CrossMark

English Language and Linguistics 21.3: 439–473. © Cambridge University Press 2016
doi:10.1017/S1360674316000204

Null subjects in Middle English

GEORGE WALKDEN
University of Manchester

and

KRISTIAN A. RUSTEN
Bergen University College

(Received 14 September 2015; revised 23 May 2016)

This article investigates the occurrence and distribution of referential null subjects in Middle English. Whereas Modern English is the textbook example of a non-null-subject language, the case has recently been made that Old English permits null subjects to a limited extent, which raises the question of what happens in the middle period. In this article we investigate Middle English using data drawn from the *Penn–Helsinki Parsed Corpus of Middle English Prose* and the new *Parsed Corpus of Middle English Poetry*, aiming to shed light on the linguistic and extralinguistic factors conditioning the alternation between null and overt subjects. Generalized mixed-effects logistic regression and random forests are used to assess the importance of the variables included. We show that the set of factors at play is similar to that found for Old English, and we document a near-complete disappearance of the null subject option by the end of the Middle English period.

1 Introduction

This article investigates the occurrence and distribution of referential null subjects in Middle English (ME). As is well known, Modern English is a non-null-subject language which in its standard varieties only allows the omission of a referential subject in 'topic drop' contexts, i.e. in main clauses which have linear verb-initial word order. In contrast, Old English (OE), the oldest documented stage of the English language, has been claimed to have been a null-subject language to one extent or another, although the issue is to some degree still controversial. While the occurrence of null subjects in OE has received increased quantitative attention in recent years, very little scholarly work has concentrated on the status of the null-subject phenomenon in ME. This is the issue we aim to address. In this introduction we outline the research context for the article and the specific questions we will be asking. Section 2 presents our methodology, and section 3 our results. Section 4 provides some discussion, and section 5 summarizes and concludes.

[1] We are grateful to audiences at ICEHL 17 (Zurich, 2012) and ICEHL 18 (Leuven, 2014), for encouraging us to pursue this research and for their thoughts on the results. Special thanks go to Elly van Gelderen, Laura Percival and Daniel Ezra Johnson. Thanks also to editor Laurel Brinton and the anonymous reviewers for *English Language and Linguistics*, who have made this a better article.

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

1.1 The research context: Old English

The possibility of referential null subjects in Old English (OE) has received a lot of attention in recent years (Rusten 2010, 2013, 2014, 2015; Coppess & Pires 2013; van Gelderen 2013; Walkden 2013, 2014, 2016). In particular, new quantitative corpus-based work has made it possible to evaluate in more detail the claims of earlier work (Pogatscher 1901; Ohlander 1943; Mitchell 1985: I, 633; van Gelderen 2000: 121–49) that null subjects were a feature of this language. A number of generalizations emerge from this work.

First, referential null subjects are a low-frequency phenomenon in all texts of all genres. In some texts, they are vanishingly rare, or even absent entirely. This is true of the Alfredian West Saxon texts as well as the works of Ælfric and Wulfstan, which together make up a large proportion of the attested OE prose corpus: this makes it understandable that Hulk & van Kemenade (1995: 245) start from the assumption that referential null subjects were not a grammatical possibility in OE. Even in the texts that permit referential null subjects more robustly, the overall relative frequency of null subjects compared to subject pronouns is never above 25 per cent (see the tables in Walkden 2013: 159–61 and Rusten 2015: 59–60).[2] The relative frequency of referential null subjects is also much higher in poetry than it is in prose (Rusten 2015: 66–7), suggesting that metrical requirements may have played a role by favouring the insertion or (more usually) deletion of an unstressed monosyllable.

Secondly, there is a strong, though not categorical, effect of clause type: in subordinate clauses, referential null subjects are less frequent than in main clauses (Walkden 2013: 163–4; Rusten 2015: 67–9).

Thirdly, person also plays a role: third-person subjects are more likely to be null than first- or second-person subjects (Berndt 1956: 65–8; Walkden 2013: 164–6; Rusten 2015: 69–71). Number (singular vs plural) does not have a consistent effect.

The effects of the linguistic variables clause type and person, in particular, suggest that – at least for some authors – referential null subjects were a native phenomenon in OE, albeit one that was dwindling in frequency and productivity: Rusten (2015: 60–2) finds that later texts feature lower frequencies of referential null subjects than earlier ones.

A further factor claimed to influence null subjects in OE is region of origin (Walkden 2013: 162–3; 2014: 183–4): Walkden suggests that Anglian or Anglian-influenced texts allow referential null subjects, but that 'pure' West Saxon texts disallow them. Rusten (2015: 65–6) problematizes this hypothesis on the grounds that the few texts which do exhibit null subjects to any meaningful degree may do so for other other reasons, e.g. by virtue of their genre or translation status. Moreover, a number of uncontroversially Anglian-influenced texts, such as the OE *Herbarium*, do not display

[2] The Old Northumbrian gloss to the Latin Lindisfarne Gospels is an exception to this, but here the status of the text as a glossal translation (Nagucka 1997) is likely to have artificially inflated the frequency of referential null subjects. See Berndt (1956) and Walkden (2016) for detailed discussion of null subjects in this text, and Rusten (2015: 63–5) on the effect of translation status more generally.

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

referential null subjects. Given the limited nature of the OE corpus, which is mostly dialectally homogeneous, it is probably not possible to draw a firm conclusion either way.

Importantly, though referential null subjects are rarer in later OE than they are in earlier texts, they have not completely vanished by the end of the period. The late OE poems investigated by Rusten (2015: 61–2) display referential null subjects at an overall average frequency of 7.5 per cent, and for one late prose text, the *Benedictine Rule*, the frequency is 5.7 per cent. This suggests that we might be able to witness the demise of the referential null subject by investigating the Middle English corpus.

1.2 The research context: Middle English

Studies of referential null subjects for ME are altogether fewer and further between than for OE.[3] Einenkel (1916), Visser (1963: 4ff) and Ohlander (1981) present a number of examples, the latter categorized according to the nature of the antecedent. Mustanoja (1960: 138–44) also discusses subject omission, including relative frequency, but does not provide quantitative evidence, and does not systematically distinguish between true referential null subjects and conjunction reduction of the kind that is grammatical in Present-day English (e.g. *The king went to Norway and* ***Ø*** *met the bishop*). Small-scale quantitative studies have been carried out by van Gelderen (2000) and Rusten (2014).

The corpus used by van Gelderen (2000) is a selection of three texts from an early version of the *Penn–Helsinki Parsed Corpus of Middle English* (PPCME; Kroch & Taylor 2000). She concludes that referential null subjects are 'highly exceptional' in Layamon's *Brut* and the *History of the Holy Rood Tree*, but 'quite frequent' in texts of the Katherine Group. She finds that, as in OE, there is a person asymmetry in favour of the third person, and also that in later ME, from the fourteenth century onwards, null referential subjects are very rare (2000: 146). Two of her examples (2000: 143) are given below as (1) and (2).[4]

(1) þa he hefde þus idon. **Ø** sende hire þus to seggen hire wil he hefde iwraht.
when he had thus done **Ø** sent her thus to say her will he had worked
'When he had done so, **he** sent to her to tell her that he had done what she wanted.'
(Katherine Group, *St Juliana*; CMJULIA,97.28)

(2) Ah **Ø** beoð se cleane ouercumen
but **Ø** are so clean.ADV overcome
'but **they** are so completely overcome'
(Katherine Group, *St Margaret*; CMMARGA,80.397)

[3] The occurrence of *non*-referential null subjects in ME has attracted a little more attention. Williams (2000) looks at existential sentences in which Present-day English would employ an expletive *there*, and relates the loss of such sentences to the loss of verb-initial declarative clauses. Zimmermann (2012) looks at the increasing use of expletives with clausal associates (both finite *that*-clauses and non-finite clauses) in OE and ME, relating it to the general rise of a canonical preverbal subject position. In this study we have taken steps to exclude non-referential subjects from consideration.

[4] In our examples the symbol **Ø** indicates an omitted subject. This notation is used purely to ease interpretation, and is not intended to suggest an analysis of null subjects as involving a particular type of empty category, nor as a claim about the structural or linear position of such a category.

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

Rusten (2014) addresses ME in passing as part of a more general overview of null subjects in the history of English. He investigates the frequency of referential null subjects in ten prose texts from the PPCME2. For his whole sample, the average relative frequency is 0.8 per cent, with *St Katherine* out in front at 3.4 per cent. On this basis, he contests van Gelderen's claim that referential null subjects were 'quite frequent' in ME texts of the Katherine Group.

From the above survey it is clear that there is scope for a more wide-ranging study of referential null subjects in ME. Rusten (2014) and van Gelderen (2000) look at only a small number of texts, all prose. Moreover, some of the effects found to condition subject omission in OE have not been investigated: we do not know whether referential null subjects in ME are conditioned by genre/text type or by clause type. We also do not know whether region of origin plays a role. We saw above that a dialect-split hypothesis is difficult to assess for OE, but it should be much less so for ME, which has been described as 'par excellence, the dialectal phase of English' (Strang 1970: 224). Syntax has traditionally been considered 'the Cinderella of Middle English dialectology' (McIntosh *et al.* 1986: 32; see also Fischer *et al.* 2000: 69), and so investigating the dialectal origins of texts which display referential null subjects is a step towards remedying this situation. In any event, these are the holes which the present study aims to fill.

2 Methodology

The work presented in this article is corpus-based and essentially quantitatively oriented. It is also exploratory in the sense that no large-scale quantitative investigation of null subjects in ME has been carried out to date. We aim to provide a substantial contrastive survey of the occurrence and distribution of overt and null referential pronominal subjects in ME, and also to analyse the data according to the variables mentioned in section 1.2 using statistical methods. Our primary technique of analysis is generalized mixed-effects logistic regression modelling. This powerful technique has become increasingly commonly used in linguistics, and through its use we aim to be able to identify and statistically assess the relative role of various linguistic and extra-linguistic variables in allowing null subjects in ME. We supplement the regression analysis by running a random forest of conditional inference trees, with variable importance assessed by using the conditional permutation accuracy importance measure proposed by Strobl *et al.* (2008). We also make restricted use of the chi-square test, with Yates' continuity correction, and Cramer's V.

Section 2.1 briefly explains the process of data collection, while section 2.2 gives additional details on the statistical analysis.

2.1 Data collection

The data presented and analysed in this article were collected by pulling all pronominal subjects, overt and null, from two ME corpora covering both prose and verse material.

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

Searches were restricted to finite clauses only. The data material forming the empirical basis for this article includes a total of 58,024 pronominal subjects drawn from 92 texts of ME prose and poetry. As noted in footnote 3, we exclusively address referential null subjects in this article. Consequently, searches for null subjects were restricted to those empty subjects labelled **pro** by the corpus annotators. We have thus excluded from consideration null non-referential subjects (labelled **exp**) and subjects elided under coordination (labelled **con**). As concerns overt subject pronouns, we strove to eliminate the overt counterparts of null non-referential subjects from the dataset, meaning that structures such as (3) do not feature in our data. Note, however, that it is not practically possible to exclude all instances of overt non-referential *hit*/*it* from the corpus searches.[5] Overt existential/presentative uses of *there* (4) are not tagged as pronouns and do not feature in our data.

(3) Is **hit** a gret woundour þat þu hast forsake styngyngge lust of body for þe
is **it** a great wonder that you have forsaken stinging lust of body for the
swete sauour of maydenhood?
sweet savour of maidenhood
'Is **it** a great wonder that you have forsaken the stinging lust of the body for the sweet savour of maidenhood? '(Aelred of Rievaulx's *De Institutione Inclusarum* (MS Vernon) (CMAELR3,33.196))

(4) And **þere** is so meche multytude of þat folk þat þei ben
and **there** is so great multitude of that folk that they are
with outen nombre
with out number
'And there is such great a multitude of that people that they cannot be counted.'
(*Mandeville's travels*, CMMANDEV,42.1053)

The prose data were extracted from the second edition of the *Penn–Helsinki Parsed Corpus of Middle English* (PPCME2) using the *CorpusSearch* 2 program (Randall *et al.* 2005–13). The PPCME2 is a treebank consisting of syntactically and morphologically parsed material from 56 ME prose texts composed between *c*. 1150 and 1500, and it comprises *c*. 1.2 million words of running text. The prose data from the PPCME2 were complemented by verse data collected by means of an exhaustive search of a pre-final version of the *Parsed Corpus of Middle English Poetry* (PCMEP; Zimmermann 2015).[6] The PCMEP is a treebank containing a substantial sample of

[5] As an example, it may be mentioned that overt expletive *it* in ambient sentences is annotationally indistinguishable from referential *it*. As a consequence, examples such as *it ne reyneth not in þat contree* 'it does not rain in that country' (*Mandeville's Travels* (CMMANDEV,29.696)) are counted among the overt subject pronouns in our data. This introduces a degree of inaccuracy. However, when we built a lexicon of all the verbs co-occurring with the pronoun *hit*/*it* in our data, we found very few tokens of common weather verbs. For example, we found only three tokens of the form *reyneth* 'rains' in clauses with *hit*/*it*, and we found no examples of the verbs *blouen* 'blow' or *snouen* 'snow' in these clauses. Thus, while there is some potential for error here, due to the high total number of overt pronouns under investigation, we are confident that the inclusion of these structures does not exert undue influence on the results.

