sciendo *Studia Anglica Posnaniensia 56 (2021): 85–119*
doi: 10.2478/stap-2021-0016

OBJECT-VERB IN EARLY MODERN ENGLISH: MODELLING MARKEDNESS[1]

JAVIER PÉREZ-GUERRA[2]

ABSTRACT

Although Verb-Object (VO) is the basic unmarked constituent order of predicates in Present-Day English, in earlier stages of the language Object-Verb (OV) is the preferred pattern in some syntactic contexts. OV predicates are significantly frequent in Old and Middle English, and are still attested up to 1550, when they "appear to dwindle away" (Moerenhout & van der Wurff 2005: 83). This study looks at OV in Early Modern English (EModE), using a corpus-based perspective and statistical modelling to explore a number of textual, syntactic, and semantic/processing variables which may account for what by that time had already become a marked, though not yet archaic, word-order pattern. The data for the study were retrieved from the Penn-Helsinki Parsed Corpus of Early Modern English (1500–1710) and the Parsed Corpus of Early English Correspondence (c.1410–1695), the largest electronic parsed collections of EModE texts. The findings reveal a preference for OV in speech-related text types, which are less constrained by the rules of grammar, in marked syntactic contexts, and in configurations not subject to the general linearisation principles of end-weight and given-new. Where these principles are complied with, the probability of VO increases.

Keywords: Early Modern English; word order; object; corpus; diachrony; multivariate analysis.

1. Introduction

Word order has always been an appealing area of research in the history of English. To give a recent example, the feature "free word order but often V2 [Verb-second] and OV [Object–Verb]" ranks first in van Gelderen's (2018: 17, Table 2.1) account of the major diachronic changes in the syntax and morphology

1 I would like to acknowledge and thank the Spanish State Research Agency and the European Regional Development Fund (grant no. PID2020-117541GB-I00), and the Regional Government of Galicia (grant no. ED431C 2021/52) for their generous financial support.
2 Universidade de Vigo. FFT. Campus Universitario. 36310 Vigo, Spain. jperez@uvigo.es

of English, with the following associated headlines per period: "OV > VO [Verb–Object]" in Early Middle English, "SV [Subject–Verb]; some V2 [Verb-second]" in Late Middle English and "loss of V2" in Early Modern English. The study presented in this article addresses an issue which illustrates well the changes in word order that English has undergone over time: the ordering of verbs and objects in predicates. My analysis focuses in particular on examples such as that in (1), in which the order of the verb and the object does not follow the standard word-order pattern of Present-Day English (PDE), in that the object is preverbal:

(1) but I know that God y=e= maker $\text{hit}_{\text{object}}$ $\text{guides}_{\text{verb}}$, (BOETHEL-E2-P1,17.134)

In earlier periods of the language, English was 'more OV' than it is now. Other studies (cf. Section 2) have shown that the use of OV peters out in Middle English. To quote Seoane (2017: 83), "Early Modern English inherits from Middle English this strong tendency towards SVO word order [and] English ceased to have a general SOV word order [...] in declarative sentences and became the SVO language it is today". Fischer et al. (2000: 139) challenge this view, however:

> One might get the impression from some of the literature that OV order disappeared completely after the Old English period. But [the data] show that OV remained possible throughout the Middle English period. In fact, it did not disappear from prose writings until the sixteenth century, while in verse it continued being used as a productive option well into the nineteenth century.

This study examines VO/OV variation in English in the transition from OV(+VO) to VO word order. In contrast to most previous studies, my analysis focuses on the Early Modern English (EModE) period, uses large multi-genre corpora rather than the more usually sampled small or genre-specific corpora, and applies a multivariate statistical model to a series of variables identified as potentially significant in the literature. My aim is to determine the factors responsible for OV in English when both options were available.

The article is organised as follows. Section 2 contains a brief history of OV and VO patterns and accounts for the main theoretical models and perspectives on the subject, including some suggestions as to the main variables at work. Section 3 describes the goals of the research, the data used, and the variables identified in the previous sections. Section 4 presents the analysis of the data and my findings, followed, finally, by a summary of the study and the main conclusions in Section 5.

2. OV and VO word order: A historical overview

The goal of this study is not to provide a theoretical explanation of the OV>VO switch in English but to identify the linguistic factors which may explain OV. This section summarises previous studies on the diachrony of OV and VO word order in English in order to compile a list of potential variables which may be modelled statistically in the sections that follow. My review of the literature therefore focuses not on the theoretical models and explanations found there,[3] but on the influence exerted on ordering choices by specific grammatical patterns and usage-related constraints.

Old English. The relative order of verbs and objects in Old English (OE) has been explored at length by Pintzuk (1999), Fischer et al. (2000), and Moerenhout & van der Wurff (2010), who claim that both OV and VO orders are attested in OE, as illustrated in (2) and (3), respectively:

(2a) OV_1: OvV:
þe æfre on gefeohte his handa wolde afylan
who ever in battle his hands would defile
'whoever would defile his hands in battle'
(Ælfric's *Lives of Saints* 25.858; Pintzuk 1999: 102)

---

[3] The main theoretical explanations for the highly marked status of OV in modern English and its unmarked status in Old English (OE) and, partially, in Middle English (ME), may be summarised as follows:
(i) The consolidation of VO is a consequence of "abrupt" reanalysis in ME (Lightfoot 1979, 1991), technically due to changes in parametrisation. This theoretical option is favoured by Pintzuk (2002, 2005), among others. Using data from the York-Toronto-Helsinki Parsed Corpus of Old English Prose (YCOE), Pintzuk proposes that the OV>VO switch is part of a general process of change in the syntax of English, which sees the language evolve over time from head-final to head-initial: OE allows both OV and VO; Early ME represents a transitional language with two clearly competing grammars, OV and VO; before the end of ME, VO finally wins out, in keeping with the postulates of Universal Grammar.
(ii) VO is the result of the loss of object movement in ME. Roberts (1997) and Biberauer & Roberts (2005), for example, claim that OE has VO grammar with 'large' overt movements which make objects land in predicate-initial position to check a strong feature. By the end of OE and in ME, such movements are lost (or become non-overt) and these more infrequent instances of OV are reanalysed as examples of object movement (or modern 'object shift'). In Roberts (1997), this reanalysis is attributed to the breakdown of the Case system: when object Case marking is no longer required and (overt) object movement ceases to be required by any kind of Case-feature checking. By Late ME, object movement/'shift' has been lost completely, according to Roberts (1997).
(iii) The drift from OV to VO is a language-external change, as argued by Trips (2002), due to Scandinavian influence in particular. To quote Trips (2002: 75), in Old Norse "VO word order was much more frequent than OV order". When Scandinavian forces invaded and gradually settled parts of Britain between the eighth and tenth centuries, the contact situation was intense and "the VO pattern was taken over".

(2b) OV_2: vOV:
He ne mæg his agne aberan
he not can his own support
'He cannot support his own' (CP 7.53.1; Moerenhout & van der Wurff 2005: 85)

(3) VO:
Ælfric munuc gret ÆDelwærd ealdormann eadmodlice.
Ælfric monk greets Æthelweard nobleman humbly
'The monk Ælfric humbly greets the nobleman Aethelweard.' (ÆGenPref 1)

Fischer & van der Wurff (2006: 185) argue that "OE verbs are usually in clause-final position", making VO a "complication", according to which "a finite verb is moved to second position in main clauses"; or, in other words, a marked[4] option of an "OV with V2" grammar. More significant in terms of this study is their observation that OV is a frequent alternative with pronominal objects, with phrasal-verb particles (when the predicates are phrasal), in subordinate clauses, and in main clauses with auxiliaries. In Section 3, I will consider variables such as the categorial status of the object (pronominal versus non-pronominal, i.e., complete or fully fledged noun phrase), the presence of phrasal-verb particles, the type of syntactic dependency of the clause containing the predicate on the matrix sentence (subordinate, main), and the presence or absence of an auxiliary in the predicate.

Middle English. OV and VO are also attested in Early Middle English (EME), with VO already more prevalent (cf. Allen 2000; Kroch & Taylor 2000; Koopman 2005; and Moerenhout & van der Wurff 2010, among others). Fischer & van der Wurff (2006: 187) remark the "steady decline" of OV in EME, a period ruled by the increasingly rigid VO word order, according to Trips (2002). Pintzuk & Taylor's (2006: 257) corpus-based frequencies reveal that "more than half the objects in Old English are preverbal, compared to only 6 percent of the objects in Middle English".

When looking at the determinants of VO and OV in EME, Kroch & Taylor (2000: 82) note that end-weight plays a significant role, especially since preverbal objects tended to be pronominal in EME (specifically in subordinate clauses). Finally, in their study on the distribution of noun-phrase objects in EME, Elenbaas & van Kemenade (2014: 164) conclude that definite/specific objects are

[4] Unlike Fischer & van der Wurff (2006), Kayne (1994: 49) treats VO as the basic (underlying) word order in English, making OV the marked alternative resulting from the leftward movement of the object from a clause-final to preverbal position.

often preverbal. I include these factors (end-weight and the presence of quantifiers and/or definite/indefinite determiners in the object phrase) in my list of analysis variables (Section 4).

According to van der Wurff (1997), Ingham (2002), and Moerenhout & van der Wurff (2010), the examples investigated in research on OV word order in Late Middle English (LME) are limited to non-literary English, and mostly to the following patterns: predicates with auxiliaries (Ingham's 2002 'embraciated' vOV constructions), predicates with negated/quantified objects,[5] and coordinated and nonfinite clauses. Other variables from these studies examined here include: object negation, coordination, and degree of finiteness (infinitive, *ing*, *ed* or finite). As regards the general frequency of OV, though Fischer et al. (2000: 162) report that "after about 1300 [...] clauses with VO order begin to vastly outnumber those with VO order", van der Wurff & Foster (1997a: 439) claim that OV may have survived "much more tenaciously than suggested", and argue that OV represented not only the survival of an earlier form or an archaism, but an information-packaging given-new function: "OV in late ME prose is anti-triggered by new objects" (van der Wurff & Foster 1997b: 147). The inclusion of given-new in the statistical model is discussed in Section 4.

