

Celtic substrate influence on the Old English copula

Luke Bradley
Undergraduate Thesis
Queen Mary, University of London

April 2011

Abstract

This paper strengthens the ‘Celtic Hypothesis’ by examining the usage of the two forms of ‘be’ in Old English dialects. Different proportions of Celtic populations are found in different OE dialect areas. We expect that areas with higher Celtic populations would exhibit a more rigid semantic distinction between the two forms. Evidence from the Old English corpus suggests that greater substrate influence is found in the North, West and Midlands, as predicted by historical, linguistic, archaeological and genetic research on Celtic and Anglo-Saxon population patterning.

Keywords Celtic Hypothesis, copula, substrates; Old English, *beon*, *wesan*

Queen Mary’s OPAL #28
Occasional Papers Advancing Linguistics

1. Introduction

This investigation aims to discover what it was that made the English language different from the West Germanic dialects that it evolved from with regard to one specific feature: the existence of two formally separate paradigms of the verb ‘to be’.¹ The semantic distinction between the paradigms will be examined below, but first let us compare the two OE present paradigms with the single paradigms of Old High German and Old Norse, two sister languages:²

‘To be’	OE ‘wesan’	OE ‘bēon’	OHG ‘wesan’	ON ‘vera’
1sg.	ic ēom	ic bēo	ih bim	ek em
2sg.	þū eart	þū bist	dū bist	þú ert
3sg.	hē is	hē biþ	ēr ist	þat er
1pl.	wē sindon	wē bēoþ	wir birum	vér erum
2pl.	gē sindon	gē bēoþ	ir birut	þér erúþ
3pl.	hīe sindon	hīe bēoþ	sie sint	þau eru

Table 1: Present paradigms of ‘to be’ in OE, OHG and ON

In OHG, as in other continental Germanic languages, the *b-* and *wes-* forms mixed freely within one single paradigm. In ON, there are no *b-* forms. Why did OE make the distinction? It was not a distinction inherited from Proto-Germanic: **beunan* still carried the future aspect of being or becoming obligatorily while **wesaną* had the meaning ‘to stay, and only secondarily ‘to be’ (Ringe 2006: 70). It was only in OE that both verbs’ primary meaning is translated as ‘to be’.

The answer to this riddle is based on a language shift framework, the ‘Celtic Hypothesis’: it is assumed that Old English could not have become so noticeably differentiated from Old Frisian, Old Saxon et al. without a sizeable substrate population that was compelled to learn a form of the language. This would have been a completely different situation from Germanic-speaking areas on the mainland, which had long since expelled or assimilated the Celtic populations. The substrate, composed of British Celts, will be seen to form a much larger proportion of Old English speakers at the time of the Norman Conquest than the original Germanic peoples who are traditionally said, following Bede, to have invaded the Isles in the fourth and fifth centuries CE.

2. Literature Review

The Old English ‘twofold paradigm of “to be”’, as Lutz (2009) calls it, has not received as much attention as other structural coincidences between Celtic and English. This may be because it did not persist into the Middle and Modern periods, unlike, for example, periphrastic use of ‘do’. Yet

¹ Abbreviations used in this paper: A: Anglian (dialect), K: Kentish (dialect), M: Mercian (dialect), OE: Old English, OHG: Old High German, ON: Old Norse, PrE: Present-day English, WS: West Saxon (dialect)

² Standardized forms are given in the table. There are, of course, many variant spellings of each form, as will be elaborated presently.

1500 to 1000 years ago there did exist the possibility of a full-fledged lexical and semantic distinction for native speakers of English, as well as for speakers of Celtic (both Insular and Continental, according to Vennemann 2010: 391), just as there does for speakers of Spanish today (*ser* and *estar*). English and Celtic exhibit what must be described as a marked semantic differentiation when compared to the single copula paradigm of Proto-Indo-European, and that they are found geographically overlapping in their early history is surely no coincidence.

We see a reluctance in traditional grammars to mention the possibility of contact in the development of a ‘twofold paradigm of “to be”’. Mitchell and Robinson (2003), in the standard manual of Old English, first mention *bið* (from *bēon*) as a ‘special future form’ of ‘to be’, but then proceed to note its gnomic use for the ‘statement of an eternal truth’. The example given is ‘*wyrd bið ful aræd*’, meaning ‘fate is quite inexorable’. If fate is eternally inexorable, then we can assume it always has been, and so *bēon* cannot be defined as a special future form of ‘to be’. We are left knowing little of the true distinction or genesis of the twofold paradigm.

In the admirably thorough 20-page section on Old Irish ‘to be’ in Thurneysen (1980), we nevertheless find no mention of possible reasons for the twofold paradigm, and no mention whatsoever of Old English and its probable influence from Celtic in this regard. We are shown formal cognates with other IE branches e.g. *attá*, the ‘ordinary present indicative’, with Latin *stāre* and Greek *ἕσταν*, but nothing is said of their semantic connections. Similarly in Penny’s (2000: 191-3, 234) standard history of Spanish we learn of the formal derivations of *ser* and *estar* from the Latin with no mention of any contact-based explanations of the semantic development itself.

