
VARIATION IN OLD ENGLISH REVISITED: A CORPUS-DRIVEN APPROACH*

S O U M I K D E Y

W I L L I A M G . S A K A S

C I T Y U N I V E R S I T Y O F N E W Y O R K

ABSTRACT Corpus linguistics can be divided into two major avenues of research — *corpus-based*: Searching a corpus based on preexisting hypotheses or intuitions, and *corpus-driven*: An unbiased search of a corpus independent of any framing hypothesis or intuition. Corpus-driven methods have been touted to be more proficient in identifying previously undocumented patterns. This article revisits the variation observed in Old English (OE) by first discussing some of the existing corpus-based studies and their findings. Next, a corpus-driven methodology of exploration based on generating and searching for all possible permutations of selected syntactic labels (*S*, *V*, *O*, *p* and *Aux.*), i.e., all possible word order patterns is presented. Finally, after applying the corpus-driven methodology to the York-Toronto-Helsinki Corpus of Old English (YCOE) and outlining some broad assumptions that are valid cross-linguistically, the word order patterns attested in YCOE are syntactically analyzed — of note is the in-depth analysis of embedded adverbial adjunct clauses with respect to CP-recursion. This study documents and presents analyses of an extensive list of word order patterns in OE and categorically verifies certain theories of OE syntax, and challenges others. To the best of our knowledge, the study presented in this article is the first corpus-driven investigation of the variation observed in OE. More generally, this study lays a foundation for future corpus-driven and corpus-based research on Old English syntax.

* We would like to thank Jason Kandybowicz for his insights on some of the thornier syntactic analyses and A. W. Strouse for his invaluable help with Old English glosses and translations. Finally, for their sagacious advice and extensive feedback, we are very grateful to three patient JHS referees who helped us find the path to this final draft. This research was supported in part by grant 6D160-00 01 from the Professional Staff Congress of the City University of New York.

1 INTRODUCTION

1.1 *Variation Observed in Old English*

There is documented variation in the word order of Old English (OE). Consider (1) and (2).

- (1) *ða godan weorc ðe he ær geworht hæfde.*
 the good work that he before made had.
 ‘the good work that he had made before.’

[CP:33.219.6.1464;¹ (Trips 2002: 76)]

- (2) *þæt he wolde geswutelian swa his digelmyse eow.*
 that he would reveal so his secrets you.
 ‘that he wanted to reveal his secrets to you in such a way.’

[ÆLS (Thomas) 166; (Haeberli 1999: 360)]

The two examples above illustrate conflicting word orders. (1) exhibits verb-final order, and (2) verb-initial order. Existing theories of OE grammar claim that Old English is: verb-final (OV) in base (Van Kemenade 1987, Haeberli 2002, 2005, Koopman 1995b), or verb-initial (VO) in base (Kiparsky 1996, Roberts 1997, Wallenberg 2009, Struik & Van Kemenade 2020), or that OE is both VO and OV in base with variation in INFL-medial and final (Pintzuk 1991, Pintzuk & Haeberli 2008) where the VO/OV and INFL-medial/final grammars are in competition (Kroch 1989, 1994). Fuss & Trips (2001), while acknowledging Pintzuk’s evidence of variation in Old English, propose that the variation is not manifested in a ‘dual base’ grammar of OE involving headedness direction in IP and VP but rather argues for headedness variation in VP alone.

In this article, we present a corpus-driven approach to the study of word order variation found in the York-Toronto-Helsinki Parsed Corpus of Old English Prose (YCOE). We first briefly discuss corpus-based vs. corpus-driven approaches and then review several existing theories of OE syntax which helped shape the linguistic assumptions we adopt, to finally derive syntactic analyses of the results of our corpus-driven search.

¹ Examples drawn from the York-Toronto-Helsinki Parsed Corpus of Old English Prose (Taylor, Warner, Pintzuk & Beths 2003, Taylor 2007) will be cited with an index which can be used to locate the example in the corpus.

1.2 *Corpus-based vs corpus-driven*

There are two main research methodologies in corpus linguistics: *corpus-based* and *corpus-driven*. Corpus-based studies begin with predefined hypotheses that come from the researcher's deductions and/or intuitions based on linguistic theory. A search is then conducted on relevant corpora with the aim of testing these hypotheses. In contrast, a corpus-driven study makes generic a priori assumptions and approaches a corpus without hypotheses in hand. As described in [Tognini-Bonelli \(2001: 84\)](#), in a corpus-driven study, the corpus *drives* the study — theoretical conclusions are made after the analysis of the results of a thorough search of the corpus. Proponents of this approach ([Hoey 2009](#), [Krishnamurthy 2008](#), [Tognini-Bonelli 2001](#), [Sinclair 2004](#), [Hunston 2002](#), [Teubert 2005](#), among others) claim that corpus-driven approaches are more successful in uncovering novel, previously unattested observations.