[6] The version of the PCMEP forming the basis for the poetic component of the present study was the most recent version as of June 2015.

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

ME verse texts composed between *c*. 1120 and 1402, and the version we use includes 36 ME poems comprising a total of 88,521 parsed words. The inclusion of verse data allows us to investigate potential genre differences in the non-expression of referential subjects in ME, but there are additional advantages to incorporating poetic data in the study. Notably, the PPCME2 displays a distinct lack of prose material covering the period *c*. 1250–1350. The data from the PCMEP allow us to offset this skewedness somewhat, since this corpus contains several poetic texts from the same period (as also noted by Zimmermann 2013). Secondly, the PPCME2 is somewhat skewed in terms of the dialectal distribution of the texts it contains: the majority of them are from the Midlands, and there is a particular lack of early prose material written in Southern and Northern dialects of ME. The PCMEP data help address this issue inasmuch as they include several early texts of southern origin (e.g. *The Grave*, *Body and Soul*, *The First Worcester Fragment*, *Poema Morale*). The lack of Northern texts does admittedly remain a problem, however.

2.2 Statistical analysis

Once the overt and null pronominal subjects had been extracted from the corpora, the data in the corpus output files were collated into dataframe format fit for statistical analysis. Each pronominal token in the dataset was subsequently enriched with information on the variables mentioned above. That is, all overt and null subjects in the dataframe were specified according to their person and number features and according to the clause type in which they occur. We also added information on the genre, period and dialect of the texts from which the tokens were drawn. This enabled us to carry out a statistical analysis which essentially is rooted in variable rule sociolinguistic research, as pioneered by Cedergren & Sankoff (1974).

The variable rule program – a generalized regression implementation capable of modelling the joint effect of a number of independent factors on a linguistic variable – has been 'employed successfully for over three decades',[7] as pointed out by Johnson (2009: 359). However, it is a significant shortcoming of this program that generalized linear models are incapable of accounting for variation which may be contingent upon membership in a specific group (see e.g. Jenset 2010: 99). For our purposes, then, each ME text can be viewed as a 'group' to which a given population of overt and null pronouns belongs. The 92 individual texts, or 'groups', can furthermore be viewed as being a source of variation in themselves. The immediate sociolinguistic parallel is the individual speaker, who may idiosyncratically exhibit a given pattern of variation with respect to a linguistic variable regardless of factors such as gender, age, class, etc. (see also e.g. Johnson 2009). The individual speaker/group/text is thus a source of variation which needs to be accounted for statistically in the same way as other factors. This can only be done by modelling the individual/text/group as a *random effect*. Random

[7] Modern versions of the variable rule program include Goldvarb X (Sankoff *et al.* 2005) and Goldvarb Lion (Sankoff *et al.* 2012).

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

effects are variables which do not exhaust the possible range of variation in the sample, whereas *fixed effects* do. Hence, NUMBER is a fixed effect, since it has only two possible levels in ME: singular and plural. TEXT is a random effect since the potential levels of this variable include more texts than those investigated in our study: the levels of this variable 'constitute only a subset of the possible categories available in the population' (Tagliamonte & Baayen 2012: 143). If a random effect is analysed statistically as if it were a fixed effect, as pointed out by Tagliamonte & Baayen (2012: 143), 'the conclusions reached will only be valid for the individuals ... sampled'. That is, 'the model does not generalize to the population of subjects and items' (Baayen 2008: 241).

As a consequence of this, this article will present the results of a generalized mixed-effects logistic regression model, computed using Rbrul (Johnson 2009), which takes into account both random and fixed effects. Thus, we model the binary response variable NULL as a function of one random effect and various fixed-effect variables. In all, we distinguish seven variables – or *covariates* – of which three are linguistic (TYPE, PERSON, NUMBER) and four are extra-linguistic (TEXT, GENRE, PERIOD, DIALECT). The variables of TEXT, PERSON and NUMBER should be self-explanatory. TYPE distinguishes between non-conjunct main, second (and subsequent) conjunct main and subordinate clauses. Following the compilers of the *Helsinki Corpus* (Rissanen *et al.* 1991), we make a distinction between four ME periods labelled 'M1' (1150–1250), 'M2' (1250–1350), 'M3' (1350–1420) and 'M4' (1420–1500). In cases where the date of composition of a text is unknown, we have relied on the manuscript date; in other cases we have classified texts according to a known period of composition. We distinguish between two genres – prose and poetry – and four main dialects: Northern, Midlands, Kentish and Southern.[8] TEXT has been modelled as a random effect, while all other variables have been modelled as fixed effects.

3 Results

3.1 Overviews by genre, period and dialect

Table 1 offers a summarizing overview of the distribution of subject pronouns and referential null subjects in the PPCME2 and PCMEP corpora by genre. The table shows that referential null subjects very much constitute a low-frequency phenomenon in ME prose and poetry. Figure 1 shows the texts in which the percentage of referential null subjects is highest; the numbers in white indicate the number of tokens of null subjects that this percentage represents for each text.

The aggregate relative frequency of 0.6 per cent null subjects in the prose is clearly lower than the corresponding aggregate frequency for null subjects of 1.3 per cent

[8] One of the texts in the PCMEP corpus, *The Land of Cokaygne*, is an early Irish English text. This text contains 23 overt pronominal subjects but no null subjects, and it has been excluded from the regression analysis as a consequence of this, as well as on the basis of it being a geographical outlier.

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

Table 1. *Distribution of pronouns vs referential null subjects in the PPCME2 and the PCMEP*

Genre	Overt	Null (%)	Total
Prose	51,965	289 (0.6%)	52,254
Poetry	5,689	81 (1.4%)	5,770
Total	57,654	370 (0.6%)	58,024

Table 2. *Distribution of pronouns vs referential null subjects in the PPCME2 and the PCMEP by period*

Period	Overt	Null (%)	Total
M1 (1150–1250)	14,636	251 (1.7%)	14,887
M2 (1250–1350)	8,332	53 (0.6%)	8,385
M3 (1350–1420)	21,217	50 (0.2%)	21,267
M4 (1420–1500)	13,469	16 (0.1%)	13,485
Total	57,654	370 (0.6%)	58,024

in Old English prose reported by Rusten (2015: 59). Additionally, it is interesting to note that the difference between prose and poetry (1.4 per cent null subjects) is almost levelled at the ME stage: Rusten (2015: 60) reported an aggregate relative frequency of 11.5 per cent null subjects in a substantial sample of OE poetic texts. This could perhaps be suggestive of a diachronic decline in the permissibility of null subjects between the two periods.[9] Even so, the overall frequencies are low enough in both periods that it may be questioned how productive the null subject phenomenon actually was, even at the OE stage.

While the difference in aggregate relative frequency between ME prose and poetry is small, it does emerge as statistically significant in a chi-square test.[10] That being said, the effect size of the association is tiny,[11] which means that this result could be due simply to the size of the sample. This indicates that we are justified in applying the more robust statistical measure of logistic regression modelling.

Table 2 gives an overview of the distribution of subject pronouns and null subjects by the *Helsinki Corpus* periods. The table shows that null subjects are more common in the M1 period, while being significantly less frequent in the subsequent periods,[12] but the effect size is tiny in all cases.[13] It is somewhat remarkable that the aggregate

[9] Although the caveat obviously needs to be added that the ME corpus does not quite represent a direct descendant of the OE corpus, since the OE texts are primarily West Saxon and the ME texts to a considerable extent are from the Midlands.

[10] $X^2 = 58.02$, $p < .0001$; Yates' chi-square with df = 1.

[11] Cramer's V = 0.032

[12] M1 vs M2: $X^2 = 45.4$, $p < .0001$; M1 vs M3 = $X^2 = 221.54$, $p < .0001$; M1 vs M4: $X^2 = 184.79$, $p < .0001$; Yates' chi-square with df = 1.

[13] M1 vs M2: 0.0446; M1 vs M3 = 0.0786; M1 vs M4 = 0.0811; Cramer's V.

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

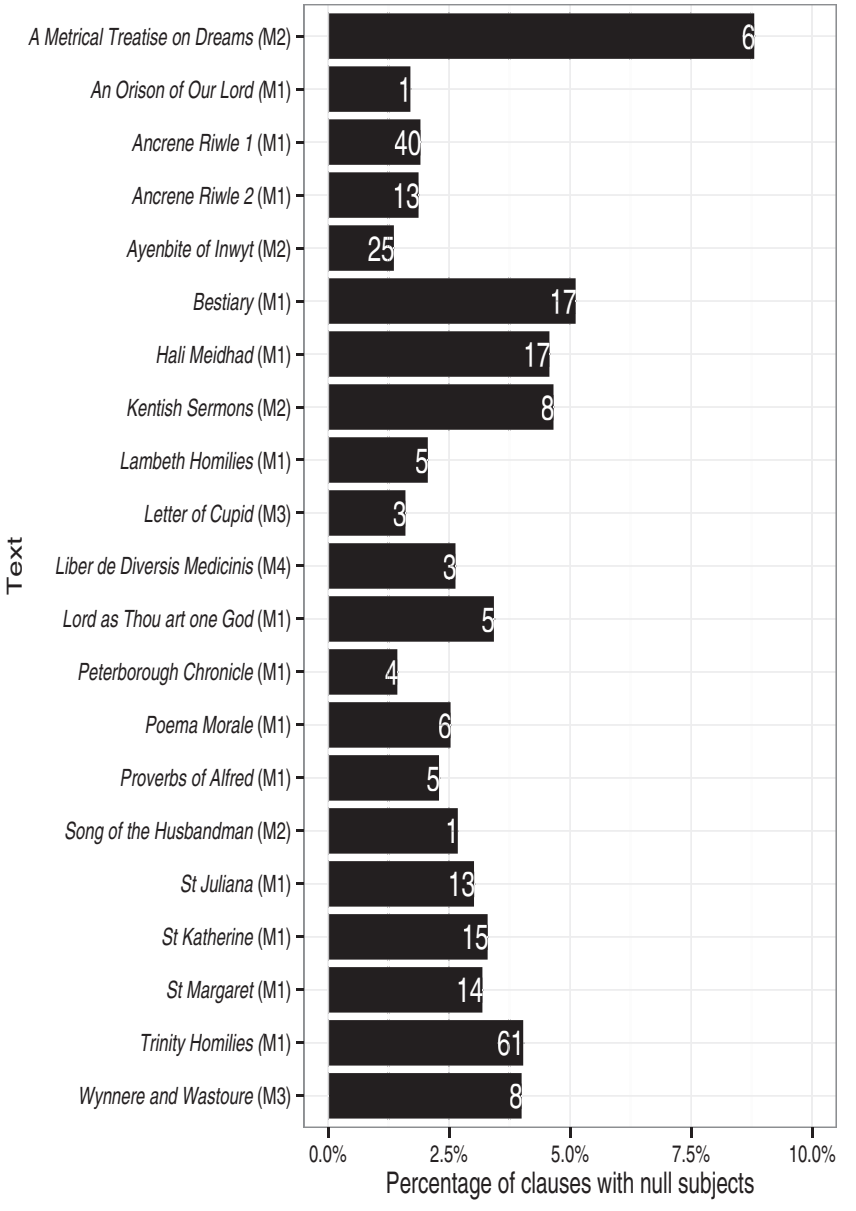


Figure 1. Null subjects by text

frequency for null subjects is as high as 1.7 per cent in the M1 period, since Rusten (2015: 61) finds only a relative frequency of 0.7 per cent in late OE prose.

