

The variable syntax of pronominal prepositional objects in Old English: Some statistical modelling problems

Rhona Alcorn, University of Edinburgh
rhona.alcorn@ling.ed.ac.uk

1. Introduction

There is no greater impediment to the advancement of knowledge than the ambiguity of words.
Thomas Reid (1710-1769), Scottish philosopher

OE personal pronouns often turn up in ‘special’ positions, i.e. where other types of functionally-equivalent arguments are not (or rarely) found. Pronominal prepositional objects are a particularly good evidential source because their special placement is always unambiguous and can be recognised atheoretically (even if their structural position cannot).

The canonical position for (non-relativized) prepositional objects is to the immediate right of the preposition: 99.9% of nominals, demonstrative pronouns and modified or coordinated personal pronouns occur there. Bare personal pronouns are also found right-of-P, e.g. (1), but also appear in a variety of ‘special’ left-of-P positions, e.g. (2).

- (1) *pronoun to the immediate right of the preposition (71.5%), e.g.:*

Ac þa hundas comon **to him**

(cogregdC,GDPref_and_4_[C]:34.310.6.4623)

‘But the dogs came to him’

- (2) a. *pronoun to the immediate left of the preposition (18.9%), e.g.:*

and se hælend sylf of heofonum com **him to**

(coalive,ÆLS_[Thomas]:13.7546)

‘and the Saviour Himself came to him from heaven’

- b. *pronoun somewhere further to the preposition’s left (9.6%), e.g.:*

ond misenlico wilddeor **him** þær comon **to**

(comart3,Mart_5_[Kotzor]:Ju2,A.6.887)

‘and various wild beasts came to him there’

Leading analyses of OE syntax predict left-of-P placement to be a freely available option (e.g. van Kemenade 1987, 1999; Pintzuk 1991). But evidence suggests it is sensitive to a range of factors (e.g. Wende 1915, Taylor 2008), making the dataset represented by (1,2) a good candidate for logistic regression.

This paper considers two types of ambiguity that could hamper such an analysis:

1. Certain prepositions are identical to certain verbal prefixes and/or adverbial particles and/or adverbs, e.g. *on*. Can we rely on the methods of preposition identification employed by the editors of the York-Toronto-Helsinki Parsed Corpus of Old English Prose (‘YCOE’) (Taylor, Warner, Pintzuk & Beths 2003)?
2. Third person personal pronouns are unambiguous for case and are rarely specially placed unless dative (as I will show). How can this effect be controlled for first and second person data, whose dative and accusative forms are identical?

Q1 is also relevant to those who use prefixes and particles (henceforth referred to as prefixes) to diagnose underlying clause structure and who rely on the YCOE for their identification. Q2 requires a method for disambiguating case-ambiguous forms.

2. Identification of prepositions in YCOE

2.1 Parsing protocol

YCOE labels: <P> for prepositions; <ADV> for adverbs; and <RP> for prefixes. A word with the *form* of a preposition is labelled <P> if the corpus eds. associated it with a complement, otherwise <RP> if on the RP list, otherwise <ADV>.¹ When orthographically attached to a verb in the base edition, an element on the RP list is *always* labelled <RP>, but with a <+> extension. Potential for two types of misanalysis:

- i. prepositions not associated with their complement → prepositions labelled <RP(+)> or <ADV> → relevant data excluded from study, e.g. (3a). See §2.2.

(3) a. and him freondlice **tospræc** (NB **tosprecan*)
(coaelive, ÆLS_[Martin]:678.6408)

‘and spoke to him companionably’

b. and him cuðlice to<P> spræcon
(cocathom2, ÆCHom_II, 22:195.161.4316)

‘and spoke to him openly’

- ii. words with the form of a preposition are wrongly associated with a complement → <P> label attached to prefixes and/or adverbs → improper data included in study, e.g. (4,5a) (purported prepositional object underlined). See §2.3.