In practice, the distinction between corpus-based and corpus-driven linguistics is not so neat due to the necessity of adopting certain linguistic assumptions in order to construct search parameters and aid in the post-search analysis (e.g., deciding whether or not to assign part-of-speech tags prior to a corpus search); see [Biber \(2009\)](#) and references there for discussion. As many have noted (e.g., [Biber 2009](#), [Lyubymova 2019](#)), one can think of the distinction as more of a scale.

Historical linguists have generally been reluctant to adopt predominantly corpus-driven methodologies, e.g., [Ringe & Eska \(2013: Chapter 1\)](#), [Campbell \(2013: 484\)](#) who oppose the adoption of quantitative methods in historical corpus linguistics. However, there has been work, e.g., [Jenset \(2013\)](#), [McGillivray \(2014\)](#), [Morley & Sift \(2006\)](#) among others, that has applied corpus-driven methodologies to the study of historical linguistics. To the best of our knowledge, all corpus studies regarding OE word order variation have been corpus-based where researchers searched a corpus guided by one or more linguistic hypotheses regarding the grammar of OE.

In this study, we complement previous corpus-based studies by employing a corpus-driven methodology to investigate word order variation in YCOE. Since we were not targeting specific linguistic phenomena, the corpus-driven search was able to document a comprehensive collection of word order patterns², some previously identified and some previously unidentified, that corroborate or challenge previous findings concerning OE.

This complementary approach of theoretical corpus-based studies with empirical corpus-driven studies mirrors productive lines of investigation in the natural and social sciences; for example some astronomers look for pat-

² Subject to certain constraints, see §3.

terns when looking at a celestial object, and others form a hypothesis and predict what patterns they will find. We would hope that future corpus-based investigations of Old English will expand and delve deeper into some of the findings we present here.

2 BACKGROUND

There is much debate about the underlying syntactic structure and movement operations in Old English. In this section, we review several theories which informed the linguistic assumptions we adopted to derive syntactic analyses of the word order patterns uncovered during our corpus-driven search. The assumptions are detailed in §3.

2.1 *Subject Position*

Modern syntactic theory suggests that subjects are base generated in Spec(VP) and might move to Spec(IP) due to the Extended Projection Principle (EPP: Chomsky 1981, 1982) or higher due to topicalization.

Most of the literature on OE (e.g., Pintzuk 1991, 1993, Pintzuk & Haerberli 2008, Haerberli 2002, 2005, Haerberli & Ihsane 2016, Biberauer & Van Kemenade 2011, Van Kemenade & Milicev 2012) among others, proposes that there are two subject positions in OE. In most natural languages, the subject is restricted to one of two subject positions: high and low. OE shows evidence of variability in the position of the subject.

For example, Haerberli (2002, 2005) identifies the higher subject position as Spec(AgrP) and the lower position as Spec(TP) with a split-INFL analysis, while Pintzuk & Haerberli (2008) posit the lower position of DP subjects³ in OE to be Spec(VP) while pronominal subjects behave as clitics and, as in most West Germanic languages, move to Spec(IP) or higher.

Subjects have also been documented to undergo postposition, i.e., a rightward movement similar to the rightward adjunction of non-pronominal NPs/DPs and PPs. Pintzuk (1991: 161-162) provides evidence of subject postposition in OE).

³ In modern syntax, all subjects and objects are DPs, even ones without an overt determiner as it is deemed as DP with an empty D. In the language used in Pintzuk & Haerberli (2008) they do not follow this modern convention and refer to DP subjects as subjects having an overt determiner. We will use DP subjects/objects in the same sense as Pintzuk & Haerberli (2008).

2.2 INFL headedness

The headedness of INFL, i.e., whether OE is INFL-medial or final, has been heavily contested. Early research advocated for both INFL-medial (Van Kemenade 1987) and INFL-final (Kiparsky 1995) analyses. Pintzuk (1991) proposes the Double Base Hypothesis providing evidence of variation in IP headedness. Consider an example of OE, (3), and two structural analyses of (3), in (4).