Figure 2 shows the effect of genre across the four periods.

Table 3 gives a corresponding survey of the distribution of subject pronouns and null subjects by dialect.

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

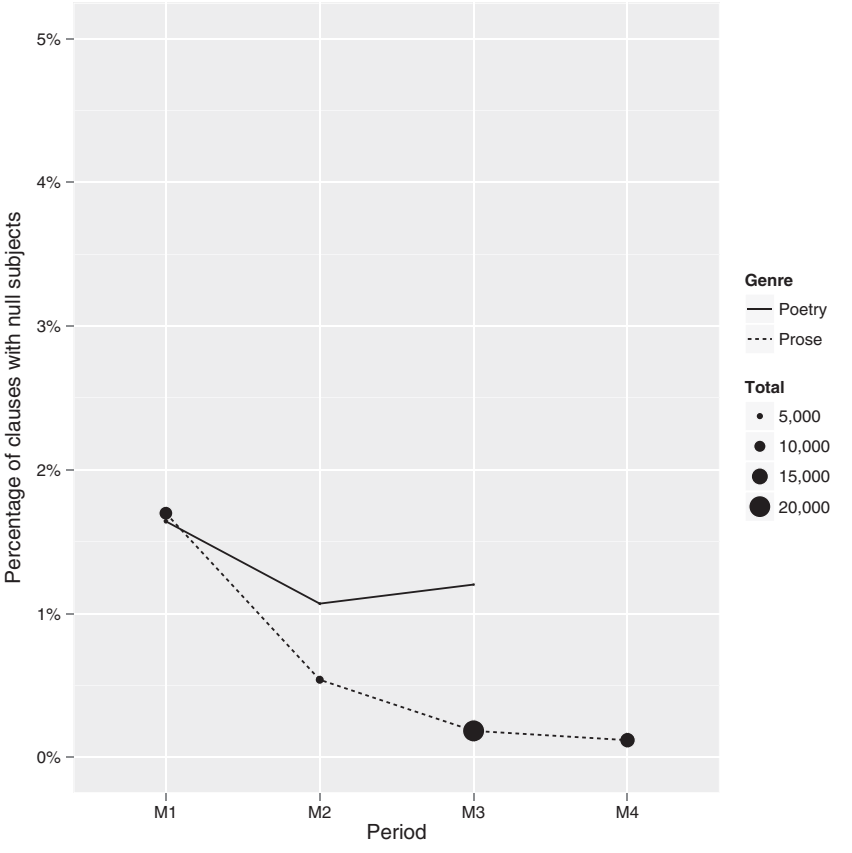


Figure 2. Null subjects by genre across time

There seems to be very little systematicity to the patterning of null subjects by dialect, as is also indicated by figure 3, which shows the development of null subject occurrence in each dialect across time. From table 3 it appears that the Kentish dialect favours null subjects the most, but this is likely to be an epiphenomenon. There are only three Kentish texts in the corpus: the *Ayenbite of Inwyt*, the *Kentish Sermons* and the *Kentish Homilies*. The latter is the only truly autochthonous text of the three, and contains only one null subject out of 224. In the Old English corpus, Kentish texts also do not contain many null subjects. As for the *Ayenbite of Inwyt* and the *Kentish Sermons*, which contain more null subjects, both are literal translations from Old French (Thomson 1907: 396; Hall 1995: 669), which was itself a null-subject language (Adams 1987). Thus the apparent favouring effect of the Kentish dialect may be an artefact of the texts available to us. Similarly, the Northern texts in the corpus, which strongly disfavour null subjects, are predominantly late. The sole Irish text, the *Land of Cokaygne*, dates from the M3 period and contains no null subjects.

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

Table 3. *Distribution of pronouns vs referential null subjects in the PPCME2 and the PCMEP by dialect area*

Dialect	Overt	Null (%)	Total
Northern	3,917	8 (0.2%)	3,925
Midlands	44,328	287 (0.6%)	44,615
Kentish	2,169	34 (1.5%)	2,203
Southern	7,217	41 (0.6%)	7,258
Irish	23	0 (0%)	23
Total	57,654	370 (0.6%)	58,024

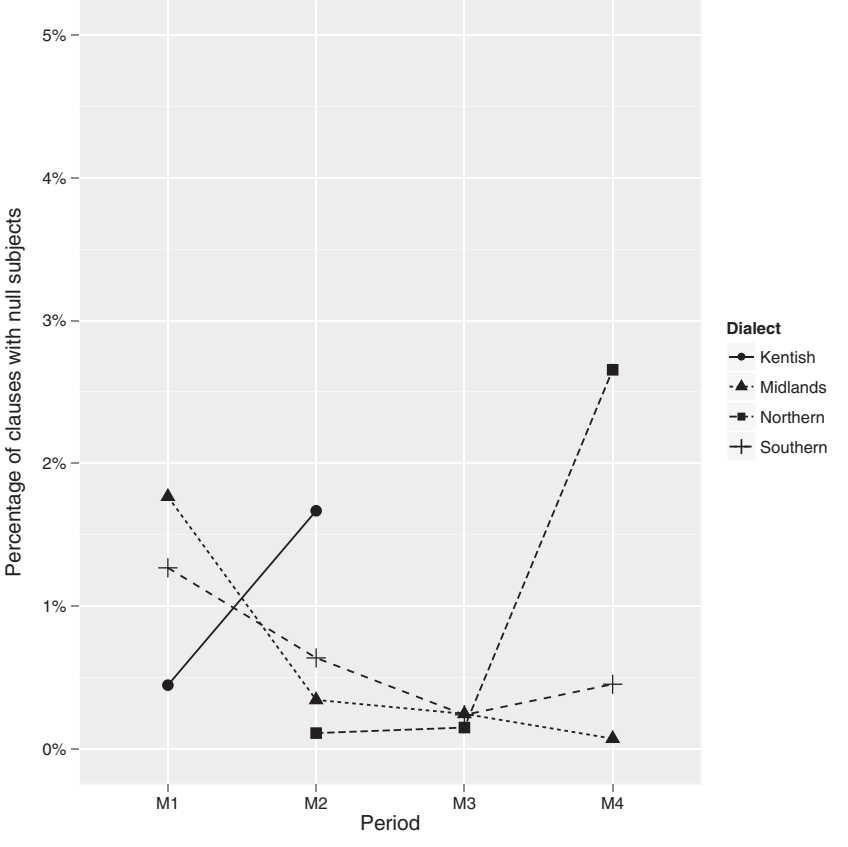


Figure 3. Null subjects by dialect across time

The dialect-split hypothesis of Walkden (2013), according to which null subjects were a robust feature of Anglian but not of West Saxon OE texts, finds no support in the ME data, then: there is simply nothing to be discerned from table 3 and figure 3. That the PPCME2, which constitutes the bulk of our data, is ill-suited to research on dialect syntax has been remarked before (e.g. by Wallage 2005: 68–9), and this

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

Table 4. *Pronouns vs referential null subjects in the PPCME2 by clause type*

Clause type	Overt	Null (%)	Total
Main	12,865	67 (0.5%)	12,932
Conjunct	13,750	109 (0.8%)	13,859
Subordinate	25,350	113 (0.4%)	25,463
Total	51,965	289 (0.4%)	52,254

is of course partly a function of the available material. Nevertheless, investigation of other sources such as those contained in the *Linguistic Atlas of Early Middle English* (LAEME; Laing & Lass 2008) corpus might be able to round out the picture as regards the crucial early periods, and this is a desideratum for future research.

3.2 Clause type

As in OE, null subjects can be observed to occur in ME non-conjunct main clauses (5), second and subsequent conjunct clauses (6) and subordinate clauses (7).

(5) þa þe hehe reue iherde þis ondswere; **Ø** bigon to wreðen swiðe.
when the high revee heard this answer; **Ø** began to rage greatly
'When the high revee heard this answer, **he** began to rage greatly.'
(*St Juliana*, CMJULIA,98.38)

(6) Also Wilferus hadde tweie holy sustres Kynedda and Kyneswida mynchouns. and **Ø**
also Wilferus had two holy sisters, Kynedda and Kyneswida nuns and **Ø**
were boþe i-buryed at Medehamstede þat now hatte Burgh;
were both buried at Medehamstede that now is-called Burgh
'Also, Wilferus had two holy sisters, the nuns Kynedda and Kyneswida, and **they** were both buried at Medehamstede, which is now called Burgh.'
(*John of Trevisa's Polychronicon* CMPOLYCH,VI,129.897)

(7) þu seist þat **Ø** on gode bileuest.
you say-2.SG that **Ø** in God believe-2.SG.
'You say that **you** believe in God.' (*Trinity Homilies*, TRINIT,23.315)

Table 4 gives an overview of the distribution of overt and null subjects in 56 texts of ME prose by clause type, and table 5 breaks this down for texts in which the relative frequency of null subjects is comparatively high.

Over the whole prose dataset, no clause type has a frequency of null subjects above 1 per cent, and the differences between the clause types are minor.[14] There is nevertheless a clear overall hierarchy: conjunct clauses favour null subjects the

[14] The difference between main and conjunct clauses is statistically significant in a chi-square contingency table ($X^2 = 6.98$, p = 0.008) but effect size is tiny (Cramer's V = 0.02). The difference between main and subordinate clauses is not statistically significant ($X^2 = 0.86$, p = 0.35).

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

Table 5. *Pronouns vs referential null subjects in various Middle English prose texts by clause type*

Text	Clause type	Overt	Null (%)	Total
Ancrene Wisse (1)	Main	568	8 (1.4%)	576
(CMANCRIW-1.psd)	Conjunct	266	11 (4%)	277
	Subordinate	1,197	21 (1.7%)	1,218
	Totals	2,031	40 (1.9%)	2,071
Ancrene Wisse (2)	Main	233	4 (1.7%)	237
(CMANCRIW-2)	Conjunct	74	4 (5.1%)	78
	Subordinate	367	5 (1.3%)	372
	Totals	674	13 (1.9%)	687
St Juliana	Main	112	6 (5.1%)	118
(CMJULIA.psd)	Conjunct	87	5 (5.4%)	92
	Subordinate	216	2 (0.9%)	218
	Totals	415	13 (3%)	428
St Katherine	Main	86	8 (8.5%	94
(CMKATHE.psd)	Conjunct	80	5 (5.9%)	85
	Subordinate	271	2 (0.7%)	273
	Totals	437	15 (3.3%)	452
Kentish Sermons	Main	41	2 (4.7%)	43
(CMKENTSE.psd)	Conjunct	23	2 (8%)	25
	Subordinate	99	4 (3.9%)	103
	Totals	163	8 (4.7%)	171
Trinity Homilies	Main	317	11 (3.4%)	328
(CMTRINIT.psd)	Conjunct	244	20 (7.6%)	264
	Subordinate	883	30 (3.3%)	913
	Totals	1,444	61 (4.1%)	1,505

most, subordinate clauses the least. The same hierarchy is found within individual texts, except *Ancrene Wisse* 1, in which subordinate clauses edge above main clauses, and in *St Katherine*, where null subjects are more frequent in non-conjunct than in conjunct main clauses. These exceptions aside, this is precisely the same effect found for Old English by Walkden (2013: 159–61) and Rusten (2015: 68), and for Old Saxon by Walkden (2014: 190); a dispreference for null subjects in subordinate clauses as opposed to main clauses is also found for Old High German by Axel (2007: 310) and for late Old Swedish by Håkansson (2013: 170, 173).

