



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

The decline of local anchoring

Citation for published version:

Los, B, Dreschler, GA, Kemenade, AV, Komen, ER & Coretta, S 2023, 'The decline of local anchoring: A quantitative investigation', *English Language and Linguistics*. <https://doi.org/10.1017/S1360674323000047>

Digital Object Identifier (DOI):

[10.1017/S1360674323000047](https://doi.org/10.1017/S1360674323000047)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

English Language and Linguistics

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



The decline of local anchoring: a quantitative investigation¹

BETTELOU LOS 

The University of Edinburgh

GEA DRESCHLER 

Vrije Universiteit Amsterdam

ANS VAN KEMENADE

Radboud University Nijmegen

ERWIN KOMEN 

Radboud University Nijmegen

and

STEFANO CORETTA

The University of Edinburgh

(Received 27 September 2022; revised 18 January 2023)

This article presents a quantitative study of the referential status of PPs in clause-initial position in the history of English. Earlier work (Los 2009; Dreschler 2015) proposed that main-clause-initial PPs in Old English primarily function as ‘local anchors’, linking a new clause to the immediately preceding discourse. As this function was an integral part of the verb-second (V2) constraint, the decline of local anchors was attributed to the loss of V2 in the fifteenth century, so that only the contrasting and frame-setting functions of these PPs remain in PDE. This article tests these hypotheses in the syntactically parsed corpora of OE, ME, EModE and LModE texts, using the Pentaset-categories (*New*, *Inert*, *Assumed*, *Inferred* or *Identity*; Komen 2011), based on Prince’s categories (Prince 1981). The finding is that *Identity* clause-initial PPs decline steeply from early ME onwards, which means the decline pre-dates the loss of V2. A likely trigger is the loss of the OE paradigm of demonstrative, which functioned as standalone demonstrative pronouns as well as demonstrative determiners, and the loss of gender marking more generally. From EModE onwards, main-clause-initial PPs that still link to the preceding discourse do so much more indirectly, by an *Inferred* link.

Keywords: English historical syntax, verb-second, demonstrative pronouns, information structure, local anchors

¹ We would like to thank Laurel Brinton and Kristin Bech, as well as an anonymous reviewer, for valuable feedback on an earlier version of this article. We also would like to acknowledge the support of the Netherlands Organization for Scientific Research (NWO), grant 360-70-370.

1 Introduction

A considerable body of work over the last decade or so has shown that word order in Old English (OE) is more flexible than in later stages of the language and is partly determined by information structure. Information structure explains both some of the variation in object placement in OE (Taylor & Pintzuk 2012a, 2012b, 2014; Struik & van Kemenade 2020, 2022) and subject placement (Bech 2001; Hinterhölzl & Petrova 2010; van Kemenade, Milićev & Baayen 2008; van Kemenade & Milićev 2012). Information structure is also a key factor in scenarios for word order change after the OE period; specifically, Los (2009) argues that the loss of V2, which resulted in a subject-initial grammar, compromised information structure to such an extent that new structures emerged and others increased in frequency to compensate for the loss: stressed-focus *it*-clefts and cross-linguistically rare passives. Los further proposes that V2 is not just a syntactic constraint, but also entails a specific use of the clause-initial position in the case of non-subjects, so-called local anchoring (Los & Dreschler 2012). Consider the clause-initial PP *of þære* in (1):

- (1) & þa sealde he heom mid his agenre handa ane trywene flascan wines fulle,
 and then gave he them with his own hands a wooden bottle of-wine full
 to þon þæt hi mihton heom þa on heora færeldre to underngeweorce habban.
 to that that they might for-themselves that in their journey for breakfast have
Of þære hi druncon, oð þæt hi to rauennam becomon.
 from that they drank, until-that they to Ravenna came
 ‘and then, with his own hands, he gave them a wooden bottle full enough for breakfast during
 their whole journey. They drank from it, and it lasted all the way to Ravenna.’ (GD_1_ [C]:9.66.12.742–743)

Of þære in (1) refers to the bottle, newly introduced in the preceding discourse and anchored ‘locally’ to the immediately preceding sentence. Note that this information in clause-initial position is not necessarily contrastive and/or emphatic, i.e. it is information-structurally unmarked. A sentence like (1) is difficult to replicate in Present-day English (PDE), as non-subjects in clause-initial position have become restricted to contrastive and frame-setting interpretations: adjuncts like *Of þære* could be unmarked themes in OE in the sense of Halliday (1967), while they are marked themes in PDE, where only subjects can be unmarked themes. Even if we assume that demonstratives as independent pronouns have more restricted functionality in PDE than in OE, this cannot by itself account for the infelicity of clause-initial *From that* in PDE, as the expanded alternative, *From that bottle*, would still be interpreted as having a specific emphasis that is not present in the OE sentence. Los (2009) argues that the loss of V2 in the fifteenth century led to a stricter mapping between syntactic function and information-structural status: subjects as the default expression of given information, including links to the preceding discourse, complements and objects as the default expression of new information. The pre-subject position, which remained available to objects and adjuncts, also acquired a specific, contrastive,

information-structural function, which was not compatible with non-contrastive local anchors. The subject appears to have taken over the local anchoring function. PDE allows subjects where V2 languages like Dutch or German would use adverbials, as is clear from e.g. Dutch–English translation studies like Hannay & Keizer (1993) or Lemmens & Parr (1995), which advise against a literal translation of examples such as (2a) (relevant items in bold):

- (2) (a) En **daarmee** was de tragedie van Bergkamp compleet.
 (b) And **with that** Bergkamp's tragedy was complete.
 (c) And that made Bergkamp's tragedy complete. (Hannay & Keizer 1993: 68)

The literal translation as in (2b) renders the Dutch clause-initial adverbial with a clause-initial adverbial in PDE, but is less felicitous than (2c), which reworks the adverbial into a subject expressing the link to the previous sentence. Similar observations can be found in comparative studies of PDE and German, another V2 language (Rohdenburg 1974: 11; Hawkins 1986: 58–61). Komen *et al.* (2014) provide quantitative evidence for this change in function of the subject in the history of English, whereas other studies connect the increased functional load of the subject to an increased and/or extended use of passives, middles and so-called permissive subjects (Los 2009, 2018; van Gelderen 2011; Los & Dreschler 2012; Dreschler 2015, 2020). An example of an extended use of the passive is the exceptional case-marking (ECM) construction with verbs of thinking and declaring like *report* in (3):

- (3) This mushroom is reported to have a lobster like flavor when cooked.
 (<http://caldwell.ces.ncsu.edu/2015/01/336308/>. Accessed 19 Oct 2015, via WebCorp, www.webcorp.org.uk)

Many of these verbs do not have an acceptable active counterpart (**They reported this mushroom to have a lobster like flavor when cooked*); Birner & Ward (2002) label the passive ECM-construction an 'information-packaging' device, its primary function being to allow discourse-old information to be expressed as a subject.

Examples of middles are (4)–(5) (relevant verb in bold):

- (4) Speaker A: It seems to be impressing our American friend.
 Speaker B: Americans **impress** easily (*Inspector Lewis*, series 2, episode 'And the moonbeams kiss the sea', quoted in Dreschler 2015: 372)
- (5) I think [Kate Middleton] **photographs** better than she **knits**. (*Graham Norton Show*, series 9, episode 1, 15 April 2011, quoted in Dreschler 2015: 372)

Note that these verbs are not used with their default valencies: their subjects are not agents but patients. The productivity of middle formation in English is particularly exemplified in (5), meaning, as it does, 'Kate Middleton is more photogenic than she is knittogenic' (in the context of an enthusiastic knitter who knits effigies of members of the royal family). Key in these developments is the increased functional load of the subject that is at least

partly the result of the loss of discourse links expressed by initial adjuncts – what we will call ‘local anchoring’.

This article will build a quantitative case for the loss of local anchoring by analysing clause-initial PPs in the history of English, annotated with quite minimal information: just their referential status. The goal is to uncover the decrease of local anchoring, and discover what restricted its use in the later periods. The extraordinary thing about these adjuncts is that even though their frequency was greatly reduced over time, they were not completely lost.

2 Background

Following van Kemenade (1987), OE has generally been analysed as a V2 language, meaning that – as in present-day languages such as Dutch and German – the finite verb occurs in second position in main clauses. Subject-finite verb inversion was categorical in some contexts such as questions, negative-initial clauses, and clauses introduced by a temporal adverb *þa* or *þonne* ‘then’. But V2 in OE differed from the V2 pattern as it is found in present-day V2 languages. Outside the categorical contexts, OE shows alternation between V2 and V3, with pronominal subjects typically found with the latter pattern (as in (6)), and nominal subjects (as in (7)) with the former (subjects in bold, finite verb underlined):

(6) *Æfter þysum wordum heo to-tær hyre gewædu*
 after these words she tore her garments
 ‘After these words she tore her garments’
 (Ælfric’s *Lives of Saints*, Eugenia, l.233)

(7) *On þam dagum wæs Alexander geboren on Crecum...*
 in those days was Alexander born in Greece
 ‘At that time, Alexander was born in Greece ...’
 (Orosius, 104.21; Haerberli 2002: 245, his 1d)

The consensus in the literature appears to be that the finite verb in OE only moves to C when the first constituent is a question-word, a negated constituent, or a member of a restricted set of adverbs (*þa* or *þonne* ‘then’, *swa* ‘so’ or *þus* ‘thus’), while other types of first constituents – like the adjuncts in (6) and (7) – trigger movement of the finite verb to a lower head (AgrS in Haerberli 2002; F in van Kemenade & Westergaard 2012: 91), with the pronominal subjects moving to the specifier of that head, while full NP subjects remain in the default subject position. This means that (6) and (7) both have the finite verb moved into the lower head (AgrS/F) rather than C, even though their surface order is different. In recent years, this variation has been reassessed in terms of its information-structural properties: the variation in (6) and (7) is to a large extent governed by the information status of the subject, with new subjects generally following the verb and given or discourse-old subjects preceding the verb (Bech 2001; van Kemenade 2012; van Kemenade & Westergaard 2012). In fact, Hinterhölzl & Petrova (2010) and Los (2012) propose that in sentences such as these, movement of

the finite verb in OE to the lower AgrS/F position may originally have arisen as a marker of information structure, separating given or background elements from new elements, while movement to the higher position, C, may have arisen as a way to mark off a focus domain. Hinterhölzl & Petrova (2010: 319) even call V2 order in OE an ‘accident’: a surface effect which happens when there is only one background element.