- (3) *þe God worhte þurh hine*
 which God wrought through him
 ‘which God wrought through him’
 [ÆLS 31.7, (Pintzuk 1991: 75)]

- (4) a. INFL-medial:
 $þe [IP God_j [I worhte_i] [VP t_j þurh hine t_i]]$
 b. INFL-final:
 $þe [IP God_j [VP t_j t_k t_i] [I worhte_i]] [PP þurh hine]_k$
 [ÆLS 31.7, (Pintzuk 1991: 75)]

Although (3) is ambiguous, Pintzuk provides other examples of clauses that can unambiguously be analyzed as INFL-medial as well as clauses that can unambiguously be analyzed as INFL-final. For example, Pintzuk notes that clauses with heavy object and subject constituents that occur before the verb and auxiliary verb are unambiguously INFL-final. Pintzuk also provides evidence for the existence of INFL-medial phrase structure in OE in subordinate clauses by analyzing the distribution of particles, pronouns, and one-syllable adverbs.⁴

Considering particles, in West Germanic languages like Modern German and Dutch, which are believed to be OV, the particle always precedes the verb. In Old English there is variation in particle position relative to the verb. Particles appear in both pre-verbal and post-verbal positions. Pintzuk analyzes this by dividing word order patterns into groups — verb-medial (the main verb precedes the object) and verb-final (where the main verb follows the object).

She notes that in verb-medial clauses without an auxiliary verb, particles appear with considerable frequency after the main verb. Particles almost never appear after the main verb in verb-final clauses. One analysis Pintzuk

⁴ Pintzuk (1991) also points to the dearth of verb projection raising in subordinate clauses as additional evidence of INFL-medial word order. However, she later revised that conjecture in Haerberli & Pintzuk (2012).

gives for verb-medial clauses is the leftward movement of V to INFL-medial over a particle and an object.

However, in verb-final clauses since particles cannot be postposed to the right, as a postposition operation is bound by a heaviness constraint, particles are rarely observed post-verbally. This mismatch of particle distribution in verb-medial and final clauses can be seen as unambiguous evidence for INFL-medial in OE.⁵

Pintzuk analyzes the distribution of pronouns and single-syllable adverbs similarly to that of particles. These elements can only appear post-verbally if they are postposed or if the verb has moved leftward over them. However, like particles, these elements are not heavy and hence do not undergo postposition. Pintzuk observes that only in verb-medial clauses do these elements occur post-verbally, making a case for INFL-medial base phrase structure.⁶

More recent work by Roberts (1997), Struik & Van Kemenade (2020, 2022), Biberauer & Roberts (2005) and others adopt Kayne (1994)'s hypothesis and assume that OE is INFL-medial with verb-final word orders derived by object scrambling to the left.

2.3 *Verb position*

In modern syntactic theory, the verb is base generated in VP and can either move to INFL to get tense and then continue to COMP in some cases or the verb can stay in situ with tense hopping down to V. In addition, in head-final languages, the verb can raise to the right (verb raising) or can take its complements with it as it raises to the right (verb projection raising).

There is a considerable amount of literature (Van Kemenade 1987, Koopman 1990, Higgins 1991, Lightfoot 1991) that claims that OE has an OV base phrase structure. Pintzuk (1991) as part of her Double Base Hypothesis proposes a variable VO/OV base phrase structure. She points out that the distribution of particles paves the way for an extended analysis of OE which allows for VO in base.

- (5) *he wolde adræfan ut anne æpeling*
 he would drive out a prince
 'he would drive out a prince'

[ChronB (T) 82.18-19 (755), (Pintzuk 1991: 180)]

⁵ Pintzuk (1991) observes that particles follow the main verb in verb-medial constructions with an auxiliary verb but attributes this word order to underlying VO base phrase structure.

⁶ Pintzuk (1991) notes that the frequency in which these elements appear post-verbally is less than that of particles due to the fact that these elements sometimes behave as clitics and move leftward and attach to the periphery of Spec(IP).

In verb-medial phrase structures with an auxiliary, the particle sometimes appears after the main verb. She points to (5) as evidence of VO base phrase structure because the main verb *adraefan* cannot move leftward due to an overt auxiliary verb *wolde* that blocks movement and, as pointed out, particles like *ut* cannot postpose.⁷ Pintzuk proposes that the VO base phrase structure can only exist in an INFL-medial structural description of an OE sentence and not in an INFL-final account.⁸ Roberts (1997), Biberauer & Roberts (2005), Elenbaas & Van Kemenade (2014), Struik & Van Kemenade (2020) propose that OE is VO in base and OV word order is produced by O moving leftward.