The restriction on referential null subjects in subordinate clauses is surprising from the perspective of present-day null-subject languages. It can be accounted for under the syntactic analyses of Håkansson (2013: 174–6) or Walkden (2014: 209–13), which share the intuition that null arguments must be discourse-licensed and that this licensing is syntactically restricted in subordinate clauses.[15] The greater

[15] The restriction is reminiscent of the so-called 'topic-drop' languages like Modern German, Dutch and Norwegian, in which topics can be omitted when they would otherwise occur in initial position in V2 clauses. This kind of topic drop cannot be the full story for Middle English, however, since examples like

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

Table 6. *Pronouns vs referential null subjects in the PCMEP by clause type*

Clause type	Overt	Null (%)	Total
Main	2,978	51 (1.7%)	3,029
Conjunct	403	8 (1.9%)	411
Subordinate	2,308	22 (0.9%)	2,330
Total	5,689	81 (1.4%)	5,770

frequency of null subjects in conjunct clauses is slightly more mysterious. However, as we shall see, most null subjects in ME have third-person reference, and the difference could therefore be purely pragmatic in origin: third-person null subjects are usually infelicitous in out-of-the-blue contexts such as the start of a discourse, and conjunct clauses as defined here are of necessity not discourse-initial. Alternatively (or additionally), there could be some syntactic basis for the difference. Faarlund (1990: 104) suggests that conjunction reduction, the process by which VPs can share a subject in Modern English examples such as *The king went to Norway and* ***Ø*** *met the bishop*, was more liberal in Old Norse, operating regardless of grammatical function or case, and this may have been true for Middle English as well. A distinct process of conjunction reduction, yielding two routes for unexpressed referential subjects in Middle English conjunct clauses, would also explain the difference in frequency: some subjects could be null because of syntactic discourse-licensing, others because of conjunction reduction. We will not attempt to determine which of these alternatives is correct. The quantitative results for Middle English poetry are also compatible with the hierarchy conjunct > main > subordinate, as shown by table 6. Again, the hierarchy is robust across the dataset but not for every individual text, as table 7 demonstrates (in fact, none of the texts in table 7 shows this hierarchy).[16]

Finally, figure 4 shows the gradual decline of referential null subjects in all clause types. It may be noted that the relative frequencies for null subjects by clause type converge at a point slightly above 0 per cent in the M4 period; thus, at the tail-end of the period dealt with here, referential null subjects are virtually non-existent regardless of clause type.

(1) and (12) contain an overt topical element in initial position. The case against topic drop is made by van Gelderen (2000: 146) for Middle English and by Walkden (2014: 201–3) for early Germanic more generally.

[16] But the difference is only statistically significant in the case of main vs subordinate clauses ($X^2 = 4.82$, p = 0.03), and effect size is tiny (Cramer's V = 0.03). The difference between main and conjunct clauses is not statistically significant ($X^2 = 0.03$, p = 0.86), and the same is true for the difference between conjunct and subordinate clauses ($X^2 = 2.38$, p = 0.12).

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

Table 7. *Pronouns vs referential null subjects in various Middle English verse texts by clause type*

Text	Clause type	Overt	Null	Total
Bestiary	Main	130	15 (10.3%)	145
(1225.Bestiary.psd)	Conjunct	34	2 (5.6%)	36
	Subordinate	150	0 (0%)	150
	Totals	314	17 (5.1%)	331
The Owl and the Nightingale	Main	510	2 (0.4%)	512
(1220.OwlNight.psd)	Conjunct	6	0 (0%)	6
	Subordinate	419	7 (1.6%)	426
	Totals	935	9 (1%)	944
Wynnere and Wastoure	Main	57	5 (8.1%)	62
(1360.WynWas.psd)	Conjunct	27	1 (3.6%)	28
	Subordinate	107	2 (1.8%)	109
	Totals	191	8 (4%)	199

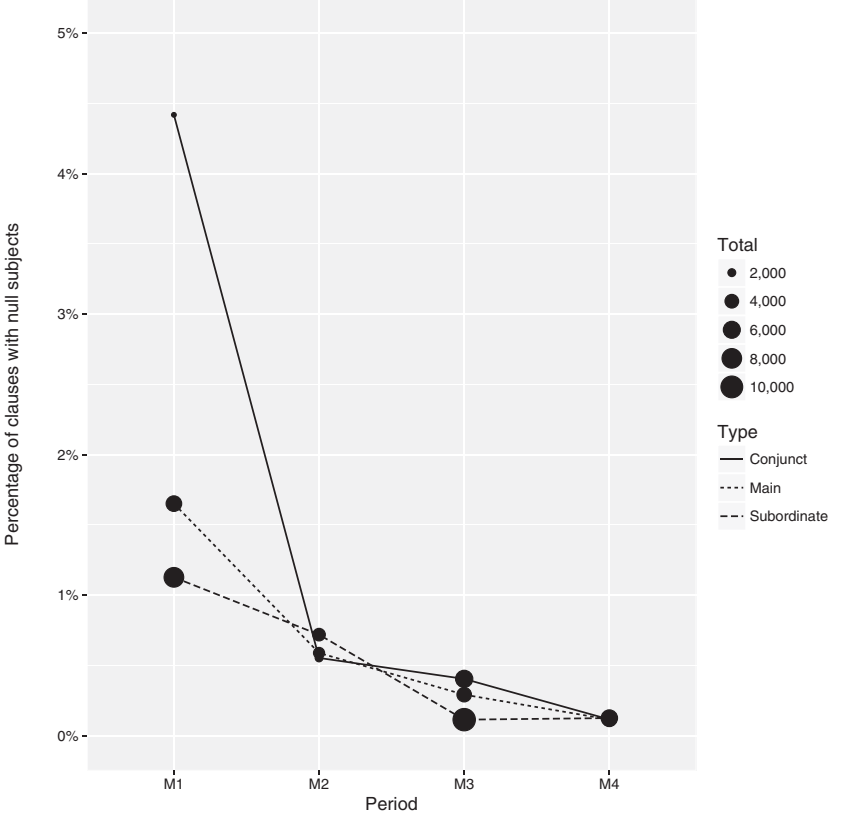


Figure 4. Null subjects by clause type across time

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

Table 8. *Pronouns vs referential null subjects in the PPCME2 by person and number*

Person	Number	Overt	Null (%)	Total
1	sg	6,666	7 (0.1%)	6,673
	pl	2,869	7 (0.2%)	2,876
2	sg	6,119	29 (0.5%)	6,148
	pl	2,741	4 (0.1%)	2,745
3	sg	24,848	200 (0.8%)	25,048
	pl	8,722	42 (0.5%)	8,764
Total		51,965	289 (0.6%)	52,254

3.3 Person and number

Similarly to OE, the ME corpora exhibit null subjects in all person and number combinations. In (8)–(11), examples are given of first- (8), second- (9) and third-person singular (10) and third-person plural (11) null subjects.

(8) þenne scheomeð me þer-wið & **Ø** fleo ham from
then shames me.DAT therewith & **Ø** flee them from
'Then I am ashamed and flee from them' (*St Margaret*, CMMARGA,79.388)

(9) Wendest **Ø** þet hit þin were
believed-2.SG **Ø** that it yours were
'Did **you** believe that it was yours?' (*Body and Soul*, BodySoul,116.4.36.FragC)

(10) **Ø** Henges dun his heaued
Ø hangs down his head
'**He** hangs his head down.' (*The Wooing of the Lord*, WooingLord,283.269.181)

(11) He sende in-to hem hounde-fleges and **Ø** ete hem;
he sent in-to them hound-flies and **Ø** ate them
'He sent unto them gadflies and **they** devoured them.'
(*The earliest complete English prose psalter*, CMEARLPS,96.4193)

Table 8 shows how person and number affect the expression of subjects in the prose, and table 9 breaks this down for a selection of texts with a relatively high proportion of null subjects.

The general pattern visible from tables 8 and 9 is that, in the first person and in the second person plural, null subjects are so vanishingly rare as to be attributable to scribal or editorial error.[17] In the third person and in the second person singular, the proportion of null subjects is somewhat higher. Exactly the same pattern can be seen in

[17] In fact, if we assume that the first clause of 3.3 involves an oblique experiencer *subject* (see Allen 1995), then the subjectless second clause can be treated as a straightforward case of conjunction reduction. For ease of retrieval, however, we have assumed throughout our quantitative investigation that subjects are nominative.

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

Table 9. *Pronouns vs referential null subjects in various Middle English prose texts by person and number*

Text	Person	Number	Overt	Null (%)	Total
Ancrene Riwle (1)	1	sg	261	2 (0.8%)	263
(CMANCRIW-1.psd)		pl	92	1 (1.1%)	93
	2	sg	220	1 (0.5%)	221
		pl	211	1 (0.5%)	212
	3	sg	874	30 (3.3%)	904
		pl	373	5 (1.3%)	378
	Totals		2,031	40 (1.9%)	2,071
Ancrene Riwle (2)	1	sg	109	0 (0%)	109
(CMANCRIW-2.psd)		pl	34	0 (0%)	34
	2	sg	69	4 (5%)	73
		pl	90	1 (1%)	91
	3	sg	254	7 (3%)	261
		pl	118	1 (1%)	119
	Totals		674	13 (1.9%)	687
St Juliana	1	sg	104	0 (0%)	104
(CMJULIA.psd)		pl	22	0 (0%)	22
	2	sg	94	1 (1.1%)	95
		pl	18	0 (0%)	18
	3	sg	112	12 (9.7%)	124
		pl	65	0 (0%)	65
	Totals		415	13 (3%)	428
St Katherine	1	sg	72	1 (1.4%)	73
(CMKATHE.psd)		pl	26	0 (0%)	26
	2	sg	99	3 (2.9%)	102
		pl	28	0 (0%)	28
	3	sg	102	9 (8.1%)	111
		pl	110	2 (1.8%)	112
	Totals		437	15 (3.3%)	452
Kentish Sermons	1	sg	12	0 (0%)	12
(CMKENTSE.psd)		pl	24	0 (0%)	24
	2	sg	9	0 (0%)	9
		pl	20	0 (0%)	20
	3	sg	44	6 (12%)	50
		pl	54	2 (3.6%)	56
	Totals		163	8 (4.7%)	171
Trinity Homilies	1	sg	127	1 (0.8%)	128
(CMTRINIT.PSD)		pl	235	6 (2.5%)	241
	2	sg	68	2 (2.9%)	70
		pl	23	1 (4.2%)	24
	3	sg	768	44 (5.4%)	812
		pl	223	7 (3%)	230
	Totals		1,444	61 (4.1%)	1,505

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

Table 10. *Pronouns vs referential null subjects in the PCMEP by person and number*

Person	Number	Overt	Null (%)	Total
1	sg	1,357	4 (0.3%)	1,361
	pl	269	0 (0%)	269
2	sg	1,097	10 (0.9%)	1,107
	pl	192	1 (0.5%)	193
3	sg	2,342	54 (2.3%)	2,396
	pl	432	12 (2.7%)	444
Total		5,689	81 (1.4%)	5,770

Table 11. *Pronouns vs referential null subjects in various ME verse texts by person and number*

Text	Person	Number	Overt	Null (%)	Total
Bestiary (1225.Bestiary.psd)	1	sg	9	0 (0%)	9
		pl	28	0 (0%)	28
	2	sg	13	1 (7.1%)	14
		pl	56	0 (0%)	56
	3	sg	208	16 (7.1%)	224
		pl	0	0 (0%)	0
	Totals		314	17 (5.1%)	331
The Owl and the Nightingale (1220.OwlNight.psd)	1	sg	258	0 (0%)	258
		pl	9	0 (0%)	9
	2	sg	291	1 (0.3%)	292
		pl	8	1 (11.1%)	9
	3	sg	323	7 (2.1%)	330
		pl	46	0 (0%)	46
	Totals		935	9 (1%)	944
Wynnere and Wastoure (1360.WynWas.psd)	1	sg	58	2 (3.3%)	60
		pl	5	0 (0%)	5
	2	sg	33	0 (0%)	33
		pl	16	0 (0%)	16
	3	sg	62	3 (4.6%)	65
		pl	17	3 (15%)	20
	Totals		191	8 (4%)	199

the poetry data, except that the difference between third and non-third person is more pronounced than in the prose. See tables 10 and 11.

The finding that null subjects are more frequent in the third person than elsewhere is not a surprise given the findings of van Gelderen (2000), as well as the research establishing this for Old English (Walkden 2013: 164–6; Rusten 2015: 69–71) and other early Germanic languages (see Rosenkvist 2009 & Walkden 2014: ch. 5 for an overview). The prevalence of null subjects in the second person singular is, however,

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

unexpected. It is not likely that verbal agreement played a role in licensing null subjects in Old English, and *prima facie* even less so for Middle English, given the widespread morphological impoverishment that this period witnessed. Nevertheless, the second person singular stood out from the paradigm for Middle English in most verbs, with its distinctive *-st* ending. A couple of examples are given below. Example (12) is from the early Southern poem *Body and Soul*, which contributes 3 of the 10 poetic examples of second-person singular null subjects despite otherwise containing no null subjects. Example (13) is from the Katherine Group text *Hali Meidhad*, which contributes 7 of the 29 prose examples in the second person singular. Taken together, the Katherine Group plus the *Ancrene Riwle* account for 16 of the 29 – though these are of course texts in which null subjects are more frequently encountered in general.

