Further details of data collection and coding protocols

1. The data
The data for this study are taken from the York-Penn-Helsinki Parsed Corpus of Old English Prose (YCOE). We include referential objects only, and although our focus is on lexical objects, for reasons which will become clear, we also look briefly at pronouns and objects containing a clause. As is usual in this type of study, we limit the data to clauses with an auxiliary verb (Aux) and a main verb (V) in order to abstract away from the effects of verb second (V2) (Pintzuk 1999, among others). While focusing on clauses with two verbs minimizes structural ambiguity (although not eliminating it entirely), it increases the variation, since the verbs can also appear in both orders, AuxV and V Aux. Thus, combining the OV/VO and the AuxV/V Aux alternations, we have four variants: O V Aux, Aux O V, V Aux O and Aux V O. Examples of the relevant orders are given in (1).

(1) a. O V Aux
   gif heo þæt bysmor forberan wolde
   if she that disgrace tolerate would
   ‘if she would tolerate that disgrace’
   (coaelive,+ALS_[Eugenia]:185.305)

   b. Aux O V
   þurh þa heo sceal hyre scippend understandan
   through which it must its creator understand
   ‘through which it must understand its creator’
   (coaelive,+ALS_[Christmas]:157.125)

   c. V Aux O
   þæt he friðian wolde þa leasan wudewan
   that he make-peace-with would the false widow
   ‘that he would make peace with the false widow’
   (coaelive,+ALS_[Eugenia]:209.315)
d. Aux V O

swa þæt heo bið forloren þam ecan life
so that it is lost the eternal life
‘so that it is lost to the eternal life’
(coaelive,+ALS_[Christmas]:144.117)

The dataset includes all the VAux clauses in the YCOE (subject to the exclusions in below). As there are approximately three times as many AuxV clauses as VAux clauses in the corpus, in order to make the sample sizes of the two clause types comparable, we use about one-third of the AuxV clauses. We have excluded the following types:

- AuxV clauses with the object before the Aux, as in (2). In these clauses the object has clearly moved leftward out of the VP, probably due to constraints different from those affecting the OV/VO alternation;
- VAux main clauses with non-overt subjects, or with the object before the subject, to avoid potential cases of topicalization, as in (3);
- pronominal (personal and demonstrative) objects, since pronominal objects usually appear pre-verbally, particularly in VAux clauses (Pintzuk 2005);
- quantified objects (including negative objects), for two reasons: first, it is difficult to determine their information status; second, it has been shown (Pintzuk and Taylor 2006) that quantified objects exhibit special syntactic behavior;
- a few additional cases where the information status of the object is unclear.

(2) O Aux V

þæt hi mine þeawas magon him secgan
that they my customs may him tell
‘that they might tell him my customs’
(coaelive,+ALS[Agnes]:313.1932)

(3) O S V Aux

and þam deadan þu ne miht eft lif forgifan
and (to) the dead you NEG can again life give
‘and you cannot give life again to the dead’
(coaelive,+ALS_[Cecilia]:327.7304)
Table 1 gives basic statistics for the total dataset by type of object.

Table 1. The dataset by object type and verb order

<table>
<thead>
<tr>
<th>Object type (complexity)</th>
<th>AuxV order</th>
<th>VAux order</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>pronominal</td>
<td>573</td>
<td>644</td>
<td>1,217</td>
</tr>
<tr>
<td>lexical</td>
<td>871</td>
<td>769</td>
<td>1,640</td>
</tr>
<tr>
<td>clausal</td>
<td>88</td>
<td>42</td>
<td>130</td>
</tr>
<tr>
<td>Total</td>
<td>1,532</td>
<td>1,455</td>
<td>2,987</td>
</tr>
</tbody>
</table>

2. Operationalizing grammatical weight

Methods for determining grammatical weight and complexity are frequently discussed in studies of word order variation. Wasow (1997: 85) lists eight different proposals found in the literature, but no single method or unit has been unanimously adopted by researchers. Measures may be categorical (e.g. a constituent is heavy if it contains more than one phonological phrase) or continuous (e.g. the number of words). Among the continuous measures that have been used for weight are number of words, nodes, maximal nodes, syllables or graphemes. Wasow (1997: 120) concludes that continuous measures are better than categorical ones; Szmrecsanyi (2004) compares three measures—word counts, node counts, and an index of syntactic complexity—and finds that all three correlate very highly. In our chapter we use the term “grammatical weight” to refer to the characteristics of the string, and “grammatical complexity” to refer to the structure. Our measure of grammatical weight is the number of maximal nodes dominated by the constituent, under the assumptions that (i) all lexical and functional categories project maximal nodes, and (ii) maximal constituents with empty heads do not count for weight. For example, the bracketed nominal phrase in (4) dominates three maximal nodes, two NPs and a PP.

