. I considered fourteen features for this study. All data were obtained from the electronic, parsed corpora YCOE (Taylor et al. 2003) and PPCME2 (Kroch and Taylor 2000*a*).

[1]The online documentation of the YCOE (Taylor et al. 2003) cites the wrong manuscript as the foundation for the edition, the badly damaged British Museum, Cotton Otho B.X, while the actual base text, indicated in table 2, is the *Beowulf* manuscript, British Museum, Cotton Vitellius A.XV, ff. 94-98.

[2]Since Hulme (1904) printed the homily as part of his discussion of the Old English *Nicodemus*, the DOE short title initialises "Nic." The YCOE (Taylor et al. 2003) lists the text under a similar abbreviation. However, it would seem more appropriate to keep the piece under a heading for homilies of unspecified occasion ("HomU"). Certainly, the YCOE genre of this text as "apocrypha" is unjustified. That the text is a homily and not an apocryphon is also evidenced by the heading of a second version of the same text ("Nic(E)", Ker 57.17), **Sermo** *in Ressurectione domini* (emphasis mine).

The criteria used for the dating of early English texts are all syntactic in nature. This design has a couple of important advantages. Firstly, changes can be measured with large numbers. In fact, several features are attested by thousands of relevant examples (see below). This should ensure high reliability. Secondly, syntactic features are less prone to distortions during the scribal transition process than, for example, phonological or morphological features. A scribe may easily adapt the phonological structure of a word to his own or a more innovative accent, level inflectional endings, substitute individual words etc. But he is less likely to spend a lot of time reflecting on possible changes in complex word order. Hence, syntax may retain more archaic characteristics of a text and can function as a probe into the date of its original exemplar while other features are more sensitive to emendation and can be used more efficiently to ascertain a manuscript date. Finally, it seems plausible that syntactic features are less variable across dialects, genres and idiolects than, for instance, the lexicon employed. The measurements will therefore be relatively robust to the noise introduced by such external factors.

1/2. IP1/IP2. The first set of features concerns the well-studied change in IP-headedness. During the course of the Old and Middle English periods, inflected verbs gradually come to precede their complements resulting in the head-initial word order familiar from Modern English. On account of the availability of various postposition processes (Kemenade 1987), the best way to measure this change is by comparing the occurence of pre- and post-verbal non-postposing elements, such as pronouns, particles, stranded prepositions and others (e.g., Pintzuk 1999; Pintzuk and Haeberli 2008). The feature was measured indiscriminately of the non-postposing diagnostic involved, but separately for root (IP1) and subordinate clauses (IP2). For root clauses, the conservative pattern is exemplified in (1a), the innovative pattern in (1b) with the particle *ut* 'out' in pre- and post-verbal position respectively.

(1) a. & þa faam of his muðe **ut eode**
and then foam of his mouth out went
'And then foam came out of his mouth'
(cobede,Bede_3:9.184.24.1845) (c. 895 A.D.)

b. ac se biscop **eode ut** hym togeanes
but the bishop went out him towards
'But the bishop went out, towards him'
(coaelhom,ÆHom_27:43.3959) (c. 1000 A.D.)

The examples in (2) illustrate the same variation in subordinate clauses with the two diagnostic elements *forð* and *on*.

(2) a. ... swa þa dagas **forð onsceortiað**
... as the days forth on shorten
'... as the days continue to grow shorter'
(comart2,Mart_2.1_[Herzfeld-Kotzor]:B.13.12) (c. 875 A.D.)

b. ... butan hi **sungon** þone lofsang **forð on**
... unless they sang the praise-song forth on
... unless they continued to sing the praise'
(coaelive, ÆLS_[Swithun]:230.4371) (c. 1000 A.D.)

3. IP3. The same change can be measured by comparing clauses with the order *nonfinite verb - finite verb* as instances of conservative, necessarily I-final headedness (3a) to clauses with the order *finite verb - nonfinite verb*, which could be innovative, I-initial or involve verb

(projection) raising (3b) (Haeberli and Pintzuk 2012). Only subordinate clauses were considered since the variation is the most pronounced in this clause type.

(3) a. ... þa he to heofonum **astigan wolde**
... when he to heaven ascend would
'... when he was about to ascend to heaven'
(coblick,HomS_46_[BlHom_11]:117.14.1481) (c. 950 A.D.)

b. ... for þan he **wolde** æfter feowertig dagen **stigen** to heofene
... for that he would after forty days ascend to heaven
'... because he would ascend to heaven after forty days'
(coeluc2,Eluc_2_[Warn_46]:47.37) (c. 1110 A.D.)

4. IP4. Yet another way of measuring the change in IP-headedness comes from clauses which, in addition to a subject, include a heavy, non-pronominal constituent. If this constituent occurs before the finite, lexical verb, the clause must be necessarily I-final (4a). If the constituent occurs after the verb, the clause is either I-initial or involves some rightward postposition process (4b) (Pintzuk 1999). The heavy constituent was defined as a nominal argument or prepositional phrase. As for the previous feature, I only considered subordinate clauses.

(4) a. ... þa ðu [**mæstne welan**] **hæfdest**
... when you most wealth had
'... when you had the greatest wealth'
(coboeth,Bo:26.58.24.1078) (c. 890 A.D.)

b. ... for ðan ðe we **habbað** [**heofenlice welan**]
... for that that we have heavenly wealth
'... because we have heavenly wealth'
(cocathom2,ÆCHom_II,_38:281.63.6343) (c. 1000 A.D.)

5. VP1. During the early English period, there is a gradual change in the headedness of VP. VP is the complement of I containing lexical material like objects and non-finite verbs. A head-initial VP can only be found under an innovative, head-initial IP, i.e. word order patterns with a head-initial VP but head-final IP are ungrammatical, such as the order **nonfinite verb - object - finite verb* (e.g., Smith 1893; Kiparsky 1996; Pintzuk 1999). Since a head-initial VP is contingent on an innovative IP setting, the frequency of the former is generally lower than the frequency of the latter. In fact, the larger part of this development takes place only in Middle English (Kroch and Taylor 2000*b*; Trips 2002). There are two different ways of measuring the change in VP headedness. Firstly, the same non-postposing elements as for IP1/IP2 can be used with a nonfinite verb. Pre-verbal diagnostic elements may or may not indicate a head-final VP, but where a diagnostic element occurs after the nonfinite verb, the clause must necessarily be V-initial. This is illustrated below with the non-subject pronoun *ðe* 'you' as a diagnostic element in pre- (5a) and post-verbal position (5b). Since diagnostic elements after nonfinite verbs occur on the whole very rarely, I did not make a difference between different diagnostics or different clause types but considered the entire material jointly.

(5) a. heofenas **ðe** synt mid gefean **ontynede**
heavens you are with companions opened
'The heavens are opened to you with your companions'
(comart3,Mart_5_[Kotzor]:Au17,A.22.1498) (c. 875 A.D.)

b. he wolde **ateon ðe** fram Drihtne
he wanted draw you from Lord
'He wanted to draw you from the Lord'
(cootest,Deut:13.10.4736) (c. 1000 A.D.)

6. VP2. Secondly, the change in VP headedness can be measured by comparing *direct object - nonfinite verb* (6a) to *nonfinite verb - direct object* orders (6b). However, this measurement is not as straight-forward as it may at first appear. Both leftwards scrambling and rightwards postposition are available in Old English and obscure the frequency of actual VO and OV structures. In order to at least minimise the risk of these process, I required the objects to immediately precede or follow their verb. Further, object positioning has been shown to be determined by the discourse status, heaviness and quantificational type of the object as well as clause type (e.g. Fischer et al. 2000; Pintzuk 2005; Pintzuk and Taylor 2009). In a recent multivariate analysis, Pintzuk and Taylor (2012, 56, table 3.12) found a factor weight of .541 for the placement of an object after a finite main verb that is exactly two words long. In other words, two-word objects occur in near free variation with respect to their positioning relative to the verb. Therefore, I decided to require all direct objects to dominate exactly two-words. While I thus controlled for the heaviness of the object, pragmatic, quantificational and clause-type effects were not considered.

(6) a. & hyt ne mihte [**þæt hus**] **astirian**
and it not could that house shake
'And it couldn't shake that house'
(cowsgosp,Lk_[WSCp]:6.48.4082) (c. 990 A.D.)
b. ... for þan ðe he ne mihte **geopenian** [**Petres digelnysse**]
... for that that he not could open Peter's secret
'... because he couldn't open Peter's secret'
(cocathom1,ÆCHom_I,_26:395.189.5122) (c. 1000 A.D.)