(4) to middes mergenes stande eastweard & bebeode hine Gode geornlice & hine gesenige cyrre hine, sungonges ymb<P>,
(colaece, Lch_II_[1]:47.1.5.1548–52)

‘mid-morning (he) should stand (facing) eastward and should offer himself earnestly to God and cross himself (and) turn himself around_{ADV} sunwise (i.e. following the sun’s path)’

(5) a. ... þæt land, þe ðu hyra fæderum fore<P> swore?
(cootest, Num:11.12.4035)

b. ... ðæt land, þe ic fore swor heora fæderum
(cootest, Num:14.22.4178)

2.2 Prepositions labelled <RP(+)> or <ADV>

2.2.1 Generally discountable instances

We can discount 94% of all 3,739 <RP>s, 87% of all 10,918 <RP+>s and 98% of all 72,111 <ADV>s as possible prepositions without reference to their semantics under the following conditions (conditions (i,ii) are definitive, while (iii,iv) are based on strong probabilities):

- i. **no lexical ambiguity**: eliminates a swathe of <ADV> elements, including the most frequently occurring, e.g. *ða*, *ðeah*, *ðonne*, *ðus*, *eft*, *forðam*, *forðon*, *her*, *nu*, *oft*, *sona*, *swa*, *swilce*, and also certain <RP> elements, e.g. *aweg*, *forð*, *up*, *ut*.
- ii. **morphological complexity**: eliminates comparative & superlative <ADV> elements, e.g. *swiðor*, *swiðost*, and those ending *-lice*.
- iii. **collocation**: the element’s clause does not have an object argument.
- iv. **word order**: the element follows an object that isn’t a bare personal pronoun, e.g. (6), or the element precedes, but isn’t adjacent to, the object argument, e.g. (7).

¹ The YCOE’s RP list comprises: *adun(e)*, *æfter*, *aweg*, (*of*)*dune*, *fore*, *forð*, *fram*, *geond*, *in*, *mid*, *nider*, *of*, *ofer*, *on*, *ongean*, *onweg*, *to*, *purh*, *under*, *up*, *ut*, *wið*, *wiðer* & *ymb(e)*. Kastovsky (1992: 376) also lists *æt*, *be*, *efen*, *eft*, *for* and *full*. Those in bold are identical to a preposition.

- (6) and actus apostolorum eac_{<ADV>} he gesette
(coaelive, ÆELS_[Mark]:150.3301)
‘and he also wrote the Acts of the Apostles’
- (7) swa Crist **oferswiþde** þæt deofol mid þisse cyþnesse
(coblick,HomS_10_[BIHom_3]:31.82.420)
‘as Christ overpowered the devil with this knowledge’

2.2.2 Remaining cases

<RP> elements: a further 2% can be eliminated on grounds that prepositional *æfter*, *mid*, *of* and *to* strongly favour dative (cf. §3.2), e.g. (8). A cursory look at the final 167 tokens suggests the number of elements labelled <RP> instead of <P> is likely to be <167 (<4.5% of total), e.g. (9).

- (8) Gif þin hand oððe þin fot þe æswicige: aceorf of_{<RP>} þæt lim
(cocathom1, ÆCHom_I,_34:470.147.6791)
‘If your hand or your foot should offend you, cut off the limb’
- (9) a. læd in_{<RP>} þas men
(cogenesiC,Gen_[Ker]:43.16.307)
‘lead in those men’
- b. Dryhtyn, Dryhtyn, læt us in_{<RP>}
(cowsgosp,Mt_[WSCp]:25.11.1739)
‘Lord, Lord, let us in’

<RP+> elements: another 1% can be discounted as they immediately follow *ne* or infinitival *to*, (Kastovsky 1992: 375), e.g. (10). Of the remaining 1,288 tokens (all of which follow a bare personal pronoun object), only 5% (N = 63) involve a reified prefixed verb that is not listed in Clark Hall (1960), e.g. (2a), and further analysis is needed to determine which are prepositions and which are adverbs.² But there may also be some prepositions among the 95% which *are* plausible prefixes, although some may remain functionally ambiguous even in context: e.g. does (11) involve *tosendan* ‘to send to’ or *sendan* ‘to send’ + a PP headed by *to*? In sum, the number of prepositions labelled <RP+> is unlikely to be more than 1,288 (11.8% of total), and individual analysis of these tokens is likely to reduce this number further.

- (10) a. and ge me ne **underfengon**
(coaelhom, ÆHom_11:441.1726)
‘and you did not receive me’
- b. ac ge seceað me to **ofsleanne**
(cowsgosp,Jn_[WSCp]:8.37.6443)
‘but you seek to kill me’
- (11) þa com þara sacerda ealdorman þe Petrus him **tosende**
(coblick,LS_20_[AssumptMor[BIHom_13]]:153.291.1924)
‘Then came the priests’ chief, who Petrus had sent to them’

² Only two of the unlisted reified prefixed verbs are associated with >2 tokens: *tosprecan* (x5) and *tobugan* (x4).