Naturally, it is arguable whether these points need to be mentioned in the above-cited works, but all are authoritative, detailed descriptions of the languages in question. Therefore, to exclude all mention of phenomena largely responsible for differentiating the languages from others in their families and making them into autonomous languages in the first place (i.e. language contact and language shift) is rather unhelpful to linguists undertaking comparative work. For all the help synchronic descriptions can be, no language exists in a vacuum. This is where recent, more contact-oriented work, particularly that which looks at the relationship between Celtic and Old (and Middle) English, comes in.

A historical problem that current scholars of the Celtic influence of English agree on is the belief, widely held until recently and still perpetuated in some influential quarters,³ that a language must show lexical influence from a proposed contact language for contact to have been historically likely. In their recent work on English-Celtic contact, Filppula et al. note that “in conditions of language shift, such as those which have characterised many parts of the British Isles for centuries, contact influences can be expected to be found in the domains of phonology and syntax rather than lexicon” (2007: 2). Lutz and Vennemann share this view and all three point to Keller 1925, a pioneering work which was the first to suggest the possible substratal influence on OE: “[D]ie altenglischen Formen und Funktionen der Wurzel **bheu*, die den

³ e.g. Baugh and Cable’s standard history of English (1993: 85)

anderen germanischen Dialekten fremd sind, entstanden im Munde und im Denken von englisch sprechenden Briten.”⁴

While Filppula et al. advance the theory that Keller’s paper went “largely unnoticed amongst Anglicists” thanks to its being written in German, Lutz believes it may be down to the attempted refutation by Flasdieck (1937), which she in turn refutes on the grounds of Flasdieck’s lack of sociolinguistic, contact linguistic, and geographical argument. The recently invigorated Celtic-OE hypothesis seems, then, to be gaining academic credibility. A review of Filppula et al., discussing the ‘received view’ that one should expect more lexical influence on OE for such a hypothesis to be true, notes the “shortcomings of such reasoning” (Stalmaszczyk 2005: 304). This historical argument that without lexical influence, one cannot demonstrate extended periods of language shift, has retained credibility only amongst traditional Germanic philologists such as Baugh and Cable, yet the standard references are not the only place that the old orthodoxy continues.

It is important to note the human factor behind the hypothesised language shift which would have given rise to the twofold paradigm in OE. This is the question of the scale of Anglo-Saxon settlement, and the degree of Celtic population movement and/or decline. In his study of Hamito-Semitic influence in Celtic, Hewitt (2009: 977) draws attention to another recent argument against the view (which has become an orthodoxy amongst proponents of the Celtic Hypothesis): that only a relatively small number of Anglo-Saxon settlers entered the British Isles. This is the crucial suggestion which would explain widespread presence of Celtic structural features due to language shift. Hewitt mentions a study by Capelli et al. (2003: 979) which adduces genetic evidence for the questionable magnitude of Anglo-Saxon incursion, and thus the possibility of mass adoption of Anglo-Saxon by the Celtic underclasses: “By analyzing 1772 Y chromosomes from 25 predominantly small urban locations, we found that different parts of the British Isles have sharply different paternal histories.”

Hewitt also describes Coates (2004) as casting doubt on the Celtic influence on Anglo-Saxon. However, in a more recent paper, Coates (2007: 16-17) undermines the anti-substrate argument. We first of all hear that “I know of no case where a political ascendancy has imposed its own language on a conquered people without a discernible impact from the language of the conquered”. This appears to ignore the ample structural evidence that has been summarised by, among others, McWhorter (2008: 163-164), who has reappraised his view of English in favour of the Celtic Hypothesis after examining what he calls the ‘modern revival school’ of the familiar names such as Klemola, Filppula, Hickey and Vennemann. Coates then goes on to describe the “exciting new work of Stephen Oppenheimer (2006)”, which “finds no support in the genetic profile of modern Britain for the idea of an incursion of Angles and Saxons in large numbers”. Although this point is mentioned in the context of a larger picture (namely the idea of a non-Celtic population being present in Britain hundreds of years before the first Anglo-Saxons arrived), it still seems to undermine the anti-language shift position. Following the above-mentioned research, we shall assume large-scale language shift over a period of centuries imposed

⁴ “The OE forms and functions of the root **bheu*, which are foreign to the other Germanic dialects, arose in the mouths and the minds of English-speaking Britons [my translation].”

on originally Celtic-speaking peoples by a new, relatively small dominant class of Germanic speakers.