(12) **Ø** Oþre beræfedest rihtes istreones
Ø others bereaved-2SG rightful-GEN wealth-GEN
'**You** robbed others of rightful wealth'
(*Body and Soul*; BodySoul,190.7.22.FragE)

(13) nawt ane for þe-seolf; ase þerf godes spuse; ah **Ø** schalt for monie oþre
not alone for yourself as needs God's spouse but **Ø** shall-2SG for many others
'not only for yourself, as God's spouse must, but **you** will [care] for many others'
(Katherine Group, *Hali Meidhad*; CMHALI,149.299)

Perhaps authors at this time were aware at some level that the distinctive ending would make it easier for their readers to identify the referent. Somewhat interestingly, second-person singular null subjects are (sporadically) observable even in Early Modern English, e.g. in William Stevenson's[18] *Gammar Gvrtons nedle* (mid sixteenth century), and even as late as in Thomas Middleton's play *A chaste maid in Cheapside* (1630). See examples (14)–(15).

(14) Seest **Ø** not how cha rent & torn, my heels, my knees & my breech
see-2.SG **Ø** not how I.am rent and torn, my heels, my knees and my breech
'Do **you** not see that I am rent and torn, my heels, my knees and my breeches? '
(*Gammar Gvrtons nedle*; STEVENSO-E1-P1,19.159)

(15) S. Walt. How dost **Ø** Iacke?
S. Walt. how do-2.SG **Ø** Jack
'S. Walt: How do **you** do, Jack?'
(*A chaste maid in Cheapside*; MIDDLET-E2-P1,10.51)

Outside the second person, number does not have a consistent effect. In the first person, plural subjects are more likely to be null than singular subjects in prose, but singular subjects are more likely to be null than plural subjects in poetry. In the third person, the picture is reversed: plural subjects are more likely to be null than singular subjects in poetry, but singular subjects are more likely to be null than plural subjects

[18] As pointed out by the compilers of the *Penn–Helsinki Parsed Corpus of Early Modern English* (Kroch *et al.* 2004), the authorship of this text is contested.

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

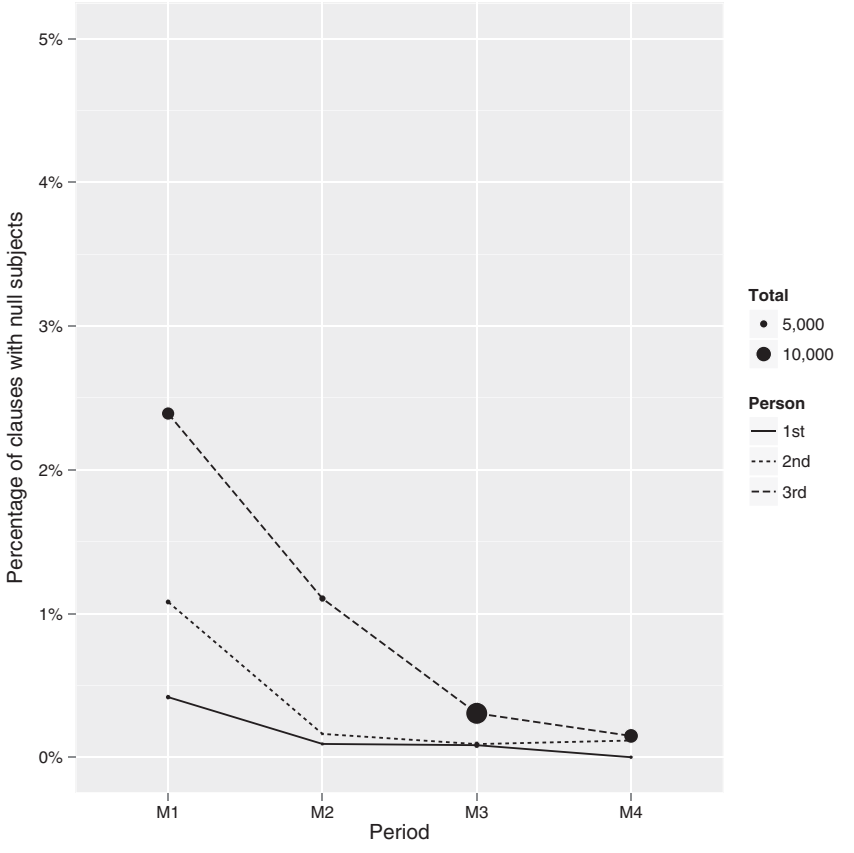


Figure 5. Null subjects by person across time

in prose. Of these effects, only the last is statistically significant at the 0.05 level in a chi-square test, perhaps as a result of the greater number of examples.[19] This finding mirrors findings for a variety of early Germanic languages, including Old English, in which number did not consistently favour or disfavour subject expression (Walkden 2014: ch. 5).

Figures 5 and 6 show the effects of person and number respectively across the four periods.

3.4 Modelling Middle English null subjects: regression analysis

In order to assess the relative importance of the variables discussed above in conditioning the occurrence of null subjects, a one-level mixed-effects logistic

[19] 1st person, prose: $X^2 = 1.772$, p = 0.1831; 1st person, poetry: $X^2 = 0.047$, p = 0.8284; 3rd person, prose: $X^2 = 8.867$, p = 0.0029; 3rd person, poetry: $X^2 = 0.164$, p = 0.6855; Yates' chi-square with df = 1 in all cases.

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

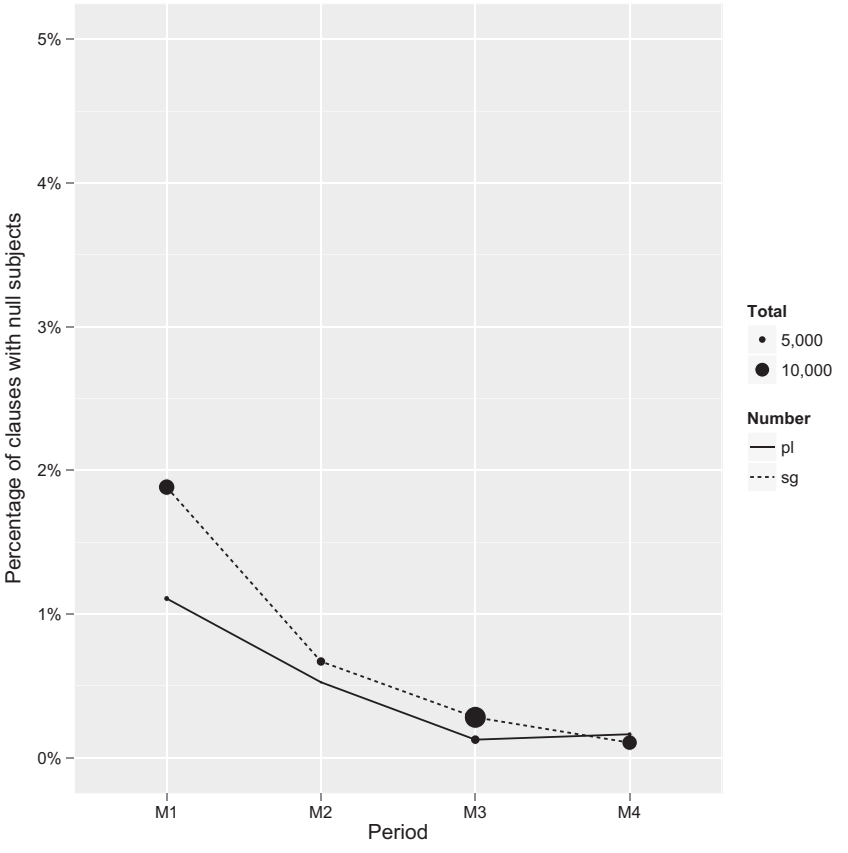


Figure 6. Null subjects by number across time

regression was carried out using Rbrul (Johnson 2009), as described in section 2.2. The results are reported in table 12. The table consists of two parts. The first rows give summary information about the number of pronominal tokens upon which the model builds as well as the degrees of freedom of the model and the overall intercept (i.e. the model baseline) and deviance (a measure of model fit to the data, i.e. a quantification of the degree to which the predictions of the model deviate from the actual data). The remainder of the table gives an analysis of the influence exerted by the fixed-effect predictors on the likelihood of having a null subject instead of an overt pronominal one in the ME data. It can be observed that the model intercept log-odds coefficient is -5.923, indicating that the likelihood of a null subject is estimated to be low across the dataset. This, of course, is entirely in line with expectations in light of the overviews given in preceding sections. The variable TEXT is treated as a random effect, as noted, for example, in section 2.2. The random intercept values for individual texts can be found in a table in the Appendix.

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

Table 12. *Results of a one-level generalised mixed-effects regression analysis*

	intercept	n	df	deviance	
	−5.923	58,001	14	3573.166	
predictors	factor	log-odds	tokens	n/n+y	centred factor weight
Type	Conjunct	0.512	14,268	0.008	0.625
(p < 0.001)	Main	0.014	15,946	0.007	0.503
	Subordinate	− 0.526	27,787	0.005	0.372
Person	3	0.886	36,635	0.008	0.708
(p < 0.001)	2	0.120	10,189	0.004	0.53
	1	− 1.007	11,177	0.002	0.268
Number	sg	0.248	42,722	0.007	0.562
(p < 0.001)	pl	− 0.248	15,279	0.004	0.438
Period	M1	1.287	14,887	0.017	0.784
(p < 0.001)	M2	0.225	8,362	0.006	0.556
	M3	− 0.574	21,267	0.002	0.36
	M4	− 0.938	13,485	0.001	0.281
Genre	Poetry	0.48	5,747	0.014	0.618
(p = 0.0158)	Prose	− 0.48	52,254	0.006	0.382
Dialect	Kentish	1.077	2,203	0.015	0.746
(p = 0.271)	Northern	− 0.138	3,925	0.002	0.466
	Midlands	− 0.251	44,615	0.006	0.438
	Southern	− 0.688	7,258	0.006	0.334

In terms of interpretation, the crucial number in table 12 is the log-odds ratio: this number indicates a favouring effect on pronoun omission if positive, and a disfavouring effect if negative. The larger the number, the larger the effect. Thus, for instance, conjunct clauses and main clauses both favour pronoun omission, while subordinate clauses disfavour it, but main clauses favour it by only a very small amount. In table 12, the factors (i.e. the individual levels of each predictor) are ordered according to their centred factor weight.

The goodness of fit of the model as measured by Nagelkerke's pseudo-R^2 is 0.459, indicating that 45.9 per cent of the observed variation can be explained by the predictors included. The fixed predictors account for 0.259 of this, while between-text variation contributes the remaining 0.2. The deviance value for the entire model is 3573.166. The deviance value given by Rbrul is 'a measure of how well the model fits the data', and 'the larger the deviance, the worse the fit' (Johnson 2010).[20] It is therefore evident that the model's fit to the data is not particularly good. Somer's D_{xy} rank correlation, a measure of the predictive capabilities of the model, is 0.768.[21] Overall, then, the explanatory capabilities of a logistic regression model taking into

[20] These citations are taken from the unpaginated Rbrul manual (www.danielezrajohnson.com/Rbrul_manual.html).

[21] A Somer's D_{xy} value of 0 would indicate randomness whereas a value of 1 would indicate a perfect fit to the data.

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

account linguistic and extralinguistic variables proposed by previous research as constituting important factors determining the occurrence of null subjects are actually not overly impressive. This state of affairs likely emerges as a reflex of the general scarcity of null subjects in the ME dataset.

3.4.1 Within-text predictors

All of the within-text predictors (TYPE, PERSON and NUMBER) emerged as significant. These predictors are all significant on the $p < 0.001$ level. In summary, then, logistic regression modelling of the TYPE variable shows that conjunct clauses have a distinct favouring effect on the occurrence of null subjects (log-odds: 0.512, centred factor weight: 0.625). Main clauses also display a favouring effect on nullness, although in this case the effect is very weak (log-odds: 0.014). The weakness of this effect is also illustrated by the centred factor weight of 0.503, which essentially indicates that there is no particular effect, whether positive or negative.[22] Subordinate clauses, on the other hand, are associated with a clear disfavouring effect on the occurrence of a null subject in the ME data (log-odds: −0.526, centred factor weight: 0.372).