7./8. SB1/SB2. In early English, non-subject pronouns frequently occur in a scrambled position in the left periphery of the clause (e.g., Kemenade 1987; Pintzuk 1996; Bergen 2003) while in Modern English, no such position is available. I measured this change by counting all instances of non-subject pronouns that are positioned along with the subject before the verb but are separated from it by another element. I contrasted this number with occurrences of non-subject pronouns that are separated from the subject and are thus closer to the verb. Root clauses (SB1) and subordinate clauses (SB2) were investigated separately. Example (7) illustrates this measurements in root clauses. (7a) shows the object pronoun *hine* 'him' in a presumably scrambled position as it occurs along with the subject in a pre-verbal position and is separated from the finite verb for instance by the adverb *þær* 'there'. Note that the design of this search query may effect an I-final headedness for the majority of the clauses investigated. As a consequence, the measurement may register a somewhat diluted effect, detecting not only pronominal scrambling but also the change in IP-headedness. Example (7b) includes the object pronoun *hine* 'him' in a non-scrambled position as it is separated from the subject, here by the finite verb itself.

(7) a. he **hine þær** to biscope gehalgode
he him there to bishop consecrated
'He consecrated him bishop there'
(cobede,Bede_2:15.146.14.1408) (c. 890 A.D.)

b. & he **halgode** **hine**
and he consecrated him
'And he consecrated him'
(cootest,Exod:20.11.3187) (c. 1000 A.D.)

The sentences in (8) exemplify the same change in subordinate clauses. Since I-final structures are less common in root than in subordinate clauses, the measurement may work more adequately in the latter than in the former context.

(8) a. ... þæt he sylfa **us hider** gesecean wolde
... that he self us hither seek would
'... that he himself would seek us here'
(coblick,HomS_46_[BlHom_11]:119.67.1519) (c. 950 A.D.)

b. ... gif þu **geþyldelice me** gehyran wylt
... if you patiently me hear want
'... if you want to listen to me patiently'
(cocathom1,ÆCHom_I,_38:514.201.7680)) (c. 1000 A.D.)

9./10. SB3/SB4. As the design of the previous search may have led to interference effects from IP-headedness, I investigated the phenomenon of pronominal scrambling with a second measurement. I considered clauses that are likely to be I-initial since the finite verb precedes a phonologically light element, which is much less likely to postpose than heavy elements. A phonologically light element was defined as any phrase dominating only a single word, including nonfinite verbs. Some of these clauses are necessarily I-initial, namely where the one-word phrase is a diagnostic element like a particle or a pronoun. Admittedly, however, a certain number of the observed clauses may not actually be I-initial but involve postposition of the light element. I then compared the number of clauses with a pre-verbal pronoun as likely instances of pronominal scrambling to the number of clauses with post-verbal pronouns. The searches were carried out separately for root (SB3) and subordinate clauses (SB4). The sentences in (9) exemplify this measurement in the first clause type. (9a) is a likely case of pronominal scrambling because light elements such as *ealle* 'all' or *leue* 'permission' follow the finite verb, which is therefore more likely than not to be in an I-initial configuration, and the pronoun *hem* 'them' occurs in pre-verbal position. Example (9b) does not involve pronominal scrambling as the pronoun *hire* 'her' is positioned post-verbally.

(9) a. & Se kyng: **hem geaf** ealle leue ham to farene
and the king them gave all leave hom to go
'And the king gave them all permission to go home'
(CMPETERB,51.295) (c. A.D. 1130)

b. He **forȝaf** **hire** ðane deað
He forgave her the death
'He absolved her from death'
(CMVICES1,111.1355) (c. 1200 A.D.)

The examples in (10) illustrate the same change in subordinate clause with a nonfinite verb as the phonologically light post-verbal element. (10a) includes a pre-verbal non-subject pronoun which plausibly receives a scrambling analysis. In contrast, (10b) shows a pronoun after the finite verb, which is therefore not scrambled.

(10) a. ... þæt ealle men **him sceoldon** gafol gyldan
... that all men him should tribute pay
'...that all men should pay tribute to him'
(coverhom,HomS_1_[ScraggVerc_5]:107.908) (c.950 A.D.)

b. ... þæt his bigleofa **moste** mid ealle **him** losian
... that his money could with all him become-lost
'... that he could lose his money altogether'
(coaelhom,ÆHom_22:609.3665) (c. 1000 A.D.)

11. SUB. Early English subjects can be placed in relatively variable positions, depending on different information-structural properties (Biberauer and Kemenade 2011). Modern English, on the other hand, has developed a canonical, pre-verbal subject position that hosts the great majority of all subjects. I investigated this change by contrasting all subordinate clauses that do not show the subject in first position immediately after the complementiser or other subordinator, such as (11a), to subordinate clauses that do show the subject in first position after the complementiser or subordinator and that immediately precede the finite verb, such as (11b).

(11) a. ... an gewrit, þær **wæron** on awritene [**ealra þara ricestena monna**
... a document where were on written of.all of.the richest men
noman]
names
'... a document which the names of the richest men were written on'
(coorosiu,Or_6:3.136.4.2863) (c.890 A.D.)

b. ... þe wanes. þer [**þi martyrdom**] **is** iwriten inne.
... the walls where your martyrdom is written in
'... the walls where an account of your martyrdom is kept'
(CMMARGA,90.557) (c. 1200 A.D.)

12. V-C. There is a set of early English root clause environments in which the finite verb is placed in a high position as evidenced by the fact that it consistently precedes pronominal subjects. These conditions are the initial adverbs *þa*, *þonne* 'then' and to a lesser degree *nu* 'now' and others, imperatives (Pintzuk 1993), negated verbs (Ingham 2005), subjunctives (Bergen 2003), as well as verb-first sentences signalling some kind of transition or contrast (Ohkado 2004). In Modern English, the finite verb does not normally occur before subjects in these contexts. A very coarse measurement was used to explore this change. I contrasted the word order *finite verb - subject pronoun* as a diagnostic for high verb placement (12a), with the order *subject pronoun - finite verb* (12b) in non-conjoined declarative root clauses.

(12) a. ðonne ne **become** **he** no utane to ðæm sæde ðære wrohte
then not became.sbjctv he not outwards to the seed of.the strife
'Then he would not have become outwardly the seed of strife'
(cocura,CP:47.357.23.2430) (c. 890 A.D.)

b. **He becom** þa to his geferum.
he came then to his companions
'He then came to his companions'
(cocathom2,ÆCHom_II,_28:222.26.4893) (c. 1000 A.D.)

13. GEN. In early English, adnominal, modifier or complement genitive phrases can precede as well as follow their head noun. I compared all instances of the conservative pattern with the noun before the genitive, as in (13a), to the innovative structure with the genitive in prenominal position, as in (13b). For a recent overview of genitives in nominal constructions, see for example Crisma (2012).

(13) a. Her endað nu [seo æftre froferboc **Boeties**]
here ends now the second joy-book Boethius'
'Here now ends the second book of consolotion of Boethius'
(coboeth,Bo:21.50.5.909) (c. 890 A.D.)

b. he awrat [ða forman **Cristes** boc]
he wrote the first Christ's book
'He wrote the first book of Christ'
(cocathom2,ÆCHom_II,_37:272.21.6135) (c. 1000 A.D.)

14. REL. There are two basic relativisation strategies in early English. Either a form of the *se*-demonstrative paradigm can be employed as a relative pronoun, as in (14a), or an indeclinable particle like *þe* is used, as in (14b). Over time, the former option declines (Mitchell 1985; Suárez-Gómez 2006). I compared the number of occurences of the two variants under the relativisation of a nominal phrase. From the Middle English period on, the indeclinable relativiser frequently takes the form *þæt* (Suárez-Gómez 2012), which happens to be homonymous with the nominative/accusative, singular, neuter form of the *se*-demonstrative. However, a mismatch in number/gender with the antecedent often identifies the word as a relative particle and the Old English corpus is tagged accordingly. Ambiguous cases of *þæt* were counted towards the innovative pattern.

(14) a. ðone rihtan Godes geleafan, **ðone** hi gefyrn awurpon
the right God's belief which they formerly rejected
'the true faith of God, which they had formerly rejected'
(cobede,BedeHead:3.16.5.74) (c. 890)

b. þone halgan geleafan **þe** we habbað to Gode
the holy belief that we have to God
'the true belief that we have in God '
(coaelhom,ÆHom_11:72.1533) (c. 1000 A.D.)

Table 3 below summarizes the development of the features across the five periods in the general early English text corpus.