3. Method

Having outlined the theoretical credibility of assuming a large Celtic population ripe for language shift, we move to the crux of this study. We know that the formal existence of the double paradigm of 'to be' was transferred into OE from early British (whether Welsh and Cornish had already become separate languages at the stage of the transfer is irrelevant, as they both preserve the distinction in their early forms). What we cannot be so sure of without further investigation is whether the *semantic* distinction, as discussed above, was preserved to the same extent in *different areas* of the Anglo-Saxon realm. The key to this investigation is that if we find the semantic distinction better preserved in the OE of certain areas of England, we can use this as evidence of greater Celtic populations in those areas. Put simply: if a majority of Celts were driven north or west early on, we expect to meet with semantic fuzziness in texts from the south and east: *b*-forms being used of a current, fleeting state of affairs, for example, contrary to the canonical use of *wes*-forms in this instance. Note that the aim is not to discover whether or not there were Celts in large numbers in England or not; this is assumed outright, as elaborated above. The aim is rather to find out whether certain areas of England had higher concentrations of those Celts, as reflected by the linguistic evidence.

Two facts make the methodological side of this investigation less than complicated. Firstly, the study can be seen as a simple exercise in novel interdisciplinarity, combining studies of early medieval population shift on the one hand and Old English dialectology on the other. As a corpus-based approach has not, to my knowledge, attempted to reconcile these two specific areas before, there is little danger of presenting unnecessary new theories or 'sexing up' unsatisfactory data. The second simplifying fact is that a trailblazing study was conducted by Matti Rissanen in 1992. Rissanen's execution and justification of corpus usage will be used as a blueprint for the present study. Rissanen's was one of the first studies to use a computerized corpus of Old English to look at a structural or syntactic feature: previous research making use of corpora had seen them as a convenient centralised repository of texts, without making the leap from manually reading through a relevant text to running automated searches on it. Wischer (2010), amongst other scholars, has capitalised on the model proposed by Rissanen in carrying out large-scale investigations of morphosyntactic features of OE, although her own treatment of the differences between *bēon* and *wesan* downplays dialectal differences to the point of irrelevance. It is hoped that the smaller sample of texts involved in this survey will give a more balanced and relevant view of the dialects, rather than including the whole gamut of West Saxon writings.

The corpus used for the research is the York-Toronto-Helsinki Parsed Corpus of Old English Prose (known at the time of Rissanen's paper as the Helsinki Corpus, and sometimes still referred to as such). The fact that this corpus was specifically designed for syntactic research and is therefore fully parsed makes it ideal to home in on the two paradigms of *bēon* and *wesan*. Other, larger corpora exist but were eschewed because they are not tagged. We do not need the full complement of OE texts because we expect some intradialectal homogeneity in usage of the

copulas: a representative sample, on the smaller side, of each dialect should suffice. Corpora of poetry were also jettisoned due to the associated risk of reflecting inaccurate usage of the two verbs: the danger of sacrificing faithfulness to actual language for the sake of metrics or alliteration. A variation of the CorpusSearch program designed to run under Windows was employed, developed by the curators of the Icelandic Parsed Historical Corpus (Wallenberg et al. 2011). This variation has the same functionality as the original.

The texts of the YTH corpus are sorted by chronology, author and, crucially, dialect. Chronology is relevant only insofar as later OE texts tended towards a normalised version of West Saxon, and so it seemed naturally to use the earlier texts available for each dialect, in accordance with standard diachronic practice. (This is also the reason behind the preponderance of texts marked ‘West Saxon’ in the corpus.) Although sorting by dialect removes the need manually to sift through texts, it also presents one with the stark reality of the imbalance of texts available for each dialect: there are only two extant poems in the Kentish dialect, for example. This is another reason why a larger corpus is not necessary. There are, for example, no representatives of Northumbrian prose in the corpus, and few Kentish samples. It thus makes little difference if we have 100 or 1000 West Saxon texts in the corpus, as long as the West Saxon sample is representative of a standard form and not a fringe form (i.e. not also showing features of other dialects). Fortunately, these mixed forms are also labelled as such by the corpus compilers (e.g. WS/A: West Saxon/Anglian), and will be taken as corresponding to geographical as well as linguistic ‘in-betweeners’, forming their own groups in the results. It is important to note that the combinations AM/K and K/AM/WS were not used, because on inspection they turned out to be from texts with separate sections in each dialect rather than displaying mixed dialect features.