Information structure also plays a role in the V2 system in another respect – and this is our main focus in this article. As Los (2009) proposes, following Halliday (1967), the subject in PDE is the unmarked theme, and adverbials and complements are only chosen as themes, i.e. as initial elements, for contrast and emphasis. The following examples illustrate typical clause-initial non-subjects in PDE (in bold).

(8) **This** I can play.

(9) **In the latest book**, Asterix and Obelix are tasked with protecting Adrenaline from the Romans, who have captured her father.

(Both examples from guardian.co.uk, accessed 25 October 2019)

In (8), the object is fronted, and is contrastive, which is appropriate, as the immediately preceding text describes failed attempts at other games. Example (9) illustrates frame-setters, which – in Krifka’s (2008: 55) definition – function to ‘set the frame in which the following expression should be interpreted’. The proposition in (9) holds for the domain of the latest book in the series. Krifka points out that an important aspect of frame-setting is that there is a sense of implied alternatives when frame-setters are used, which means that they are contrastive or, in Krifka’s terms, focused. This fits example (9): it is in this book and not the earlier books that Asterix and Obelix are protecting Adrenaline. Fronted objects and frame-setters both represent cases where there is a clear motivation for deviating from the unmarked option of a subject as theme. Biber *et al.*’s (1999: 772) extensive corpus data on adverbials provide further support for this observation that non-subjects are marked: even though adverbials can occur in initial position, they most commonly occur in final position, and much less frequently in initial position. It is especially circumstance adverbials that are ‘very marked’ in initial position (1999: 803); example (10) illustrates the use of such a circumstance adverbial, with the next sentence indicating a contrasted option (both adjuncts in bold). Stance adverbials, as in (11), and linking adverbials especially, as in (12), are more frequent in initial position.

(10) **In many cultures**, the practice of abstaining entirely from animal produce has an established history: with their belief systems rooted in nonviolence, many Rastafarians, followers of Jainism and certain sects of Buddhism have been swearing off meat, fish, eggs and dairy for centuries. **In large swathes of the west**, though, public awareness of what veganism actually entails has been sketchy.

(11) **Needless to say**, Johnson wants to push ahead, without installing safeguards.

(12) **In contrast**, it’s unusual for disabled actors to be hired to play characters who aren’t defined by their disability [...].

(All examples from guardian.co.uk, accessed 28 October 2019)

In OE, however, non-subjects in initial position are more versatile than they are in PDE, syntactically as well as information-structurally. In particular, fronted objects and clause-initial PPs are less restricted. Speyer (2010) identifies a type of object fronting of a demonstrative pronoun in OE which – unlike PDE object fronting – is anaphoric; it does not present new information and is not contrastive, but just serves as a link to the preceding discourse, which has introduced the antecedent, *se frofergast* ‘the Holy Spirit’.

- (13) **Þone** asende se Sunu
 this sent the son
 ‘The son sent this one’ (coaelhom, Ehom_9:114.1350; Speyer 2010: 38)

Dreschler (2015) and Bech (2014) provide similar examples for clause-initial PPs:

- (14) **In ðisses cninges rice** se eadiga ærcebiscep Scs Laurentius forðferde
 in this.GEN king-GEN reign the blessed archbishop Saint Laurentius departed
 ‘In this king’s reign, the blessed archbishop St Laurentius died’
 (Bede 116; Bech 2014: 523)

The fronted object in (13) and the PP in (14) are typical examples of local anchoring: there is a local link to the immediately preceding discourse. Crucially, these local anchoring phrases in initial position appear to be information-structurally unmarked, in the sense that they do not necessarily receive any special emphasis or indicate contrast or focus.

Some aspects of the decline of local anchoring have been addressed in earlier work. Los & Dreschler (2012) and Dreschler (2015) present data that show a decrease in adverbials and objects in first position, while there is an increase in clause-initial subjects from the OE period through to the Late Modern English (LModE) period. The decline of clause-initial PPs that encode given rather than new information from late Middle English (ME) onwards is charted by Pérez-Guerra (2005: 357ff.). The discourse-linking aspects of the initial position in earlier periods is studied by Los & Dreschler (2012), Bech (2014) and Dreschler (2015), who all show a decline in clause-initial prepositional phrases with anaphoric elements or based on information status. While these earlier studies lend support to the overall hypothesis about the decline of local anchoring, the nature of the problem – involving information structure rather than syntax – makes it difficult to pinpoint which period is particularly significant. Manual information-structural annotation leads to small datasets; scholars also differ in how they operationalize the relevant information-structural notions and work on different texts. The present study aims to contribute to the testing of the hypothesis by presenting data from a large-scale corpus study with a single consistent system for information status coding.

3 Methodology

We will investigate PPs that occur as initial constituents in main clauses, using the syntactically parsed corpora of historical English: *The York–Toronto–Helsinki Parsed*

Corpus of Old English Prose (YCOE; Taylor *et al.* 2003); *The Penn–Helsinki Parsed Corpus of Middle English*, 2nd edition (PPCME2; Kroch & Taylor 2000); *The Penn–Helsinki Parsed Corpus of Early Modern English* (PPCEME; Kroch *et al.* 2004) and *The Penn–Helsinki Parsed Corpus of Modern British English* (PPCMBE; Kroch *et al.* 2010). The earlier subperiods of the Middle English corpus, the PPCME2, are somewhat problematic: many of the texts in the first subperiod, M1 (c.1150–1250), are copies or adaptations of OE texts, or have such a complex history that the dating of the manuscript is unlikely to bear any correspondence to the actual date of composition (see e.g. De Bastiani 2022: 121–2 for discussion and further references), while for the second Middle English subperiod, M2 (c.1250–1350), there is very little textual material. This notorious M2 data gap is unfortunate, as this appears to be an important period for syntactic change (see e.g. Truswell *et al.* 2019: 21–2).

The hypothesis is that OE will have higher frequencies than later periods of NPs inside clause-initial PPs that present ‘given’ information by referring back to a referent in the immediately preceding discourse, and lower frequencies of NPs that are ‘new’ information, such as frame-setters that demarcate the context in which the following proposition is true. To test the hypothesis, all clause-initial PPs of the Penn–Helsinki corpora were annotated following the Pentaset-annotation scheme (Komen, Los & van Kemenade 2023), marking the NP inside the PP as having a previously mentioned referent or not; and if yes, what the relationship is between the NP and that referent. The Pentaset categories (figure 1) build on Prince’s (1981) categories, with some streamlining.

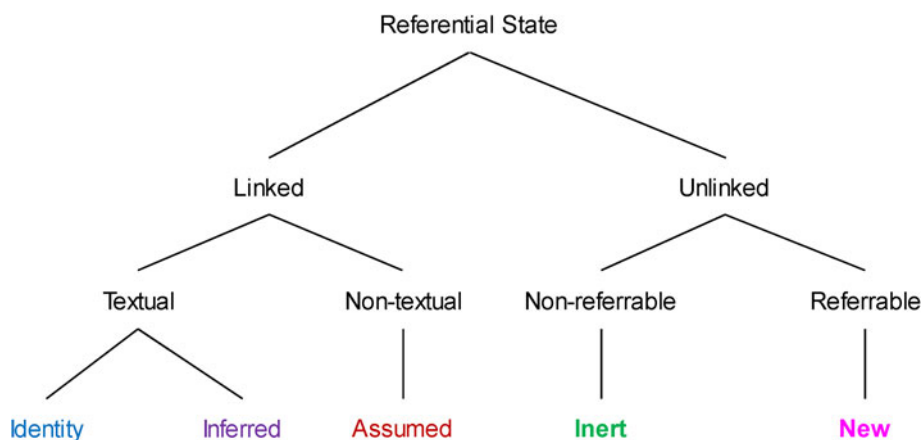


Figure 1. The referential state primitives in the Pentaset (Komen, Los & van Kemenade 2023)

The first distinction is between NPs with and without antecedents, i.e. linked or unlinked. If unlinked, the model allows for a category *inert* for NPs that do not in fact refer to or introduce a referent, as they function as attributes of other entities (like *a doctor* in *She is a doctor* or *She trained as a doctor*) – they are discursively inert. Inert items are most typically bare nouns. In the context of a historical corpus, we can also

think of NPs inside PPs that are in the process of grammaticalization, like *cause* inside *be cause* ‘because’, or *stead* inside *in stead* ‘instead’ as discursively inert. Any NP that is unlinked but does refer to or introduce a referent is given the status *new*.