Most of the literature on OE supports the theory that OE has obligatory V-to-I movement, e.g., Pintzuk (1991) proposes that INFL needs to be overtly realized in OE. The literature on the rise of *do*-support in Late Middle English suggests that Early Middle English (and consequently OE) only had obligatory V-to-I movement until affix hopping began to compete with it. Haerberli & Ihsane (2016) use the position of adverbs in a diachronic setting to argue for the same. The word order *S V Adv* (evidence of V moving to INFL over the adverb) declines in Middle English as time progresses, whereas the distribution of *S Adv V* increases (V in situ). However, the word order *S Adv V* is ambiguous and, by itself, is not evidence *against* V-to-I movement. In fact, Haerberli & Ihsane (2016: 8) state that the word order *S Adv V* is predominant in OE; however, they point out that it is not clear whether the surface order is a result of INFL-final underlying structure. Elenbaas & Van Kemenade (2014: 158-159) present certain word orders in which, according to their analysis, V is in base position and does not move to INFL. Thus, apart from the narrative of the rise of *do*-support in a diachronic setting in Middle English, there is no unambiguous evidence against the possibility that, in certain word orders, the verb could remain in base position in OE. We explore this possibility later in the paper.

2.4 Object Position

Universally, objects are base generated as the complement of the main verb in VP and can move both leftward and rightward through topicalization, scrambling or a postposition operation. Pintzuk & Taylor (2006) present postposi-

⁷ Kroch & Taylor (2000), Pintzuk & Taylor (2006), Pintzuk & Haerberli (2008) also assume that verb particles, like pronouns and stranded prepositions, are prosodically light elements.

⁸ The variations accounted for by Pintzuk's Double Base Hypothesis are INFL-medial VO, INFL-medial OV, and INFL-final OV. It is important to note that Pintzuk did not find any evidence of INFL-final VO phrase structure. INFL-final VO is believed to be rare in natural languages. (Steele 1975, Dryer 1992, Sheehan, Biberauer, Roberts & Holmberg 2017 among others, cf. Fuss & Trips 2001, who point out that Kandoka-Lusi, a dialect of Kaliai-Kove, contains such word orders.)

tion of objects as a movement that: (i) is bound by heaviness, (ii) moves the object to the right of the verb, (iii) is only observed in non-negative (positive) objects as shown in (6) where the non-negative object *ece edlean on Godes rice* has been postposed. In addition, [Pintzuk & Taylor \(2006\)](#) also claim that postposition is unambiguously observed in INFL-final clauses in OE.

- (6) *he þæs habban sceal ece edlean on Godes rice.*
 he of-that have must eternal reward in God's kingdom.
 'he must have eternal reward of that in God's kingdom.'
 [cowulf, WHom_7:161.501, ([Pintzuk & Taylor 2006: 262](#))]

[Van Kemenade \(1987\)](#), [Tomaselli \(1995\)](#) and others claim that in OE, pronominal objects behave much like clitics (cf. [Koopman 1997](#)). [Pintzuk \(1991\)](#) postulates that pronominal clitics attach to the right periphery of IP. [Pintzuk & Taylor \(2006\)](#) investigate different types of objects — negative, positive, and quantified, and examine different movements associated with objects that generate the variation in object position observed in OE.

[Pintzuk & Taylor \(2006\)](#) observe that objects can undergo leftward movement (prepose), i.e., in VO contexts, objects can move leftward over V. They observe that most examples of object preposing are instances of quantified and negative objects and conjecture that scrambling in VO contexts is restricted to these types of objects.⁹ [Pintzuk & Haeberli \(2008\)](#) also acknowledge preposing of certain objects (negative or pronominal) in OE and attribute its cause to scrambling to a position between Spec(IP) and INFL (high scrambling) or a position between INFL and VP (low scrambling). However, the claim that object scrambling is subject to specific constraints on object types is contested by [Elenbaas & Van Kemenade \(2014\)](#).

2.5 Topic Position

The topic position is generally assumed to be Spec(CP) for matrix clauses. For embedded clauses, a popular theory that accounts for topicalization to Spec(CP) is *CP-recursion* ([Authier 1992](#), [Watanabe 1992](#), [de Haan & Weerman 1986](#), [Lasnik & Saito 1992](#) among others).

CP-recursion occurs only in certain environments — most generally in subordinate clauses governed by certain matrix verbs ([Iatridou 1991](#), [Iatridou & Kroch 1992](#), [Van Kemenade 1997](#), [de Haan & Weerman 1986](#), [Vikner 2001](#),

⁹ This is in line with [Kroch & Taylor \(2000\)](#) who also claim that leftward scrambling of an object is only possible in VO grammars with quantified and negative objects (see also [Haider 2005, 2007](#)).

2017). CP-recursion can also exist in ‘*peripheral-adverbial clauses*’ (Haegeman 2012, Frey 2012, Alshammari 2018 among others) which are adjunct subordinate clauses not governed by the matrix verb and “add a comment on the event of the main clause without directly linking to its meaning” (Miyagawa 2017: 7).