Early Modern English. Early Modern English (EModE) is especially revealing as regards the patterning of predicates, since from LME onwards word order in English changes from a model ruled by verb-second syntax and information structure, to a more syntacticised solution according to which unmarked subjects are preverbal and generally sentence-initial, and unmarked objects are postverbal, independently of their informative status (cf., among others, Fischer 1992: 371, Van Hoorick 1994: 53, and Bybee 2015: 185 on the syntacticisation of English word order). Research on word order in EModE has been carried out by van der Wurff & Foster (1997a), Moerenhout & van der Wurff (2005: 187), and Fischer & van der Wurff (2006), among others. In short, the claim here is that in the period 1500–1550 "OV survives productively" (van der Wurff & Foster 1997a:

---

5 Ingham (2002) finds approximately 90 percent of the OV clauses to contain negated objects, leading him to argue, from the standpoint of generative grammar, that Neg movement of the object to Spec of NegP, i.e., between Infl and VP, was possible at this stage of the history of English. Elsewhere, he also claims (2000: 34) that Neg movement is no longer available in PDE, whereas in LME it was a form of A'-movement and was an optional alternative. As pointed out by an anonymous reviewer, Pintzuk & Taylor (2006: 271) contend in their study on the change from OV to VO in ME that positive, negative, and quantified objects can underlyingly occur pre- or postverbally and are eligible for different syntactic derivations (positive objects may postpose from an underlying preverbal position in VO clauses, negative objects may prepose from the underlying postverbal position, and quantified objects are subject to both types of movement).

448), and from 1550 onwards it falls away and becomes "exceptional" (Rissanen 1999: 267), this exception referring to its use in poetry texts. Van der Wurff & Foster's (1997a) study on OV order in sixteenth-century English shows that only 42 percent of the OV examples in their corpus contain pronominal objects, which contradicts the assumption that OV obeys given-new, and opens up the possibility of decoupling OV and information status. I will return to their claim that "the association between OV and pronominal objects seems to be lost in the course of time" (1997a: 451) in Section 4 when analysing the connection between VO/OV, given-new and the categorial status of objects.

Late Modern English onwards. There are no in-depth studies on OV in later periods of the history of English. In view of the scarcity of examples after EModE, van der Wurff & Foster (1997b) argue that OV is an archaism in Late Modern (LModE) and PDE, while Takizawa's (2012) empirical analysis of OV sentences with predicates governed by the verb *make* identifies only 79 examples among the 520 million words in the Bank of English corpus of PDE.

The literature reviewed in this section points to a number of possible factors to explain word order in predicates with objects, and also indicates the possible reasons for the success of VO in the history of English, especially after the fixation of word order in the language.

3. The case study: Goals, data, and variables

This section presents the justification for this research (Section 3.1), together with a description of the data used, the retrieval process (Section 3.2), and the variables identified from the review of the literature in the previous section (Section 3.3). This information will pave the way for the statistical analysis presented in Section 4.

3.1. Rationale and goals

This study looks at OV as a marked ordering of the predicate in the recent history of English, and builds on the previous work in this area described in Section 2 above. As pointed out in Section 1, there are three main reasons why more research is needed in this area. Firstly, whereas most of the literature on OV(/VO) focuses on OE and ME, the scope of this study is limited specifically to the EModE period, i.e., after the fixation and syntacticisation of English word order, so the statistical results are not biased by the operation of word-order rules different from those which apply in PDE. Also, as Table 1 below shows, the

number of examples in LModE is so small (3 examples in total) that statistical treatment of VO/OV for this period is not feasible. The research question that has not yet been asked, therefore, is: what forces shaped OV in EModE, when VO had already become consolidated as the unmarked ordering alternative?

Secondly, variation between VO and OV can now be analysed using data from large, multi-genre corpora, whereas previous studies of OV were typically based on small or genre-specific corpora (e.g., letters). The approach used here represents a substantial improvement on earlier work in this area, since text type has been shown to influence the distribution of OV, as in Foster & van der Wurff's (1995) analysis of the frequencies of OV in poetry versus other text types (or genres) for the fourteenth and fifteenth centuries:

- six times more frequent in poetry by 1340
- ten times more frequent in poetry by 1400
- twenty times more frequent in poetry by 1470.

Thirdly, previous empirical analyses of VO and OV are essentially qualitative in nature, based on raw data or, at best, on normalised frequencies. As announced in Section 1, in this study I approach VO/OV variation by applying a widely accepted statistical multivariate analysis of a large database which should shed light on the factors that trigger each pattern in EModE.

3.2. Data

The data were retrieved primarily from two corpora: the Penn-Helsinki Parsed Corpus of Early Modern English (PPCEME; Kroch et al. 2004), which comprises 1,737,853 words from the Helsinki directories of the Penn-Helsinki Parsed Corpus of Early Modern English plus two supplements, and the Parsed Corpus of Early English Correspondence (PCEEC, E1+E2+E3 periods; CEEC Project Team 2006), containing 1,775,310 words. To contextualise the data for EModE, I also looked at VO/OV in (Late) Modern English (1700–1914) using the Penn Parsed Corpus of Modern British English (PPCMBE; Kroch et al. 2010), which comprises 948,895 words.

PPCEME, PCEEC, and PPCMBE are parsed corpora, with almost identical parsing conventions based on part-of-speech and syntactic tagsets couched within the Principles-and-Parameters framework. I used the parsed (.psd) corpus files and retrieved significant instances of VO and OV by means of CorpusSearch 2 (Randall 2008). A simplified version of the search query used to identify clauses with OV predicates is shown below:

(4) node: IP*
query: ((IP* idoms *SBJ)
AND (IP* idoms *OB*|CP-THT|CP-QUE)
AND (IP* idoms VA*|VB*|BA*|BE*|DA*|DO*|HA*|HV*)
AND (*SBJ precedes
VA*|VB*|BA*|BE*|DA*|DO*|HA*|HV*)
AND (*SBJ precedes *OB*|CP-THT|CP-QUE)
AND (*OB*|CP-THT|CP-QUE precedes
VA*|VB*|BA*|BE*|DA*|DO*|HA*|HV*))

The query in (4) retrieves clauses (IPs[6]) containing subjects, objects, and verbal groups which are immediate constituents of the clause and subject to the following conditions: firstly, subjects precede verbal groups; secondly, subjects precede objects; and thirdly, objects precede verbs. Nominal objects are parsed as (*)OB(*) in the corpus, so the alternative tags CP-THT and CP-QUE have been necessary here in order to detect clausal objects not parsed as *OB* in the corpora, such as those underlined in *Noam (that) it was going to rain in Cambridge announced* and *Noam when it is going to rain in Cambridge asked*, respectively. Alternative tags have also been necessary in order to identify the different possible verbal forms, such as BE, DO, HV, and VB, and participles, such as BA (of *be*),[7] DA (of *do*), HA (of *have*), and VA (of other verbs).

To guarantee precision, the resulting database was subjected to extensive manual revision. The list of examples below contains some of the results retrieved by the queries which were excluded from the final database, together with a few misparsed instances: *me-thinks* (and variants), as in (5)–(7), in which the italicised pronouns are parsed as objects of the verbs; catenative constructions such as those in (8)–(10), in which the accusative pronouns in italics are parsed not as the subjects of the embedded nonfinite clauses but as the objects of the main verbs; predicates comprising two objects, one preverbal (OB1) and one postverbal (OB2), the latter following the subject within the same IP, as in (11); and clearly formulaic examples, such as (12) and (13):

---

6 The node IP* retrieves IP-ABS (absolute clauses), IP-IMP (imperative clauses), IP-INF (infinitival clauses), IP-MAT (matrix clauses), IP-PPL (participial clauses), IP-SMC (small clauses), and IP-SUB (subordinate clauses).

7 *Be* is included in the query so that word-order choices such as *John a doctor is* versus *John is a doctor* can also be added to the model. The parser labels predicative complements in copulative predicates as *OB*, so these examples can be retrieved by the query once *be* has been added to the list of verbs immediately dominated by the matrix clausal node. The corpus compilers explain that "[w]hen the second NP in a copular construction is coreferential with the subject (THAT WOMAN IS HILARY CLINTON) or predicated of the subject (SALLY IS MY DOCTOR), it is labelled NP-OB1" (https://www.ling.upenn.edu/hist-corpora/annotation/index.html).

(5) it will serue to make a wiser man than you a foole, *me* thinks. (ARMIN-E2-H,45.390)
(6) the whyche *me* semeth generally good, (FITZH-E1-P1,10.23)
(7) I can when *me* lust make him sory and sad, (UDALL-E1-P1,L33.53)
(8) He would not let *me* be in peace, (HARLEY-E2-H,5.117)
(9) And for that intent he cam to Bethlem in hys owne persone to se *them* take Downe. (TORKINGT-E1-P2,47.22)
(10) A small thing might make me *all* in the grounde to throwe. (UDALL-E1-H,L.297.214)
(11) The greatest worth [that $Ø_{obj}$ fortunes guiftes woorthyest can giue], be such as in abondant sorte to wicked folkes do hap. (BOETHEL-E2-P1,35.486)
(12) W=th= my intire love and saluts to thee and my daughter, I remaine Thyne, till death us p=t=, Richard Haddock. (RHADDSR-1670-E3-P2,12.67)
(13) But I neuer reade it yet as Gode me saue. (UDALL-E1-P2,L847.115)

The raw and normalised frequencies of OV in the EModE and LModE database are set out in Table 1:

Table 1. OV/VO raw/normalised frequencies and percentages

		OV raw[8]	VO raw (max.)	OV %	words	OV nf (/1,000w)
EModE1	1500–1569	186			877,015	0.21
EModE2	1570–1639	67			1,539,138	0.04
EModE3	1640–1710	20			1,097,010	0.02
EModE	*1500–1710*	*273*	*76,014*	*0.36%*		
LModE1	1700–1769	2			298,764	0.01
LModE2	1770–1839				368,804	0.00
LModE3	1840–1914	1			281,327	0.00
LModE	*1700–1914*	*3*	*41,071*	*0.007%*		

Figure 1, which displays the normalised frequencies for OV from the fourteenth century onwards, is based on both my own data for EModE and LModE, and data from some of the studies mentioned above:[9,10]

[8] All examples from Bible text types were discarded since they do not necessarily reflect adjustment to current syntactic conventions.
[9] Foster & van der Wurff (1995) for 1330–1380, and Moerenhout & van der Wurff (2010) for 1378–1400, 1421–1442, and 1442–1479.
[10] To facilitate comparison between this and previous studies, the baseline for the normalised frequencies is the number of words (and not, e.g., the number of IPs or verbs, which would be methodologically more plausible).