Feature	850-910	910-990	990-1100	1100-1200	1200-1230
IP1	12.9% (150/1161)	28.5% (419/1470)	35.9% (1125/31300)	33.4% (108/323)	77.9% (353/453)
IP2	3.8% (103/2696)	7.7% (116/1514)	9.6% (329/3426)	29.5% (88/298)	80.0% (412/515)
IP3	38.6% (759/1964)	43.8% (369/842)	50.1% (1151/2297)	79.4% (297/374)	88.4% (334/378)
IP4	31.7% (1132/3570)	37.2% (851/2283)	47.3% (2482/5248)	66.8% (342/512)	79.4% (431/543)
VP1	2.0% (28/1413)	2.8% (21/751)	6.0% (101/1673)	13.2% (51/387)	56.0% (289/516)
VP2	28.8% (204/708)	30.1% (118/392)	47.9% (667/1392)	57.2% (151/264)	83.5% (298/357)
SB1	33.3% (187/562)	59.8% (438/732)	75.1% (1143/1522)	77.3% (92/119)	96.5% (381/395)
SB2	10.8% (137/1270)	26.2% (142/541)	40.2% (478/1188)	83.1% (270/325)	93.8% (120/128)
SB3	25.1% (82/327)	42.0% (89/212)	53.1% (518/976)	76.5% (310/405)	91.9% (205/223)
SB4	25.4% (76/299)	62.5% (115/184)	69.8% (375/537)	71.8% (252/351)	89.8% (123/137)
SUB	62.6% (1224/1955)	67.5% (747/1106)	74.8% (2148/2873)	81.5% (379/465)	80.1% (492/614)
V-C	49.3% (2030/4120)	58.7% (2138/3645)	74.3% (5210/7010)	83.2% (636/764)	79.1% (721/911)
GEN	65.0% (5845/8999)	80.8% (4478/5543)	84.9% (10871/12807)	96.3% (1012/1051)	98.6% (1250/1268)
REL	61.5% (3168/5153)	77.9% (2555/3280)	78.4% (6916/8824)	96.2% (1631/1695)	99.9% (1854/1856)

Table 3: Development of feature values across the five periods in the general early English text corpus

4 Probabilistic text classification

Probabilistic classification techniques attempt to identify to which of a set of classes a new observation belongs on the basis of quantifiable properties drawn from observations whose class-membership is known. In the context of the present study, the period of the undated early English texts is to be ascertained on the basis of the text-internal criteria found in the general early English corpus. I will first describe the functioning of the probabilistic classifier I used for this task - a Naïve Bayes classifier. Next, the accuracy of the model will be evaluated. Finally, I will present the results of the classification process for the undated early English texts.

4.1 Naïve Bayes classifier

I used a classification method called Naïve Bayes for the dating of the undated early English texts. A naïve Bayes classifier is a convenient classification model for the task at hand because it is simple and hence easy to train and use, can handle missing values, is quite robust to irrelevant features and isolated noise, and has frequently been proven to work well for real-world text classification tasks, such as document categorization or spam filtering.

A Naïve Bayes classifier model for continuous data simply consists of means and standard deviations of the variables observed for each of the possible classes. In the present case, means and standard deviations of the values for the fourteen syntactic features for each of the five periods were determined, based on the texts from the general early English text corpus. This is illustrated for feature IP1 with the histogram at the left-hand side of figure 1. It shows the frequency counts of the feature values observed in the texts for each of the five periods. Means and standard deviations can be calculated from this information. From the means and standard deviations, Naïve Bayes calculates normally-distributed probability density functions of the features for each class. The rescaled histogram at the right-hand side of figure 1 shows the normally-distributed probability density function of feature IP1 for the five periods.

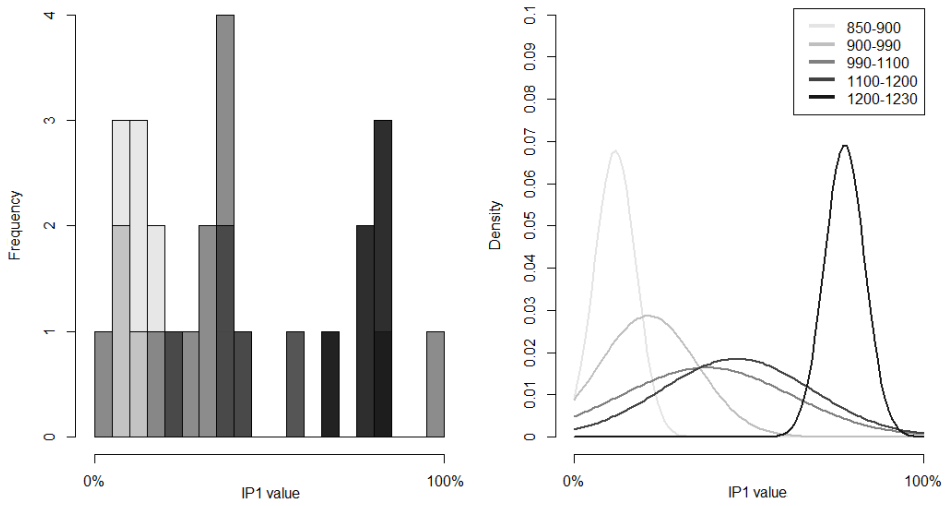

Figure 1: Histogram and probability density function assuming normal distribution for feature IP1 in the general early English text corpus

The classification technique is naïve because it assumes that the features used in the classification process are independent of each other. In principle, this assumption seems to be

well-justified in my data: there is no obvious reason to be concerned that the value obtained for, say, the realization of the relativiser is dependent on, for instance, pronominal scrambling etc. However, some features are explicitly taken to measure the same underlying change (e.g. IP1-IP4). Even though the independence assumption may thus be violated in some cases, this should not negatively impact the overall performance of the model too greatly because it has been shown that Naïve Bayes classifiers are quite robust to such violations (Domingos and Pazzani 1996).

The classifier is Baysian in that it adopts Bayes' theorem for the calculation of the probability that an instance belongs to a particular class, which, applied to the dating problem, reads as follows:

$$(15) \quad P(period|criteria) = \frac{P(period) \times P(criteria|period)}{P(criteria)}$$

For each undated early English text, I calculated the five posterior probabilities (one for each period) that a text belongs to a particular period given the criteria measured for this text. The probabilities sum to 1 and the period with the largest probability is selected as the most likely date of composition for the text.

These posterior probabilities are calculated as follows. First, the prior probabilities of the five periods are determined. The priors are taken simply from the input frequencies of the observations from each period. For example, 10 out of 43 texts in the general early English corpus were composed in period 1 - hence the prior for period 1 is $10/43 = 0.23$.

Next, to get the numerator of Bayes' theorem, the period prior is multiplied with the likelihood that one would find in an undated text a vector of features with the particular values as they were measured given a period. Such a likelihood of a period on an individual feature value corresponds to the height of the normally-distributed probability density curves at the specific point of the value. For example, an undated text might show an IP1 value of 15%. This would correspond to a likelihood of 5.9% for period 1, 2.6% for period 2, 1.1% for period 3, 0.6% for period 4 and ≈0% for period 5, as can be gauged from the probability density curves shown at the right hand side of figure 1. The overall likelihood of a period i on all criteria is then simply calculated as the product of the individual k to n criterion probabilities given the period.

$$(16) \quad P(criteria|period_i) = \prod_{k=1}^{n} P(criterion_k|period_i)$$

Finally, the numerator is divided by the a priori probability of the criteria for an undated text. The criteria prior is calculated as the sum of all likelihoods multiplied by the period priors. In other words, the prior of the feature values in an undated text corresponds to the probability of these feature values given period 1 times the prior of period 1, plus the probability of these feature values given period 2 times the prior of period 2, ... plus the probability of these feature values given period 5 times the prior of period 5. Therefore, the probability of the criteria will be identical for all periods and hence ultimately irrelevant for the decision process.

I calculated the posterior probabilities conditioned on the five periods for the undated early English texts with the statistics software *R* using the Naïve Bayes function included in package e1071.

4.2 Accuracy

In order to evaluate the accuracy of the model, I retained some textual material from the general Old English text corpus for testing. Since the purpose of this study is text classification, I tried

to identify coherent stretches of text - an entire chapter from a book, a complete sermon from a collection of homilies, the manuscript material attributed to one scribe etc. In this way, the test-material was philologically similar to the actual undated texts. This methodology is also justified by the fact that the lengths of the texts differ widely. For some very short texts, the extraction of a fixed percentage of words would have led to unclassifiably short text files. Other texts are very long so that they fruitfully allow the separation of more than just one classifiable test excerpt.