(4) [DP [D his] [NP [N tocyme] [PP [P on] [DP [D e] [NP [N worulde] ] ] ] ] ]

   his arrival in world
   ‘his arrival in the world’

   (coaelive,+ALS[Peter's_Chair]:97.2341)

3. Operationalizing complexity

Wasow (2002: 32–35) shows that it is possible in PDE Heavy-NP Shift to detect an independent effect of complexity in addition to an incremental effect of weight. In an experimental context, he
finds that NPs of equal weight (in words) are more likely to be shifted if they dominate a verb (i.e. a clause). Gries (2003) for PDE Particle Movement and Bies (1996) for ENHG object post-position find that not only clausal objects but also complex ones (conjoined or post-modified) have higher than expected rates of post-verbal position given their length. The full complexity scale we use for objects is: pronominal, simple (unmodified, pre-modified or with a pre- or post-nominal genitive), illustrated in (5), complex (coordinated or post-modified by a PP, an adjectival phrase or an appositive), illustrated in (6), and clausal (dominating a clause). Although Germanic and Romance languages display distinct weight effects for unmodified, pre-modified, and post-modified constituents (Roland Hinterhölzl, p.c.), our data do not show significant differences within the “simple” category.

(5) **simple objects**

a. **unmodified**

[æt hì sceoldon vfel mid gode forgylдан]

that they should evil with good repay

‘that they should repay evil with good’

(cocathom1,+ACHom_I_26.392.130.5063)

b. **pre-modified with determiner/possessive pronoun/adjective or a combination**

i. **determiner (or possessive pronoun)**

His discipuli woldon eac [æt fòk] fedan

his disciples would also the people feed

‘His disciples would also feed the people’

(cocathom1,+ACHom_I_12.277.51.2230)

ii. **adjective**

[æc hì wurdon betæhte hærrihte engelicum bosmum]

but they were committed straightaway angels bosoms

‘but they instantly were committed to the bosoms of angels’

(cocathom1,+ACHom_I_5.220.101.980)

iii. **determiner + adjective**

Unwilles [wæ glon forleosan þa hwilwendlican god]

Unwillingly we may lose the temporary goods

‘Against our will we may lose the temporary goods’
c. **pre-nominal genitive**

þæt we sceoldan ures Drihtnes wundor oncnawan
that we should our Lord’s wonders know
‘so that we should know our Lord’s wonders’

(d. **post-nominal genitive**

for þon wit habbað oferhleoðred þæt gemære uncræs leohtes
because we have over-spoken the limit our light
‘because we have spoken beyond the limit of our light’

(6) **complex objects**

a. **coordinated**

& on eallum þingum we sceolon healdan sibbe and annysse
and in all things we shall hold peace and unity
‘and in all things we shall hold peace and unity’

(b. **post-modified by PP**

Ða ða se Simeon hæfde gewitegod þas witegunge be Criste
when the Simeon had prophesied this prophecy about Christ
‘when Simeon had prophesied this prophecy about Christ’

(c. **post-modified by an adjective phrase**

& cwæð ðæt hi sceoldon habban sunu him gemæne
and said that they should have son them common
‘and said that they should have a son common to them’

(d. **post-modified by an appositive**

forþon þe he ne wolde þy ærran geare gehyran hone arwyrðan fæder Ecgberht
because he would not listen to the venerable father Ecgberht
‘because the previous year he would not listen to the venerable father Ecgberht’
(cobede,Bede_4.27.358.8.3599)

4. Operationalizing the given/new distinction

There is a vast literature addressing the assignment of information status to entities within utterances. We have followed the common line in studies of this sort (e.g. Arnold et al. 2000; Bech 2001; Gries 2003) and used the binary distinction of given vs. new. However, defining these notions on the basis of discourse mention alone is clearly inadequate, since an entity does not have to be mentioned in order to be accessible to or identifiable by the hearer. Instead, we divide the data into given and new entities primarily on the basis of insights drawn from the work of Birner (2006) building on Prince (1981) and Gundel, Hedberg, and Zacharski (1993).