If linked, the model makes a further distinction between referents mentioned or merely implied in the previous discourse. Those with textual antecedents are separated into cases of *identity* or cases with the status *inferred*. In cases of *identity*, the NP and its antecedent refer to the same referent; an example would be *þære* ‘that’ in (1), referring to *flasce* ‘bottle’ of the previous discourse. The status *inferred* is more indirect in that there is no exact match with a previously mentioned referent, but the referent’s identity can nevertheless be inferred from an evoked schema – once *a car* is mentioned, we can talk about *the driver*; once *a house* is mentioned, we can talk about *the windows*, in which case *driver* and *windows* would be linked to the earlier mentions of a car and a house, respectively, with the status of the link marked as *inferred*. This is Prince’s category of *inferrable*. A subtype of inferrable referents are ‘containing inferrables’ (Prince 1981), which can be inferred from information within the same NP they occur in. The NP *ðisses cyninges rice* ‘this king’s reign’ in (14) is an example, since the head noun *rice* is one of the possible inferences that can be made from the referent *ðisses cyninges* which has an antecedent in the previous discourse. Referents that cannot be linked to any referent mentioned in the previous discourse, whether with a status of *identity* or *inferred*, but have an extra-textual antecedent that can nevertheless be assumed to be in the common ground, are labelled as *assumed*.

The five statuses as a set (the Pentaset) inform an annotation scheme that minimizes ambiguity and hence promotes inter-rater agreement. The Pentaset-annotation scheme allows for a degree of (semi-)automatic annotation: NPs as subject or object complements can receive the status *inert* on the basis of their syntactic function; first- and second-person pronouns are always ‘given’ in the situational context of any discourse, so they receive the status *assumed*. In direct speech within a narrative, however, where they have referents in the discourse outside that direct speech, they will be linked to those referents.

Even in cases where automatic detection of information-structural statuses is not possible, the fact that the annotation scheme is added to a corpus that has already been enriched with syntactic information means that any of Prince’s (1981) categories that are not included in the Pentaset, like her ‘brand new anchored’ versus ‘brand new unanchored’ statuses, can still be retrieved. ‘Brand new unanchored’ refers to cases like *ane trywene flascan wines fulle* ‘a wooden bottle full of wine’ in (1); in our annotation scheme, *flasce* will not receive any of the three *linked* statuses (*identity*, *inferred* or *assumed*) or the *inert* status, so will be *new*. The NP contains a postmodifier (*wine fulle*) but this is not an anchor in the sense of Prince because it will not have received a *linked* status, either. Prince’s ‘brand new anchored’, on the other hand, which would be the status of, for example, *guy* in her example *a guy I work with* (Prince 1981: 236) will be retrievable because of the combination of the information statuses (*new* for *guy*, *identity* for *I*) and its syntactic annotation: *guy* is anchored because the following postmodifier, the relative clause, contains an anchor (*I*) with a *linked* status (*identity*).

The (semi-)automatic annotation tries to keep any appeal to syntactic function or morphological form to a minimum, because categories cannot be assumed to have the same functionality over time; for instance, demonstrative pronouns double as demonstrative determiners in OE, while these two morphological categories have distinct information-structural functions in PDE (see e.g. Gundel, Hedberg & Zacharsky 1993).

Note that the Pentaset-annotation scheme does not employ information-structural notions like topic, focus or contrastiveness, and does not directly mark a PP as frame-setting or as containing a local anchor. Such notions require manual (as opposed to semi-automatic) annotation, which means that only small samples of text can be investigated. While notions of contrast or emphasis are clearly relevant to charting the development of local anchoring, they are difficult to operationalize in the absence of native-speaker judgements. There is also no (complete) consensus in the literature about how to define these notions; and it has become clear that even where there appears to be a consensus, as in the case of aboutness topics, these can be much more difficult to identify in languages other than PDE; Cook & Bildhauer (2011, 2013), for instance, report on annotation experiments of German, where the presence of local anchors and pronominal subjects in the same clause gave annotators two options for 'topic', leading to very low inter-rater scores.

The labelled bracketing files of the Penn–Helsinki parsed corpora were converted to xml format, and coreference information was added to the xml corpora. This annotation was done by means of CESAX (Komen 2011, 2012, 2013), which employs a semi-automatic algorithm to resolve antecedent identity automatically where possible. When in doubt, CESAX asks for user input, offering a pool of potential antecedents, evaluated against a ranked set of constraints. An evaluation of the performance of CESAX is provided in Komen (2012). CESAX automatically processed 54 per cent of the 3,083 NPs in the LModE text investigated, 5 per cent of which were found to be erroneous.² The human annotator agreed with about 40 per cent of the remaining suggestions, choosing other options for the remaining 60 per cent. The total success rate of the algorithm was 72 per cent (Komen 2012). The vast majority of the main-clause-initial PPs for this study were annotated by two annotators (Los and Komen), and a sample of 1,677 NPs within such PPs for the Early Modern English (EModE) part of the corpus were compared, using CESAX's inter-rater agreement calculator. Cohen's Kappa was 0.86 for antecedent agreement, and 0.95 for referential type agreement, which demonstrates that a consistent annotation was achieved.

Four texts (*Herbarium*, *Lacnunga*, *Leechdoms* and *Medicina de quadrupedibus*) were excluded because they mostly contain formulaic clause-initial PPs, which function as section headings ('Against gout:') and are not a constituent of the following clause.

² The reason for choosing an LModE text was to facilitate a comparison with a manual annotation experiment of CESAX's precursor CESAC (see again Komen 2012 for details), and this earlier experiment had been done with LModE so as to have a reasonable number of annotators. The expertise required to read OE or ME texts meant that the vast majority of those texts were annotated by a single annotator.

4 Results

The results of our corpus investigation are given in [table 1](#).³

Table 1. *The Pentaset status of NPs in main clause-initial PPs*

Period	New	Inferred	Inert	Assumed	Identity	TOTAL
OE	1,555	1,139	967	157	2,367	6,185
ME	1,211	1,015	351	72	1,419	4,068
EModE	2,021	1,174	421	118	1,208	4,942
LModE	431	1,124	776	128	538	2,997

With these high numbers, it is feasible to not just look at the four periods but to drill down to the individual centuries, as in [figure 2](#) – although we need to bear in mind the problems of coverage in the first two subperiods of ME signalled in section 3.

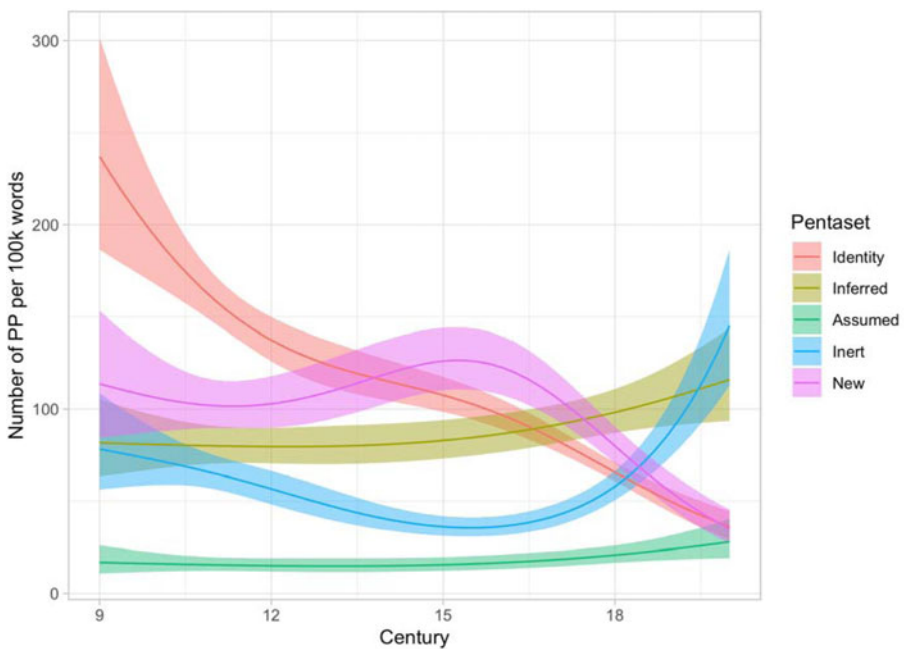


Figure 2. The Pentaset status of NPs in main-clause-initial PPs

³ The data can be accessed at this stable URL: <https://datashare.ed.ac.uk/handle/10283/4753>; data and code of [figures 2](#) and [3](#) can be found at <https://github.com/stefanocoretta/2020-penta>

For [figure 2](#), we fitted a generalized additive mixed-effects model to the number of prepositional phrases, using a Poisson distribution. We included the following terms (in parentheses, an explanation of how the term contributes to the model): Pentaset (Identity, Inferred, Assumed, Inert, New) as a parametric term (average number of prepositional phrases according to Pentaset), a smooth over century (9th, 10th, 11th, 12th, 13th, 14th, 15th, 16th, 17th, 18th, 19th) by Pentaset (change in number of prepositional phrases over century by Pentaset) and a by-text factor smooth over century (to account for variations between texts). An offset term was also included to account for the fact that length (in words) differed across texts. The reported estimates are the number of prepositional phrases assuming a text length of 100k words.

Our hypothesis about the diachronic development predicts a difference in the ratios between ‘given’ NPs (i.e. NPs with Pentaset statuses of *identity*, *inferred* and *assumed*) versus ‘new’ NPs within main-clause-initial PPs between all periods (as the rate of ‘given’ should show a persistent decline from OE onwards). This prediction is borne out for *identity*, the ‘heartland’ of the local anchor (like *Of þære* in (1)), as [figure 2](#) shows a clear decline, confirming the hypothesis that main-clause-initial PPs become less referential. The decline is not evenly spread over all prepositions, as [figure 3](#) shows, which contains the twelve most frequent prepositions over time, labelled by their PDE counterparts (so that *with* also includes OE *mid*, and *in* includes OE *binnan*).