As for person, it may be observed that null subjects are favoured in the third person (log-odds: 0.886, centred factor weight: 0.708), but disfavoured in the first person (log-odds: −1.007, centred factor weight: 0.268). The effect of second person, although positive (log-odds: 0.120), is more or less neutral – i.e. it exerts no particular favouring or disfavouring influence on the nullness of subject pronouns – as shown by the centred factor weight of 0.53. These results are more or less in conformance with expectations following the tabularized presentation of the data by person in section 3 (compare tables 8 and 10).

Less obvious from the tables above is the fact that the influence of the NUMBER variable is also statistically significant: singular number demonstrates a favouring effect (log-odds: 0.248) on subject-pronoun nullness as compared to plural number (−0.248). However, the log-odds are distinctly weak in both cases, and the centred factor weight of 0.562 (singular) and 0.438 (plural) are both very close to 0.5, a value signifying a neutral effect. It is therefore likely that the significant result is due to interaction with PERSON.[23] This issue will be dealt with in section 3.5.

3.4.2 Between-text predictors

Of the between-text predictors, PERIOD was significant at the $p < 0.001$ level, and GENRE was significant at the $p < 0.02$ level. DIALECT did not emerge as a significant predictor. As concerns PERIOD, the pattern demonstrated by the Rbrul analysis seems consistent with a picture where the permissibility of null subjects declines over time, as also shown e.g. in figures 5–6: M1 favours null subjects the most (log-odds: 1.287, centred factor weight: 0.784). The positive effect of the M2 period is close to the neutral 0.5 mark in terms of centred factor weight, while M3 and M4 have negative log-odds, demonstrating progressively disfavouring effects. This is entirely in conformance

[22] As pointed out by Johnson (2009: 360–1), 'a factor weight of 0.5 is equivalent to no effect'.
[23] Thanks to an anonymous reviewer for bringing this to our attention.

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

with expectations in light of the distribution of subject pronouns and null subjects by period presented in table 2: although rare at all stages of ME, null subjects are significantly more frequent in the M1 period than in the later M2–M4 periods. At a log-odds coefficient of 1.287, M1 is the strongest of all the modelled effects.

As for GENRE, poetry can be observed to favour nullness whereas prose disfavours it. This is not entirely unexpected since aggregate relative frequencies for null subjects – while certainly low – are more than twice as high in the poetry (1.4%) as in the prose (0.6%). Finally, as also noted above, Walkden's dialect-split hypothesis finds no support in the ME data subject to analysis here.

3.5 Modelling Middle English null subjects: random forests

It is a potential problem of the analysis above that the robustness of mixed-effects models may decrease when percentages of the investigated phenomenon are low, since small changes in proportions in such cases can result in large changes in logits.[24] Consequently, confidence in the results could be strengthened by investigating the ME data by means of a different statistical measure. Thus, to supplement the regression model above, we built a random forest of conditional inference trees using the *cforest* function in the *party* package (Hothorn *et al.* 2015).[25] Then, we assessed variable importance by means of the conditional permutation accuracy importance measure proposed by Strobl *et al.* (2008).

For our purposes, there are two distinct advantages to running a random forest. First, random forests operate entirely differently from logistic regression models (see e.g. Strobl *et al.* 2009 for an introduction to the mechanics), and hence, if similar results are attained, our confidence in the regression model would be strengthened. Second, random forests are more capable of handling complex interactions than regression models are. This is relevant since, as noted above, the variables of PERSON and NUMBER appear to interact with one another in the regression model.

Figure 7 gives a dotplot showing the conditional permutation variable importance for a random forest taking into account the fixed-effects predictors which emerged as significant in the regression model in section 3.4.[26] The variable DIALECT was left out since null subjects displayed no particular distributional tendency in table 3 and since DIALECT did not emerge as significant in the regression model. The dashed vertical line is drawn at the location of the least important variable, and the further to the right a variable is, the more important it was judged by the model.

It can be observed that the random forest reinforces the conclusions of the regression model in some key areas. Most notably, the random forest identifies PERIOD as the most important variable in terms of predicting omission of a subject pronoun. As recalled, the M1 period demonstrated the highest log-odds (1.287) and the highest

[24] We would like to thank one of the anonymous reviewers for pointing this out to us.

[25] Seed = 147, ntree = 1000 and ntry = 5. The results remained consistent over several runs using different parameters.

[26] The plot was created in R using the *varimp* function in the *party* package (Hothorn *et al.* 2015).

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

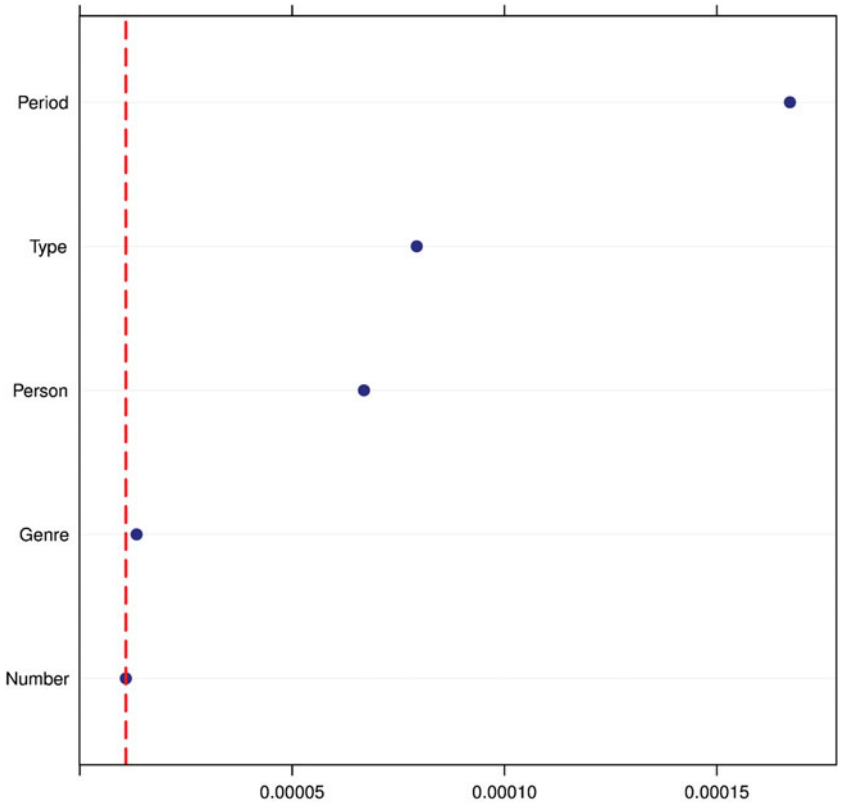


Figure 7. (Colour online) Conditional permutation variable importance for a random forest incorporating the significant fixed-effects predictors of the regression model in table 12

centred factor weight (0.784) of all the variables in the regression model. Similarly to the regression model, TYPE and PERSON emerge as important variables, while GENRE and NUMBER are the least important predictors in the random forest. Furthermore, the random forest serves to disentangle the potential interaction of PERSON and NUMBER, with the former emerging as clearly more important than the latter. This is consonant with the regression model in that the log-odds values achieved for the levels of PERSON (third person: 0.886, second person: 0.120, first person: −1.007) are more extreme than those achieved for NUMBER (singular: 0.248, plural: −0.248). Also as recalled, contrary to the centred factor weight for the third person, the centred factor weight for both levels of the NUMBER variable hovered around 0.5, essentially indicating no effect.

4 Discussion and analysis

As stated in section 1.1, the possibility of referential null subjects in OE has received a substantial amount of quantitatively based attention in recent years. This article has

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

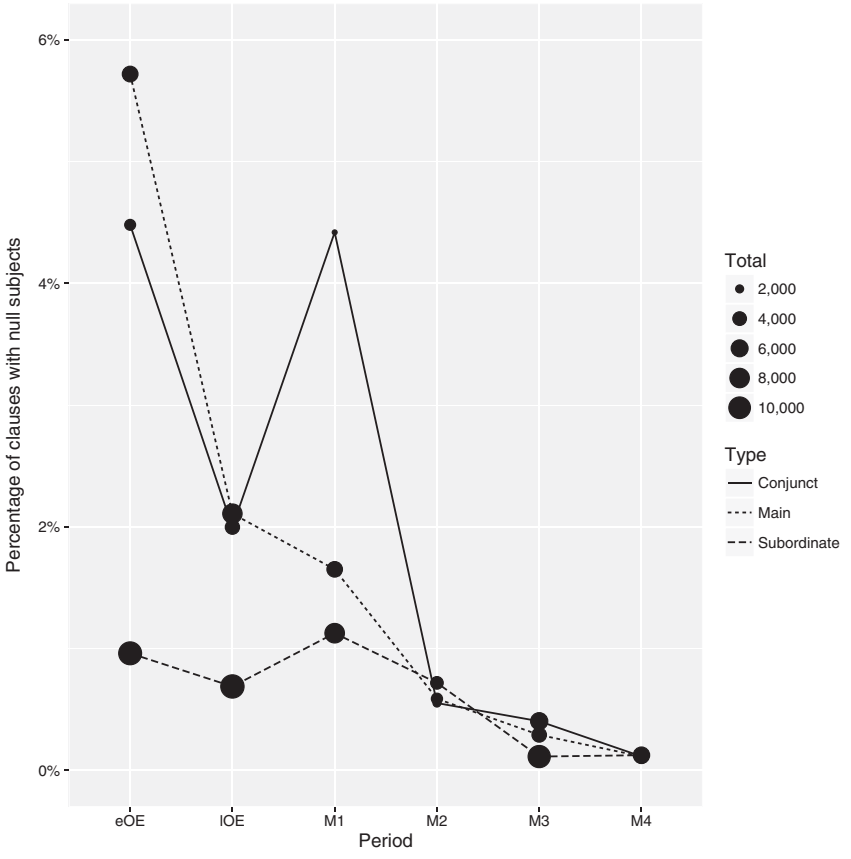


Figure 8. Null subjects by clause type in Old and Middle English

supplemented previous work on OE with a large-scale explorative and quantitative investigation of the occurrence and distribution of referential null subjects in ME prose and poetry. On the basis of a considerable amount of data, then, it can be concluded that null subjects constitute a marginal phenomenon in terms of rates of occurrence in both OE and ME. This is illustrated in figure 8, which gives an overview of the decline of the null subject phenomenon across early Old English (eOE, before 950), late Old English (lOE, after 950) and the four ME periods by clause type.[27]

The figure shows that a relatively sharp decline in relative frequencies for null subjects takes place in both main clauses (from < 6% to 2%) and conjunct clauses (*c*. 4.5% to *c*. 2%) from eOE to lOE. It should be noted, however, that the high frequencies in the eOE period are largely due to data from the poetry. Null subjects

[27] The OE data are based on the data in Rusten (2015). It should be mentioned here that the data material forming the basis for Rusten's OE figures is much more restricted in size than the ME data dealt with here. See Rusten (2016) for a statistical analysis of null subjects based on a much larger amount of data.