The following 50 texts were randomly selected for testing: Period 1: (1) *Bede* book 2, ch. 1-7, (2) *Bede* book 2, ch. 8-16, (3) *Bede* book 5, ch 1-12, (4) *Boethius*, ch. 1-10, (5) *Cura Pastoralis*, ch. 1-6, (6) *Cura Pastoralis*, ch. 7-12, (7) *Martyrology* part 1, (8) an excerpt from *Martyrology* part 5, entries for December 31- March 23, (9) *Marvels of the East* in its entirety, (10) *Orosius*, beginning up to book 1, ch.1 , (11) *Orosius* book 6, (12) *Soliloquy* book 3, (13) the beginning of *Soliloquy*, from preface up to book 1, ch.25. Period 2: (14) *Anglo-Saxon Chronicle*, Parker, scribe 2, (15) *Benedictine Rule*, from beginning up to book 4, chapter 17, (16) *Blickling Homilies* no. 19, entitled 'St. Andrew,' (17) *Blickling Homilies* no. 3, 4, 5 on Lent, (18) later parts of *Chrodegang*, from chapter 69 to the end, (19) *Vercelli Homilies* no. 11-13, (20) *Vercelli Homilies* no. 19-21, (21) the Gospel of John from the *West-Saxon Gospels*. Period 3: (22) Ælfric's *Catholic Homilies I*, no. 5, 6, (23) Ælfric's *Catholic Homilies I*, no. 33-36, (24) Ælfric's *Catholic Homilies I*, no 36-40, (25) Ælfric's *Catholic Homilies II*, no. 14, (26) Ælfric's *Catholic Homilies II*, no. 44, (27) Ælfric's *Lives of Saints*, beginning up to 'Julian and Basilissa,' (28) Ælfric's *Lives of Saints*, 'Exaltation of the Holy Cross,' (29) Ælfric's *Supplemental Homilies* no. 18, (30) Ælfric's *Supplemental Homilies* no. 23, 24, (31) Ælfric's *Letter to Wulfsige*, (32) Ælfric's *Old Testament* translation of Genesis, (33) Ælfric's Genesis Prologue and Epilogue, (34) Ælfric's *Temporibus*, (35) *Wulfstan Homilies* no. 1b, 2, 3, (36) Wulfstan's *Institutes of Polity* in its entirety, (37) *Apollonius* in its entirety, (38) the first 450 tokens of *Byrhtferth's manual*. Period 4: (39) The second continuation of the Peterborough *Anglo-Saxon Chronicle*, 1132-1154, (40) *In Festis Sancti Marie* in its entirety, (41) *Lambeth Homilies* from p. 145 to the end, (42) *Trinity Homilies* no. I-IV, (43) *Trinity Homilies* from p. 173 to the end, (44) *Vices and Virtues* from p. 131 to the end. Period 5: (45) *Ancrene Riwle* part 2, from p. 228 to the end, (46) *Hali Meidhad*, from p. 156 to the end, (47) *St. Juliana* from p. 94 to p. 102, (48) *St. Katherine* from p.40 to the end, (49) *St. Margaret*, from p.53 to p. 64, (50) *Sawles Warde*, from p. 165 to p. 172.

Unfortunately, it was not easily possible to calculate the number of words of the newly created test texts because of the presence of morpho-syntactic annotation in the text files. I therefore counted the number of sentence tokens, rather than words, in the training and test texts. The aim was to retain 20% of the overall text material for testing. Table 4 below shows the number of tokens in the training and test sets.

	Training	Test	% Test
Period 1:	16,245	4,046	19.94
Period 2:	13,591	3,512	20.53
Period 3:	35,220	8,409	19.27
Period 4:	4,159	1,104	20.98
Period 5:	4,747	1,133	19.27

Table 4: Number of tokens in the training and test material

I then trained a Naïve Bayes classifier model on the training texts and used it to predict the period of the 50 test texts. The confusion matrix below summarizes the results of this evaluation.

	actual				
	1	2	3	4	5
predicted 1	10	4	2	0	0
2	2	2	1	0	0
3	1	2	13	0	0
4	0	0	1	6	1
5	0	0	0	0	5

Table 5: Confusion matrix of test texts

Overall 36 out of 50 texts were classified correctly, which corresponds to an accuracy of 72%. However, the performance of the classifier across the five periods is markedly different. Middle English texts are virtually always recognized (11/12 texts, 92%). Likewise, early and late Old English texts are largely classified in a satisfactory way (10 out of 13 texts, 77% and 13 out of 17 texts, 76% respectively). In contrast, only 2 out of 8 texts from the 10$^{\text{th}}$ century were predicted accurately (25%). Thus, the classifier may perform poorly for period 2.

The main reasons for the poor model performance on 10$^{\text{th}}$ century texts appears to be the transitional nature of this period, sandwiched between early and late Old English. As a consequence, the feature values often do not stratify the Old English periods sharply enough for the second period to emerge as a clearly distinct class. Furthermore, the composition of some of the test texts may in fact fall into periods 1 and 3. Those texts would thus actually have been classified correctly. For example, certain *Blickling homilies* may have been composed before the 10$^{\text{th}}$ century; the *West-Saxon Gospels* were written at the very end of period 2 etc. The text collections as a whole may be more representative of 10$^{\text{th}}$ century English than their individual components.

4.3 Results

I trained the final Naïve Bayes classifier model on the whole general early English text corpus. Since the entire text material was used, accuracy might have slightly improved. I then used the model to predict the period of the undated texts. The results of this study are shown in table 6. Grey cells indicate the period of a text for which the classifier returned the highest probability; light-grey cells mean that the classifier did not converge on one unambiguous result.

5 Discussion

All of the undated texts appear to be of Old rather than Middle English origin; they were all grouped into periods 1 to 3. The fact that none of the texts was categorized as period 5 validates the Naïve Bayes classifier model since this category was a mere control group. I will now discuss the results for the individual texts in turn.

Based on the probabilistic text classification, *St. Chad* appears to be a very early text. This finding is consistent with Vleeskruyer's (1953, 70) assessment, who independently argued for an early date based on lexical features. For example, 12.9% of all clauses with a non-postposing

Text	Periods	1	2	3	4	5
1. Saint Chad	1-2	1.000	0.000	0.000	0.000	0.000
2. Saint Christopher	1-2	0.629	0.362	0.009	0.000	0.000
3. Mary of Egypt	1-3	0.999	0.001	0.000	0.000	0.000
4. Saint Euphrosyne	1-3	0.024	0.905	0.071	0.000	0.000
5. Saint Eustace	1-3	0.006	0.001	0.992	0.000	0.000
6. Seven Sleepers	1-3	0.381	0.619	0.000	0.000	0.000
7. Holy Rood-Tree	1-3	0.001	0.999	0.000	0.000	0.000
8. James the Greater	1-4	0.000	0.002	0.898	0.100	0.000
9. Saint Neot	3-4	0.022	0.689	0.289	0.000	0.000
10. Saint Margaret C	3-4	0.007	0.952	0.041	0.000	0.000
11. Saint Margaret T	1-3	0.000	0.000	1.000	0.000	0.000
12. Herbarium	1-3	0.064	0.936	0.000	0.000	0.000
13. Quadrupedibus	1-3	0.077	0.004	0.919	0.000	0.000
14. Leechbook	1-2	0.556	0.295	0.149	0.000	0.000
15. Lacnunga	2-3	0.103	0.000	0.897	0.000	0.000
16. Nicodemus	1-3	0.998	0.000	0.002	0.000	0.000
17. Vindicta	1-3	0.047	0.948	0.005	0.000	0.000
18. De Ascensione	1-3	0.000	0.000	0.999	0.001	0.000
19. Distichs of Cato	2-3	0.016	0.983	0.001	0.000	0.000

Table 6: Results of the classification of undated early English texts

diagnostic show I-initial phrase structure in period 1 (IP1), and in *St. Chad* even fewer, 8.7%, (2 out of 23) instances do. In period 1, 61.5% of all relative clauses are formed with a demonstrative of the *se*-paradigm (REL); *St. Chad* is even more traditional in this respect, 38.9% (14 out of 36) of all relatives featuring the conservative variant etc. Therefore, Vleeskruyer's suggested date of composition, 850-75 A.D., seems to me to be a perfectly appropriate estimate.

St. Christopher must be a relatively early, 9th or 10th century text (Bately 1988) and the probabilistic text classification is in agreement with this. However, the classifier did not return an unambiguous result; both periods 1 and 2 remain an option. The probability for the earlier period is higher and might therefore be preferred.