Because the Kentish contingent of texts was so small, at only 1765 words, it was felt that representatives of the other dialects should not dwarf this sample, although they should be still be reasonably large, in order to increase the chances of coming across more varied forms of the copula. The aim was to take 2-3 texts from each dialect or dialect mixture with the sum total coming to around 10,000 words, although a deviation occurred in the case of West Saxon/Anglian Mercian due to the lack of suitably short texts. As well as Kentish, West Saxon, West Saxon/Anglian Mercian and West Saxon/Anglian, a sample of the unknown dialects was taken to act as a control group, since they were assumed to show the least marked dialectal features. The deviations in sample size were flattened by averaging the results out to the number of tokens that would occur in a 2000 word text, a round figure close to the size of the smallest sample. It was also deemed important to select only texts that were not translations from Latin originals, as far as this was possible. Early vernacular translations of texts written in classical languages tended to be over-faithful to the syntax of the originals, and syntactic effects of this nature would have been undesirable. This fact is well-established, especially for Germanic, notably thanks to the Gothic Bible translations of Ulfila: “Biblical Gothic was probably rather different from the everyday language since word-order, idiom and syntax followed the patterns of Ulfila’s Greek model” (Heather 1997: 178).

Some technical issues are involved in the search for *bēon* and *wesan*. Firstly, it is well-known that there was nothing like a standardised spelling for Old English. In fact, if we were to

see signs of a standard orthography emerging,⁵ we would have reason to be suspicious of the texts that are precisely the most useful to this investigation, in that they might not authentically represent the dialects in which they are supposed to be written. This requires awareness of the many spelling variations throughout the paradigms. <y> substitutes for <i> in many of the *wes*-forms; *eth* and *thorn* replace each other freely in the *b*-forms; <ae> and <a> jostle for position, for example in *waes* and *waeron* (the singular and plural preterites of the *wes*-forms). The first person singular of *bēon* alone is attested in four different spellings. If certain variants were to be overlooked it would be disastrous, particularly as they are better represented in certain dialects. A well-known example is the Northumbrian predilection for <o> in cases of short /a/ (although in this corpus that specific issue does not arise). This kind of orthographical uncertainty is an issue with any dead language. Every spelling variation was thus checked and listed before the search was carried out, and the results may be consulted in Appendix II (note that <+t> represents both <þ> and <ð>). Importantly, searches were run on present forms of the verb only, as preterites (*wæs* and the like) did not exhibit a twofold paradigm.

A second technical issue, also pressing but at a deeper level than the orthography, is the semantic value of ‘to be’ in the sentences. What we are investigating is ‘to be’ as copula, not as auxiliary, nor as an element of the ‘resolved tenses’, as Mitchell and Robinson (2001: 110) call them. They draw attention to the original nature of the past participle, or perfect, construction in OE. Whereas the participle nowadays has fully verbal status and forms with ‘have’ are true auxiliary constructions (forms with ‘to be’ also exist in archaic constructions such as ‘I am come’), the picture was somewhat different in OE. The participle here still carried its original, adjectival status. Pertejo (2003: 143), in a study using the same corpus but focussed on participial constructions, gives the following example:

- (1) þonne beo we sittende be þæm wege, swa se blinda **dyde**
 ‘Then we should be sitting at the way-side, as the blind man **did**’.

The *dyde* here helps us to ascertain that *sittende* is used in a present participial verbal construction (with *beo* as auxiliary) rather than as an adjectival participle, but this is a help rarely afforded us in the canon. It can often be difficult to distinguish between a pure copular usage of *bēon* or *wesan* and an auxiliary usage; for safety’s sake, participial constructions have therefore been omitted in this study. The fact that *weorðan*, a lexically and semantically separate verb that could nevertheless be used passively or in certain copular constructions, falls under the remit of BE (see Appendix II) in the corpus is not helpful, because it means that one cannot easily call up all copula instances of *bēon* and *wesan*. “All forms of BEON, WESAN, and (GE)WEOR+DAN are labelled with BE tags regardless of meaning” (Taylor 2003). Though beyond the scale of this work, it would be useful to develop a way of semantically tagging the differences between usage of *bēon* and *wesan* in the corpus. Minor issues such as *beo* (first person singular) also masquerading as a preposition, to wit a spelling variant of *be* (‘by’), were removed manually.

⁵ As we do in 10th and 11th century texts, especially poetry, though this cannot compare to a centralised, officially standardised language.

4. Discussion of Data⁶

Table 2 shows the average occurrences, per 2000-word text, for the paradigms of the two forms of ‘to be’ in the 5 dialectal groups. (Spellings have been standardized.)

	WS	WS/A	WS/AM	K	?
bēo					
bist			0.07		
biþ	0.92	10.9	17.79	21.53	13.39
bēoþ		8.68	0.42		0.62
beonde (pp.)					
beon (inf.)	1.1	0.67	1.2	2.27	2.8
beo (subj.)	1.29	0.89	0.07	3.4	
beo (BEPH) ⁷		0.22	0.14		
beo (BEP)	0.18				0.31
beon (subj.)	0.74	0.22	0.14	4.53	
ēom			0.42		0.93
eart			0.35		1.24
is	13.84	12.9	11.25	27.2	12.45
sindon	0.55	0.44			0.93
wesende (pp.)					
wesan (inf.)		0.22			
si (subj.)	0.37	1.11	0.14	3.4	2.18
sin (subj.)			0.35	1.13	2.18
Word Count in Samples	10836	8990	28440	1765	6426

Table 2: Results of textual analysis

⁶ Abbreviations used in Table 2: BEPS: present subjunctive, BEPH: ambiguous subjunctive/imperative, BEP: ambiguous subjunctive/indicative, BE: infinitive, BEPI: present indicative, inf: infinitive, K: Kentish, pp: present participle, subj: subjunctive, WS: West Saxon, WS/A: West Saxon/Anglian, WS/AM: West Saxon/Anglian-Mercian

⁷ Ambiguous. See Appendix II.