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

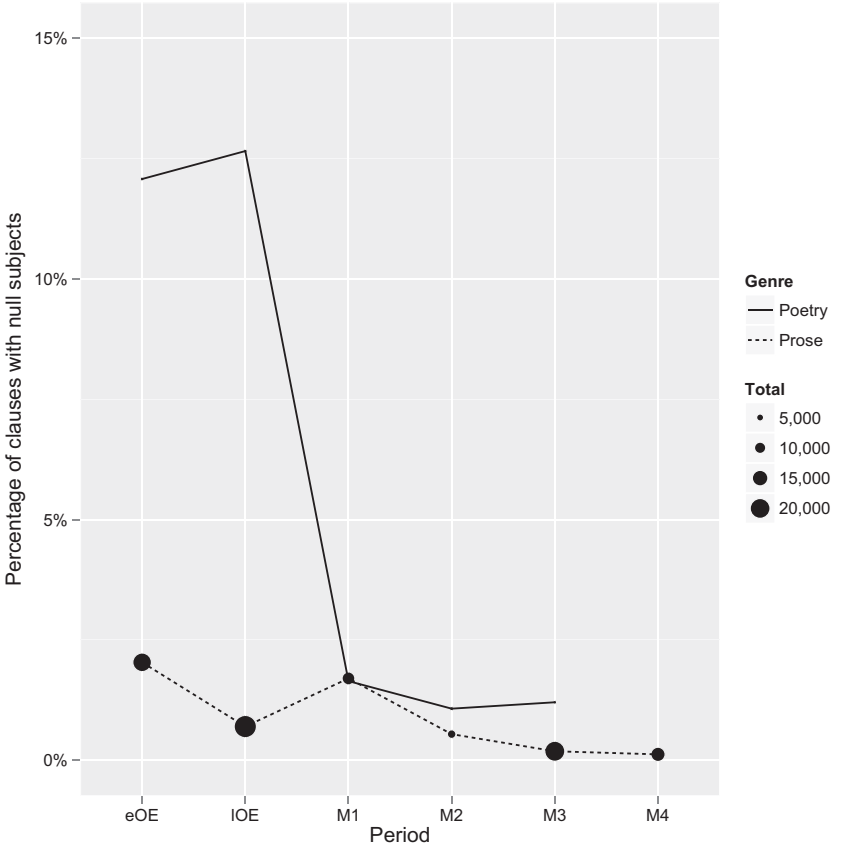


Figure 9. Null subjects by genre in Old and Middle English

also become rarer in subordinate clauses in the same period. The decline continues steadily in main clauses, while there is a minor spike from lOE to M1 in subordinate clauses, and a more notable one in conjunct clauses. After M1, frequencies decline in all clause types, and by the M4 period – as also noted in section 3.2 – null subjects occur at a rate marginally above 0 per cent regardless of clause type.

Figure 9 gives a corresponding survey of the decline of null subjects across OE and ME by genre. In eOE and lOE, aggregate frequencies for null subjects are much higher in poetry than in prose. Frequencies for null subjects in both genres converge at *c*. 2 per cent in the M1 period and remain relatively constant for the verse texts until the M3 period (after which point there are no more poetic data in our dataset), while null subjects are virtually non-existent at the same stage in prose, and remain so for the duration of the ME period. This means that null subjects in OE can be seen mainly as a feature of the poetry,[28] which is somewhat problematic for any account

[28] Indeed, an anonymous reviewer points out that the broad strokes of the overarching OE–ME diachronic pattern documented here 'appears to be largely driven by poetry'.

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

seeking to describe that language as a canonical pro-drop language (van Gelderen 2013). In this context, it is interesting to note that Mitchell (1985: II, 992–3), in a section dedicated to differences between the syntax of prose and poetry, mentions 'clauses with unexpressed personal pronoun subjects' as a feature that helps 'to achieve compression and to give the poetry its characteristic texture'. When the ME data are considered in isolation, however, the genre effect is not as pronounced, as indicated both by the marginal differences in relative frequency as well as in figure 7. ME poetry is substantially different from OE alliterative poetry (see e.g. Fischer et al. 2000: 32–3), and this could be taken as reinforcing the impression that the earliest null subjects indeed are primarily a feature of OE poetry.

As concerns the loss of null subjects, as pointed out by Walkden (2013: 158), it is not necessarily to be expected that null subjects should occur at frequencies of exactly 0 per cent at a stage where such subjects are illicit: quantitative work carried out by researchers at the University of Pennsylvania has found that 'well-established generalizations in a language are violated in naturally-occurring usage at a low, relatively constant rate of about 1%' (Santorini 1992: 609, fn. 17).[29] Extremely sporadic omission of an unstressed pronominal subject could be viewed as constituting a violation of a 'well-established' generalization in a grammar requiring explicit pronominal subjects. At the same time, there can be no doubt that 'there are often stubborn relic forms left over "after" language change' (Denison 2003: 59).[30] By means of exemplification, Denison mentions 'the remaining irregular plurals in Modern English' which persist even though 'nearly all other nouns have gone over to a single regular pattern', and 'the fossil syntax of such PPs as *the whole world over* after the general loss of post-positional prepositions' (2003: 59). ME null subjects may be a comparable case. Therefore, we do not wish to accurately specify a period in our data in which null subjects are ungrammatical. However, as demonstrated by figure 5, first- and second-person pronouns are null at relative frequencies of $< 1.1\%$ in all ME periods. From the M2 period, that frequency is just slightly above 0 per cent. Third-person subject pronouns are null at a frequency of *c*. 2.5 per cent in the M1 period, but the figure drops to *c*. 1 per cent in the M2 period and to < 0.5 per cent after that. This means that M1 is the only period in which a subset of the null subjects occur at frequencies above a well-established quantitative cut-off point for grammaticality. At the least, then, there appears to be little reason to claim that null subjects persist to any quantitatively notable degree beyond the M1 period.[31]

In conclusion, constructions with overt pronominal subjects are clearly the productive ones at all stages of ME, including M1; more than 97.5 per cent of all third-person subject pronouns are overtly realized in this period. Thus, while we in this article have identified a number of linguistic and extra-linguistic variables favouring

[29] See also Santorini (1989), Pintzuk (1991) and Bies (1996).

[30] See also Lass' (1997: 382) notion of *residue*.

[31] Moreover, at least some late examples may be instances of 'diary drop', which is available even in Present-day English in the appropriate register (Haegeman 1990; Haegeman & Ihsane 1999; Weir 2008).

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

null subjects to a statistically significant degree (i.e. third-person reference, singular number, occurrence in conjunct clauses and the genre of poetry), it is difficult to construct a predictively satisfactory model accounting both for the distributional facts and for the state of affairs that subject pronouns are exceedingly rarely null in ME. That is, even though null subjects tend to be third person and even though null subjects tend to occur in main/conjunct clause environments in both OE and ME, null subjects occur at such low rates in both periods (see also Rusten (2016), which gives a statistical analysis of null subjects in 181 OE texts and text excerpts) that it is impossible to predict realistically when a third-person subject pronoun occurring in a main or conjunct clause may be null. This recalls Mitchell (1985: I, 633), who notes that overt and null pronominal subjects occur apparently haphazardly in 'what appear to us parallel situations' in OE. Nevertheless, null subjects in ME behave in a similar way to those in OE, and also to those in other early Northwest Germanic languages (see Walkden 2014 for a cross-Germanic perspective). This would suggest that the variables we have identified here did have special status in sanctioning null subjects at a (poetry aside, largely unattested) stage of English where null subjects presumably were more frequent.

As shown in section 3, then, the most salient finding is that null subjects are overwhelmingly dispreferred in ME as compared to the overt variant, irrespective of clause type, person, number, period, genre and dialect. Our main purpose in this article has been to present the relevant ME data and analyse them statistically. We will note, however, that descriptively, at least, the distribution we observe can be captured by positing two probabilistically weighted grammars in competition, in the sense of Kroch (1989, 1994, 2001): one which permits null subjects, and one which does not. What we witness in the ME and even OE periods may be the tail end of an S-curve in which, by the attested period, the null-subject grammar has dropped to a very low frequency. This ties in with the tentative conclusion of Walkden (2014: 224–6) that null subjects were more widely available in (unattested) earlier stages of Germanic. Moreover, the within-text effects of person and clause type might be diachronically constant effects in the sense of Kroch's (1989) Constant Rate Hypothesis, though we have not attempted to fit a logistic curve or establish the strength of these contextual effects, as the attested evidence is just too meagre.

5 Summary

We have presented a quantitative investigation of the occurrence and distribution of referential null subjects in Middle English prose and poetry. Supplementing previous inquiries into null subjects in Old English, we have aimed to address several gaps in existing knowledge. The article has given a wide-ranging survey of 58,024 null and overt pronominal subjects in 92 Middle English prose and verse texts. On the basis of these data, we have investigated some of the effects found to condition subject omission in Old English. These include the linguistic effects of clause type, person and number and the extralinguistic effects of genre/text type, period and dialect. We have

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

assessed the significance of our findings and the relative importance of each of the investigated variables using generalized mixed-effects logistic regression modelling and random forests.

Referential null subjects are very rare in Middle English. Even so, a number of statistically significant factors favouring subject omission were identified. As in Old English, Middle English null subjects are favoured in the third person, as well as in conjunct clauses and to a lesser degree in main clauses. These facts could be captured if it is assumed that null subjects are discourse-licensed, as in Walkden (2013) and Håkansson (2013), and also if conjunction reduction operated regardless of grammatical function or case, as suggested for Old Norse by Faarlund (1990). Also as in Old English, poetry permits null subjects to a greater extent than prose. Moreover, apart from the fact that null subjects occur more frequently in poetry, it could be observed that such subjects are similarly distributed across variables such as clause type, person and number in both genres. A modest, but clear, diachronic decline could be observed from early to late Middle English.

Finally, we noted that the distribution we observe can be captured descriptively by positing two grammars in competition, in the sense of Kroch (1989, 1994, 2001). The patterns observed over the attested Old and Middle English texts suggest that we are witnessing the very end of an S-curve, and that in the Middle English period the null subject option peters out completely, in favour of the rigid requirement for an overt subject that we see today.

Authors' addresses:

School of Arts, Languages and Cultures
University of Manchester
Oxford Road
Manchester M13 9PL
United Kingdom
george.walkden@manchester.ac.uk

Department of English
Bergen University College
Inndalsveien
PO Box 7030
5020 Bergen
Norway
kristian.andersen.rusten@hib.no

References

Adams, Marianne. 1987. From Old French to the theory of pro-drop. *Natural Language and Linguistic Theory* 5, 1–32.

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

Allen, Cynthia. 1995. *Case marking and reanalysis: Grammatical relations from Old to Early Modern English*. Oxford: Oxford University Press.
Axel, Katrin. 2007. *Studies on Old High German syntax: Left sentence periphery, verb placement and verb-second*. Amsterdam: John Benjamins.
Baayen, R. Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
Berndt, Rolf. 1956. *Form und Funktion des Verbums im nördlichen Spätaltenglischen*. Halle: Max Niemeyer.
Bies, Elizabeth A. 1996. Syntax and discourse factors in Early New High German: Evidence for verb-final word order. MA thesis, University of Pennsylvania.
Cedergren, Henrietta & David Sankoff. 1974. Variable rules: Performance as a statistical reflection of competence. *Language* 50, 333–55.
Coppess, Emily & Acrisio Pires. 2013. The residue of syntactic change: Partial pro-drop in Old English. Paper presented at the LSA Summer Institute's Diachronic Syntax Workshop, Ann Arbor.
Denison, David. 2003. Log(ist)ic and simplistic S-curves. In Raymond Hickey (ed.), *Motives for language change*, 54–70. Cambridge: Cambridge University Press.
Einenkel, Eugen. 1916. *Geschichte der englischen Sprache II: Historische Syntax*. Strasbourg: Karl J. Trübner.
Faarlund, Jan Terje. 1990. *Syntactic change: Toward a theory of historical syntax*. Berlin: Mouton de Gruyter.
Fischer, Olga, Ans van Kemenade, Willem Koopman & Wim van der Wurff. 2000. *The syntax of Early English*. Cambridge: Cambridge University Press.
Gelderen, Elly van. 2000. *A history of English reflexive pronouns: Person, self and interpretability*. Amsterdam: John Benjamins.
Gelderen, Elly van. 2013. Null subjects in Old English. *Linguistic Inquiry* 44, 271–85.
Haegeman, Liliane. 1990. Understood subjects in English diaries. *Multilingua* 9, 157–99.
Haegeman, Liliane & Tabea Ihsane. 1999. Subject ellipsis in embedded clauses in English. *English Language and Linguistics* 3, 117–45.
Håkansson, David. 2013. Null referential subjects in the history of Swedish. *Journal of Historical Linguistics* 3(2), 155–91.
Hall, Joseph. 1995. *Selections from Early Middle English 1130–1250*, part I. Oxford: Clarendon.
Hothorn, Torsten, Kurt Hornik, Carolin Strobl & Achim Zeileis. 2015. *Party: A laboratory for recursive partytioning*. http://cran.r-project.org/web/packages/party/index.html.
Hulk, Aafke & Ans van Kemenade. 1995. V2, pro-drop, functional projections and language change. In Adrian Battye & Ian G. Roberts (eds.), *Clause structure and language change*, 227–56. Oxford: Oxford University Press.
Jenset, Gard B. 2010. A corpus based study of the evolution of *there*: Statistical analysis and cognitive interpretation. PhD thesis, University of Bergen.
Johnson, Daniel Ezra. 2009. Getting off the *GoldVarb* standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass* 3, 359–83.
Johnson, Daniel Ezra. 2010. *Rbrul manual*. Online. www.danielezrajohnson.com/Rbrul_manual.html/.
Kroch, Anthony. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1, 199–244.
Kroch, Anthony. 1994. Morphosyntactic variation. In Katherine Beals (ed.), *Proceedings of the 30th annual meeting of the Chicago Linguistic Society*, 180–201. Chicago: Chicago Linguistic Society.