The four non-Ælfrician texts from a manuscript otherwise associated with Ælfric's writings, *Mary of Egypt*, *Saint Euphrosyne*, *Saint Eustace and his Companions* and *Seven Sleepers* were grouped into periods 1, 2, 3 and an indeterminate period respectively. This may be a surprising result in so far as one might have expected the common transmission of the texts to translate into a common provenance as well. It is of course conceivable that the classifier functioned inaccurately here. On the other hand, independent appraisals of the texts parallel my findings: *Mary of Egypt* has been criticised for its poor Latin translation (Magennis 1986; Magennis 2002). This is a characteristic often associated with early Old English texts (e.g., Sisam 1953, 85). The probabilistic text classification did indeed assign the text to period 1. *Saint Euphrosyne* has been linked to the 10th century Benedictine reform movement (Donovan 1999) and that is also the period that the text classifier determined for this saint's life. The genre of *St. Eustace*, "practically, a secular and romantic story" (Skeat 1881-1900, 452) may point to a late date since fictional romances (most notably, *Apollonius*) first appear during this period. Therefore, the result of the text classifier - period 3 - converges once again with an

independent argument for the same date. Finally, *Seven Sleepers* has been argued to originate in late Old English,"in a late Anglo-Saxon environment of pragmatic literacy, administrative vitality and urban and economic sophistication" (Cubitt 2009, 1048). Moreover, one passage in the text describes periodic recoinage of currency that must have occurred after King Edgar's monetary reforms in 973 (ibid.: 1023). *Seven Sleepers* may therefore have been composed shortly after that date. Further, Metrical Charm 3, dating perhaps to the late 10$^{\text{th}}$ century, refers to the Legend of the Seven Sleepers (Bonser 1945), which may suggest that the two texts were composed in the same period. These arguments point to a late 10$^{\text{th}}$ century provenance of the legend. Even though the probabilistic classifier did not deliver an unambiguous result, it did return the highest probability for period 2. All in all then, the classification results may in fact have reliably detected the date of origin for these texts.

In contrast to many of his contemporaries, Napier (1894) had assumed that the *History of the Holy Rood-Tree* had been composed much earlier than its time of attestation in a 12$^{\text{th}}$ century manuscript. Indeed, a version in an early 11$^{\text{th}}$ century manuscript was subsequently discovered (Ker 1940; Colgrave and Hyde 1962). The results of the probabilistic classification support this finding as the text was assigned ambiguously to periods 2 or 3, the former with a higher probability. It is therefore possible that the text was composed even earlier than is now commonly assumed, perhaps in the mid to late 10$^{\text{th}}$ century. One argument in favour of the plausibility of this conjecture is the fact that the tradition of likening the Holy Cross to a tree pre-dates the early 11$^{\text{th}}$ century, as demonstrated, for example, by the much older poem *The Dream of the Rood*.

James the Greater, a text whose date of composition it is notoriously difficult to obtain reliable information for, was assigned to period 3 with a probability of c. 90%. In the absence of any other scholarly argumentation, the 11$^{\text{th}}$ century would thus appear to be the preferred date of composition for this text.

Scholars of the *Life of St. Neot* often assume a post-Conquest date of composition for the text because it has been transmitted in a 12$^{\text{th}}$ century manuscript (see the list of scholars in Godden 2010, 197, fn. 15). However, Goddon demonstrates quite conclusively that this date is unwarranted. There is "not a single word in the text that was not current in tenth century English or earlier, and one might expect a text composed in the twelfth century to betray some signs of the changes in lexis that are generally evident at that date. [...] Nor is there anything in the phrasing or syntax which might suggest a post-Conquest date" (ibid.: 202). The probabilistic text classification supports the assertion that *St. Neot* was written considerably earlier than the 12$^{\text{th}}$ century. It returned periods 2 or 3 as plausible dates of composition. Despite a higher probability for period 2, I find myself in agreement with Goddon's (ibid.) analysis showing that the author of *St. Neot* was familiar with Ælfric and Wulfstan's writings, thus postdating them. Period 3 seems to be the correct period for the text.

The only texts for which my findings are in direct contradiction to standard scholarly opinion are the two Old English saint's lives of *St. Margaret*. Clayton and Magennis (1994) maintain that the composition of the Corpus version falls into a period "not long before that of the [12$^{\text{th}}$ century] manuscript in which it is preserved." (ibid. 70). This passion would therefore post-date the text in Tiberius, which is extant in a mid 11$^{\text{th}}$ century manuscript. In contrast, the Naïve Bayes classifier model assigned the Corpus passion to period 2, but the Tiberius version to period 3, exactly inverting the relative ordering of the two saint's lives. Magennis and Clayton give three arguments in favour of a late origin of the Corpus text: (i) the "insistent and developed use of the language of love" (ibid. 70), (ii) the use of *bread* in the sense 'bread' [which] is extremely rare in Old English" (ibid. 104), and (iii) the occurrence of the lexical

item *seagnt* and the Middle English indefinite pronouns *men* and *me* (ibid. 106). Their first argument seems to me to be the strongest. It is indeed true that a great number of Margaret's utterances reveal her personal love for God and that such formulations in direct speech are virtually absent in other Old English saint's lives. An example is shown below.

(17) Drihten leof, æfre ic þe lufode
Lord dear, always I you loved
'Beloved God, I have always loved you'
(comargaC,LS_14_[MargaretCCCC_303]:5.16.63)

However, the hypothesis that such extracts bear testimony to a 12th century change in worshiping practice is hard to test on account of the lack of comparable texts from the same period. As a consequence, the strong language of love is as isolated in the early 12th century as it would be a hundred years earlier. The second argument in support of a late composition of the corpus text seems less convincing. While it is clear that the word *bread* 'bread' is rare in Old English, it is by no means unattested. Furthermore, it must already have been used in its Modern English sense 'bread' as is evidenced, for example, by the fact that it glosses the Latin word *panis* 'bread' in the Rushworth Gospel Glosses (Ker No. 292, 10th century).

(18) ... iuxta locum ubi manducauerunt panem ...
... neh ðær stowe ðer gietun þæt **bread** ...
... near of.the place where ate that bread ...
'... near the place where they ate the bread ...'
JnGl (Ru) (6.23) (DOE short title, chapter and line in DOE edition)

The single incidence of *bread* (mentioned twice in immediately adjacent tokens) could therefore simply reflect a rare Old English occurrence of the word from the 11th or even 10th century. Finally, the arguments from spelling are, in the context of the present study, irrelevant. As the authors themselves note, "it is often difficult to distinguish in our text between authorial and scribal contributions" and "we cannot be sure that these reduced forms [Middle English *men, me* < Old English *man* 'one'; Middle English *seagnt* < Old English *sanct* 'saint'] were what the original translator, as opposed to the later scribe, wrote" (ibid. 106). In other words, the innovative spellings may reflect the onset of Middle English pronunciation traditions but not the date of the original text.

In principle, it is possible for a text transmitted in a 12th century manuscript to have originated at a much earlier period (see discussion of *St. Chad* and *St. Neot* above) and the syntactic evidence suggests that this is indeed the case here. The Corpus version of *St. Margaret* is more conservative than the average of period 3 in terms of IP and VP headedness, pronominal scrambling, subject placement and V-to-C movement. Solely the genitive and relative clause features accord well with a date of composition in the 11th century but even these criteria do not justify placing the text in period 4. The manifestation of feature IP1 provides a good example. Only 1 out of 55 relevant main clauses in the Corpus version place a diagnostic element in post-verbal position (1.8%). Sentence (19a) illustrates the majority pattern. In contrast, there are post-verbal diagnostic elements in 19 out of 61 relevant main clauses in the Anglo-Saxon Chronicle from the entries for 1100 to the end of the second continuation in 1154 (31.1%) - i.e. for the time span that Clayton and Magennis set for the composition of the Corpus version of *St. Margaret*.[3] An innovative example from the *Chronicle* is given in (19b).

[3] $\mathcal{X}^2$, df=1, p<0.001

(19) a. and he hit **up** anam
and he it up took
'And he took it (=the child) up'
(comargaC,LS_14_[MargaretCCCC_303]:3.4.15)

b. Sume he iaf **up**
some he gave up
'He gave up some (=of the castles)'
(CMPETERB,58.572) (c. 1140 A.D.)

The discrepancy in syntactic feature values between the Corpus version and the average of definitively early 11th century texts, taken together with the fact that the Corpus text "is itself a copy, and we do not know at how many removes it is from the original composition" (ibid. 106), therefore points to a date of composition much earlier than the 12th century. As a compromise between Clayton and Magennis's analysis and the probabilistic text classification, I would suggest placing the Corpus version in the middle of the 11th century and thus assign both versions of *St. Margaret* to period 3.

I will now turn my attention to the Old English medical texts. Firstly, the *Herbarium* was classified as a 10th century text. This finding accords well with scholarly opinion: The text must be based on an unidentified Latin original. While there are various manuscripts of the Latin *enlarged Herbarium* (de Vriend 1984, lv-lxi), only one - a 10th century manuscript at Montecassino Italy - has illustrations similar to those found in the Old English Cotton manuscript. Thus, it stands to reason that the Old English version was translated at a time when illustrations in its Latin *vorlage* became fashionable, i.e. quite possibly the 10th century (Meaney 2000). Furthermore, the three major Old English copies of the text all come from the early to mid 11th century and show only minor variations. Under the assumption that greater variation among different witnesses of the same text indicates that more time has passed since the composition of their earliest common node in the stemma, it follows that the composition of the original Old English *Herbarium* falls into a period that is not too far removed from the early - mid eleventh century (Arsdall 2002, 76). A late 10th century origin, which would only require a few decades of transmission history before the surviving manuscripts, certainly seems compatible with this reasoning.