It was important to form a clear picture of the relative population levels of Celtic speakers in each of the dialect areas analysed. A combination of traditional theory, largely based on the mapping-out of historical British population movement presented in Jackson 1953, and more recent genetic theory drawing on the work of Oppenheimer (see above), with some consideration of perspectives from archaeology (Arnold 1997), was employed to this end. The word ‘relative’ in ‘relative population levels’ must be borne in mind. This is due to the fact that, as noted above, there is no consensus amongst scholars as to whether large Celtic populations ‘stayed behind’ at all, let alone what their specific numbers were. We are able, however, to sketch out a hierarchy of relative Celtic presence in each of our dialect areas.

Evidence agrees that we expect to find a greater and earlier Germanic concentration in Anglian than in Mercian areas (cf. Weale et al. 2002:1012), and that Kent played host to the earliest and greatest amount of Celtic displacement (e.g. Jackson 1953: 208ff.). Given the proximity of the West Saxon area to Wales, which retained large Celtic populations and did not see large-scale Germanic dominance until after Kent and Anglia, but before Mercia, we may set up a hierarchy of Kent>Anglia>Wessex>Mercia. Our hierarchical labelling of textual provenance, *mutatis mutandis*, will K>WS/A>WS>WS/AM. Although the last dialect mixture shows features from three other dialects, its place at the end of the hierarchy is appropriate, as two of the three were spoken in heavily Celtic areas.

Interestingly, Kentish employment of *b-* and *wes-* subjunctive forms is equally high, whereas in other dialects one form tends to be better represented than the other. This may suggest a lack of differentiation between the two in Kentish. There is a notable difference in the distribution of third person singular forms: whereas Kentish seems to heavily favour *wes*-forms, WS/AM conversely shows many more *b*-forms. As the *wes*-form was the one which persisted into PrE, this suggests a lack of rigorous distinction between the two forms in less Celtic areas and early obsolescence of the *b*-forms in this position. The WS/A evidence for this form also supports this supposition, with relatively balanced attestation.

A notable phenomenon is the presence of tokens marked BEP/BEPH, which denote ambiguous forms. While these were attested in WS, WS/A and WS/AM, they are absent in Kentish. This may be coincidental, but it seems reasonable to suppose that it points to an early preference for the *wes*-forms, avoiding the semantic nuances of the *b*-forms which would have been introduced by Celtic language shifters on the model of their own lexicon. One might hypothesise, on the basis of this Kentish trend, that the *wes*-forms were moving towards domination at the time of the migrations; it was the Celtic areas which introduced and/or preserved a semantic distinction. Although this distinction was eventually lost as well, the forms that became standard were in some cases the *b*-forms, as is well-known from modern dialects of Northern England, East Anglia, etc; this does not occur in Standard English.

5. Conclusion

While the methods of this analysis seemed sound and likely to produce interesting results, and did so to an extent, the textual samples were not always sufficient in quality or quantity for the questions to be answered. One issue was the lack of ‘fringe’ dialects in the corpus: there was a lack of Northumbrian evidence, and so swathes of potentially interesting forms exhibiting Celtic influence were absent. There was also a notable dearth of data from Kentish, which was regrettable, given that this dialect was where most of the starting-points for interesting trends (read: lack of Celtic substrate influence) came from.

We may nevertheless conjecture, based on the restricted analysis that these data allowed, that some form of the original hypothesis would hold. More specifically, we posit that Kentish dialects would have shown less careful distinction between the semantics of *b-* and *wes-* forms, and that Mercian and possibly West Saxon dialects would have shown less mixing between the two paradigms. A larger textual sample would be necessary to demonstrate this, however. The decision to use only samples from the corpus in the first place appears a reasonable one; some interesting inferences could be made using a Kentish sample 5 times smaller than the 10,000-word target. The inescapable conclusion is that diversity of dialects rather than sample size was an issue here, although a search over the full range of OE texts would no doubt produce more reliable results.

It is also worth noting that no amount of automated corpus searching can overcome the setback noted above, that of the three possible translations of ‘to be’, both copular and auxiliary/participial. The lack of semantic information contained in the corpus unquestionably obscured some of the results. It is hoped, however, that the Celtic Hypothesis will remain a viable avenue for research and that this survey was justified in its aim of using the semantics of Old English to prove something fundamental about the populations that make up the whole history of the British Isles, namely where language shift occurred (i.e. everywhere except the Southeast and East), where wholesale replacement occurred (i.e. the Southeast and East), and the nature of the continuum of variation between the two.