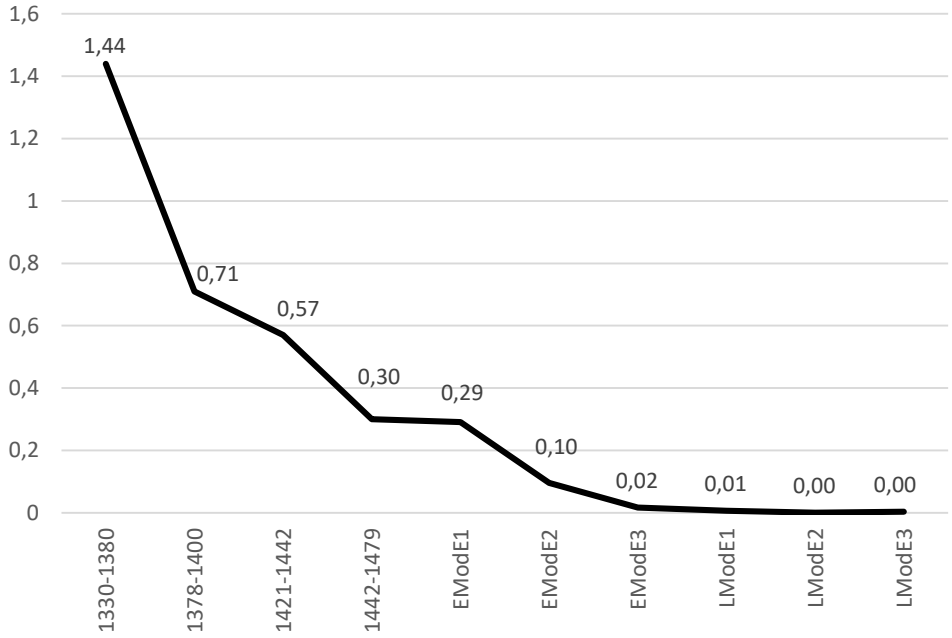


Figure 1. Normalised OV frequencies (/1,000w) from the 14th century onwards

Constrained to the scope (OV in EModE) and goal of the study (to identify the linguistic forces responsible for OV in EModE), the resulting data set comprised 273 examples of OV and 76,014[11] examples of VO. Since the addition of 76,014 VO examples to the database proved impracticable, I applied the R (R Core Team 2017) function 'n.for.survey' ('epiDisplay' package, Chongsuvivatwong 2018) *n.for.survey(p=.08, delta=.02, popsize=76,287, alpha=0.05)* in order to ascertain the minimal size of the database in terms of statistical significance. The minimal sample size for the survey of a complete population (*popsize*) of 273 OV plus 76,014 VO examples (76,287 in total) was determined as 700 OV-plus-VO instances, based on the following criteria: the estimated probability (*p*) of 0.08, and a very conservative margin of error of the

[11] The maximum number of VO examples in the EModE corpora was determined using the following CorpusSearch query:
((IP* idoms *SBJ)
AND (IP* idoms *OB*|CP-THT|CP-QUE)
AND (IP* idoms VA*|VB*|DA*|DO*|HA*|HV*)
AND (*SBJ precedes VA*|VB*|DA*|DO*|HA*|HV*)
AND (*SBJ precedes *OB*|CP-THT|CP-QUE)
AND (VA*|VB*|DA*|DO*|HA*|HV* precedes *OB*|CP-THT|CP-QUE))

estimate (*delta*) of 0.02, i.e., of the difference between the estimated prevalence and the width of one side of the percent confidence level defined by *alpha* (an *alpha* value of 0.05 implies that the confidence level is 95%). The final data set for this study comprised 928 examples in total, 273 of them examples of OV.

3.3. Variables

The preliminary list of variables in (14) contains all the potential determinants of VO/OV variation mentioned or analysed in the works cited in Section 2:

(14) textual variables:
- genre
- (sub)period

linguistic variables:
- pattern
- co-occurrence with auxiliaries
- discontinuity between object and verb (or verb and object)
- clause-initial verbal form
- presence of a particle (in a phrasal verbal group)
- verbal finiteness
- main/subordinate/coordinated clause
- (c/)overt subject
- subject length
- object length
- category of object (pronoun, noun phrase, clause, other)

semantic, discourse-related variables:
- quantified objects
- negated objects

The following paragraphs explain briefly the process of coding each variable and the decisions taken in order to obtain convincing statistical evidence from the data. The first variable, 'Genre', encompasses the values listed in the right-hand column in Table 2. Given the small number of examples in the database, in particular of OV, I adopted a drastically simplified version of the taxonomy of text types in Culpeper & Kytö (2010). In my revised taxonomy, the different text types recognised by the corpus compilers are classified into two main categories, 'writ' and 'speech', comprising text types within the writing-based/purposed/like family and the speech-based/purposed/like group, respectively. As shown in Table 2, the 'speech' group also comprises texts included in the 'drama-comedy' category. In order to assess the proportion of examples containing verse or at least rhyming couplets, I conducted a

qualitative analysis of OV examples in the 'drama-comedy' category, and found only 14 potential cases of OV due to rhythmic variation, which cannot be said to distort the conclusions reported in Section 4. Philosophy texts are listed in a separate category since they contain features of both formal written texts and speech-based/purposed language owing to the inclusion of the dialogues from Boethius' *De Consolatione Philosophiae* in the corpus. The particular behaviour of the 'phil' category is explained in more detail in Section 4.

Table 2. Genre

writ	writing-based/purposed/like	Educ-treatise
		History
		Law
		Science-medicine
		Science-other
		Travelogue
		Biography-auto
		Biography-other
		Fiction
speech	speech-based/purposed/like	Diary-priv
		Drama-comedy
		Letters-priv
		Letters-non-priv
		Proceeding-trials
		Sermon
phil	philosophy	

The variable 'Period' classifies the examples diachronically into three subperiods: EModE1 1500–1569 ('E1'), EModE2 1570–1639 ('E2'), and EModE3 1640–1710 ('E3').

'Pattern' categorises the (surface) clause design. The list of patterns is extensive and here I include only the most representative types. The following naming conventions have been used to represent the main clause constituents: 'S' for subject, 'V' for lexical verb, 'v' for auxiliary, 'O' for object, '0' for covert subject, and 'X' for other.

VO patterns

(i) SVO:

(15) SVO: but [the Trinity]$_S$ keep$_V$ you$_O$. (APLUMPT-E1-H,185.85)
(16) SvVO: when he$_S$ was$_v$ building$_V$ [that admirable worke of his tombe]$_O$ (ARMIN-E2-H,46.410)
(17) SVXO: He$_S$ had$_V$ [no sooner]$_X$ [the liberty of his tongue]$_O$, (DELONEY-E2-P2,51.297)
(18) SvVXO: but by her cheeks you$_S$ might$_v$ find$_V$ guilty$_X$ Gilbert$_O$ (ARMIN-E2-P2,39.298)
(19) SvXVO: [the middle letter]$_S$ doth$_v$ alwayes$_X$ signifie$_V$ [the Angle propounded]$_O$, (BLUNDEV-E2-P2,57V.18)
(20) SvXvVO: that I$_S$ shoulde$_v$ thus$_X$ haue$_v$ refused$_V$ [the oth]$_O$. (MORELET2-E1-H,506.44)
(21) SvXVXO: And if [any one]$_S$ shall$_v$ throughly$_X$ weigh$_V$ [in his Mind]$_X$ [the Force and Energy of the one and of the other]$_O$, (BOETHPR-E3-H,191.376)
(22) SXvVO: I$_S$ truly$_X$ can$_v$ accuse$_V$ you$_O$ of none. (THOWARD2-E2-P2,101.55)
(23) SXVXO: And in this yere [the kynge]$_S$ [at the Request of the duke of Orleaunce]$_X$ sent$_V$ [ouer the foresayd duke]$_X$ [his sone]$_O$ (FABYAN-E1-H,174V.C2.196)

(ii) with inverted subjects:

(24) VSO: Ford. Has$_V$ Page$_S$ [any braines]$_O$? (SHAKESP-E2-P1,49,C1.876)
(25) vSVO: And thus do$_v$ [the best Divines]$_S$ expound$_V$ [the Place]$_O$. (JUDALL-E2-P2,1,175.312)
(26) vSVXO: L. C. J. Did$_v$ [my Lady Lisle]$_S$ ask$_V$ you$_X$ [that Question]$_O$? (LISLE-E3-P2,4.118.337)
(27) vSXVO: should$_v$ we$_S$ therefore$_X$ judg$_V$ [those who retain their Sight]$_O$ to be blind also? (BOETHPR-E3-H,183.330)

(iii) subjectless:

(28) 0VO: and 0 saw$_V$ [great danger]$_O$ on both hands: (BURNETCHA-E3-P1,2,171.260)
(29) 0vVO: and 0 will$_v$ emploie$_V$ [all other meanes possible]$_O$, (EDMONDES-E2-H,394.23)