Medicina de Quadrupedibus, which co-occurs with the *Herbarium* in all known manuscripts, was assigned to period 3. The most compehensive discussion of this text has been offered by de Vriend (1984), whose analysis I will use as a basis to evaluate the plausability of my classification.

First of all, de Vriend suggests that both *Quadrupedibus* as well as the *Herbarium* were translated in Northumbria during the eighth century. This date is certainly much too early to be in accordance with the manifestation of the syntactic feature values that I observed for these texts. However, de Vriend allows for either a gradual modernizing of the language due to frequent copying or a completely new translation at a later date (ibid: xliii). Therefore, I will assume that, from a syntactic point of view, the points of origin of these texts are properly placed in the 10th or 11th centuries. Secondly, for the translation of *Quadrupedibus*, de Vriend maintains that the earlier Northumbrian version detectably influenced the now extant, basically West-Saxon copies. He lists three arguments in favour of some Anglian influence on *Quadrupedibus*: The text shows (i) a significantly higher frequency of unmutated and unsyncopated 1st and 2nd person indicative ending of strong verbs than is normal in West-Saxon texts (ibid. lxxi), (ii) a small number of words which are mainly associated with Anglian texts (e.g. *nænig* 'none') (ibid. lxii-lxiii), and (iii) a high frequency of literally translated Latin participle constructions

(ibid. lxxiii-lxxiv). In my opinion, if valid, these arguments merely show that *Quadrupedibus* has an Anglian element, but not that its date of compositon is correspondingly early. For instance, the text might have been translated in the 10th or 11th centuries by a Northumbrian scribe who was trained in the West-Saxon standard. De Vriend lists only one argument in support of the earliness of the translation: In *Quadrupedibus*, 1 *pening* 'penny, pennyweight' corresponds to 1 dragma or $\frac{1}{3}$ of a dragma. In the *Herbarium*, 1 *pening* is equal to only $\frac{1}{4}$, $\frac{1}{5}$ or even $\frac{1}{6}$ of a dragma (ibid. lxxxii). A pennyweight seems to be greater in the former than in the latter text. This might indicate that the extant version of *Quadrupedibus* is older than *Herbarium* if the *pening* was re-valued at a lower level in the late Old English period. However, it is difficult to assess the validity of this reasoning on account of the erratic nature of the Old English record of weights and measurements. It is possible that no exact equivalent between a penny and a dragma ever existed. Dragma could simply have been interpreted as 'some small quantity of silver' and hence been translated in variable ways. To give just one example of the complexity of the Anglo-Saxon conversion system, the Northern Rushworth Gospel glosses 'dragmas decem' '10 dragmas' as *fif sceattas teasiðum* '5 tithe-coins' (DOE: MtGl (Ru) 15.8). The Bosworth-Toller Dictionary states that the *sceatt* is worth a West-Saxon penny (Bosworth and Toller 1898, 827). Consequently, a penny should correspond to two dragmas - a very different result from what is found in the medical texts. All in all then, the argument from the unit *pening* is not unproblematic. Finally, de Vried offers some indications that might support the assumption that *Quadrupedibus* actually post-dates the *Herbarium*. There are some Latin manuscripts that contain only the *Herbarium*, some which include the *Herbarium* as well as *Quadrupedibus*, but there are none that only include *Quadrupedibus* to the exclusion of the *Herbarium*. Therefore, the Old English text is based either on a *vorlage* that contained both the *Herbarium* and *Quadrupedibus* together, or the *Herbarium* on its own. If the second scenario is correct, it would follow that *Quadrupedibus* was supplemented following the original translation of the primary medical text (de Vriend 1984, xlii-xliii). What this discussion shows is the enormous difficulty in finding the best fit for a date of composition for *Quadrupedibus* because of conflicting evidence. I would suggest that the different positions be balanced so that the text will be assigned to a period that is neither as late as my classification suggested nor as early as de Vriend believed. In other words, the common ancestor from which all three 11th century manuscripts indubitably descended and which already contained both the *Herbarium* as well as *Quadrupedibus* (ibid. xliii) could be regarded as having originated together in the late 10th century, i.e. period 2.

The third medical text, *Bald's Leechbook*, could not be assigned to one period unambiguously. The text comprises three collections of medical remedies that scholars agree were composed at different times. I therefore ran the Naïve Bayes classifier on the three subparts separately. The results are presented in table 7.

Text	1	2	3	4	5
Leechbook Part 1	0.979	0.021	0.000	0.000	0.000
Leechbook Part 2	0.996	0.004	0.000	0.000	0.000
Leechbook Part 3	0.051	0.018	0.931	0.000	0.000

Table 7: Results of the classification for the three sections of *Bald's Leechbook*

When the three subparts are considered separately, the ambiguity disappears and each part can be assigned clearly to one period. Moreover, the new findings correspond very well to scholarly assessments of the *Leechbook*. Parts 1 and 2 are likely to have been written within

King Alfred's educational reform programme and indeed the classification procedure assigns period 1 the highest probability for these parts. The chapters in part 3 were probably added later and, again, the text classification is in accordance with this analysis, returning period 3 as the most likely point of origin for this part (Nokes 2004). Hence, the three parts of *Bald's Leechbook* should be treated separately in studies of diachronic syntax. Parts 1 and 2 should be regarded as early Old English while part 3 appears to be more representative of late Old English.

The results for the last medical text, *Lacnunga*, are unsurprising. It is uncontroversial that it must have been composed relatively late (Arsdall 2002) and, specifically, after the *Herbarium*. The proabilistic text classifications suggest that, as expected, *Lacnunga* belongs to period 3.

The Old English apocrypha, *Gospel of Nicodemus* and *Vindicta Salvatoris*, were assigned to periods 1 and 2 respectively. It is remarkable that the exact Latin *vorlage* manuscript is known from which these Old English texts were translated, a 9th century compilation from Saint-Omer, Flanders, France. During any point between the late ninth and the mid-eleventh century, the manuscript must have left the continent and come to England. The texts were then translated into Old English. Later, the manuscript went back to its original place of origin (Cross 1996*a*, 3). If the assignment of *Nicodemus* to period 1 is accurate, it would follow that the transferral of the manuscript from the Continent to England did in fact happen early, i.e. already in the late ninth century. In theory, this is a possible scenario. For example, two of the many Anglo-Flemish connections listed by Cross and Crick (1996, 17-21) are early enough to provide a plausible context for the journey of the manuscript to England: "The West Saxon royal family was connected by marriage with the comital house of Flanders in the later ninth century. One of Alfred's leading advisers was a former monk of Saint-Bertin [=St. Omer]" (ibid. 17). However, there are more convincing arguments against such a view. Firstly, relations between England and Flanders stengthen during the 10th and 11th centuries, which makes a relocation of manuscripts more likely during those periods. For instance, Cross and Crick (ibid.) mention that "Edmund, Alfred's grandson, is supposed to have installed monks from Saint-Bertin [=St. Omer] at the monastery of St. Peter's, Bath, in 944. In 961-2 Adelulf, abbot-elect of Saint-Bertin [=St. Omer], was sent by the count of Flanders, Arnulf, on a mission to England." Secondly, there are annotations in the manuscript by no fewer than four Anglo-Saxon hands, mainly from the middle of the 11th century, but the earliest possibly dating to the abbacy of Odbert (986-1007) (ibid. 21). If the manuscript had been in England since c. 900, one would probably expect annotations by Anglo-Saxon hands to appear earlier. Thirdly, while it is clear that the manuscript pre-dates the three surviving copies of the Old English text by more than one remove (ibid. 17), the differences between the translations do not seem great enough to justify the assumption that the original exemplar dates back further than a century. Fourthly, the manuscript may have been in England for only a relatively short period of time, perhaps as a loan (ibid.: 20). If the manuscript had arrived in the 9th century and stayed there at least until the second half of the 11th, the plausible loan hypothesis would become highly improbable. Fifth, one can appeal to an argument from authority: Allen (1939, 60) asserts that "the language of Ms C. [Ker No. 20, from the third quarter of the eleventh century] should be dated around 1000" but certainly not 900. Marx (1997, 125) believes that the translation of Nicodemus could have been made "possibly as early as the mid tenth century" but not before that time. Finally, there is good reason to believe that *Nicodemus* and *Vindicta* were translated together. In fact, in his detailed study of the style of the two texts, Orchard (1996, 130) concludes that "both the Old English *Gospel of Nicodemus* and the *Avenging of the Saviour* [=*Vindicta*] are the products not simply of a single scriptorium [...], but of a single

author." Hence, the divergence in the classification results between the two apocrypha casts serious doubt on the correctness of the assigned periods for at least one of them. The assessment of the model accuarcy above showed that the classifier frequently underperforms on period 2, assigning texts to periods 1 or 3 that actually belong into the 10th century. Therefore, it might make more sense to assign the text classified as belonging to period 1 into period 2 than the other way around. All in all then, there appears to be ample reason to date both Old English apocrypha to the same period and allocate them to the mid to late 10th century, i.e. period 2.