Although there are disparate analyses of CP-recursion, the general consensus in the literature is that embedded topicalization to Spec(CP) due to CP-recursion cannot occur in *wh*-clauses (Iatridou 1991, Iatridou & Kroch 1992, Watanabe 1992 among others). It is also not viable in ‘*central adverbial*’ clauses — which are temporal adjunct clauses that relate to the time of the event of the main clause (Haegeman 2012, Frey 2012, Alshammari 2018 among others).

Pintzuk (1991) made the claim that the topic position in OE is Spec(IP). Kroch & Taylor (1997) revisit Pintzuk’s analysis and introduce ‘CP-V2’ and ‘IP-V2’ languages: CP-V2 languages allow verb-second (V2) order only in those subordinate clauses with a specific structure (CP-recursion structure as discussed in de Haan & Weerman 1986, Iatridou & Kroch 1992) while IP-V2 languages show V2 word order in a broad range of subordinate clauses (Diesing 1990, Santorini 1989, 1992, Rögnvaldsson & Thráinsson 1990). Kroch & Taylor (1997) propose that OE is like Icelandic & Yiddish with CP-V2 in matrix clauses and IP-V2 in subordinate clauses and propose Spec(IP) as a topic position to accommodate topic positions outside of CP-recursion environments in embedded clauses (cf. Van Kemenade 1997).

3 METHODOLOGY AND ASSUMPTIONS

Corpus linguistics seeks to identify patterns within words or from strings of words or phrases. Corpus-based studies search corpora to identify specific patterns to corroborate a linguistic hypothesis. Corpus-driven studies, on the other hand, seek to extract patterns from corpora without targeting specific patterns — the deductive process of forming hypotheses and determining the strength of evidence for those hypotheses begins after the search.

3.1 *Word order patterns*

A variety of techniques to conduct a corpus-driven study exist. One of the most common techniques is the use of *n*-grams: a sequence of *n* words with a fixed order.¹⁰ For example in (5) *he wolde adræfan ut anne æpeling* there are

¹⁰ For purposes of exposition, we construct *n*-grams of words. However, *n*-grams could be constructed out of any unit that exists in the corpus either in the raw data or its annotations (e.g., letters, morphemes, syllables, etc.) The use of *n*-grams of part of speech tags is the focus of

four 3-grams (or *trigrams*):

- *he wolde adræfan,*
- *wolde adræfan ut,*
- *adræfan ut anne,*
- *ut anne æþeling.*

By counting the frequency of the n-grams, one can identify groups of words that often occur next to each other. One problem with using n-grams of words is that when n is greater than 3, the n-gram frequencies become less informative. N-grams of four, five, or six words will likely occur a small number of times in a corpus and many will occur only once, especially in relatively small corpora.

This study uses a technique which can capture re-occurring patterns of longer length. Rather than using actual words in the corpus, we use the annotation in YCOE to label words with *S*, *V*, *O*, *Aux*, and *p* which correspond to subject, verb, object, auxiliary verb and particle respectively. We then create word order patterns out of those labels.¹¹ So, (5) would correspond to the word order pattern *S Aux V p O*. By having a more general constraint where multiple words from the corpus can fall under a each of these labels, these word order patterns are effective in the discovery of relatively long and informative sequences.

We conducted a corpus-driven exploration of the York-Toronto-Helsinki Parsed Corpus of Old English Prose (a 1.5 million word syntactically annotated corpus) employing word order patterns, aimed at investigating the reported variation in OE. CorpusSearch 2 (Randall 2004)¹² was used to implement search queries. Unlike previous corpus-based studies of OE, our search queries were not designed to identify occurrences (or not) of specific constructions. Rather, we took a set of syntactic labels and generated *all* word order permutations, i.e., a set of all possible n-grams that could be constructed

this section.

11 This approach is closely related to using part-of-speech grams (*PoS-grams*). A PoS-gram is a string of part-of-speech categories (or *PoS tags*) of a specific length (Stubbs 2007: 91). In our study we use PoS tags for particles, auxiliary verbs and verbs, but using a higher-level syntactic label for nouns yields more information. Consider ‘*Mary John kissed*’. Its PoS tags would be *N N V* whereas using syntactic labels we have *O S V*.