Although a more detailed account of the variation must wait until a future investigation, we have a few suggestions for some of the fluctuations found in [figure 3](#). The high rates of

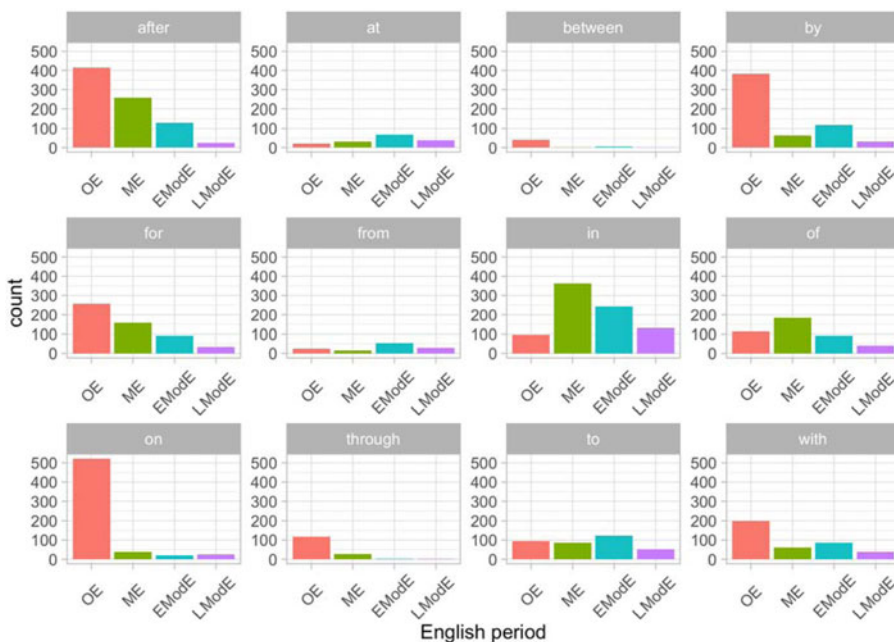


Figure 3. *Identity* with the most frequent prepositions

on in OE should be regarded in the light of the low rates of *in*, as a large share of the semantic field of ‘in’ was expressed by *on* at that period; OE *wip* means ‘against’ rather than ‘with’, but *wip* is rare in clause-initial position (only one instance of *identity*; twelve further instances with the status *new* occur in the medical texts that were excluded (*Herbarium*, *Lacnunga*, *Leechdoms* and *Medicina de quadrupedibus*) because these clause-initial PPs have heading-like functions and are followed by imperative rather than declarative clauses; see the end of section 3); so the majority of the *identity*-PPs included in the table for ‘with’ contain *mid* rather than *wip*. The preposition *by* does not mark ‘demoted’ agents in long passives in OE (which uses *fram* in this function); for an overview of the development of *by*, see Cuyckens (2002: 262–3). On inspection, the little bump in EModE for *identity* with this preposition links up with a finding from another EModE corpus discussed in Los & Lubbers (forthcoming), who note that 1700 appears to be something of a watershed, as the later texts in their corpus observe the flow of given to new information more diligently than the earlier texts, and are more geared to using the long passive in order to manoeuvre new information, expressed in the *by*-phrase, into clause-final position. In addition to a number of ‘demoted’ agents in clause-initial *by*-phrases, other items contributing to the EModE bump in figure 3 are of the type exemplified in (15) (*by*-phrase in bold):

- (15) **by this letter** it aperreth how carefull the lord president was to have the rebels thorowly prosecuted (perrott-e2-p1.d.1.p.1.s.138)

The previous context, which is the content of the relevant letter in full, does not contain anything to suggest that *by this letter* is used contrastively, in terms of evoking alternatives; this is a link to the previous discourse without any special marking, i.e. a proper local anchor, equivalent to a phrasing like *this letter shows* in PDE.

The category *inferred* holds its own and even increases slightly; so within the category ‘given’, there is a shift between the proportions of *inferred* and *identity*. *Inert* goes up from EModE onwards, reflecting the higher frequencies of PPs like *for example* and *in faith* in EModE, and of PPs like *at night*, *of course*, and *in fact* in LModE. *Assumed* PPs are a marginal phenomenon in all periods and do not show clear patterns of change through time, perhaps with the exception of a small increase between EModE and LModE.

Figure 2 also shows that the share of *new* referents gradually increases until EModE. We expect that frame-setters contain mostly new information, and with the decline of local anchors, new referents indeed become more prominent, and constitute a larger share of main-clause-initial PPs. A look at the data shows, however, that the bulge in EModE is largely due to the fact that the corpus contains a number of private diaries that structure their entries around time adverbials, bumping up the numbers of cases with *after* (*after private prayer*), *at* (*at noon*) and *in* (*in the morning*). Two-thirds of all new *after*-PPs, for instance, are from one particular diary (*Diary of Lady Margaret Hoby*, 1599–1605). A second factor is that the algorithm for *inert* looks for bare nouns, so that LModE *at last* is recognized as an *inert* PP but EModE *at the last* is not, nor is *in the meane while*. Even though some of the data around 1600 are somewhat skewed, the conclusion must be that there was indeed a shift in the information-structural status

of the clause-initial adjuncts at least until EModE: they were more likely to encode discourse-links rather than frame-setters in the earlier periods. Ultimately, however, *new* PPs decline, too.

The hypothesis links the decline of local anchors to the loss of V2 in the fifteenth century, and it predicts a clear and significant decline from ME to EModE, with fewer *identity* PPs and more *inferred* and *new* PPs. However, if V2 was the trigger, the significant decline of local anchors from early ME onwards is unexpected.

5 The functionality of standalone demonstratives

5.1 Introduction

We will turn our attention in this section to the initial decline of local anchoring in the transition from OE to ME, and tentatively link it to the loss of an articulated demonstrative paradigm, which had the dual function of independently used demonstrative pronouns and demonstrative determiners. We will only consider the pronominal use here, and call them ‘standalone demonstratives’ in what follows. The paradigm is given in [table 2](#).

Table 2. *The demonstrative paradigm in OE*

	Singular			Plural
	masculine	feminine	neuter	all genders
nom. sg	se	sēo	þæt	þa
gen. sg	þæs	þære	þæs	þāra
dat. sg	þæm	þære	þæm	þæm
acc. sg	þone	þā	þæt	þā
instr. sg	þy, þon			

It is important to note that the paradigm marks case, number and gender. The fact that the demonstrative could be used as an independent pronoun as well as a demonstrative determiner in OE significantly enhanced its referential potential; unlike case, which is determined at the level of the clause, grammatical gender is a stable feature of the noun, and persists from one clause to the next (cf. example (1) above, where *flasce* ‘bottle’, a feminine noun, is referred to by a standalone feminine demonstrative, two clauses down). The paradigm was lost early on in the transition from OE to ME.

The decline of local anchors is a combination of two factors: the decline of main-clause-initial adjuncts as a means of creating links to the previous discourse, a function that has been taken over by the subject (Los 2018), and a change in the referential functionality of standalone demonstratives (see next section). Standalone demonstratives were ultimately not replaced by an alternative system, although there

were contenders (*the same*, *there*+preposition, and personal pronouns, particularly the innovative use of (*h*)*it*); we will discuss these in section 5.3.

5.2 The antecedents of standalone demonstratives: from NP to stretch of discourse

This section zooms in on two prepositions in our corpus, *in* and *on*. *In* was selected because it is highly frequent in all periods and stable in its semantics, primarily denoting a container; *on* was selected because of the semantic overlap with *in* in Old English (see also the brief discussion of figure 3 above).

The antecedents of *this/these* and *that/those* in a main-clause-initial *in*-PP shows that there is a dramatic decline in standalone demonstratives with NP antecedents from ME onwards. Table 3 gives the numbers of standalone demonstratives and how many of them are true local anchors.

Table 3. Main-clause-initial *in*- and *on*-PPs in the corpus with standalone demonstratives: numbers of local anchors (identity – local – NP antecedent) (before slash) versus total numbers of such PPs (after slash)

Period	this	that	these	those	Totals
		with <i>in</i>			
OE	0/1	11/18	0/0	4/5	15/22
ME	1/15	3/10	2/2	0/0	6/27
EModE	3/15	1/2	1/1	0/0	5/18
LModE	0/14	0/2	3/3	0/0	3/19
		with <i>on</i>			
OE	3/5	31/56	0/0	3/7	37/68
ME	0/0	0/0	0/0	0/0	0/0
EModE	0/1	0/0	0/0	0/0	0/1
LModE	0/0	0/0	2/2	0/0	2/2
Totals	7/51	46/86	8/8	7/12	68/157

The default standalone demonstrative in OE as complement of main-clause-initial *in* and *on* is the *se* demonstrative, whose paradigm we presented in table 2. After OE, the numbers of standalone demonstratives with NP antecedents plummet, even more so if we realize that all the instances of *in these* in EModE and LModE, as well as one of the two instances in ME, refer to the same biblical sentence, here represented by its LModE incarnation:

- (16) Now there is in Jerusalem by the sheep gate a pool, which is called in Hebrew Bethesda, having five porches. **In these** lay a multitude of them that were sick, blind, halt, withered (erv-new-1881.d.1.p.1.s.331)