Our new category includes:
- referentially new objects
- new discourse referents (in the sense of Karttunen 1976): a short-term discourse referent that fails to establish a permanent referent, but is or could be referred back to with a pronoun within a limited domain (If a man buys a donkey, he will beat it)
- “bridging” inferrable (Birner 2006): the referent is anchored to an already mentioned referent but the referent is not accessible without the anchor (frequently cases of alienable possession)

Our given category includes:
- previously mentioned entities
- shared/cultural knowledge
- situationally evoked entities: the referent forms part of the situation of the discourse, and is therefore “known” to both the speaker and the hearer
- “elaborating” inferrables (Birner 2006): the referent is anchored to an already mentioned referent and the referent is accessible without the anchor (frequently body parts or other cases of inalienable possession)

(7) Object types coded as new
a. referentially new
Be ðam sagað Sanctus Arculfus
about that said Saint Arculfus
þæt he gesawe medmicle cirican butan Bethlem þære ceastré
that he saw little church outside Bethlehem the city
‘St. Arculfus says about this that he saw a little church outside the city of Bethlehem’
(Mart_5_[Kotzor].Se30_A.3.1906)

b. new discourse referent
Ðeah þe hwa wille her on life habban gode dagas, he ne mæg hi her findan
Yet whoever will here in life have good days, he NEG can them here find
‘Yet whoever will have good days here in life, he cannot find them here.’
(coaelive,+ALS[Ash_Wed]:82.2748)

c. bridging inferable:
ðæt is, ðæt ic sette minne renbogan on wolcnum
that is, that I set my rainbow in clouds
‘that is, that I set my rainbow in the clouds’
(coocest,Gen.9.13.388)

(8) Object types coded as given
a. previously mentioned
& of ðæs treowes wæstme þe is on middan neorxnawange,
and of the tree’s fruit which is in middle paradise
God bebead us, ðæt we ne æton, ne we ðæt treow ne hrepodon ði læs ðe we swelton.
God bid us that we NEG eat nor we the tree NEG touched lest we die
‘and of the fruit of the tree which is in the middle of Paradise, God bid us that we may not eat,
nor may we touch the tree lest we die.’
(coocest,Gen:3.3.123)

b. shared/cultural knowledge
Ond þeah þe wærgcweodole Godes rice gesittan ne mægen
And although those-that-curse God’s kingdom occupy NEG may
‘And although those that curse may not occupy God’s kingdom . . .’
(cobede,Bede_4:27.356.26.3595)
c. *situationally evoked*

[Context: Katherine has been matched against fifty debaters.]

\[\text{ÞV qð ha keiser nauest nawt pis strif rihtwisliche idealet}\]

You said she Emperor NEG-have not this contest fairly matched

‘Emperor, she said, you have not matched this contest fairly . . .’

(CMKATHE,30.187)

d. *elaborating inferable*

\[\text{Healdað mine bebodu } \& \text{ mine domas}\]

Keep my commands and my judgements

\[\text{ðæt ge libbon eower lif butan ælære sorhge.}\]

so-that you may-live your life without any sorrow

‘Keep my commands and my judgements so you may live your life without any sorrow’

(Lev:25.18.3864)

5. The interaction between weight/complexity and given/new

It is sometimes suggested (cf. Niv 1992, quoted in Arnold et al. 2000) that weight/complexity and information status are not independent: since newer constituents are on average heavier than given ones, the weight effect can be reduced to the information status effect. Arnold et al. (2000), Wasow (1997, 2002), Wasow and Arnold (2003) and Gries (2003) have addressed this issue for PDE constructions: they demonstrate that for categories other than pronouns, weight and information status are highly correlated but independent. The same is true for OE simple objects. Table 2 shows that for each weight in maximal node counts, new objects always have a higher frequency of VO order than given ones, although the difference is only significant at the lower end of the weight continuum.
Table 2. Frequency of VO order by maximal node count for given and new simple objects

<table>
<thead>
<tr>
<th>Nodes</th>
<th>Given objects</th>
<th></th>
<th>New objects</th>
<th></th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%VO order</td>
<td>N</td>
<td>%VO order</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>16%</td>
<td>499</td>
<td>26%</td>
<td>162</td>
<td>**</td>
</tr>
<tr>
<td>3</td>
<td>27%</td>
<td>329</td>
<td>37%</td>
<td>203</td>
<td>*</td>
</tr>
<tr>
<td>4</td>
<td>37%</td>
<td>101</td>
<td>41%</td>
<td>63</td>
<td>NS</td>
</tr>
<tr>
<td>5</td>
<td>41%</td>
<td>32</td>
<td>67%</td>
<td>21</td>
<td>NS</td>
</tr>
<tr>
<td>6+</td>
<td>67%</td>
<td>6</td>
<td>80%</td>
<td>5</td>
<td>NS</td>
</tr>
</tbody>
</table>

** = $p < .01$; * = $p < 0.05$

References