12 CorpusSearch is a Java program written by Beth Randall, Ann Taylor, and Anthony Kroch as part of a project at the University of Pennsylvania directed by Anthony Kroch that supports research in corpus linguistics. “It is useful both for the construction of syntactically annotated corpora and for searching them.”

using those labels. Note that this set of n-grams is constructed programmatically in an exhaustive manner uninformed by any linguistic theory. For example, if the set of tags were *S*, *V* and *O*, the generated set of word order patterns would be *S O V*, *S V O*, *V S O*, *V O S*, *O V S* and *O S V*.¹³

We then retrieved all instantiations of each word order pattern in the corpus which rendered a count of how often the word order pattern was present. As mentioned, we limited our tags to: *S*(ubject), *Aux*(iliary verb), *O*(bject), *V*(erb), and *p*(article). These tags have served as ‘*diagnostic elements*’ in previous studies (see §2). For purposes of explication, we present results in four groups of word order patterns of increasing syntactic complexity:

- i. word orders consisting of *S*, *V* and *O*.
- ii. word orders consisting of *S*, *V*, *O* and *p*.
- iii. word orders consisting of *S*, *V*, *Aux* and *O*.
- iv. word orders consisting of *S*, *V*, *Aux*, *O* and *p*.

After generating a list of all instantiations in the corpus of word order patterns and their counts within each group, the word orders that exist in the corpus were analyzed with an eye to uncover unambiguous evidence of specific OE grammatical structure (e.g., IP and VP head directionality).

3.2 *Syntactic label assignment*

The YCOE corpus employs a rich set of labels for NPs¹⁴ as well as for verbs and auxiliaries. We accept only NPs labeled as nominative in the corpus to be subjects; all NPs other than nominative NPs are evaluated as objects¹⁵ and are excluded from our search.

13 Our n-grams, of course, are nothing more than a set of ordered labels that linguists have used for decades to discuss word order patterns.

14 In the text, we will frequently refer to DPs. YCOE does not provide additional tags for DPs. DPs in YCOE are represented as recursive NPs with NP dominating D.

15 It should be noted that although PPs can serve as objects, their status is ambiguous since they can also function as adjuncts. Unlike NPs, PPs in the YCOE are not tagged with any functional role (except resumptive). Hence, we omit them as possible objects in our search. Of course, this lowers the counts of objects that exist in YCOE in the searches described in the following sections and indeed we might not uncover some patterns that do exist. Analysing the entirety of PPs that exist in YCOE by hand is intractable. What *is* gained by our approach is the retrieval of a large breadth of word order patterns that exist in YCOE. See section 6 for discussion.

We ignored sentences with NPs that dominate empty tags (**con**, **exp** & **pro** tags) and traces (YCOE marks traces to identify their origin)^{16,17} as well as sentences with NPs tagged as adjuncts.

When addressing tags for verbs and auxiliaries, in cases where there is a single verb, we use any of the existing verb tags except the auxiliary, modal verbs, and infinitival ‘TO’ in the corpus to identify the main verb (i.e., *verb*, *verb participle*, *be*, and *have*). In cases where there is both an auxiliary verb and a main verb, we use the main verb tags (*verb* and *verb participle* tags) for the main verb and the rest of the verb tags (*auxiliary*, *be*, *have*, *modal*) for the auxiliary.¹⁸

3.3 Search and Post-processing

Using the YCOE tags to effectively tag the YCOE with our labels (S, V, O, Aux and p), we searched the YCOE corpus, as described above, for instances of all possible word order permutations generated from our set of labels (minimally containing a subject, verb, and object). After this initial search, we conducted some additional post-processing to assure that the retrieved word orders accurately represented linguistic patterns that could be found in the Old English texts that comprise the YCOE.

- i. For all word orders for which nine or more instances were found in the corpus:
- ii. Ignore all phrases that have a prepositional phrase (see footnote 15).
- iii. Ignore all IPs that dominate more than one object or subject, or more than one auxiliary or main verb.
- iv. Ignore all phrases that have a comma in a word order pattern. This was done to ensure that the retrieved word order was not obscured by it.¹⁹

16 Some genitive NPs that are part of the subject can be displaced and misidentified as an object giving a pseudo word order.

17 We do, however, assume movement with corresponding traces and empty categories in the analyses in subsequent sections. This is standard in corpus-driven research — discover ‘raw’ unstructured sequences in a corpus then subsequently determine what underlying linguistic factors give rise to those sequences.

18 Note that ambiguous cases where it is unclear which verbal element is an auxiliary and which is the main verb were removed from consideration (e.g., when there are multiple *have* tags).

19 “*John, he is wonderful*” is an example where the comma demarcates the focus element, and serves a syntactic function.