The two cases of *on these* in LModE refer to people, a reminder that the plural standalone demonstrative, unlike the singular, can still refer to human referents in PDE (Huddleston & Pullum *et al.* 2002: 1504). The antecedents of the four remaining local anchors proper in EModE, all with *in*, are presented in (17)–(20) (antecedent and demonstrative in bold):

- (17) **an old red cappe** with corners....and in **that** she appeared in the glasse
(gifford-e2-h.d.1.p.1.s.281)
- (18) **the 4th tale** ... in **this** is showne how flattery feedes them (armin-e2-p1.d.1.p.1.s.366)
- (19) **a neat, but small Dwelling** ... in **this** we designed to have worn some days away
(fryer-e3-p1.d.2.p.2.s.123)
- (20) **a very bright and strong reflection**... in **this** i could perceive the image of the window pretly well (hooke-e3-h.d.1.p.1.s.11)

In ME, there are no examples with *on*; there are five clear examples with *in* (if we discard the one instance of a variant of (16)):

- (21) **The bridde degre**....For in **þat** comeþ þe grace of þe Holy Goost
the third degree for in that comes the mercy of the Holy Ghost
doun in-to a soule
down into a soul
'The third degree ... that is the stage at which the mercy of the Holy Ghost descends into a soul' (cmhilton.d.4.p.4.s.107)
- (22) **Pacyens**..., for in **þat** schal ge kepyn gowr sowle.
patience ... for in that shall you keep your soul
'Patience ... for in that must you possess your soul.' (cmkempe.d.1.p.1.s.947)
- (23) The vij **asshynges** ...: In **þis** we praye God to delyvere vs from euyll
the 7th prayer ... in this we pray God to deliver us from evil
'The 7th prayer ... This is the prayer in which we pray God to deliver us from evil'
(cmroyal.d.1.p.1.s.73)
- (24) **charite**; in **þat** hynges all
charity in that hinges all
'charity; on that everything depends' (cmrollep.d.9.p.9.s.755)
- (25) **kueade creftes**. Ine **þise** / zenegeþ moche uolk.
witchcraft in these sin many people
'witchcraft. This is something in which many people sin.' (cmayenbi.d.38.p.38.s.758)

The majority of standalone demonstratives in ME, EModE and LModE, however, do not refer to an NP antecedent but to a larger entity; an example is (26) from ME, where *ine þet* refers to behaving badly in church:

- (26) and huanne þe ssolest / yhere his messe / oþer his sermon /
and when you should hear his mass or his sermon
at cherche: **þou iangledest** / **and bourdedest** /

at church you chatter and jest
to-uor god. and **ine þet** / þu bere him / litel worþssipe.
 before God and in that you show him little respect
 ‘and when you should be attending to his mass or his sermon in church, you chatter and jest
 before God and in that you show him little respect’ (cmayenbi.d.18.p.18.s.303)

By contrast, the thirty-seven proper local anchors in OE with singular *that* with *in* or *on* refer to places (a dwelling, a province, a camp, a monastery, a church, a vineyard, the Pantheon), but also to an angelic vision, to a sacrifice, to ale, to prayer, to a horse, to a day or a year, and to (a personification of) sorrow. Of the instances where the referent is not local, the reference is to a place (a monastery, an altar), but also to books and to a point made in a debate. Of the plural standalone demonstratives, five refer to groups of people that have just been introduced, analogous to PDE *among those*, while other single instances refer to candelabras, visions, horses and elephants. The one plural referent that is not local is again a place (churches). The standalone demonstratives that do not have NP antecedents either refer cataphorically (two with *this*, nine with *that*), or to a person’s conduct or act (like (26)), an event, or to a quoted text.

A second very marked difference between OE on the one hand and ME, EModE and LModE on the other hand is genre; in ME, EModE and LModE, hardly any of the instances of standalone demonstratives occur in narrative texts, while they do in OE. This explanation could be the difference in antecedents: places in OE, and stretches of discourse in the later periods. This ties in with various observations that have been made with respect to PDE versus Modern German texts: in both languages, personal pronouns generally continue existing topics whereas demonstratives are topic shifters, as is demonstrated in (27), from Becher (2010: 1313):

- (27) [Modern computers]_i can perform [different tasks]_k. They_i / #_k / These #_i / _k . . .
- (a) # They_k include mathematical calculations, text processing . . .
 - (b) These_k include mathematical calculations, text processing . . .
 - (c) They_i can solve mathematical equations, process textual data . . .
 - (d) # These_i can solve mathematical equations, process textual data . . .

Modern computers and *different tasks* are both plural, so in theory, *they* and *these* could refer to either. Using *these*, however, most felicitously refers to the new information of the previous sentence, i.e. *different tasks*, rather than to the subject *modern computers*, and hence shift the topic from the existing one (the computers) to a new one (different tasks). Where PDE and German (and OE) differ is that German (and OE) can use standalone demonstratives as subjects to topic-shift to a new singular human referent, which PDE cannot do (Huddleston & Pullum *et al.* 2002: 1504). Instead, singular standalone demonstratives as subjects in English have developed a clear text-structuring function, where the reference is to the entire previous sentence or

stretch of discourse (so an entity of a higher order than a referent) rather than a particular focused constituent. An EModE example from our corpus is (28):

- (28) As he [John Wilmot, 2nd Earl of Rochester] told me, for five years together he was continually Drunk not all the while under the visible effect of it, but his blood was so inflamed, that he was not in all that time cool enough to be perfectly Master of himself. **This** led him to say and do many wild and unaccountable things: By **this**, he said, he had broke the firm constitution of his Health, that seemed so strong, that nothing was too hard for it. (burnetroc-e3-h)

Instead of topic-shifting, singular *this* ‘rather establishes a new attention focus by shifting the addressee’s attention to the state of affairs expressed in the preceding sentence’ (Becher 2010: 1312); cf. *this* in (28), with both instances referring to a period of excessive alcohol consumption, as described in the first sentence.

Even though this textual use of *this* gives the proximal demonstrative an important, possibly even central role in textual cohesion (Consten *et al.* 2007: 83), Becher’s (2010) comparison of the concluding paragraphs of a number of PDE and German texts of a similar genre shows that demonstratives are nevertheless much more frequent in German than personal pronouns, which is the reverse of the situation in PDE – in PDE, demonstratives connect stretches of discourse, whereas in German, particularly in the form of pronominal adverbs (prepositional phrases built on *da* ‘there’, like *damit* ‘with that’; see also the discussion of ME *perfor* in (33) below), they connect clauses, using the clause-initial position made available by a V2 syntax.

5.3 Alternatives for standalone demonstratives

If NP antecedents become problematic for singular demonstratives, what alternative expressions are available to restore that functionality? We saw that the referential functionality of a PP like *from that* in the PDE translation of example (1) could be ameliorated by spelling out the referent: *from that bottle*. Spelling out the referent, as *by this tale* in (29), appears as an alternative to the standalone demonstrative in (30).

- (29) **By this** ye may se that he that wyll lerne no good by example |nor good maner to hym shewyd is worthy to be taught with open rebukes. (1526: merrytal-e1-p2 60.16)
- (30) **By this tale** a man may well p~ceyue that they that be brought vp without lernyng or good maner shall never be but rude and bestely all though they haue good naturall wyttys. (1526: merrytal-e1-p1 2.13)

In this particular case, standalone *this* can also be argued to be referring to a stretch of discourse (i.e. the tale), which means that any reduced functionality is not necessarily a problem here; we find that the ratio of the frequencies of *by this* versus *by this tale* introducing the moral of the tale is 11:18 in this text.

An interesting development, and relevant as a potential response to the deficient functionality of standalone demonstratives in PPs, is the emergence of various types of ‘deictic strengtheners’ after ME. One type is *the same* in (31)–(32):

- (31) then you shall make a very strong brine of water and salt, and in **the same** you shall boile a handfull or two of Saxifrage (markham-e2-p2, 2,116.228-9)
- (32) havynge an noneste man with me, whoo had a foreste byll on hys bake, and with **the same** he cute downe a greate sorte of brakes (mowntayne-e1-h, 211.292-3)

Note that the function of *the same* is identical to that of a standalone demonstrative in earlier periods – it refers to an NP antecedent.

Another new development that could have replaced standalone demonstratives is pronominal adverbs, like *perfor* ‘therefor’ in (33):

- (33) Thanne was Silvery pope iii. zere. And for he cursed a byschop then was Silverius pope three year and because he cursed a bishop pei clepid Anthemius for heresie, **perfor** þe emperour, at instans of they called Anthemius for heresy for-that the emperor on instigation of his wif, exiled and killid him in a ylde þei clepe Ponciane his wife exiled and killed him on an island they call Ponza
 ‘Then Silverius was pope for three years. And because he excommunicated a bishop they called Anthemius for heresy, the emperor, on the instigation of his wife, exiled him to, and killed him on, an island they call Ponza’ (Capgrave-72.1214)

The pronominal adverb *perfor* ‘for that’ is referential, unlike PDE *therefore*, and refers back to the clause following the first *for* (underlined): ‘Because he had excommunicated Anthemius for committing heresy, (for that reason) the emperor exiled and killed him’ or ‘It was because he had excommunicated Anthemius for committing heresy that the emperor exiled and killed him’.⁴ Pronominal adverbs with *there* are coded as PPs in the corpus and are included in the clause-initial PPs annotated for this study, but they are marginal in terms of frequency and window of emergence and disappearance (end of ME/beginning of EModE).⁵ Neither pronominal adverbs nor the use of *the same* caught on as a replacement for standalone demonstratives.