(30) 0VXO: and 0 kepe$_{V}$ close$_{X}$ [such matters]$_{O}$. (LATIMER-E1-H,38L.351)

OV patterns

(i) OV:

(31) SOV: This profe I trow may serue, though I$_{S}$ [no word]$_{O}$ spoke$_{V}$. (STEVENSO-E1-H,54.218)
(32) SOXV: God$_{S}$ [all Rules]$_{O}$ [by goodnes]$_{X}$ order$_{V}$ (BOETHEL-E2-P2,71.256)
(33) SXOV: who$_{S}$ [for like faulte out of the citie]$_{X}$ [the name of kings]$_{O}$ abolisshed$_{V}$. (BOETHEL-E2-P1,34.464)
(34) SXOXvV: And Goodlucke$_{S}$ [I dare sweare,]$_{X}$ [your witte]$_{O}$ therin$_{X}$ would$_{v}$ low$_{V}$. (UDALL-E1-P2,L1563.786)

(ii) vOV:

(35) SvOV: alledging that he$_{S}$ hath$_{v}$ nothing$_{O}$ done$_{V}$, (WOLSEY-E1-H,2.2,21.17)
(36) SvOXV: I$_{S}$ shall$_{v}$ hir$_{O}$ [no more]$_{X}$ see$_{V}$. (UDALL-E1-H,L.1111.442)
(37) SvXOvV: We$_{S}$ should$_{v}$ therat$_{X}$ [such a sporte and pastime]$_{O}$ haue$_{v}$ founde$_{V}$, (UDALL-E1-P2,L1563.780)
(38) SXvOV: Here [Martin luther]$_{S}$ [for his shrewed brayne]$_{X}$ wyll$_{v}$ [some thyng]$_{O}$ wrastell$_{V}$ agaynst vs. (FISHER-E1-P2,337.68)

(iii) vOV with inversion:

(39) vSOV: C. Cust. Will$_{v}$ ye$_{S}$ [my tale]$_{O}$ breake$_{V}$? (UDALL-E1-P2,L1469.671)
(40) vSOXV: T. Trusty. Do$_{v}$ you$_{S}$ [that part]$_{O}$ wel$_{X}$ play$_{V}$ (UDALL-E1-P2,L1594.797)
(41) vSXOV: So shall$_{v}$ we$_{S}$ pleasantly$_{X}$ [bothe the tyme]$_{O}$ beguile$_{V}$ now, And eke dispatche all our workes ere we can tell how. (UDALL-E1-H,L.297.196)

(iv) subjectless:

(42) 0OvV: nor also 0 none$_{O}$ can$_{v}$ haue$_{V}$. (MORERIC-E1-P1,32.135)
(43) 0OXV: and 0 hym$_{O}$ [myserably in his Chaumbre]$_{X}$ slewe$_{V}$ (FABYAN-E1-H,170R.C1.85)
(44) 0vOV: But I woulde be auenged in the meane space, On that vile scribler, that 0 did$_{v}$ [my wowyng]$_{O}$ disgrace$_{V}$. (UDALL-E1-H,L.1145.493)

(45) 0XOV: And 0 [by and by]$_X$ them$_O$ opened$_V$, euen as they were before, (STEVENSO-E1-H,14.147)
(46) 0XvOV: ich trust 0 soone$_X$ shalt$_v$ it$_O$ see$_V$ (STEVENSO-E1-P1,33.539)

In view of the scarcity of examples in some of the subcategories, and also to avoid undesirable collinearity effects with other variables, the system of classification for 'Pattern' variables was subsequently simplified and reduced to a single variable, as explained at the end of this inventory.

The dichotomous 'Auxiliary' variable encodes the presence or absence of an auxiliary verbal segment in the verbal group.

'Continuous' is used in examples in which lexical material (encoded as 'X' in the taxonomy of patterns) is found between the verbs and the objects (in VO constructions) or vice versa (in OV examples). In other words, this variable is associated with either (v)VXO or OX(v)V patterns.

'Verb-first', also with only two possible (positive or negative) values, tags verb-first constructions, found mostly in interrogative and exclamative clauses, but also in instances of subject-verb inversion, as triggered by an emphatic or negative sentence-initial word (*only*, *so*, etc.), for example.

The variable 'Particle' applies in examples such as (47),[12] which contain particles in phrasal verbal groups (e.g., *away* in this example). As observed in previous studies (see Section 2), particle placement can be a significant predictor of object position:

(47) And there was a Justice of peace had taken away much of frends goods: (FOX-E3-P2,109.140)

The 'Finiteness' variable encodes the status of the clause in relation to how finite/nonfinite the matrix verb is. The options are: finite clause (the most frequent in the data), infinitive clause (illustrated in (48)), and *ing* clause (in (49)). No VO or OV *ed* clauses were attested in the corpora:

(48) And thus I desyre our Lorde to have you in his moste gratious tuytion. (GCROMW-E1-P1,209.9)
(49) The Priest and the Tanner seeing the Taylor, mused what hee made there: (DELONEY-E2-P1,16.253)

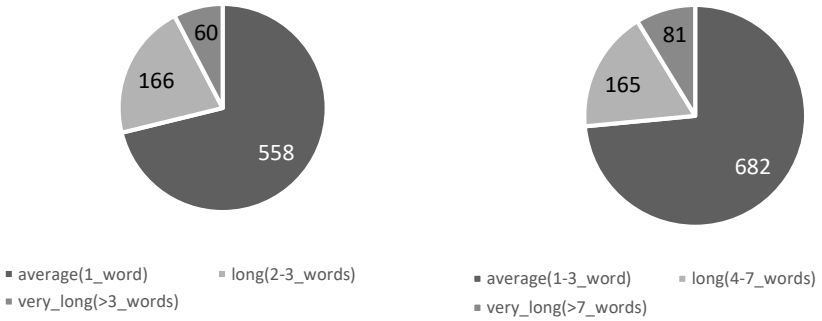
[12] Elenbaas (2007, 2013) suggests that in examples such as (47) the verb moves from the post-object/particle position, the underlying order being OV, to the pre-object/particle position, to surface as a VO predicate. In this study, the examples are categorised according to their surface design, (47) being a clear example of VO order.

The 'Main_sub(ordinate)' variable concerns the syntactic connection between the example and its host sentence. More specifically, it identifies: main clauses (the vast majority of the examples in the data), subordinate clauses (illustrated in (50)), and clauses coordinated with other units within the same sentence (as in (51)):

(50) for I thinke so God me mende, This will proue some foolishe matter in the ende. (UDALL-E1-P2,L751.17)
(51) "Then that is the top of felicitie, that stowtly rules & 0 gently all disposith." (BOETHEL-E2-P2,71.264)

The overt/covert subject dichotomy is accounted for by the '(C/)Overt subject' variable, with the 'covert' option referring to (subjectless) patterns containing a '0' subject.

The length of the subject ('Subj_length') and that of the object ('Obj_length') are also taken as variables. Measurement methods vary between studies. In his influential work on verb-particle organisation, Gries (2003: 83–84) codes for the number of syllables as well as the number of words, concluding that both measures yield very similar results, with the number of syllables proving a slightly better predictor of the ordering choice. Similarly, in his study on utterance length, Yaruss (1999: 339) reports that "there were very strong, positive, significant correlations [...] among measures of length in words, syllables, morphemes, and clausal constituents". Szmrecsányi (2004) compares metrics based on words, syntactic nodes in phrase-markers, and complexity counts, and concludes, as before, that "determining length in words [...] is by all means one that is nearly as accurate as the most sophisticated and cognitively, conceptually, or even psychologically 'more real' methods" (2004: 1038). Shih & Grafmiller (2011) demonstrate in their discussion of genitive and dative alternations that "the number of words [...] can act as a sufficient proxy for [...] 'weight' [length]". In this study, subject and object length is measured in words. The resulting numerical values were ordinalised as follows to facilitate their subsequent factorisation. For 'Subj_length', I established three levels: 'average' length (1 word, yielding 558 examples), 'long' (2–3 words, 166 examples) and 'very long' (more than 3 words, 60 examples). As regards objects, the following categories were established: 'average' length (1–3 words, 628 examples), 'long' (4–7 words, 165 examples) and 'very long' (more than 7 words, 81 examples). The relative weight of each category in the database is illustrated in Figure 2:



Subj_length Obj_length

Figure 2. Subject and object length

The category of the object was deemed to merit its own variable. The predictor 'Object' comprises the following object types: non-*wh* pronominal objects (such as *me*, *I*, *mine*), complete (fully fledged) noun phrases with a covert or ellipsed head noun, and other cases, including clausal and *wh* objects. *Wh* pronominal objects are excluded from the 'pronominal' category because their sentence-initial position is triggered by the syntax of the sentence and is consequently not subject to variation. Instead, examples with *wh* objects have been included in the 'other' option of the variable 'Object'.

Quantified objects ('Quantif_obj') are classed as: definite, indefinite (including nominals that can host determiners and those that surface with no determiner), cardinal or ordinal, depending on the type of determiner occurring in the noun-phrase objects.[13]

When the object has been negated linguistically, e.g., by a negative determiner such as *no* (as in (52)) or a negative adverb, it is regarded as 'negated' within the variable 'Neg_obj'. Other examples are classified as 'non-negated'. It must be stressed that only linguistic negation, not semantic negation, has been taken into consideration.

---

[13] While the definite and indefinite categories of the variable 'Quantif_obj' coincide broadly with, for example, Huddleston & Pullum's (2002: 356–357) definiteness classification of objects into definite (definite article *the*, demonstrative determinatives *this*/*that*, personal determinatives *we*/*you*, and universal determinatives *all*/*both*) and indefinite (indefinite article *a*/*an*, distributive determinatives *each*/*every*, existential determinatives *some*/*any*, cardinal numerals *one*/*two*, disjunctive determinatives *either*/*neither*), cardinals and ordinals were initially kept separate to allow for their specific contribution to the statistical model to be assessed by the regression analysis.