For all word orders for which eight or fewer instances were found in the corpus, we manually:

- i. Ignore all phrases that contain PPs that intervene between the OE words tagged with the syntactic labels: *S*, *V*, *O*, *aux* and *p* of the word order pattern. Phrases with (untagged) phrase final PPs where the PPs can be manually identified as adjuncts (not objects) were left in our counts. For example, (24) contains an irrelevant prepositional adjunct which was manually identified but (24) is still considered in our counts.
- ii. Ignore all phrases that have an intervening comma between the word order tokens. If the comma exists before or after the target word order, the function of the comma was examined to see if it obscured the retrieved word order. If so, the phrase was ignored.

3.4 Syntactic Assumptions

After our comprehensive word order pattern search of the corpus, we undertook a detailed syntactic analysis of the phrases and word orders we retrieved. In order to do so, we adopted some assumptions that are importantly based on universal principles, i.e., the assumptions are not specific to OE, they (mostly) hold for all natural languages (e.g., we don't assume a priori that OE is verb-initial or verb-final, or that V obligatorily raises to I).²⁰

We list them in some detail below:

1. We assume that the topic position is in Spec(CP) for matrix clauses. However, in subordinate clauses, we assume the topic position to be Spec(IP) in non-CP-recursion environments and Spec(CP) in CP-recursion environments.²¹ We adopt the definition of CP-recursion environment in subordinate clauses to be non-wh clauses governed by a matrix verb or a peripheral-adverbial clause, and subsequently non-

²⁰ We do adopt one non-universal assumption — that COMP is somewhere to the left of IP i.e., OE is COMP-initial/medial. The assumption is not universal since there exist comp-final languages (e.g., Japanese). This assumption and the way COMP functions in general are so standard in studies of OE and West Germanic grammar that we omit description or discussion in the main text.

²¹ Salvesen & Walkden (2017) put forth an analysis of V2 in Old English and Old French which contests the topic IP analysis of OE. We do not take a stand for or against such a proposal and entertain the possibility of a topic IP analysis under the framework proposed by Kroch & Taylor (1997).

CP-recursion environments are defined as *wh*-islands or central-adverbial clauses (see discussion in §2.5).²²

2. The subject is base generated in Spec(VP) and can occupy Spec(IP) due to the Extended Projection Principle (Chomsky 1981) or be topicalized.
3. Auxiliaries/main verbs are base generated in VP and may move to INFL or COMP (through INFL) or remain in VP.
4. Postposition is an extraposition operation in which heavy NPs/DPs/PPs are adjoined to IP.
5. Verb projection raising carries all/some complemental elements along with the verb out of the VP projection and leaves others stranded in their base position. Verb raising/projection raising is possible only in INFL-final OV languages.²³
6. Particles:
 - a. are light elements that do not postpose or scramble.
 - b. can topicalize²⁴ and undergo leftward movement above the subject.
 - c. can only move with the verb in cases of verb projection raising.²⁵

²² Hooper & Thompson (1973) provide classes of verbs whose predicates allow/disallow embedded topicalization. Salvesen & Walkden (2017) consider these classes of verbs in a study of embedded V2 in OE and Walkden & Booth (2020) coin the term '*viaduct verbs*' for the verbs that allow for embedded topicalization in their predicates. However, cross-linguistic analysis by Miyagawa (2017) shows that the analysis put forth by Hooper & Thompson (1973) doesn't hold entirely for Japanese and Spanish. We take a conservative approach in our study in line with Iatridou (1991), Vikner (2001), Iatridou & Kroch (1992) and others — we allow for possible CP-recursion in all cases of a matrix verb governing the subordinate clause.

²³ See Den Besten & Edmondson (1983), Evers (1975, 1981), Haegeman (1988), Haegeman & Riemsdijk (1986), Kroch & Santorini (1991), Salzmann (2011) among others, for formal analyses of verb (projection) raising in Germanic languages. No position is taken here on the derived structures of verb raising and verb projection raising. We simply take the behavior of verb/projection raising as akin to postposition for the purposes of this study. See Pintzuk (1996) for a similar treatment.

²⁴ See Trotzke & Quaglia (2016) for an analysis of particle topicalization in German and Elenbaas & Van Kemenade (2014) for evidence of particle topicalization in OE.

²⁵ See particle pied piping in Wurmbrand (2000), Pintzuk & Haerberli (2008) for particle movement with the verb — also called '*partial verb projection raising*' in a survey by Salzmann (2011) on verb projection raising.

- d. are generated in the complement of V and can move freely within the complement.²⁶
7. Complements of the verb can scramble/prepose leftward. We do not take a stance as to the *exact* landing site(s) of scrambling/preposing (e.g., Pintzuk & Taylor 2006; cf. Zimmermann 2011), however, following Pintzuk & Haerberli (2008) we consider two possible scrambling sites for our analysis, (i) high: somewhere between Spec(IP) and INFL, (ii) low: somewhere between INFL and Spec(VP).