Another alternative to standalone demonstratives that refer to NPs is third-person pronouns, as they are also referring elements with NP antecedents. The share of

⁴ The loss of referentiality of *therefore* in English is discussed in Los (2009) and Los & Komen (2012), citing Ball (1991); it also surfaces in Becher’s (2010: 1328) discussion of the *inertness* of PDE *therefore* compared to a German pronominal adverb like *deshalb* (example (i) from Becher, slightly adapted):

- (i) [A discovers B smoking]
 (a) Deshalb gehst Du immer auf den Balkon!
 (b) *Therefore you are always going on the balcony! (Literal translation)
 (c) So that’s why you are always going to the balcony! (Idiomatic translation)

Such main-clause-initial pronominal adverbs routinely require translations in PDE with pseudo-clefts, which unpacks the pronominal adverb into an adjunct-function (a *why*-clause in end-focus position) and a linking-function (demonstrative in subject position), as in (ic).

⁵ See also Lenker (2007, 2010) and Nevalainen & Raumolin-Brunberg (2017: 62) on the rise and fall of *thereof*, which was in competition with *its* and *of it*. *Its* was itself also an innovative form, replacing the neuter genitive *his* (presumably because *his* was increasingly associated with masculinity after the loss of gender and avoided in the neuter). *Of it* as a local anchor is similarly an ME innovation; see (34) below.

pronouns (all persons) in the total of local anchors goes up from 22 per cent in OE to 44 per cent in ME (see [table 4](#)).

Standalone demonstratives decline quite abruptly from OE to ME (from 951 to 379 occurrences per million words) while personal pronouns slightly increase (from 264 to 301 occurrences per million words); after which both standalone demonstratives and personal pronouns decline. In spite of the availability of personal pronouns, [table 4](#) shows that the personal pronouns do not step into the breach left by the loss of standalone demonstratives as local anchors, and cannot be regarded as a strategy to compensate for the local anchoring decline.

Investigating personal pronouns as an alternative to standalone demonstratives does uncover another quirk that aligns OE with modern Dutch and German: the absence of neuter personal pronouns in main-clause-initial PPs. Local anchors in main-clause-initial position in Dutch and German can only contain strong forms of neuter pronouns, which means demonstrative forms. In Dutch, the neuter personal pronoun, which takes the form of a pronominal adverb with *er-* when the pronoun finds itself in the complement of a preposition, cannot occur there; instead, we get the demonstrative form, a pronominal adverb with proximate *hier-* ('here') or distal *daar-* ('there') (see Broekhuis & Corver 2016: 1249; Travis 1984).

Hit is not found as the complement of main-clause-initial prepositions in OE, which is reminiscent of the situation in Dutch and German. The prepositions predominantly govern the dative, the neuter form of which would have been *him*, but none of the instances of that form in main-clause-initial PPs in the OE part of the corpus refer to non-human referents,⁶ apart from one postposition, & *him of afeol* '(he) him off fell' (cobede, Bede.4017), where *him* refers to a horse (a neuter noun), which is not likely to be in Spec,CP (Haerberli 2002). Another non-human antecedent picked up by a personal pronoun inside a main-clause-initial PP is one instance of feminine *hyre* 'her' referring to a city: *on hyre ne belæfde nane lafe cuce* 'in her not remained anyone alive' (cootest, Josh 10.28.5480). Prepositions like *þurh* 'through' that govern the accusative are robustly attested with the demonstrative *þæt*, but not with *hit*.

⁶ There are a number of cases of *him* referring to organs in the human body in EModE, in one text (*The anatomie of the bodie of man*, by Thomas Vicary 1548):

- (i) And this veine that commeth from Venakelis, entreth into the hart at the right Ventrikle, as I sayde before; and **in him** is brought a great portion of the thickest blood to nourishe the hart with (vicary-e1-h.d.2.p.2.s.168)
- (ii) The Lyuer is a principal member, and official; and of his first creation, spermatike; complete in quantitie of blood, of **him self** insensible, but by accident **he** is sensible, and **in him** is made the seconde digestion, and is lapped in a Senowy pannicle. (vicary-e1-h.d.3.p.3.s.321)

Although this text has been included in the EModE corpus, it derives from a much earlier text: 'In 1577, some fifteen years after his death, the surgeons at St Bartholomew's Hospital published a work entitled *The anatomie of mans body*, which they attributed to Thomas Vicary. In fact, as first shown by J F Payne in 1896, this work was very similar to a manuscript (MS 564) now in the possession of the Wellcome Library...; This manuscript is a fifteenth-century (c.1475) copy of an earlier text written in ME around 1392 by an anonymous London surgeon, who copied the work of earlier writers' (Thomas 2006: 235).

(*H*)*it*, the neuter nominative/accusative singular personal pronoun, in main-clause-initial PPs represents a ME innovation; the first example in the Parsed Penn–Helsinki Corpora of a main-clause initial PP containing *it* is *c.* 1350:

- (34) and **of it** comes stryfes and contekes schamefull, and
 and of it comes strife and dissent shameful and
 dyuerse wordes, and deuyouse and wikked sclandirs
 unkind words and devious and wicked slanders
 ‘and of it [wrath, the third deadly sin] comes strife and shameful dissent, and unkind words
 and devious and wicked slanders’ (cmedthor.d.1.p.1.s.216; The mirror of St Edmund,
 Thornton ms)

There are six examples of *it* in local anchors in the fifteenth century, from three texts. The numbers rise slightly in early ME (nine in E1, five in E2 and ten in E3) but never really take off.

Table 4 appears to show that personal pronouns could have emerged as an alternative expression for anaphoric links to NP antecedents after singular demonstratives started to refer to longer stretches of discourse rather than NPs, but this did not halt the decline in the overall numbers of such local anchors. This decline must have been caused by changes in the restrictions on the availability of non-subjects to link to the preceding discourse. The rise of (*h*)*it* to express a local anchor may have been triggered by the loss of referential functionality of the demonstrative, but also points to a change in the first position itself, whose close relationship with the demonstrative paradigm – witness its insistence on strong demonstrative forms – and with other deictic elements like *then*, *there* and *so* (see e.g. Los & van Kemenade 2018), broke down in ME.

5.4 In-PPs and the rise of containing inferrables

This section turns to the rise of the category *inferred* that we saw in figure 2. Within the category of *inferred*, of particular importance in this rise are ‘containing inferrables’ (Prince 1981; see also above, section 3). Containing inferrables typically contain a possessive pronoun, a demonstrative or a postmodifier (a genitive NP or a PP) linking to a known referent; an example is *in dïsses cyninges rice* ‘in this king’s reign’ in (14) above. In later periods, however, these containing inferrables become increasingly less ‘given’, as their modifiers – typically postmodifiers – make them identifiable from scratch; the determiner *the* can mark nouns that are new to the discourse, signalling to the reader/hearer that there is a postmodifier that makes that noun identifiable (cf. *the money that she had earned over the summer*). What we see in the periods after ME is that the link with the previous discourse becomes more tenuous in the category *inferred-containing*.

Examples of how *inferred-containing* modifiers can make NPs identifiable are given in (35)–(39) below. The head nouns inside these PP are the near-synonymous *place/stead/room*, with their modifiers referring to previous holders of certain offices or posts (modifiers in bold):

Table 4. *Personal and demonstrative pronouns as complement of P in main-clause-initial PPs*

NP type	OE	nf per million words	ME	nf per million words	EModE	nf per million words	LModE	nf per million words
DEM	1,427 (78%)	951	439 (56%)	379	447 (74%)	249	147 (67%)	154
PERS	396 (22%)	264	348 (44%)	301	159 (26%)	87	72 (33%)	76
TOTAL	1,823		787		606		219	

- (35) in **his** steade sir nicholas bacon, knight, was made lord keepour of the great seale of england, a man of greate diligence and ability in his place, whose goodnesse preserved his greatnesse from suspicion, envye and hate. (EModE, hayward-e2-p1.d.1.p.1.s.46)
- (36) in steade **of bonner**, edmund grindall was made bishopp of london (EModE, hayward-e2-p1.d.1.p.1.s.194)
- (37) in the room **of the lord chancellor**, they would have placed one watson a priest, absurd in humanity and ignorant in divinity. (EModE, raleigh-e2-p1.d.1.p.1.s.73)
- (38) In the room **of the unwarlike troops of Asia, which had most probably served in the first expedition**, a second army was drawn from the veterans and new levies of the Illyrian frontier, and a considerable body of Gothic auxiliaries were taken into the Imperial pay. (LModE, gibbon-1776.d.1.p.1.s.374)

All of these have the status of *inferred containing*. In (35)–(36), the link is to the previous discourse; the same is true for (37), but only because the Lord Chancellor happens to be mentioned at some earlier point in this long text; in terms of textual analysis, the NP is primarily identifiable because readers can be assumed to know that their government has such a post. Contrast these examples with (38), from LModE, where the postmodifier is not only much longer, but the link with the previous discourse, *the first expedition*, is also much more implicit, and requires some effort on the part of the reader to recover. The ‘first expedition’ can be interpreted as the disastrous campaign described in the previous paragraph, but the disaster had not been attributed to the character or provenance of the soldiers but to the climate, the terrain, and other factors, so *the unwarlike troops of Asia* are ‘new’; nevertheless, the *in-PP* is *inferred containing* because a link can be made to this earlier expedition.