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

Kroch, Anthony. 2001. Syntactic change. In Mark Baltin & Chris Collins (eds.), *The handbook of contemporary syntactic theory*, 699–729. Malden, MA: Blackwell.
Kroch, Anthony, Beatrice Santorini & Ariel Diertani. 2004. *Penn–Helsinki Parsed Corpus of Early Modern English*. www.ling.upenn.edu/hist-corpora/PPCEME-RELEASE-2/index.html.
Kroch, Anthony & Ann Taylor. 2000. *Penn–Helsinki Parsed Corpus of Middle English Prose*. www.ling.upenn.edu/hist-corpora/PPCME2-RELEASE-3/index.html.
Laing, Margaret & Roger Lass. 2008. *The linguistic atlas of Early Middle English*, version 3.2. www.lel.ed.ac.uk/ihd/laeme2/laeme2.html.
Lass, Roger. 1997. *Historical linguistics and language change*. Cambridge: Cambridge University Press.
McIntosh, Angus, M. L. Samuels & Michael Benskin. 1986. *A linguistic atlas of Late Middle English*. Aberdeen: Aberdeen University Press.
Mitchell, Bruce. 1985. *Old English syntax*, 2 vols. Oxford: Clarendon.
Mustanoja, Tauno F. 1960. *A Middle English syntax*, part 1: *Parts of speech*. Helsinki: Société Néophilologique.
Nagucka, Ruta. 1997. Glossal translation in the Lindisfarne Gospel according to Saint Matthew. *Studia Anglica Posnaniensia* 31, 179–201.
Ohlander, Urban. 1943. Omission of the object in English. *Studia Neophilologica* 16, 105–27.
Ohlander, Urban. 1981. Notes on the non-expression of the subject pronoun in Middle English. *Studia Neophilologica* 53, 37–49.
Pintzuk, Susan. 1991. Phrase structures in competition: Variation and change in Old English word order. PhD thesis, University of Pennsylvania.
Pogatscher, Alois. 1901. Unausgedrücktes Subjekt im Altenglischen. *Anglia* 23, 261–301.
Randall, Beth, Anthony Kroch & Ann Taylor. 2005–13. *CorpusSearch 2*. Online. http://corpussearch.sourceforge.net/.
Rissanen, Matti, Merja Kytö, Leena Kahlas-Tarkka, Matti Kilpiö, Saara Nevanlinna, Irma Taavitsainen, Terttu Nevalainen & Helena Raumolin-Brunberg. 1991. *Helsinki Corpus of English Texts*. www.helsinki.fi/varieng/CoRD/corpora/HelsinkiCorpus/.
Rosenkvist, Henrik. 2009. Referential null subjects in Germanic: An overview. *Working Papers in Scandinavian Syntax* 84, 151–80.
Rusten, Kristian A. 2010. A study of empty referential pronominal subjects in Old English. MA thesis, University of Bergen.
Rusten, Kristian A. 2013. Empty referential subjects in Old English prose – a quantitative analysis. *English Studies* 94(8), 970–92.
Rusten, Kristian A. 2014. Null referential subjects from Old to early Modern English. In Kari E. Haugland, Kevin McCafferty & Kristian A. Rusten (eds.), *'Ye whom the charms of grammar please': Studies in English language history in honour of Leiv Egil Breivik*, vol. 4, 249–70. Oxford: Peter Lang.
Rusten, Kristian A. 2015. A quantitative study of empty referential subjects in Old English prose and poetry. *Transactions of the Philological Society* 113(1), 53–75.
Rusten, Kristian A. 2016. Empty referential subjects in Old English. PhD thesis, University of Bergen.
Sankoff, David, Sali A. Tagliamonte & Eric Smith. 2005. *Goldvarb X: A variable rule application for Macintosh and Windows*. http://individual.utoronto.ca/tagliamonte/goldvarb.html.
Sankoff, David, Sali A. Tagliamonte & Eric Smith. 2012. *Goldvarb Lion*. http://individual.utoronto.ca/tagliamonte/goldvarb.html.

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

Santorini, Beatrice. 1989. The generalization of the verb-second constraint in the history of Yiddish. PhD thesis, University of Pennsylvania.
Santorini, Beatrice. 1992. Variation and change in Yiddish subordinate clause word order. *Natural Language & Linguistic Theory* 10, 595–640.
Strang, Barbara M. H. 1970. *A history of English*. London: Methuen.
Strobl, Carolin, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin & Achim Zeileis. 2008. Conditional variable importance for random forests. *BMC Bioinformatics* 9, 307.
Strobl, Carolin, James Malley & Gerhard Tutz. 2009. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods* 14, 323–48.
Tagliamonte, Sali A. & R. Harald Baayen. 2012. Models, forests and trees of York English: *Was*/*were* variation as a case study for statistical practice. *Language Variation and Change* 24, 135–78.
Thomson, Clara L. 1907. Later Transition English: Legendaries and chroniclers. In Adolphus W. Ward & Alfred R. Waller (eds.), *The Cambridge history of English literature*, vol. 1: *From the beginnings to the cycles of romance*, 374–401. Cambridge: Cambridge University Press.
Visser, Frederik Theodor. 1963. *An historical syntax of the English language*, vol. 1. Leiden: E. J. Brill.
Walkden, George. 2013. Null subjects in Old English. *Language Variation and Change* 25, 155–78.
Walkden, George. 2014. *Syntactic reconstruction and Proto-Germanic*. Oxford: Oxford University Press.
Walkden, George. 2016. Null subjects in the Lindisfarne Gospels as evidence for syntactic variation in Old English. In Julia Fernández Cuesta & Sara M. Pons Sanz (eds.), *The Old English glosses to the Lindisfarne Gospels: Language, author and context* (Buchreihe der Anglia), 237–54. Berlin: Mouton de Gruyter.
Wallage, Phillip. 2005. Negation in early English: parametric variation and grammatical competition. PhD thesis, University of York.
Weir, Andrew. 2008. Subject pronoun drop in informal English. MA thesis, University of Edinburgh.
Williams, Alexander. 2000. Null subjects in Middle English existentials. In Susan Pintzuk, George Tsoulas & Anthony Warner (eds.), *Diachronic syntax: Models and mechanisms*, 164–87. Oxford: Oxford University Press.
Zimmermann, Richard. 2012. Variably overt and empty expletives with finite (and non-finite) clauses in early English. Paper presented at the International Congress of English Historical Linguistics 17 (ICEHL 17), Zurich.
Zimmermann, Richard. 2013. The Geneva Corpus of Middle English Poetry: Its construction and possible applications. www.old-engli.sh/myres%20documents/GeCMEP.pdf.
Zimmermann, Richard. 2015. *The Parsed Corpus of Middle English Poetry*. www.pcmep.net/index.php.

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

Appendix

Table A1 *Random intercepts by text*

text	intercept	tokens	n/n+y	centred factor weight
WynWas	2.465	199	0.04	0.914
CMTHORN	2.069	113	0.027	0.877
TreatDreams	2.026	68	0.088	0.872
CMHALI	1.747	370	0.046	0.838
CMTRINIT	1.59	1,505	0.041	0.815
CMMARGA	1.546	436	0.032	0.809
CMKENTSE	1.54	171	0.047	0.808
CMKATHE	1.539	452	0.033	0.808
CMJULIA	1.406	428	0.03	0.786
CMHORSES	1.387	239	0.008	0.783
LetterCupid	1.209	185	0.016	0.751
CMPOLYCH	1.166	1,445	0.005	0.743
CMGREGOR	1.152	883	0.005	0.74
Husbandman	0.938	37	0.027	0.697
CMANCRIW-2	0.904	687	0.019	0.69
CMANCRIW-1	0.885	2,071	0.019	0.686
CMFITZJA	0.844	145	0.007	0.677
BirdFoFe	0.732	89	0.011	0.652
CMMIRK	0.723	2,798	0.004	0.65
Bestiary	0.716	331	0.051	0.648
CMEDVERN	0.654	660	0.003	0.634
PoemaMo	0.628	235	0.026	0.628
CMLAMB1	0.617	240	0.021	0.625
CMHILTON	0.584	200	0.005	0.618
CMBRUT3	0.553	1,935	0.004	0.61
CMEDMUND	0.479	211	0.005	0.593
DavyDreams	0.425	73	0.014	0.579
LordOneGod	0.369	145	0.034	0.566
CMVICES4	0.339	384	0.003	0.558
CMSAWLES	0.276	205	0.01	0.543
CMBENRUL	0.233	1,051	0.002	0.532
ThruNight	0.225	100	0.01	0.53
CMAYENBI	0.183	1,808	0.014	0.52
DisMaryCross	0.173	134	0.007	0.517
DameSirith	0.172	236	0.004	0.517
FoxWolf	0.113	166	0.006	0.502
HowHearMass	0.102	327	0.003	0.499
CMPETERB	0.09	275	0.015	0.496
IacoIose	0.088	313	0.006	0.496
OrisonLord	0.064	58	0.017	0.49
CMMALORY	0.042	3,510	0.001	0.484
CMEDTHOR	− 0.021	785	0.001	0.469
LofsongLord	− 0.041	88	0.011	0.464

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press

Table A1 *Continued*

BodySoul	– 0.042	222	0.014	0.463
ProvAlf	– 0.048	216	0.023	0.462
Fridesw	– 0.055	117	0.009	0.46
CMCAPSER	– 0.067	68	0	0.457
CMINNOCE	– 0.102	119	0	0.449
OrisFiveJoys	– 0.107	33	0	0.447
CMLAMBX1	– 0.112	1,038	0.008	0.446
CMEQUATO	– 0.122	118	0	0.444
OwlNight	– 0.14	944	0.01	0.439
CMASTRO	– 0.15	145	0	0.437
WorcFrag	– 0.151	5	0	0.436
CMGAYTRY	– 0.165	168	0	0.433
CMROYAL	– 0.207	272	0	0.423
TheGrave	– 0.217	14	0	0.42
CMMANDEV	– 0.222	2,003	0.001	0.419
WooingLord	– 0.234	174	0.011	0.416
CMREYNES	– 0.236	205	0	0.416
CMROLLEP	– 0.259	983	0.002	0.41
LofsongLady	– 0.269	24	0	0.408
CMJULNOR	– 0.291	329	0	0.402
CMCTPARS	– 0.315	1,087	0.001	0.397
CMAELR4	– 0.355	537	0	0.387
CMREYNAR	– 0.408	630	0	0.375
CMKEMPE	– 0.412	4,423	0.001	0.374
HarrowHell	– 0.42	164	0.006	0.372
PatNost	– 0.434	127	0.008	0.369
Simonie	– 0.448	220	0.005	0.365
CMBOETH	– 0.454	495	0	0.364
CMOTEST	– 0.462	436	0	0.362
CMSIEGE	– 0.471	448	0	0.36
CMVICES1	– 0.504	1,554	0.004	0.352
CMNTEST	– 0.527	843	0	0.347
LittleSerm	– 0.53	39	0	0.346
LoveRon	– 0.543	98	0.01	0.344
CMCTMELI	– 0.606	836	0	0.329
Maximian	– 0.613	110	0	0.328
Ureisun	– 0.627	106	0	0.325
SirCleges	– 0.654	282	0	0.319
CMCLOUD	– 0.661	812	0	0.317
CMPURVEY	– 0.727	977	0	0.303
ElevenPains	– 0.917	68	0	0.265
CMAELR3	– 0.934	711	0	0.261
CMEARLPS	– 0.938	2,367	0.001	0.261
CMROLLTR	– 1.093	825	0	0.232
CMCAPCHR	– 1.173	2,193	0	0.218
CMWYCSER	– 1.224	2,167	0	0.209
CMKENTHO	– 1.556	224	0.004	0.16
CMORM	– 2.216	2,234	0	0.089

https://doi.org/10.1017/S1360674316000204 Published online by Cambridge University Press