References

- Arnold, Chris. *An Archaeology of the Early Anglo-Saxon Kingdoms*, 2nd edition. London: Routledge.
- Baugh, Albert C. & Thomas Cable. 1993. *A History of the English Language*, 4th edition. London: Routledge.
- Capelli, Cristian, Nicola Redhead, Julia K. Abernethy, et al. 2003. A Y chromosome census of the British Isles. *Current Biology* 13, pp. 979–84.
- Coates, Richard. 2004. *Invisible Britons: the view from linguistics*. Paper read at the conference ‘Britons and Angles in Anglo-Saxon England’, Manchester, UK, 14–16 April 2004.
- Coates, Richard. 2007. *Linguistic light on the birth of England*. Inaugural professorial lecture: University of the West of England, Bristol, UK.
- Filppula, Markku, Juhani Klemola & Heli Paulasto. 2008. *English and Celtic in contact*. London: Routledge.
- Flasdieck, Hermann M. 1937. *Das altgermanische Verbum Substantivum unter besonderer Berücksichtigung des Altenglischen*. *Englische Studien* 71, pp. 321–49; 1938. *Nachtrag*. *Englische Studien* 72, pp. 158–60.
- Heather, Peter. 1997. *Literacy and power in the migration period*. In: Bowman, A. K & G. Woolf (eds.), *Literacy and Power in the Ancient World*. Cambridge: Cambridge University Press, pp. 177–197.
- Hewitt, Steve. 2009. *The Question of a Hamito-Semitic Substratum in Insular Celtic*. In: *Language and Linguistics Compass* 3/4, pp. 972–995.
- Jackson, Kenneth. 1953. *Language and History in Early Britain*. Edinburgh: Edinburgh University Press.
- Keller, Wolfgang. 1925. *Keltisches im englischen Verbum*. In: *Anglica: Untersuchungen zur englischen Philologie*, vol. 1. Leipzig: Mayer and Müller, pp. 55–66.
- Lutz, Angelika. 2008. *Celtic Influence in Old English and West Germanic*. *English Language and Linguistics* 13.2, pp. 227–249.
- McWhorter, John. 2009. *What else happened to English? A brief for the Celtic hypothesis*. *English Language and Linguistics* 13.2, pp. 163–191.
- Mitchell, Bruce & Fred C. Robinson. 2003. *A Guide to Old English*. Oxford: Blackwell.
- Oppenheimer, Stephen. 2006. *The Origins of the British: A Genetic Detective Story*. London: Constable and Robinson.
- Penny, Ralph. 2002. *A History of the Spanish Language*. Cambridge: Cambridge University Press.

- Pertejo, Paloma Núñez. 2003. Adjectival participles or present participles? On the classification of some dubious examples from the Helsinki Corpus. *Sederi* 13, pp. 141-153.
- Ringe, Don. 2006. *From Proto-Indo-European to Proto-Germanic*. Cambridge: Cambridge University Press.
- Rissanen, Matti. 1992. Computers are useful – for aught I know. In: Colman, Fran (ed.), *Evidence for Old English*, *Edinburgh Studies in the English Language* 2, pp. 155-168.
- Stalmaszczyk, Piotr. 2005. Review of Markku Filppula, Juhani Klemola & Heli Pitkänen (eds.) 2002. *The Celtic Roots of English*. *Journal of Sociolinguistics* 9/2, pp. 304-7.
- Taylor, Ann. 2003. *York-Toronto-Helsinki Parsed Corpus of Old English Prose Part-of-speech Reference Manual*. Available at: <http://wwwusers.york.ac.uk/~lang22/YCOE/doc/annotation/YcoePos.htm>. Accessed 27/04/2011.
- Thurneysen, Rudolf. 1980. *A Grammar of Old Irish*. Dublin: The Dublin Institute for Advanced Studies.
- Trask, R.L, 1995. Origin and relatives of the Basque language: review of the evidence. In: Hualde, J.I., et al. (Eds.), *Towards a History of the Basque Language*. Amsterdam: John Benjamins, pp. 65–77.
- Vennemann, Theo. 2010. Contact and Prehistory: The Indo-European Northwest. In: Raymond Hickey (ed.), *The Handbook of Language Contact*. Chichester: Wiley-Blackwell, pp. 380-405.
- Wallenberg, J, A. Ingason, E Sigurdsson & E. Rögnvaldsson. 2011. *Icelandic Parsed Historical Corpus (IcePaHC)*. Version 0.4. < http://www.linguist.is/icelandic_treebank>. Accessed 27/04/2011.
- Weale, Michael, Deborah Weiss et al. 2002. Y-chromosome evidence for Anglo-Saxon mass migration. *Molecular Biology and Evolution* 19(7), pp. 1008-1021.
- Wischer, Ilse. 2010. On the use of *beon* and *wesan* in Old English. In: Lenker, U, J. Huber & R. Mailhammer (eds.), *English Historical Linguistics 2008*. Amsterdam: John Benjamins, pp. 217-236.