The classifier assigned the unspecified homily *De Ascensione* to period 3. While it is difficult to present a comprehensive evaluation of this result on account of the small amount of scholarship done on this piece, there are at least some arguments supporting the classification. Most importantly, if the finding is correct, the text must post-date the Old English apocrypha. This would be a desirable conclusion because the homily uses a source text which narrates Christ's descent to Hell and the only other text known to do so is the Old English *Nicodemus*. Furthermore, the finding is in accordance with Hulme's (1904, 590) speculation that the text was composed in the 11th century.

The final undated text is *Distichs of Cato* and it was assigned to period 2. The oldest and most reliable copy of the text, a manuscript from Cambridge Trinity College, dates to c. 1100 (Ker no. 89). This is also the copy that is included in the YCOE electronic text file and hence the one that was classified here. It is in principle possible that a text that is attested only at the end of the 11th century has its roots at a much earlier point in time. Indeed, this is exactly what scholars have suggested for the Old English *Distichs*. Cox (1972) finds that most proverbs (perhaps 76 out of 81) were originally composed between the mid 10th to mid 11th century (cited in Hollis and Wright 1992, 18). Goldberg explicitly dated the original exemplar to the second half of the 10th century (cited in Hollis and Wright 1992, 18, 29). Even Treharne (2003), who seems to prefer a later date, points out that the first item of the *Durham Proverbs*, copid in the mid 11th century, is a fragment of *Distichs* no. 23 and concedes that "[i]t is possible from this argueably meagre evidence that an Old English translation of the *Dicts of Cato* existed in the tenth or earlier eleventh century" (ibid. 468). Thus it is certainly not inconceivable that the syntax of *Distichs* should accord well with other 10th century texts. Period 2 seems to me to be a plausible classification result for this text.

To sum up, most findings from the probabilistc text classification converge straight-forwardly with more traditional arguments for the date of composition of a text exemplar. Some of the results were adjusted slightly taking scholarly views into account; others provide the only evidence currently available for the point of origin of a text.

6 Conclusion

In this paper, I presented a probabilistic classification study that grouped early English texts into their respective periods of origin. I hope to have achieved two main goals.

Firstly, I attempted to offer dates of composition for nineteen undated early English texts relative to a system of five periods, spanning roughly the time between 850 and 1230 A.D. One might wish to adopt the suggested points of origin for these texts until further evidence is discovered that allows for their re-appraisal. In particular, I believe that the texts investigated here can fruitfully be employed in future diachronic studies, especially on early English syntax, where reliance on manuscript dates alone may result in the detection of incoherent developments.

Secondly, I tried to demonstrate that statistical classification based on syntactic features is

a viable method for chronological text ordering to begin with. The underlying rationale for the methodology employed in this paper assumed that syntax is relatively resilient to adaptations introduced by scribes during the copying process. Syntax may thus provide signals of the style of the original composition and its date. Of course, there can be no denying that this method does not automatically discover the true date of composition of a text but that it should rather be regarded as a fitting technique. In this respect, probabilistic text classification does not substitute or even compete with traditional dating methods. The latter must take precedence over the former. However, the fact that traditional arguments and the results from the Naïve Bayes classifier model frequently coincided indicates the soundness and reliability of the general methodology. Probabilistic text classification can supplement and strengthen philological approaches to exemplar dating.

References

Allen, T. P. (1939), *A Critical Edition of the Old English Gospel of Nicodemus: Old English and Latin texts with English introduction'*, University Microfilms, Ann Arbor Press, Houston.

Arsdall, A. v. (2002), *Medieval Herbal Remedies: The Old English Herbarium and Anglo-Saxon Medicine'*, Routledge, London.

Bately, J. (1988), 'Old english prose before and during the reign of alfred', *Anglo-Saxon England* **17**, 93–138.

Bergen, L. v. (2003), *Pronouns and Word Order in Old English: With Particular Reference to the Indefinite Pronoun Man*, Routledge, New York.

Biberauer, T. and Kemenade, A. v. (2011), 'Subject positions and information-structural diversification in the history of english', *Catalan Journal of Linguistics* **10**, 17–69.

Bonser, W. (1945), ''the seven sleepers of ephesus in anglo-saxon and later recipes', *Folklore* **56**, 254–6.

Bosworth, J. and Toller, T. N. (1898), *An Anglo-Saxon Dictionary: Based on the Manuscript Collections of the Late Joseph Bosworth*, Clarendon Press, Oxford.

Clayton, M. and Magennis, H. (1994), *The Old English Lives of St Margaret*, Vol. 9 of *Cambridge Studies in Anglo-Saxon England*, Cambridge University Press, Cambridge.

Cockayne, O. (1865), *Leechdoms, Wortcunning, and Starcraft of Early England*, Longman, London.

Colgrave, B. and Hyde, A. (1962), ''two recently discovered leaves from old english manuscripts', *Speculum* **37.1**, 60–78.

Crisma, P. (2012), Triggering syntactic change: Inertia and local causes in the history of english genitives, *in* Jonas et al. (2012), pp. 198–216.

Cross, J. E. (1986), The latinity of the ninth-century old english martyrologist, *in* P. E. Szarmach, ed., 'Studies in Earlier Old English Prose', State University of New York Press, Albany, pp. 3–9.

Cross, J. E. (1996*a*), Introduction, in *Two Old English Apocrypha and Their Manuscript Source: The Gospel of Nichodemus and The Avenging of the Saviour, with contributions by Denis Brearley, Julia Crick, Thomas Hall and Andy Orchard'* Cross (1996*b*), pp. 275–299.

Cross, J. E. and Crick, J. (1996), The manuscript: Saint omer, bibliotheque municipale, 202, *in* Cross (1996*b*), pp. 10–35.

Cross, J. E., ed. (1996*b*), *Two Old English Apocrypha and Their Manuscript Source: The Gospel of Nichodemus and The Avenging of the Saviour, with contributions by Denis Brearley, Julia Crick, Thomas Hall and Andy Orchard'*, Cambridge University Press, Cambridge.

Cubitt, C. (2009), ''as the lawbook teaches': Lawbooks and urban life in the anonymous old english legend of the seven sleepers', *English Historical Review* **124**, 1021–49.

de Vriend, H. J. (1984), *The Old English Herbarium and Medicina de Quadrupedibus*, EETS o.s. 286, Oxford University Press, London.

Dobson, E. J. (1972), *The English text of the Ancrene riwle edited from B.M. Cotton ms. Cleopatra C vi'*, EETS o.s. 267, Oxford University Press, London.

Domingos, P. and Pazzani, M. (1996), 'Beyond independence: Conditions for the optimality of the simply bayesian classifier', *Proceedings of the Thirteenth International Conference on Machine Learning* pp. 105–12.

Donovan, L. A. (1999), *Woman Saint's Lives in Old English Prose*, D.S. Brewer, Cambridge.

Drout, M. D. C. (2004), 'Re-dating the old english translation of the enlarged rule of chrodegang: The evidence of the prose', *The Journal of Germanic Philology* **103.3**, 341–368.

Fischer, O., Kemenade, A. v., Koopman, W. and Wurff, W. v. d. (2000), *The Syntax of Early English'*, Cambridge University Press, Cambridge.

Förster, M. (1920), 'Der inhalt der altenglischen handschrift vespasianus d. xiv', *Englische Studien* **54**, 46–48.

Förster, M. (1933), 'Lokalisierung und datierung der altenglischen version der chrodegang regel', *Sitzungsberichte der Bayerischen Akademie der Wissenschaften, Jahrgang 1933* pp. 7–8.

Godden, M. (2010), 'The old english life of st. neot and the legends of king alfred', *Anglo-Saxon England* **39**, 193–225.

Goolden, P. (1958), *The Old English 'Apollonius of Tyre'*, Oxford University Press, London.

Gretsch, M. (2004), *The Intellectual Foundations of the English Benedictine Reform*, Vol. 25 of *Cambridge Studies in Anglo-Saxon England*, Cambridge University Press, Cambridge.

Haeberli, E. and Pintzuk, S. (2012), Revisiting verb (projection) raising in old english, *in* Jonas et al. (2012), pp. 219–238.

Hall, J. (1920), *Selections from Early Middle English 1130-120. Part II: Notes*, Clarendon, Oxford.