One further note can be made about the loss of local anchoring in the context of these inferrables. The PP *in (his) steade*, as in (35), like some other local anchor-PPs (*inside, because, at last*), ultimately lexicalized as PDE *instead* (Tabor & Traugott 1998; Lewis 2011), which no longer requires a modifier – the link with a referent in the previous discourse now has to be recovered from the context by the hearer/reader without the aid of an explicit anaphor: ‘The appearance in the 18th century of *instead* alone as an adverbial again can be seen as information compression: the replaced item is now ellipted, to be recovered from the context The host of the *instead* in each case is the salient alternative, so that *instead* without *of* becomes associated with high information salience’ (Lewis 2011: 427–8). Dutch would require a specific link in the form of a pronominal adverb (*in plaats daarvan* ‘instead of that’) and German would require a demonstrative pronoun (*stattdessen* ‘instead-of-that’). Becher’s (2010) comparison of textual cohesion strategies in PDE and German texts refers to this phenomenon as ‘explicitation’, and provides further examples: PDE just has *involved* in a sentence like *researchers are still far from working out all the processes involved*, where German insists on adding an explicit anaphor: *daran beteiligt* ‘in that involved’; similarly, PDE would just have a relatively simple example where German has *Ein relativ einfaches Beispiel hierfür*

‘a relatively simple example of this’ (Becher 2010: 1330). *Instead* has become an adverb in PDE rather than a PP, and hence more acceptable in main clause-initial position.

5.5 Discussion

We can conclude that there is a clear break in the use of standalone demonstratives between OE and ME, long before the loss of V2. In Los & van Kemenade (2018), we speculate that the loss of the *se* paradigm and the loss of gender marking are related: grammatical gender allows demonstratives to refer to specific antecedents; of the nominal inflectional categories, gender and number, but not case, facilitate inter-clausal referent-tracking. We saw in section 5 that the singular demonstrative pronouns as local anchor drop in early ME; they are increasingly referring not to NP antecedents but to stretches of discourse. The shift to personal pronouns, particularly to the innovative *it*, has not led to a recovery, the main reason being the increasing marginality of non-subjects in first position to express links to the previous discourse. It is telling that Hasselgård (2010) only gives half a page to local anchor adjuncts in her monograph on adjuncts in English, and then with examples such as (39) – an adverb with an implicit link (in bold; cf. the discussion of *instead* in the previous section):

(39) A great cast-iron beam protruded through an opening high up in the building. **Inside** was the engine – his engine. (W2F-007) (Hasselgård 2010: 80)

In Los & van Kemenade (2018), we speculate that the loss of V2 meant the loss of a multifunctional first position (multifunctional in terms of information-structural status as well as in terms of syntactic function), which worked in tandem with an articulate, gendered, demonstrative pronoun paradigm to enable unmarked links to the immediately preceding discourse. The finer-grained analysis of this article shows that the loss of local anchors started earlier, and its timing suggests that it was due to the loss of that gendered paradigm, and later reinforced by – and possibly kickstarting – the loss of V2. In this light, it is significant that Allen’s (2022) data offer concrete support for a link between the loss of gender and the loss of demonstratives as reliable discourse reference trackers (Allen 2022: 126–7).

6 Conclusion

This article reports on a study of all main clause-initial PPs in the suite of the syntactically parsed Penn–Helsinki Corpora of OE, ME, EModE and LModE texts. All the NPs within these PPs were annotated with information about the antecedent: whether there was one, and if so, its position in the text; and the status of the link in terms of Pentaset-categories (*New, Inert, Assumed, Inferred* or *Identity*), in turn based on Prince’s categories (Prince 1981). The aim was to investigate the discourse status of the PP, in particular whether it was a local anchor (referring back to the immediately preceding discourse) or something else, like a frame-setter (forward scoping instead of backward linking). This

article has focused on the ‘given’ categories (anchors), reserving frame-setters for future research. Local anchors decline, and there is a shift to personal pronouns instead of demonstratives; the use of the pronoun *it* in local anchors is an EModE innovation. The hypothesis was that V2 created a slot for a first non-subject constituent that was particularly suited to host local anchors, and that the loss of V2 in the fifteenth century should also mean the end of local anchors. As it turned out, the decline starts earlier, in ME, although the exact timing of the decline may be difficult to pinpoint in view of the M2 data gap. Los & van Kemenade (2018) speculate that the decline in functionality of the deictic system is a consequence of the loss of a gendered paradigm, and this could also be the cause of the changing frequencies we see in ME. The connection between deictic elements and the first position made available by V2 was broken, and the loss of V2 may have further promoted the decline of local anchors. The unmarked way to establish links to the previous discourse became restricted to the subject.

The decline of local anchors must be considered in the light of the development of new ways to structure discourse in the history of English, after the loss of functionality in deictic expressions and their place in a V2 architecture had compromised earlier strategies to make textual connections. PDE still allows local anchors, albeit at very low frequencies, but with personal pronouns, most notably *it*, in what appears to be a quite restricted function.

The large-scale study conducted here focused on local anchors, but has also thrown up other questions, particularly about the transition into LModE, which does not continue many of the earlier trends; *new* main clause-initial PPs – presumably frame-setters – are down, for instance. This and other questions have to be left to future research.

Authors' addresses:

*Department of Linguistics and English Language
School of Philosophy, Psychology and Language Sciences
The University of Edinburgh
Room 2.07 Dugald Stewart Building
3 Charles Street
Edinburgh EH8 9AD
United Kingdom
B.Los@ed.ac.uk*

*Department of Language, Literature & Communication
De Boelelaan 1105
Amsterdam 1081
Netherlands
gea.dreschler@vu.nl*

*Department of Modern Languages and Cultures
Radboud University Nijmegen*

Postbus 9103
6500 HD Nijmegen
Netherlands
ans.vankemenade@ru.nl

Department of Language and Communication
Radboud University Nijmegen
Postbus 9103
6500 HD Nijmegen
Netherlands
erwin.komen@ru.nl

Department of Linguistics and English Language
School of Philosophy, Psychology and Language Sciences
The University of Edinburgh
Room 2.07 Dugald Stewart Building
3 Charles Street
Edinburgh EH8 9AD
United Kingdom
s.coretta@ed.ac.uk

References

- Allen, Cynthia. 2022. Pronominally headed relative clauses in early English. *English Language and Linguistics* 26(1), 105–32.
- Ball, Catherine N. 1991. The historical development of the *it*-cleft. PhD dissertation, University of Pennsylvania.
- Bech, Kristin. 2001. Word order patterns in Old and Middle English: A syntactic and pragmatic study. PhD dissertation, University of Bergen.
- Bech, Kristin. 2014. Tracing the loss of boundedness in the history of English: The anaphoric status of initial prepositional phrases in Old English and Late Middle English. *Anglia* 132(3), 506–35.
- Bech, Kristin & Kristine Eide (eds.). 2014. *Information structure and syntactic change in Germanic and Romance languages*. Amsterdam: John Benjamins.
- Becher, Viktor. 2010. Differences in the use of deictic expressions in English and German texts. *Linguistics* 48(4), 1309–42.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman grammar of spoken and written English*. London: Longman.
- Birner, Betty & Gregory Ward. 2002. Information packaging. In Huddleston & Pullum *et al.* (eds.), 1363–447.
- Broekhuis, Hans & Norbert Corver. 2016. *Syntax of Dutch*, vol. 3: *Verbs and verb phrases*. Amsterdam: Amsterdam University Press.
- Consten, Manfred, Mareile Knees & Monika Schwarz-Friesel. 2007. The function of complex anaphors in texts: Evidence from corpus studies and ontological considerations. In Monika Schwarz-Friesel, Manfred Consten & Mareile Knees (eds.), *Anaphors in text*, 81–102. Amsterdam: John Benjamins.
- Cook, Philippa & Felix Bildhauer. 2011. Annotating information structure: The case of topic. In Stefanie Dipper & Heike Zinsmeister (eds.), *Beyond semantics: Corpus-based investigations of*