Appendix I: OE Texts that form the corpus, Helsinki Corpus filenames, dialects and word counts

Key to dialects

A: Anglian

AM: Anglian Mercian

K: Kentish

WS: West Saxon

?: Dialect not given

Filename	Name of Text	Dialect	Word Count
coadrian.o34	Adrian and Ritheus	WS	1092
coaelhom.o3	Ælfric, Supplemental Homilies	WS	62669
coaelive.o3	Ælfric's Lives of Saints	WS	100193
coalcuin	Alcuin De virtutibus et vitiis	?	5549
coalex.o23	Alexander's Letter to Aristotle	WS/A	7271
coapollo.o3	Apollonius of Tyre	WS	6545
coaugust	Augustine	?	103
cobede.o2	Bede's History of the English Church	WS/A	80767
cobenrul.o3	Benedictine Rule	WS	20104
coblick.o23	Blickling Homilies	WS/A	42506
coboeth.o2	Boethius' Consolation of Philosophy	WS	48443
cobyrtf.o3	Byrhtferth's Manual	WS	10243
cocanedgD	Canons of Edgar (D)	?	1765
cocanedgX	Canons of Edgar (X)	WS	2118
cocathom1.o3	Ælfric's Catholic Homilies I	WS	106173
cocathom2.o3	Ælfric's Catholic Homilies II	WS	98583
cochad.o24	Saint Chad	WS/AM	2659
cochdrul	Chrodegang of Metz, Rule	?	18386
cochristoph	Saint Christopher	?	1426

cochronA.o23	Anglo-Saxon Chronicle A	WS	14583
cochronC	Anglo-Saxon Chronicle C	?	22463
cochronD	Anglo-Saxon Chronicle D	?	26691
cochronE.o34	Anglo-Saxon Chronicle E	WS	40641
cocura.o2	Cura Pastoralis	WS	68556
cocuraC	Cura Pastoralis (Cotton)	?	2119
codicts.o34	Dicts of Cato	WS	2180
coducu1.o1	Documents 1 (O1)	AM/K	1753
coducu2.o12	Documents 2 (O1/O2)	K	253
coducu2.o2	Documents 2 (O2)	K/AM/WS	1857
coducu3.o23	Documents 3 (O2/O3)	AM	679
coducu3.o3	Documents 3 (O3)	WS	7171
coducu4.o24	Documents 4 (O2/O4)	AM	193
coeluc1	Honorius of Autun, Elucidarium 1	K	1512
coeluc2	Honorius of Autun, Elucidarium 1	?	583
coepigen.o3	Ælfric's Epilogue to Genesis	WS	965
coeuphr	Saint Euphrosyne	WS	3658
coeust	Saint Eustace and his companions	WS	5271
coexodusP	Exodus (P)	WS	1096
cogenesisC	Genesis (C)	WS	5224
cogregdC.o24	Gregory's Dialogues (C)	WS/AM	91553
cogregdH.o23	Gregory's Dialogues (H)	WS	25593
coherbar	Pseudo-Apuleius, Herbarium	WS/A	22213
coinspolD.o34	Wulfstan's Institute of Polity (D)	WS	2530
coinspolX	Wulfstan's Institute of Polity (X)	?	4896
cojames	Saint James	?	1659

colacnu.o23	Lacnunga	WS/A	7099
colaece.o2	Leechdoms	WS/A	34727
colaw1cn.o3	Laws, Cnut I	WS	2386
colaw2cn.o3	Laws, Cnut II	WS	4761
colaw5atr.o3	Laws, Æthelred V	WS	1228
colaw6atr.o3	Laws, Æthelred VI	WS	2096
colawaf.o2	Laws, Alfred	WS	3314
colawafint.o2	Alfred's Introduction to Laws	WS	1966
colawger.o34	Laws, Gerefa	WS	751
colawine.ox2	Laws, Ine	WS	2755
colawnorthu.o3	Northumbra Preosta Lagu	WS	1330
colawwllad.o4	Laws, William I, Lad	WS	220
coleofri.o4	Leofric	WS	1017
colsigef.o3	Ælfric's Letter to Sigefyrth	WS	1648
colsigewB	Ælfric's Letter to Sigeward (B)	WS	3665
colsigewZ.o34	Ælfric's Letter to Sigeward (Z)	WS	10420
colwgeat	Ælfric's Letter to Wulfgeat	WS	2460
colwsigeT	Ælfric's Letter to Wulfsige (T)	WS	319
colwsigeXa.o34	Ælfric's Letter to Wulfsige (Xa)	WS	3336
colwstan1.o3	Ælfric's Letter to Wulfstan I	WS	4544
colwstan2.o3	Ælfric's Letter to Wulfstan II	WS	4036
comargaC.o34	Saint Margaret (C)	?	4196
comargaT	Saint Margaret (T)	?	3661
comart1	Martyrology, I	?	1300
comart2	Martyrology, II	?	4391
comart3.o23	Martyrology, III	WS/AM	25781
comarvel.o23	Marvels of the East	WS/A	1891
comary	Mary of Egypt	WS	8181
coneot	Saint Neot	?	2003
conicodA	Gospel of Nicodemus	?	8197