Henel, H. (1943), 'Byhrtferth's preface. the epilogue of his manual?', *Speculum* **18.3**, 288–302.

Hill, J. (2009), Ælfric: His life and works, *in* H. Magennis and M. Swan, eds, 'A Companion to Ælfric', Brill's companions to the Christian tradition, Brill, Leiden, chapter 2, pp. 35–66.

Hollis, S. and Wright, M. (1992), *Old English Prose of Secular Learning*, D.S. Brewer, Cambridge.

Hulme, W. H. (1904), 'The old english gospel of nicodemus', *Modern Philology* **1.4**, 610–614.

Ingham, R. (2005), 'The loss of neg v - c in middle english', *Linguistische Berichte* **202**, 171–206.

Jonas, D., Whitman, J. and Garrett, A., eds (2012), *Grammatical Change: Origins, Nature, Outcomes*, Oxford University Press, Oxford.

Kemenade, A. v. (1987), *Syntactic case and morphological case in the history of English*, Foris, Dordrecht.

Ker, N. R. (1940), 'An eleventh-century old english legend of the cross before christ', *Medium aevum* **9**, 84–85.

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

Kiparsky, P. (1996), the shift to head-initial vp in germanic, *in* T. Höskuldur, S. D. Epstein and S. Peter, eds, 'Studies in Comparative Germanic Syntax, 2', Kluwer, Dordrecht, pp. 140–179.

Kroch, A. and Taylor, A. (2000*a*), *Penn-Helsinki Parsed Corpus of Middle English*, http://www.ling.upenn.edu/hist-corpora/PPCME2-RELEASE-3 (Accessed 10 April 2013), 2 edn, Department of Linguistics, University of Pennsylvania.

Kroch, A. and Taylor, A. (2000*b*), Verb-object order in early middle english, *in* S. Pintzuk, G. Tsoulas and A. Warner, eds, 'Diachronic Syntax, Models and Mechanisms: Proceedings of DIGS 5', Oxford University Press, Oxford, pp. 132–163.

Laing, M. (2000), 'never the twain shall meet': Early middle english - the east-west divide, *in* I. Taavitsainen, T. Nevalainen, P. Pahta and M. Rissanen, eds, 'Placing Middle English in Context', Walter de Gruyter, Berlin, pp. 97–124.

Magennis, H. (1986), 'Contrasting features in the non-ælfrician lives in the old english lives of saints', *Anglia* **104**, 316–348.

Magennis, H. (1991), 'The anonymous old english legend of the seven sleepers and its latin source', *Leeds Studies in English* **22**, 43–56.

Magennis, H. (2002), *The Old English Life of St. Mary of Egypt: An Edition of the Old English Text with Modern English Parallel-Text Translation*, Exeter University Press, Exeter.

Marx, W. C. (1997), 'the gospel of nicodemus in old english and middle english, *in* Z. Izydorczyk, ed., 'The Medieval Gospel of Nicodemus: Texts, Intertexts and Contexts in Western Europe', Medieval & Renaissance Texts & Studies 158, Arizona State University Press, Tempe, pp. 207–260.

Meaney, A. (2000), 'The practice of medicine in england about the year 1000', *The Society for the Social History of Medicine* **13.2**, 221–237.

Milett, B. (1996), *Ancrene Wisse, the Katherine Group, and the Wooing Group*, Vol. 2 of *Annotated Bibliographies of Old and Middle English Literature*, D.S. Brewer, Cambridge.

Mitchell, B. (1985), *Old English Syntax*, Clarendon, Oxford.

Morris, R. (1874-1880), *The Blickling Homilies of the Tenth Century. From the Marquis of Lothian's Unique Mansuscript, A.D. 971*, EETS o.s. 58, 63 & 73, Trübner & Co., London.

Napier, A. S. (1888), 'Ein altenglisches leben des heiligen chad', *Anglia* **10**, 131–56.

Napier, A. S. (1894), *History of the holy Rood-tree: a twelfth century version of the cross-legend with notes on the orthography of the Ormulum and a middle English compassio Mariae*, EETS o.s. 103, Paul, Trench, Trübner & Co., London.

Nokes, R. S. (2004), 'The several compilers of balds leechbook', *Anglo-Saxon England* **33**, 51–76.

Ohkado, M. (2004), 'On the structure and function of v1 constructions in old english', *English Studies* **85.1**, 2–16.

Orchard, T. N. (1996), The style of the texts and the translation strategy, *in* Cross (1996*b*), pp. 105–130.

Pintzuk, S. (1993), 'Verb seconding in old english: Verb movement to infl', *The Linguistic Review* **10**, 5–35.

Pintzuk, S. (1996), 'cliticization in old english, *in* 'Approaching Second: Second Position Clitics and Related Phenomena', CSLI Press, Stanford, CA, pp. 366–400.

Pintzuk, S. (1999), *Phrase Structures in Competition: Variation and Change in Old English Word Order*, Garland, New York.

Pintzuk, S. (2005), 'the syntax of objects in old english, *in* M. Batllori, M.-L. Hernanz, C. Picallo and F. Roca, eds, 'Grammaticalization and parametric change', Oxford University Press, Oxford, pp. 251–266.

Pintzuk, S. and Haeberli, E. (2008), 'Structural variation in old english root clauses', *Language Variation and Change* **20**, 367–407.

Pintzuk, S. and Taylor, A. (2009), 'the loss of ov order in the history of english, *in* 'The Handbook of the History of English', Wiley-Blackwell, Oxford, pp. 249–278.

Pintzuk, S. and Taylor, A. (2012), 'the effect of information structure on object positioning in old english: A pilot study, *in* A. Meurman-Solin, M. J. López-Courso and B. Los, eds, 'Information Structure and Syntactic Change in the History of English', Oxford University Press, Oxford, pp. 47–65.

Sawyer, P. H. (1968), *Anglo-Saxon Charters: an Annotated and Bibliography*, Royal Historical Society, London.

Scheil, A. (2000), 'Bodies and boundaries in the old english life of st. mary of egypt', *Neophilologus* **84**, 137–156.

Scragg, D. (1992), *The Vercelli Homilies*, EETS o.s. 300, Oxford University Press, New York.

Sisam, C. (1951), 'The scribal tradition of the lambeth homilies', *The Review of English Studies, New Series* **2.6**, 105–113.

Sisam, K. (1953), *Studies in the History of English Literature*, Oxford University Press, Oxford.

Skeat, W. W. (1874), *The Gospel according to Saint Luke in Anglo-Saxon and Northumbrian Versions synoptically arranged: with collations exhibiting all the readings of all the MSS.*, Vol. 3 of *The Anglo-Saxon Gospels*, Cambridge University Press, Cambridge.

Skeat, W. W. (1881-1900), *Ælfric's Lives of Saints*, EETS o.s. 76, 82, 94 & 114, Oxford University Press, London.

Smith, A. C. (1893), 'The order of words in anglo-saxon prose', *Publications of the Modern Language Association of America* **8**, 210–244.

Suárez-Gómez, C. (2006), *Relativization in Early English (950-1250): the Position of Relative Clauses*, PeterLang, Bern.

Suárez-Gómez, C. (2012), 'The consolidation of þat as an invariable relativizer in this history of english', *Nordic Journal of English Studies* **11.1**, 79–107.

Swanton, M. (1996), *The Anglo-Saxon Chronicle*, Routledge, New York.

Taylor, A., Warner, A., Pintzuk, S. and Beths, F. (2003), *The York-Toronto-Helsinki Corpus of Old English Prose (YCOE)*, http://www-users.york.ac.uk/ lang22/YCOE/YcoeHome.htm (Accessed 10 April 2013), Oxford Text Archive.

Treharne, E. M. (2003), 'The form and function of the twelfth-century old english 'dicts of cato", *The Journal of English and Germanic Philology* **102.4**, 465–485.

Trips, C. (2002), *From OV to VO in Early Middle English*, John Benjamins, Amsterdam.

Vleeskruyer, R. (1953), *The Life of Saint Chad: An Old English Homily*, North-Holland, Amsterdam.

Waite, G. (2000), *Old English Prose Translations of King Alfred's Reign*, Vol. 6 of *Annotated Bibliographies of Old and Middle English Literature*, D.S. Brewer, Cambridge.

Wallach, L. (1955), 'Alcuin on virtues and vices: A manual for a carolingian soldier', *The Harvard Theological Review* **48.3**, 175–195.

Warner, R. D. N. (1917), *Early English Homilies from the Twelfth Century MS. Vesp. D. xiv*, EETS o.s. 152, Oxford University Press, Oxford.

Whitelock, D. (1942), 'Archbishop wulfstan, homilist and statesman', *Transactions of the Royal Historical Society* **24**, 25–45.