(52) M. Mery. Nay fayth ye shall promise that he shall no harme haue, (UDALL-E1-H,L.1179.505)

Following completion of this preliminary list of variables, a series of simplifications were introduced to reduce the number of predictors. Regarding the variable 'Pattern', the following actions were taken to correct the significant collinearity between it and other variables:

- The segment 'vV' was simplified to 'V' in each pattern, to avoid collinearity with the variable 'Auxiliary'.
- '0' in the subjectless examples was eliminated as a distinctive pattern constituent. In other words, in the simplified classification, no explicit distinction is made between subjectless examples and examples with overt subjects, in order to avoid collinearity with the variable '(C/)Overt subject'.
- Verb-first patterns (interrogatives, exclamatives, inversions) were no longer treated as distinct patterns, to avoid interaction with the variable 'Verb-first'.
- 'VXO' in patterns with lexical material occurring between the lexical verb and the object was simplified to 'VO', to avoid interaction with the variable 'Continuous'.
- 'SX' was reduced to 'S'. Since this research focuses on the segment comprising the verbal form up to (and including) the object in a VO construction, and the object up to the verb in a VO example, the occurrence of lexical material preceding the segment 'verb ... object' and 'object ... verb' is not relevant to the study.

These simplification processes reduced the list of possible patterns to just five – OV, VO, vOV, vXOV and vXVO – covering all of the examples in the data, based on the conditions outlined above. This list was then subjected to a further process of simplification, once again with the aim of avoiding collinearity with other variables:

- 'OV' was disregarded as a distinct pattern because of one-to-one collinearity with one of the values of the response variable (OV).
- For the same reason, 'VO' was not considered a value of the variable 'Pattern' because it expresses one of the two options of the response variable (VO).
- 'vOV' was reduced to 'OV' owing to collinearity with the variable 'Auxiliary' and then removed from the list since it coincided with the response variable.

This left me with just two levels for the variable 'Pattern', namely vXOV and vXVO, which exhibit partial collinearity with the response variable but which both contain a distinctive feature that does not interact with the other variables, namely the occurrence of lexical material between the auxiliary and the lexical verbal form. This is not an instance of discontinuity, as accounted for by the variable 'Continuous', since, as defined above, only the existence of lexical material between the verb and the object, or vice versa, determines the discontinuous status of the segment, which is not the case in either vXOV or vXVO. Once collinearity with the response variable was neutralised in these patterns, the constructions could be successfully added to the model in reference to a single variable with two options: intervening material following v ('mat') – vXVO, vXOV – and lack of intervening material following v ('no_mat'). In other words, following the elimination of collinearity effects with other variables, the whole list of patterns was reduced to one variable: 'Int_mat'.

The variables 'Particle' and 'Verb-first' could not be included in the statistical model because of the low number of occurrences in one level. The raw figures are shown below in (53) and (54):

(53)

Particle	no particle	particle
OV	272	1
VO	625	30

(54)

Verb-first	verb-first	verb-*first
OV	270	3
VO	630	25

Some variables had to be recoded owing to the lack of data:

- Since only one *ed* clause was detected in the database, the 'Finiteness' variable was encoded as a dichotomous predictor with the value: 'finite', for examples with inflected verbal forms, and 'nonfinite', comprising the infinitive, *ing* and *ed* levels.
- The levels 'very long' and 'long' in 'Obj_length' were combined and relabelled as 'long' because of the scarcity of examples of the former (see (55)).
- Because of the lack of 'other' OV examples in my database (see (56)), the levels of the variable 'Object', which reflects the syntactic category of objects, were reduced to two, namely 'pro(nominal)' and 'non_pro(nominal)', the latter subsuming the original levels 'NP' and 'other'.

- All of the non-definite (mainly 'cardinal' and 'indefinite') objects in the VO examples were added to the 'indefinite' level of the predictor 'Quantif_obj' since no examples of the OV pattern were found for the other categories of this variable, as shown in (57):

(55)

Obj_length	average	long	very long
OV	252	18	3
VO	430	147	78

(56)

Object	NP	other	pro
OV	150	0	123
VO	429	92	134

(57)

Quantif_object	definite	cardinal	indefinite
OV	178	0	95
VO	215	23	95

Collinearity between the level 'subjectless' of the variable '(C/)overt subject' and the level '0' of 'Subj_length' was resolved by discarding the former.

The final revised list of variables is presented below in (58):

(58) response variable ('Var'): VO, OV
textual variables:
- Period
- Genre

linguistic variables:
- Int(ervening)_mat(erial)
- Auxiliary
- Continuous
- Finiteness
- Main_sub(ordinate, coordinated)
- Subj(ect)_length
- Obj(ect)_length
- Object (category)

semantic, discourse-related variables:
- Quantif(ied)_obj(ect)
- Neg(ated)_obj(ect)

The number of examples in the smaller level of the response variable (273 OV instances) in relation to the number of predictors (12) is perfectly in keeping with the restrictions imposed by regression analysis. The summary of the data is shown in Table 2, where frequency is the criterion used to establish the reference levels for each predictor.

Table 2. Summary of the data per predictor

Var	Period	Genre	Int_mat	Auxiliary	Continuous	Finiteness
vo: 655	E1: 368	writ: 338	no_mat: 866	no_aux: 518	continuous: 796	finite: 848
ov: 273	E2: 324	phil: 85	mat: 62	auxiliary: 410	discontinuous: 132	nonfinite: 80
	E3: 236	speech: 505				

Main_sub	Subj_length	Object	Obj_length	Quantif_obj	Neg_obj
main: 336	average: 558	non_pro: 671	average: 682	indefinite: 535	non-neg: 882
coord: 172	long: 166	pro: 257	long: 246	definite: 393	negated: 46
sub: 420	subjectless: 144				
	very long: 60				

4. Analysis of the data

This section reports the statistical analysis of the data to determine the relative weights of the predictors within a multivariate model. A fixed-effects binomial logistic regression analysis was applied to the data using the function 'glm' in the 'stats' package (R Core Team 2017) and 'lrm' in the 'rms' package (Harrell 2019) in R, the results of which are summarised in Table 3 below. The lack of significant mutual collinearity of the data was corroborated using the functions 'alias' ('MASS' package, Venables & Ripley 2002) and 'vif'[14] (Variance Inflation

[14] VIF values:

	GVIF	Df	GVIF^(1/(2*Df))
Period	1.274139	2	1.062439
Genre	1.533674	2	1.112841
Int_mat	1.051046	1	1.025205
Auxiliary	1.234607	1	1.111129
Continuous	1.077003	1	1.037788
Finiteness	1.139830	1	1.067628
Main_sub	1.372618	2	1.082399
Subj_length	1.700642	3	1.092535
Object	1.239955	1	1.113533
Obj_length	1.149710	1	1.072245
Quantif_obj	1.193391	1	1.092425
Neg_obj	1.159316	1	1.076716

Factor; 'car' package, Fox & Weisberg 2011). Backward stepwise (*step(emode_glm, direction="backward"*; 'MASS' package) revealed that the difference between the explanatory power of the models with and without the predictor 'Continuous', as denoted by their respective AIC (Akaike Information Criterion) values, was not statistically significant (AIC values: 611.79 and 609.93 before and after step, respectively; Pr(>Chi)=0.7083). On that basis, I have used the reduced model after step and discarded the predictor 'Continuous', resulting in the model represented in Table 3 below.[15]

The fact that the model's residual deviance of 575.93 is lower than the degrees of freedom (911) indicates a lack of overdispersion of the data. The model's goodness of fit is also corroborated by the Hosmer-Lemeshow-Cessie test, which gave a *p* value of 0.055. Finally, both the C(oncordance) and Nagelkerke R^2 discrimination indices (0.926 and 0.635, respectively) provided by 'lrm' reveal that the model is very good at explaining the variation (C>0.9 indicates the model's outstanding fit and predictive power, and R^2>0.4 its plausibility) and, consequently, adequate to the research question.

In what follows, predictor levels are first grouped according to the degree of success of each of the response variable alternatives (VO and OV), as predicted by the odds ratios (OR). The analysis of the data concludes with a comprehensive interpretation of the trends revealed by the success/failure odds.

Table 3 provides a summary of the results of the regression model and the interpretation of the odds and ORs in terms of the probability of success of either VO or OV:

[15] glm(var ~ period + genre + int_mat + auxiliary + finiteness + main_sub + subj_length + object + obj_length + quantif_obj + neg_obj, data=emode, family=binomial)

Table 3. Logistic regression and odds (Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1)

	Estimate	Std. Error	z value	Pr(>|z|)		Odds/OR	
(Intercept)	-3.8921	0.4307	-9.037	<2e-16	***	0.0204	
Period:E2	-1.49	0.2531	-5.888	3.92e-09	***	0.2253	The probability of success of OV with respect to the variable's reference level is 0.18; therefore the probability of VO should be 0.82.
Period:E3	-2.3445	0.3266	-7.178	7.09e-13	***	0.0958	The probability of success of OV with respect to the variable's reference level is 0.09; therefore the probability of VO should be 0.91.
Genre:phil	3.7566	0.412	9.118	<2e-16	***	42.802	The probability of success of OV with respect to the variable's reference level is 0.98; therefore the probability of VO should be 0.02.
Genre:speech	1.8681	0.3065	6.096	1.09e-09	***	6.4759	The probability of success of OV with respect to the variable's reference level is 0.87; therefore the probability of VO should be 0.13.
Int_mat:mat	-2.124	0.6186	-3.434	0.000595	***	0.1195	The probability of success of OV with respect to the variable's reference level is 0.11; therefore the probability of VO should be 0.89.
Auxiliary:auxiliary	1.0525	0.234	4.497	6.88e-06	***	2.8648	The probability of success of OV with respect to the variable's reference level is 0.74; therefore the probability of VO should be 0.26.
Finiteness:nonfinite	-1.5234	0.5491	-2.775	0.005528	**	0.2179	The probability of success of OV with respect to the variable's reference level is 0.18; therefore the probability of VO should be 0.82.
Main_sub:coord	-0.6104	0.3366	-1.814	0.069742	.	0.5431	The probability of success of OV with respect to the variable's reference level is 0.35; therefore the probability of VO should be 0.65.