<ADV> **elements:** 93% of the remainder are instances of *eac*, *ær* or *siððan* and these can mostly be ruled out as prepositions on basis of object case and/or on the assumption that the correct part-of-speech is usually self-evident, e.g. (12–14).³ The other 7% of cases (N = 93, or 0.1% of all <ADV> elements), would require individual consideration but are sufficiently few to conclude there is no significant problem of prepositions labelled as adverbs.

- (12) a. a. Nu wylle we eow secgan sume Petres wundra, [...] Nu wylle we eac<ADV> eow secgan hu he sume dæg eode to þam Godes temple
(coaelive, ÆELS[Peter's_Chair]:23.2272–6)
‘Now we wish to tell you some of Peter’s miracles, [...] Now we **also** wish to tell you how on a certain day he went to God’s temple’
- b. & þær wearð Sigulf ealdormon ofslægen, & Sigelm ealdormon, & Eadwold cynges ðegen, & Cenulf abbod, & Sigebreht Sigulfes sunu, & Eadwald Accan sunu, & monige eac<P> him
(cochronA-2b,ChronA_[Plummer]:905.11.1187)
‘and there were killed Ealdorman Sigewulf, and Ealdorman Sigehelm, and Eadwod the king’s thane, and Abbot Cenwulf, and Sigeberht, son of Sigewulf, and Eadwold, son of Acca, and many **besides** them’
- (13) a. ... for þon ðe hy ær<ADV> Mul forbærndon
(cochronD,ChronD_[Classen-Harm]:694.1.88)
‘because they had **previously** burned Mul’
- b. ... ure Hælend cwæð ær<P> his þrowunge
(coaelhom, ÆHom_5:136.771)
‘our Lord said **before** His martyrdom’
- (14) a. ... swilce man siwige ane bytte, and blawe hi fulle windes, and wyrce siððan<ADV> an þyrl
(coaelive, ÆELS_[Cecilia]:315.7296–7)
‘as if one should sew up a bladder, and blow it full of wind, and **then** make a hole’
- b. Ðis wæs gedon syððon<P> ure Drihtnes acennednesse deccc lxxii ...
(cochronE-INTERPOLATION,ChronE_[Plummer]:963.71.1438)
‘This was done 972 years **after** our our Lord’s birth’

2.3 Prefixes/adverbs labelled <P>

None of the 87,400 <P> elements parsed as the governor of a (pro)nominal object can be generally discounted as either adverb or prefix on grounds of word order, and only 2% be discounted on the basis of word form (e.g. *toforan*, *uppan*). The other 98% therefore require individual consideration, taking account of questions such as:

- does <P> have the form of an adverb? If so, would such treatment work in context?
- does <P> have the form of a prefix? If so, is that <P> + verb combination attested as prefixed verb? If so, and if <P> does not immediately precede the verb, is there evidence that this prefixed verb is separable?
- could <P>’s purported object ‘belong’ to the verb, either in its simple form or assuming <P> as its prefix? [Verb valency? If transitive, is another object argument already present?]
- could <P>’s purported object be an ‘ungoverned’, or ‘ethic’, dative (Colman 1991: 78–99)?

With >85,500 tokens to check—and no obvious way of defining a representative and manageable sample—I see no way to gauge the likely (in)frequency of this type of error.

³ Mitchell (1985: §3832) questions whether *siððan* ever functions as preposition. Clark Hall (1960) associates it with late West Saxon.

3. Controlling for case effects

3.1 The problem

Univariate analysis suggests case effects:

	N	Left-of-P
dative	6,391	38%
accusative	978	4%
genitive	25	0%
<i>Total</i>	<i>7,394</i>	

Table 1: Special placement by case (3rd person data)

Case effects can't be controlled for 1st/2nd person data as they are acc./dat. ambiguous, so these data should be excluded from a logistic regression to avoid bias. However, their exclusion would: a) reduce dataset by 30% (> loss of power); b) prevent modelling of person effects (Alcorn forthcoming), cf. Table 2.

	N	Left-of-P
1 st person	1,724	10%
2 nd person	1,391	10%
3 rd person	7,394	38%

Table 2: Special placement by person

Instead: analyse dat. data only. As 98.5% of specially placed case-unambiguous data are dative, a model which ignores acc. and gen. data should be no less powerful than one which doesn't.