- pragmatic and discourse phenomena* (Bochumer Linguistische Arbeitsberichte 3), 45–56. Bochum: Ruhr-Universität Bochum, Sprachwissenschaftliches Institut.
- Cook, Philippa, & Felix Bildhauer. 2013. Identifying ‘aboutness topics’: Two annotation experiments. *Dialogue & Discourse* 4(2), 118–41.
- Cuyckens, Hubert. 2002. Metonymy in prepositions. In Hubert Cuyckens & Günter Radden (eds.), *Perspectives on prepositions* (Linguistische Arbeiten 454), 257–66. Tübingen: Max Niemeyer Verlag.
- De Bastiani, Chiara. 2022. The reanalysis of VO in the history of English: Evidence for a language-internal account. In Bettelou Los, Claire Cowie, Patrick Honeybone & Graeme Trousdale (eds.), *English historical linguistics: Change in structure and meaning*, 115–36. Amsterdam: John Benjamins.
- Dreschler, Gea. 2015. Passives and the loss of verb second: A study of syntactic and information-structural factors. PhD dissertation, Radboud University Nijmegen. Utrecht: LOT.
- Dreschler, Gea. 2020. ‘Fifty pounds will buy me a pair of horses for my carriage’: The history of permissive subjects in English. *English Language and Linguistics* 24(4), 719–44.
- Gelderen, Elly van. 2011. Valency changes in the history of English. *Journal of Historical Linguistics* 1(1), 106–43.
- Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69, 274–307.
- Haeblerli, Eric. 2002. Observations on the loss of verb second in the history of English. In C. Jan Wouter Zwart & Werner Abraham (eds.), *Studies in comparative syntax: Proceedings from the 15th Workshop on Comparative Germanic Syntax*, 245–72. Amsterdam: John Benjamins.
- Halliday, Michael A. K. 1967. Notes on transitivity and theme in English: Part 2. *Journal of Linguistics* 3(2), 199–244.
- Hannay, Mike & Evelien Keizer. 1993. Translation and contrastive grammar: The grammatical versus the communicative strategy. *Toegepaste Taalwetenschap in Artikelen* 45, 65–88.
- Hasselgård, Hilde. 2010. *Adjunct adverbials in English*. Cambridge: Cambridge University Press.
- Hawkins, John. 1986. *A comparative typology of English and German*. London: Croom Helm.
- Hinterhölzl, Roland & Svetlana Petrova. 2010. From V1 to V2 in West Germanic. *Lingua* 120, 315–28.
- Huddleston, Rodney & Geoffrey K. Pullum et al. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Kemenade, Ans van. 1987. *Syntactic case and morphological case in the history of English*. Dordrecht: Foris.
- Kemenade, Ans van. 2012. Rethinking the loss of verb second. In Nevalainen & Traugott (eds.), 822–34.
- Kemenade, Ans van, Tanja Milićev & R. Harald Baayen. 2008. The balance between syntax and discourse in Old English. In Maurice Gotti, Marina Dossena & Richard Dury (eds.), *English historical linguistics 2006: Selected papers from the Fourteenth International Conference on English Historical Linguistics*, 3–21. Amsterdam: John Benjamins.
- Kemenade, Ans van & Tanja Milićev. 2012. Syntax and discourse in Old and Middle English word order. In Dianne Jonas, Andrew Garrett & John Whitman (eds.), *Grammatical change: Origins, nature, outcomes*, 239–54. Oxford: Oxford University Press.
- Kemenade, Ans van & Marit Westergaard. 2012. Syntax and Information Structure: Verb Second variation in Middle English. In Meurman-Solin, López-Couso & Los (eds.), 87–118.
- Komen, Erwin. 2011. *Cesax: Coreference editor for syntactically annotated XML corpora*. Nijmegen: Radboud University Nijmegen. <https://cls.ru.nl/staff/ekomen/software/Cesax/>
- Komen, Erwin. 2012. Coreferenced corpora for information structure research. In Jukka Tyrkkö, Matti Kilpiö, Terttu Nevalainen & Matti Rissanen (eds.), *Outposts of historical corpus linguistics: From the Helsinki Corpus to a proliferation of resources* (Studies in Variation,

- Contacts and Change in English 10). Helsinki: Research Unit for Variation, Contacts, and Change in English. <https://varieng.helsinki.fi/series/volumes/10/komen/>
- Komen, Erwin. 2013. Predicting referential states using enriched texts. In Francesco Mambrini, Marco Passarotti & Caroline Sporleder (eds.), *Proceedings of the third workshop on annotation of corpora for research in the humanities* (ACRH-3), 49–60. Sofia: The Institute of Information and Communication Technologies, Bulgarian Academy of Sciences.
- Komen, Erwin, Rosanne Hebing, Ans van Kemenade & Bettelou Los. 2014. Quantifying information structure changes in English. In Bech & Eide (eds.), 81–110.
- Komen, Erwin, Bettelou Los & Ans van Kemenade. 2023. The Pentaset: Annotating referential state primitives. Unpublished MS.
- Krifka, Manfred. 2008. Basic notions of Information Structure. *Acta Linguistica Hungarica* 55 (3–4), 243–76.
- Kroch, Anthony, Beatrice Santorini & Lauren Delfs. 2004. *Penn–Helsinki Parsed Corpus of Early Modern English*, 1st edn, release 3. www.ling.upenn.edu/hist-corpora/PPCEME-RELEASE-3/index.html
- Kroch, Anthony, Beatrice Santorini & Ariel Dierani. 2010. *Penn Parsed Corpus of Modern British English*. www.ling.upenn.edu/hist-corpora/PPCMBE-RELEASE-1/index.html
- Kroch, Anthony & Ann Taylor. 2000. *Penn–Helsinki Parsed Corpus of Middle English*, 2nd edn. www.ling.upenn.edu/hist-corpora/PPCME2-RELEASE-4/index.html
- Lemmens, Marcel & Tony Parr. 1995. *Handboek voor de vertaler Nederlands-Engels*. Groningen: Wolters-Noordhoff.
- Lenker, Ursula. 2007. *Forhwi* ‘because’: Shifting deictics in the history of English causal connection. In Ursula Lenker & Anneli Meurman-Solin (eds.), *Connectives in the history of English*, 193–227. Amsterdam: John Benjamins.
- Lenker, Ursula. 2010. *Argument and rhetoric: Adverbial connectors in the history of English*. Berlin: Mouton de Gruyter.
- Lewis, Diana M. 2011. A discourse-constructural approach to the emergence of discourse markers in English. *Linguistics* 49(2), 415–43.
- Los, Bettelou. 2009. The consequences of the loss of verb-second in English: Information structure and syntax in interaction. *English Language and Linguistics* 13(1), 97–125.
- Los, Bettelou. 2012. The loss of verb second and the switch from bounded to unbounded systems. In Meurman-Solin, López-Couso & Los (eds.), 21–46.
- Los, Bettelou. 2018. ‘Permissive’ subjects and the decline of adverbial linking in the history of English. In Hubert Cuykens, Hendrik De Smet, Liesbet Heyvaert & Charlotte Maekelberghe (eds.), *Explorations in English historical syntax*, 23–49. Amsterdam: John Benjamins.
- Los, Bettelou & Gea Dreschler. 2012. The loss of local anchoring: From adverbial local anchors to permissive subjects. In Nevalainen & Traugott (eds.), 859–71.
- Los, Bettelou & Ans van Kemenade. 2018. Syntax and the morphology of deixis: The loss of demonstratives and paratactic clause linking. In Marco Coniglio, Andrew Murphy, Eva Schlachter & Tonjes Veenstra (eds.), *Atypical demonstratives: Syntax, semantics and pragmatics*, 127–58. Berlin and Boston, MA: De Gruyter Mouton.
- Los, Bettelou & Erwin Komen. 2012. Clefts as resolution strategies after the loss of a multifunctional first position. In Nevalainen & Traugott (eds.), 884–98.
- Los, Bettelou & Thijs Lubbers. Forthcoming. The development of prose style in English horse manuals: Interpreting stylistometric findings. In Peter Petr  (ed.), *A Festschrift for Hubert Cuykens* [working title].
- Meurmann-Solin, Anneli, Maria Jos  L pez-Couso & Bettelou Los (eds.). 2012. *Information Structure and syntactic change in the history of English*. New York: Oxford University Press.
- Nevalainen, Terttu & Helena Raumolin-Brunberg. 2017. *Historical sociolinguistics: Language change in Tudor and Stuart England*, 2nd edn. Abingdon: Routledge.

-
- Nevalainen, Terttu & Elizabeth C. Traugott (eds.). 2012. *The Oxford handbook of the history of English*. New York: Oxford University Press.
- Pérez-Guerra, Javier. 2005. Word order after the loss of the verb-second constraint or the importance of early Modern English in the fixation of syntactic and informative (un-)markedness. *English Studies* 86, 342–69.
- Prince, Ellen. 1981. Toward a taxonomy of given-new information. In Peter Cole (ed.), *Radical pragmatics*, 223–55. New York: Academic Press.
- Rohdenburg, Gunther. 1974. *Sekundäre Subjektivierungen im Englischen und Deutschen: Vergleichende Untersuchungen zur Verb- und Adjektivsyntax* (PAKS-Arbeitsbericht 8). Bielefeld: Cornelson-Velhagen and Klasing.
- Speyer, A. 2010. *Topicalization and stress clash avoidance in the history of English*. Berlin: De Gruyter Mouton.
- Struik, Tara & Ans van Kemenade. 2020. On the givenness of OV word order: A (re)examination of OV/VO variation in Old English. *English Language and Linguistics* 24(1), 1–22.
- Struik, Tara & Ans van Kemenade. 2022. Information structure and OV word order: A phase-based approach. *Journal of Comparative Germanic Syntax* 25(1), 79–114.
- Tabor, Whitney & Elizabeth Closs Traugott. 1998. Structural scope expansion and grammaticalization. In Anna Giacalone Ramat & Paul J. Hopper (eds.), *The limits of grammaticalization*, 229–72. Amsterdam: John Benjamins.
- Taylor, Ann & Susan Pintzuk. 2012a. Rethinking the OV/VO alternation in Old English: The effect of complexity, grammatical weight and information status. In Nevalainen & Traugott (eds.), 199–213.
- Taylor, Ann & Susan Pintzuk. 2012b. Verb order, object position and information status in Old English. *York Papers in Linguistics*, series 2, 12a, 29–52.
- Taylor, Ann & Susan Pintzuk. 2014. Testing the theory: Information structure in Old English. In Bech & Eide (eds.), 53–77.
- Taylor, Ann, Anthony Warner, Susan Pintzuk & Frank Beths. 2003. *The York–Toronto–Helsinki Parsed Corpus of Old English Prose*. Electronic texts and manuals available from the Oxford Text Archive.
- Thomas, Duncan P. 2006. Thomas Vicary and *The Anatomie of Mans Body*. *Medical History* 50(2), 235–46.
- Travis, Lisa. 1984. Parameters and effects of word order variation. PhD dissertation, Massachusetts Institute of Technology.
- Truswell, Robert, Rhona Alcorn, James Donaldson & Joel Wallenberg. 2019. A parsed linguistic atlas of Early Middle English. In Rhona Alcorn, Joanna Kopaczyk, Bettelou Los & Benjamin Molineaux (eds.), *Historical dialectology in the digital age*, 19–38. Edinburgh: Edinburgh University Press.