Main_sub:sub	0.6538	0.2451	2.667	0.007648	**	1.9228	The probability of success of OV with respect to the variable's reference level is 0.66; therefore the probability of VO should be 0.34.
Subj_length:long	0.2092	0.2891	0.724	0.469200		1.2326	The probability of success of OV with respect to the variable's reference level is 0.55; therefore the probability of VO should be 0.45.
Subj_length: subjectless	2.0536	0.3544	5.795	6.83e-09	***	7.7959	The probability of success of OV with respect to the variable's reference level is 0.89; therefore the probability of VO should be 0.11.
Subj_length:verylong	1.2705	0.4126	3.079	0.002075	**	3.5626	The probability of success of OV with respect to the variable's reference level is 0.78; therefore the probability of VO should be 0.22.
Object:pro	1.295	0.2491	5.199	2.00e-07	***	3.6509	The probability of success of OV with respect to the variable's reference level is 0.78; therefore the probability of VO should be 0.22.
Obj_length:long	-1.2581	0.3194	-3.938	8.20e-05	***	0.2841	The probability of success of OV with respect to the variable's reference level is 0.22; therefore the probability of VO should be 0.78.
Quantif_ obj:definite	1.868	0.2311	8.084	6.26e-16	***	6.4753	The probability of success of OV with respect to the variable's reference level is 0.87; therefore the probability of VO should be 0.13.
Neg_obj:negated	3.0113	0.4797	6.278	3.43e-10	***	20.313	The probability of success of OV with respect to the variable's reference level is 0.95; therefore the probability of VO should be 0.05.

The results show a greater probability of success of the response variable reference level VO when (see the corresponding effects plot in Figure 3; 'effects' package, Fox & Weisberg 2018):

- (sub)period is E2 or E2
- intervening material occurs between v and VO, that is, when the pattern has been categorised as vXVO
- verb is nonfinite
- object has been classed as 'long' in reference to its length.

By contrast, the predictor levels whose odds indicate greater success of OV are:

- speech and philosophy genre or text type
- verbal group containing auxiliary
- subordinate clause
- subjectless clause
- clause with very long subject
- pronominal object
- definite object
- negated object.

The following results in terms of frequency and syntactic context confirm VO as the unmarked linearisation option from EModE onwards. Firstly, the ORs from E1 to E2/E3 reflect an increase in the success probability of VO during the EModE period. Secondly, VO is more probable in nonfinite clauses, which reflects its widespread use in both unmarked and marked morphosyntactic contexts, and its deviation from the tendency reported for LME, when OV was favoured in nonfinite clauses (cf. Section 2). Thirdly, preference for OV is observed in constructions which obey the principles governing linearisation in contemporary English, namely end-weight (Quirk et al. 1985: 1398) and given-new (Vallduví & Engdahl 1996: 460). As regards the former, the compliance of VO with end-weight is corroborated by the ORs for the predictors 'Int_mat', 'Obj_length', 'Auxiliary' and 'Object'.

That the degree of success of the VO alternative increases when intervening material is present between v and V in an vXVO/vXOV predicate is in keeping with the principle of end-weight. In detail, the occurrence of a constituent (X) after the auxiliary in an vX[VO|OV] pattern increases the length of the preverbal (i.e., pre-V) part of the predicate and thus makes the pattern deviate from end-weight. If, in addition, the object is also preverbal (vXOV), then the pre-V part will be larger and, in consequence, this OV alternative will be more disrespectful with regard to end-weight. The preference for VO thus betters the compliance of the pattern with end-weight when intervening material is present in the predicate.

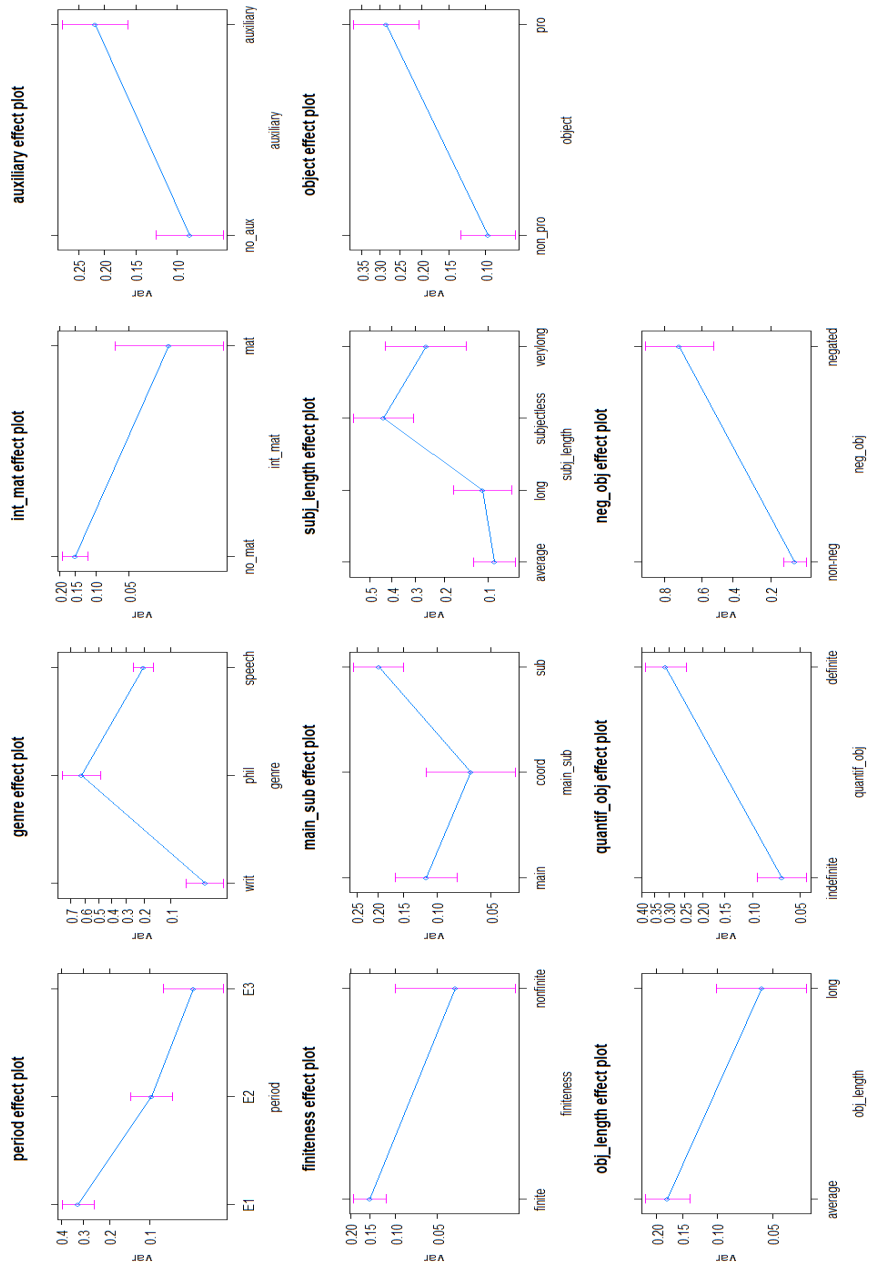


Figure 3. Effects plot

The positive influence exerted by the occurrence of auxiliaries on the selection of the VO pattern also supports end-weight. The scenarios with auxiliaries are: vVO and vOV when the pattern is VO, and OvV with OV predicates. The OR here indicates that OV is the successful option when the auxiliary is present; in other words, the success probability of OvV is greater than that of vVO/vOV in examples containing auxiliaries. However, the effect of the interaction between the predictors 'Auxiliary' and 'Obj_length' in the model[16] reveals that OV is specifically more successful in constructions with auxiliaries that contain short/average objects, i.e., in patterns which conform to the design 'Subject – short/average Object – Auxiliary – lexical Verb'. Since the length of the object has a decisive impact on the linearisation of the object and the verb, the preference for OV here demonstrates that the variation is subject to end-weight. The success of OV is also facilitated when the object is pronominal, which reveals that when the object is not long and/or pronominal, VO is not dominant since the resulting clausal design respects end-weight to a lesser extent than with longer objects.

VO/OV has also been claimed to be subject to another major principle of linearisation, given-new. The review of the literature in Section 2 revealed that, at least at some stages in the history of English, VO/OV has been constrained by this principle. Since the informative load of the linguistic units is not annotated in the corpora, I have not been able to perform a detailed analysis of the connection between the response variable and the type of (given or new) information conveyed by the objects (or by other major clausal constituents). However, considering that most pronouns are informationally given (i.e., familiar, known, inferable), the higher probability of success of the OV level with pronominal objects in the database, as shown in Table 3 above,[17] may be interpreted as indicative of the compliance of examples containing pronominal objects with the given-new principle.

In the absence of semantic annotation in my database, this study has tried to address the information structure by examining both pronominal and definite objects as part of the options covered by the variable 'Quantif_object', the typical function of definite objects being, to quote Elenbaas (2013: 501), "to refer back

---

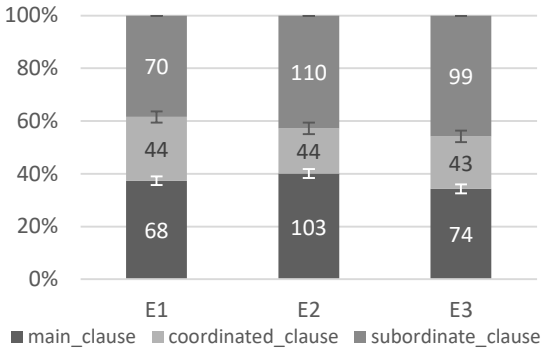
16 The lack of statistical significance (Pr(>Chi)=0.8501) in the difference between the models with and without interaction between the predictors 'object length' and 'presence of auxiliaries' (*glm(var ~ period + genre + int_mat + auxiliary*obj_length + finiteness + main_sub + subj_length + object + quantif_obj + neg_obj, data=emode, family=binomial)*) indicates that the effect of these predictors is additive; in other words, these factors separately do have a positive effect on the model.