3.2 Inferring case

Table 3 identifies eight prepositions which rarely occur with a clear accusative object.

	<u>3rd p. personal pronouns</u>			<u>Nominals, DPs, etc.</u>			
	N	Dat.	Acc.	N	Dat.	Acc.	Amb.
<i>to</i>	2,499	99.8%	0.2%	17,665	69.9%	1.5%	28.6%
<i>mid</i> ⁴	895	94.7%	5.3%	14,579	80.7%	1.6%	17.6%
<i>be</i>	257	100%	0%	4,223	79.1%	0.6%	20.3%
<i>fram</i>	245	100%	0%	3,030	80.5%	1.0%	18.5%
<i>betweenan</i> ⁵	201	100%	0%	27	88.9%	3.7%	7.4%
<i>æfter</i>	167	99.4%	0.6%	2,723	80.1%	1.2%	18.7%
<i>of</i>	130	100%	0%	6,506	79.7%	1.4%	18.9%
<i>æt</i>	91	100%	0%	2,760	57.4%	3.0%	39.5%

Table 3: Dative-favouring prepositions

On the basis of the proportions in Table 3, I assume most 1st/2nd person objects of these eight prepositions are dative. Statistical analysis can therefore proceed on the 90% of 3rd person tokens that are dative and the 61% of 1st/2nd person tokens that are governed by *to*, *mid* (except those in *Bede*—see fn. 4), *be*, *fram*, *betweenan*, *æfter*, *of* or *æt*.

Consequences: a) bias effects of case ambiguities are avoided; b) person effects can be measured; c) some loss of model power through exclusion of some (unidentifiable) 1st/2nd person dat. data—but, still, a good trade-off.

⁴ *mid* + acc. 'is generally agreed to be Anglian ... It occurs with any regularity in only *Bede* and the poetry. But even in these, the dative is more common' (Mitchell 1985: §1195). 45% of accusative objects of *mid* in Table 3 are from *Bede*.

⁵ Where the object is not a personal pronoun, a nasal-less form of 'between', e.g. *betweoh*, is more usual (Kitson 1996: 28).

References

- Alcorn, Rhona. Forthcoming. *to him vs. him to*: The variable syntax of pronominal prepositional objects in Old English. *English Language and Linguistics*.
- Clark Hall, John R. 1960. *A concise Anglo-Saxon dictionary*, 4th edn. [With a supplement by Herbert D. Meritt.] Toronto, Buffalo and London: University of Toronto Press in association with the Medieval Academy of America.
- Colman, Fran. 1991. What positions fit in? In Dieter Kastovsky (ed.) *Historical English syntax*, 51–102. Berlin: Mouton de Gruyter.
- Kastovsky, Dieter. 1992. Semantics and vocabulary. In Richard M. Hogg (ed.) *The Cambridge history of the English language, vol. 1: The beginnings to 1066*, 290–408. Cambridge: Cambridge University Press.
- Kemenade, Ans van. 1987. *Syntactic case and morphological case in the history of English*. Dordrecht: Foris.
- Kemenade, Ans van. 1999. Sentential negation and clause structure in Old English. In Ingrid Tieken-Boon van Ostade, Gunnel Tottie, & Wim van der Wurff (eds.) *Negation in the history of English*, 147–65. Berlin: Mouton de Gruyter.
- Kitson, Peter. 1996. The dialect position of the Old English Orosius. *Studia Anglica Posnaniensia* 30, 3–35.
- Mitchell, Bruce. 1985. *Old English syntax*, 2 vols. Oxford: Clarendon Press.
- Pintzuk, Susan. 1991. *Phrase structures in competition: Variation and change in Old English word order*. Ph.D. dissertation, University of Pennsylvania.
- Taylor, Ann. 2008. Contact effects of translation: Distinguishing two types of influence in Old English. *Language Variation and Change* 20(2), 341–65.
- Taylor, Ann, Anthony Warner, Susan Pintzuk & Frank Beths. 2003. *The York-Toronto-Helsinki parsed corpus of Old English prose*.
<http://www-users.york.ac.uk/~lang22/YCOE/YcoeHome.htm>
- Wende, Fritz. 1915. *Über die nachgestellten Präpositionen im Angelsächsischen*. Berlin: Mayer & Müller. [Palaestra 70.]