	(A)		
conicodC	Gospel of Nicodemus (C)	?	4629
conicodD	Gospel of Nicodemus (D)	?	1798
conicodE	Gospel of Nicodemus (E)	?	1588
coorosiu.o2	Orosius	WS	51020
cootest.o3	Heptateuch	WS	59524
coprefcath1.o3	Ælfric's Preface to Catholic Homilies I	WS	1035
coprefcath2.o3	Ælfric's Preface to Catholic Homilies II	WS	223
coprefcura.o2	Preface to the Cura Pastoralis	WS	831
coprefgen.o3	Ælfric's Preface to Genesis	WS	1399
copreflives.o3	Ælfric's Preface to Lives of Saints	WS	373
coprefsolilo	Preface to Augustine's Soliloquies	?	441
coquadru.o23	Pseudo-Apuleius, Medicina de quadrupedibus	WS/A	4276
corood	History of the Holy Rood-Tree	?	6920
cosevensl	Seven Sleepers	WS	9143
cosolilo	St. Augustine's Soliloquies	WS	15856
cosolsat1.ox4	Solomon and Saturn I	WS	2046
cosolsat2	Solomon and Saturn II	?	1235
cotempo.o3	Ælfric's De Temporibus Anni	WS	5495
coverhom	Vercelli Homilies	WS	45674
coverhomE	Vercelli Homilies (E)	?	4463
coverhomL	Vercelli Homilies (L)	?	1986
covinceB	Saint Vincent (Bodley 343)	?	728
covinsal	Vindicta Salvatoris	?	3655
cowsgosp.o3	West-Saxon Gospels	WS	71104

cowulf.o34	Wulfstan's Homilies	?	28768
------------	---------------------	---	-------

Appendix II: Material used to represent each dialect, and list of searches conducted on this material

WS: coducu3.o3, colsigewB: 10836 words total (not Latin translations)

WS/A: colacnu.o23, comarvel.o23: 8990 words total (Latin translations)

WS/AM: comart3.023, cochad.o24: 28440 words total (Latin translations)

K: coducu2.o12, coeluc1: 1765 words total (not Latin translations)

?: cosolsat2, covinceB, coverhomE: 6426 words total (not Latin translations)

BEPS: present subjunctive

BEPH: ambiguous subjunctive/imperative

BEP: ambiguous subjunctive/indicative

BE: infinitive

BEPI: present indicative

Search object	Info	WS	WS/A	WS/AM	K	?
bio	1sg.ind/subj.	2 BEPS				
beo	“	5 BEPS, 1 BEP	4 BEPS, 1 BEPH	1 BEPS, 2 BEPH	3 BEPS	1 BEP
biom	1sg.ind.					
beom	“					
bist	2sg.			1 BEPI		
byst	“					
bis	“					
bi+t	3sg.	5 BEPI	40 BEPI	245 BEPI		43 BEPI
by+t	“		9 BEPI	8 BEPI	19 BEPI	
bio+t	pl.		1 BEPI			
beo+t	“		38 BEPI	6 BEPI		2 BEPI
bi+ton	“					
bio+ton	“					
bi+tun	“					
bio+tun	“					
beonde	pres.part.					
bionde	“					
beon	inf./pl.subj.	6 BE, 4	3 BE, 1	17 BE, 2	2 BE, 4	9 BE

		BEPS	BEPS	BEPS	BEPS	
bion	“					
eom	1sg.			6 BEPI		3 BEPI
eam	“					
am	“					
eart	2sg.			5 BEPI		4 BEPI
ear+t	“					
ar+t	“					
is	3sg.	75 BEPI	58 BEPI	160 BEPI	24 BEPI	40 BEPI
sind	pl.	2 BEPI				
sint	“	1 BEPI				1 BEPI
sindon	“		2 BEPI			2 BEPI
sindun	“					
aron	“					
earon	“					
arun	“					
sie	sg.subj.		3 BEPS	1 BEPS	3 BEPS	7 BEPS
si	“	2 BEPS	2 BEPS	1 BEPS		
sien	pl.subj.				1 BEPS	7 BEPS
sin	“			5 BEPS		
wesende	pres.part.					
wesan	inf.		1 BE			