17 The lack of significant interaction between the predictors 'object length' and 'object type' suggests that objects other than those classified as average in terms of length may be pronominal and may, in consequence, convey given information.

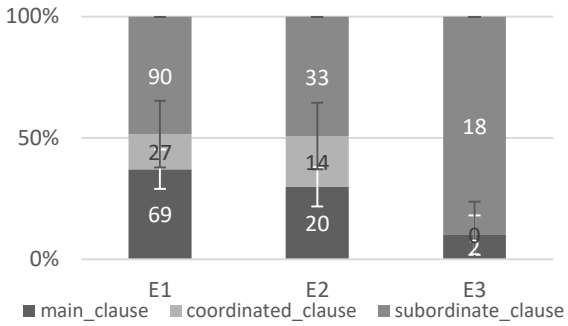
to an NP earlier in the discourse, rather than introducing a new referent". The ORs corresponding to these levels suggest that the probability of success of OV increases with definite objects, thus reinforcing the plausibility of given-new as a driving force of object-verb linearisation in modern English.

Regarding the information structure, one would assume that most negated objects contain new information and, consequently, should be placed in the final or at least the postverbal position. However, Table 3 above reveals that OV is favoured by negated objects. Ingham (2002: 302) explains this seemingly contradictory finding by arguing that in EModE (and in fact, as early as Late Middle English) only "a slight majority of embraciated negated object nouns [i.e., in OV predicates] had received no prior mention". The fact that many negated objects convey given information explains the preference for OV and indicates compliance with the principle of given-new.

The statistical model also confirms the following findings regarding the OV alternative. Firstly, the probability of success of OV increases in 'speechy' text types and genres, which are more spontaneous and less constrained by syntactic and/or processing rules, as predicted by the ORs for the 'speech' and 'philosophy' levels of the 'Genre' variable (see Section 3.3 on the speech-like character of the philosophy texts in the corpora). Secondly, as reported above in relation to the ORs for the levels of the predictors 'Int_mat', 'Obj_length', 'Object' and 'Auxiliary', OV is favoured in contexts which do not conform to the rules of modern linearisation in English. As regards the variable 'Subj_length', for example, the greater success of OV with very long subjects reflects the primacy of OV where end-weight does not apply. Far from agreeing with end-weight, therefore, the clausal design 'very long Subject – Object – Verb' illustrates the opposite principle, with longer initial constituents and shorter clause-final units. Thirdly, OV is attested in constructions which are not syntactically unmarked. A first example of this is the success of OV in subordinate clauses. The OR for the predictor 'Main_sub' shows that OV is more probable when the clause is subordinate than when it is a main clause. In fact, the trend towards VO word order in contexts of subordination increases over time, as shown in Figure 4.



VO



OV

Figure 4. Period and syntactic status of clauses

Whereas the proportions of main/coordinated/subordinate contexts that illustrate VO show no statistically significant variation during EModE, 18 out of the 20 examples of subordinate contexts in the subperiod 'E3' are OV (90 percent), compared with the proportions of OV in 'E1' and 'E2' of 48.38 and 49.25 percent, respectively. This demonstrates the very low productivity of OV by the end of EModE in main/coordinated clauses, i.e., in clausal constructions that are more frequent and therefore unmarked.

The success of OV in syntactically marked contexts is also supported by its productivity in subjectless clauses. The model reveals that VO is preferred in nonfinite clauses and that OV is more successful with covert subjects. While this could be understood as contradictory, the interaction of the predictors 'Finiteness'

and 'Subj_length' is only statistically significant at the .0919 level in the model.[18] The model shows that greater success of VO is predicted by the levels 'main' and 'coord' of the variable 'Main_sub', whose variation is not statistically significant. At the same time, ANOVA reveals a lack of significant correlation between the levels of the variables 'Subj_length' and 'Finiteness'. These findings indicate the likelihood of OV in subjectless finite sentences not necessarily attested in contexts of coordination, which highlights the marked grammatical status of OV as the option selected in EModE constructions which are not subject to the rules that govern the syntax of Present-Day English.

The analysis in this section of the multivariate model and the trends revealed by measuring the association between predictors and levels shows that the regularisation of VO in modern English is corroborated by its increased frequency during the EModE period, its adoption in not only unmarked but also marked, formerly VO-excluding syntactic constructions (e.g., nonfinite clauses), and its selection in constructions that respect the linearisation principles of end-weight and given-new. By contrast, OV has been shown to be more probable in text types and genres that are less constrained by syntactic or processing rules, in syntactic contexts that are not unmarked (e.g., subordinate and subjectless clauses) and in constructions not subject to end-weight or given-new.

5. Summary and concluding remarks

The aim of this study was to investigate Object-Verb linearisation in the recent history of English, specifically since the fixation of the principles ruling word order in Modern and Present-Day English. The review of the literature in Section 2 revealed that the order of objects and verbs in OE is triggered by morphosyntactic variables: not only by the category of the objects (pronominal versus noun phrases) but also by the syntactic dependency type (main/coordinated versus subordinate) of the clauses. In ME, determinants such as end-weight, quantification and definiteness of the objects come into play alongside the morphosyntactic forces operative in OE, with the further addition of textual factors in LME.

Previous empirical studies on OV, and also a preliminary search of this ordering in a corpus of LModE, reveal that OV is statistically marginal in EModE and a mere archaism by PDE. In light of the scarcity of data in LModE onwards, this study focuses on the forces favouring OV in EModE. My analysis diverges

[18] Model 1: var ~ period + genre + int_mat + auxiliary + finiteness + main_sub + subj_length + object + obj_length + quantif_obj + neg_obj; Model 2: var ~ period + genre + int_mat + auxiliary + obj_length + finiteness + main_sub * subj_length + object + quantif_obj + neg_obj; Pr(>Chi) = 0.09198.

from previous accounts mainly in terms of data and methodology. My data are taken from the largest electronic parsed collections of EModE texts (approximately 3.5 million words in total), and a logistic regression analysis of the data makes it possible to model the variables and the levels with greater explanatory power. The study analyses 928 examples of OV and VO sentences and classifies them according to a final taxonomy of 11 variables which instantiate systematic grammar factors such as period, genre, syntactic pattern, presence of auxiliaries, verbal finiteness, main/coordinated/subordinate clausal status, object category, subject length, object length, quantified object, negated object, and additional intervening constituents.

The results discussed in Section 4 reveal that the adoption of VO as the successful linearisation solution in modern English predicates has been progressive and has been favoured by contexts which are textually and syntactically unmarked and compliant with the principles of end-weight and given-new. Four specific conclusions can also be drawn. The first relates to the lesser effect of 'speechy' text types on VO: in fact, the more 'speech-related' and spontaneous the text types (diaries, drama, letters, trials, sermons, and philosophy texts), the greater the frequency of OV. Secondly, as far as factors evincing syntactic markedness are concerned, the probability of VO has been shown to increase in marked syntactic contexts where OV prevailed in previous periods, as in the case of nonfinite clauses. Syntactically marked constructions such as subordinate (versus main/coordinated) and subjectless clauses (versus those with overt subjects) were found to be most suited to the OV word order. Thirdly, end-weight was found to be a key triggering force here. On the one hand, OV is more probable with shorter and pronominal objects, and with verbal groups containing auxiliaries, whereas VO is preferred when lexical material intervenes between the verb and the object. The occurrence of intervening material and a preverbal object in the predicate increases the length of the pre-V component of the construction and the resulting design does not comply with end-weight, since the preverbal segment is longer than the predicate-final verbal section. On the other hand, clauses with very long subjects, which are constructionally incompatible with given-new, usually conform to the OV word order. Finally, the model shows that linearisation is also subject to given-new, with OV more probable when the objects are pronominal and definite, i.e., when they prototypically convey given information.

The greater probability of success of VO word order in textually and syntactically unmarked contexts in EModE and the progressive growth of this trend from 1500 to 1710 reveal that, once the grammar of English prioritised VO as the unmarked linearisation choice in all types of contexts, OV was left relegated to a strategy of (marked) topicalisation or stylistic fronting. In other words, this study has suggested that the increased probability of VO in nonfinite predicates in the later periods under

investigation, and the compliance of most notably the VO examples with the general linearisation principle of end-weight, are indicative of the progressive regularisation and consolidation of VO as the unmarked alternative already in Early Modern English. By contrast, the greater probability of OV predicates specifically in contexts of subordination, in sentences that flout end-weight and in speech-related text types reflects the residual status of VO as an option of predicate linearisation by the end of the EModE period.

A number of questions remain for future research. As regards the influence exerted by given-new on VO/OV, a fine-grained qualitative analysis of the database examples and their linguistic contexts may confirm the actual impact of given-new.[19] It also remains to be investigated whether verb type has a bearing on the preference for specific syntactic dependencies and, by extension, on VO/OV. In this respect, Kempen & Harbusch (2017) have studied the association between verb frequency and the main-clause status in spoken and written PDE and found that linguistic claims are often 'main-clause biased' because of the "overrepresentation of a small-set of high-frequent verbs in main clauses" (p.117).

REFERENCES

PRIMARY SOURCES

CEEC Project Team. 2006. *Parsed Corpus of Early English Correspondence, parsed version*. Annotated by Ann Taylor, Arja Nurmi, Anthony Warner, Susan Pintzuk & Terttu Nevalainen. University of York & University of Helsinki.

Kroch, Anthony, Beatrice Santorini & Lauren Delfs. 2004. *Penn-Helsinki Parsed Corpus of Early Modern English*. Department of Linguistics, University of Pennsylvania.

Kroch, Anthony, Beatrice Santorini & Ariel Diertani. 2010. *Penn Parsed Corpus of Modern British English*. Department of Linguistics, University of Pennsylvania.

---

19 Without detracting from the importance of analysing the connection between word order and information structure, I would like to emphasise the higher (statistical) significance of linguistic over informative factors in constructions such as those investigated in this paper. In her studies on the order of particles and objects in predicates governed by phrasal verbs, Elenbaas (2007, 2013) investigates, among other variables (e.g., morphosyntactic, phonetic), the kind of information expressed by the particles (transparent versus non-transparent meaning) and the objects ('given', including Prince's (1981) textually/situationally inferrable and evoked values, and 'new', for brand-new and unused objects). Even though she finds that given objects are increasingly attested in the V-O-particle order and new objects increasingly appear in the V-particle-O order from ME to EModE, Elenbaas concludes that morphosyntactic variables (e.g., length of object) have a stronger impact on particle-verb word order than object givenness or definiteness. This claim also is supported by Gries' (2003: 150) findings on particle placement in PDE.

SECONDARY SOURCES

Allen, Cynthia. 2000. Obsolescence and sudden death in syntax: The decline of the verb-final order in early Middle English. In Ricardo Bermúdez-Otero, David Denison, Richard M. Hogg & C. B. McCully (ed.), *Generative theory and corpus studies: A dialogue from 10 ICEHL*, Mouton de Gruyter. 3–26. DOI: 10.1515/9783110814699.3

Biberauer, Theresa & Ian Roberts. 2005. Changing EPP parameters in the history of English: Accounting for variation and change. *English Language and Linguistics* 9(1). 5–46. DOI: 10.1017/S1360674305001528

Bybee, Joan. 2015. *Language change*. Cambridge University Press. DOI: 10.1017/CBO9781139096768

Chongsuvivatwong, Virasakdi. 2018. *epiDisplay: Epidemiological Data Display Package*. http://medipe.psu.ac.th/epicalc

Culpeper, Jonathan & Merja Kytö. 2010. *Early Modern English dialogues. Spoken interaction as writing*. Cambridge University Press.

Elenbaas, Marion. 2007. *The synchronic and diachronic syntax of the English verb-particle combination*. PhD thesis. Utrecht: LOT.

Elenbaas, Marion. 2013. Motivations for particle verb word order in Middle and Early Modern English. *English Language and Linguistics* 17(3). 489–511. DOI: 10.1017/S1360674313000130

Elenbaas Marion & Ans van Kemenade. 2014. Verb particles and OV/VO in the history of English. *Studia Linguistica* 68(1). 140–167.

Fischer, Olga. 1992. Syntax. In Norman Blake (ed.), *The Cambridge history of the English language. Vol. 2: 1066–1476*, Cambridge University Press. 207–408. DOI: 10.1017/CHOL9780521264754.005

Fischer, Olga & Wim van der Wurff. 2006. Syntax. In Richard Hogg & David Denison (eds.), *A history of the English language*, Cambridge University Press. 109–198. DOI: 10.1017/CBO9780511791154.004

Fischer, Olga, Ans van Kemenade, Willem Koopman & Wim van der Wurff. 2000. *The syntax of early English*. Cambridge University Press. DOI: 10.1017/CBO9780511612312

Foster, Tony & Wim van der Wurff. 1995. The survival of object-verb order in Middle English: Some data. *Neophilologus* 79. 309–327. DOI: 10.1007/BF00999787

Fox, John & Sanford Weisberg. 2011. *An R companion to applied regression*. SAGE.

Gelderen, Elly van. 2018. *Analyzing syntax through texts. Old, Middle, and Early Modern English*. Edinburgh University Press.

Gries, Stefan Thomas. 2003. *Multifactorial analysis in corpus linguistics: A study of particle placement*. Continuum.

Harrell, Frank E., Jr. 2019. rms: Regression Modeling Strategies. http://biostat.mc.vanderbilt.edu/rms

Hoorick, Bart Van. 1994. *Pragmatic positions and the history of English word order*. University of Amsterdam.

Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge University Press. DOI: 10.1017/9781316423530

Ingham, Richard. 2000. Negation and OV order in Late Middle English. *Journal of Linguistics* 36(1). 13–38. DOI: 10.1017/S0022226799007963

Ingham, Richard. 2002. Negated subjects and objects in 15th-century nonliterary English. *Language Variation and Change* 14(3). 291–322. DOI: 10.1017/S095439450214302X

Kayne, Richard S. 1994. *The antisymmetry of syntax*. The MIT Press.
Kempen, Gerard & Karin Harbusch. 2017. Frequential test of (S)OV as unmarked word order in Dutch and German clauses. A serendipitous corpus-linguistic experiment. In Hilke Reckman, Lisa Lai-Shen Cheng, Maarten Hijzelendoorn & Rint Sybesma (eds.), *Crossroads semantics. Computation, experiment and grammar*, John Benjamins. 107–123. DOI: 10.1075/z.210.07kem
Koopman, Willem. 2005. Transitional syntax: Postverbal pronouns and particles in Old English. *English Language and Linguistics* 9(1). 47–62. DOI: 10.1017/S136067430500153X
Kroch, Anthony S. & Ann Taylor. 2000. Verb-object order in early Middle English. In Susan Pintzuk, George Tsoulas & Anthony Warner (eds.), *Diachronic syntax. Models and mechanisms*, Oxford University Press. 132–163.
Lightfoot, David W. 1979. *Principles of diachronic syntax*. Cambridge University Press.
Lightfoot, David W. 1991. *How to set parameters: Arguments from language change*. The MIT Press.
Moerenhout, Mike & Wim van der Wurff. 2005. Object-verb order in early sixteenth-century English prose: An exploratory study. *English Language and Linguistics* 9(1). 83–114. DOI: 10.1017/S1360674305001553
Moerenhout, Mike & Wim van der Wurff. 2010. Remnants of the old order: OV in the *Paston Letters*. *English Studies* 81(6). 513–530. DOI: 10.1076/enst.81.6.513.9184
Pintzuk, Susan. 1999. *Phrase structures in competition: Variation and change in Old English word order*. Garland.
Pintzuk, Susan. 2002. Verb-Object order in Old English: Variation as grammatical competition. In David W. Lightfoot (ed.), *Syntactic effects of morphological change*, Oxford University Press. 276–300.
Pintzuk, Susan. 2005. Arguments against a universal base: Evidence from Old English. *English Language and Linguistics* 9(1). 115–138. DOI: 10.1017/S1360674305001565
Pintzuk, Susan & Ann Taylor. 2006. The loss of OV order in the history of English. In Ans van Kemenade & Bettelou Los (eds.), *The handbook of the history of English*, Blackwell. 249–278.
Prince, Ellen F. 1981. Toward a taxonomy of given-new information. In Peter Cole (ed.), *Radical pragmatics*, Academic Press. 223–256.
Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. Longman.
R Core Team. 2017. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org
Randall, Beth. 2008. *CorpusSearch 2 users guide*. http://corpussearch.sourceforge.net/CS-manual/Contents.html
Rissanen, Matti. 1999. Syntax. In Roger Lass (ed.), *The Cambridge history of the English language, Vol. 3: Early Modern English 1476–1776*, Cambridge University Press. 187–331. DOI: 10.1017/CHOL9780521264761.005
Roberts, Ian. 1997. Directionality and word order change in the history of English. In Ans van Kemenade & Nigel Vincent (eds.), *Parameters of morphosyntactic change*, Cambridge University Press. 397–427.
Seoane, Elena. 2017. Chapter 5: Syntax. In Laurel Brinton & Alexander Bergs (eds.), *The history of English. Vol, 4: Early Modern English*, Mouton de Gruyter. 68–88. DOI: 10.1515/9783110525069-005

Shih, Stephanie & Jason Grafmiller. 2011. Weighing in on end weight. Paper delivered at the 85th Annual Meeting of the Linguistic Society of America. Pittsburg, PA, 9–11 January.

Szmrecsányi, Benedikt M. 2004. On operationalizing syntactic complexity. In Gérard Purnelle, Cédrick Fairon & Anne Dister (eds.), *Le poids des mots: Proceedings of the 7th International Conference on Textual Data Statistical Analysis, March 10–12, 2004, Vol. 2*, Presses universitaires de Louvain. 1032–1039.

Takizawa, Naohiro. 2012. A corpus-driven functional analysis of the SOV construction in Present-Day English. Paper delivered at the 4th International Conference on Corpus Linguistics.

Trips, Carola. 2002. *From OV to VO in Early Middle English*. John Benjamins. DOI: 10.1075/la.60

Vallduví, Enric & Elisabet Engdahl. 1996. The linguistic realization of information packaging. *Linguistics* 34. 459–519. DOI: 10.1515/ling.1996.34.3.459

Wurff, Wim van der. 1997. Deriving object-verb order in late Middle English. *Journal of Linguistics* 33(2). 485–509. DOI: 10.1017/S0022226797006531

Wurff, Wim van der & Tony Foster. 1997a. Object-verb order in 16th century English: A study of its frequency and status. In Raymond Hickey & Stanisław Puppel (eds.), *Language history and linguistic modelling: A festschrift for Jacek Fisiak on his 60th birthday, Vol. 1*, Mouton de Gruyter. 439–453. DOI: 10.1515/9783110820751.439

Wurff, Wim van der & Tony Foster. 1997b. From syntax to discourse: The function of object-verb order in Late Middle English. In Jacek Fisiak (ed.), *Studies in Middle English linguistics*, Mouton de Gruyter. 135–156. DOI: 10.1515/9783110814194.135

Venables, W. N. & B. D. Ripley. 2002. *Modern applied statistics with S*. Springer. DOI 10.1007/978-0-387-21706-2

Yaruss, J. Scott. 1999. Utterance length, syntactic complexity, and childhood stuttering. *Journal of Speech, Language, and Hearing Research* 42(2). 329–344. DOI: 10.1044/jslhr.4202.329