In the following sections, we turn to a discussion of the results of our queries and the linguistic analyses that follow.

4 INFL HEADEDNESS AND VO/OV

From this point forward, when OE examples are given, Old English words or constituents are annotated with their corresponding label (e.g., [*his agene sceap*]_O where the subscript denotes that ‘*his agene sceap*’ is the object). Note that an entire clause is presented, which, in some examples, has certain OE words or constituents that are irrelevant for our analysis (e.g., adverbs, adjectives, clause final adjuncts, etc.); they are presented without token annotation.

4.1 Evidence from phrases with a Subject, Object and Verb (without a particle or an auxiliary verb)

Although ambiguous in all cases, the evidence we can obtain from phrases without particles or auxiliaries, listed in Table 1, helps to demonstrate the range of variation that is possible in Old English syntax.

The six possible word order patterns with Subject, Verb, and Object and their analysis are enumerated in Table 1 and discussed below.

- *Subject Verb Object*: In this word order, the subject could or could not have been moved to a specifier position above INFL (due to either EPP or topicalization; see Assumptions 1 and 2), which leaves the possibility of the verb moving leftward out of VP. This word order could

²⁶ This has been called Particle Movement (see Gries 1999), e.g., ‘John picked *up* the book’, ‘John picked the book *up*’ — word order inside the complement is free with respect to the particle *up*. The same can be extended to an OV language, i.e., both *O p* and *p O* are licensed. Pintzuk & Haerberli (2008) define this type of particle movement in OE as ‘*clause internal leftward movement*’. However, they attribute this type of movement to long-range scrambling outside the VP. We believe that such an account is unnecessary since this type of movement is seen cross-linguistically and can be better explained by particle movement inside the VP.

Word order	Count
S O V	6133
S V O	3662
O S V	881
V S O	825
O V S	382
V O S	132

Table 1 Word orders without particles and auxiliary verbs and their number of occurrences in YCOE

also be analyzed as a product of object postposition with the position of the verb uncertain. In either case, there is no way to ascertain whether this word order is OV or VO or INFL-medial or INFL-final. (7) is an instance of *S V O* word order.

- (7) *ac se heofonlica Fæder gescipð þone lichaman*
 ac [se heofonlica Fæder]_S [gescipð]_V [þone lichaman]_O
 but the holy Father created this body
 ‘but the holy Father created this body’
 [coaelhom,+AHom_2:232.363]

- *Subject Object Verb*: A verbal complement (O) can only move to three structural landing sites: (i) Spec position above INFL by topicalization, (ii) as an adjunct to IP by postposition, and (iii) scrambled left within IP. In this word order, given a VO base, the only movement possible for O is to scramble from its base position (i.e., after the verb), since the object is sandwiched between the subject and the verb. However, O could also be in situ (or scrambled) given an OV base. Turning to the position of the verb, V could be in base VP, or it could have been raised/moved to the right. In either case, there is no way to establish if the pattern is VO or OV or whether it is INFL-final or INFL-medial. (8) is an instance of *S O V* word order.

- (8) *ac hi hit afundon*
 ac [hi]_S [hit]_O [afundon]_V
 but they it found
 ‘but they found it’
 [coaelhom,+AHom_1:192.113]

- *Object Verb Subject*: Subjects can undergo a postposition or stay in low or high subject position. This word order could be evidence for an OV grammar with the *Subject* postposed or in low subject position with the *Object* being topicalized and the *Verb* moving to I. In general, there is no clear indicator as to where the subject is. The object might have topicalized, and the verb could have moved leftward to INFL-medial, COMP-medial or could have stayed in base in the VP or moved to INFL-final with the subject postposed. Hence, this word order yields no information about VO/OV or INFL-medial/final underlying base phrase structure. (9) is an instance of *O V S* word order.

(9) *ac hi forbærnde Iosue*
 ac [hi]_O [forbærnde]_V [Iosue]_S
 but them burned Iosue
 ‘but Iosue burned them’
 [cootest,Josh:11.11.5525]

- *Object Subject Verb*: In this word order, the surface position of the object is before the subject, indicating that the object has either topicalized or scrambled. This word order gives us no information about VO/OV or INFL-medial/final base structure as the position of the verb cannot be disambiguated between being in-situ in VP or moved to INFL-final or INFL-medial. (10) is an instance of *O S V* word order.

(10) *& mildheortnesse & eaðmodnesse he lærde*
 & [mildheortnesse & eaðmodnesse]_O [he]_S [lærde]_V
 & mild-heartedness & humility he taught
 ‘& he taught mild-heartedness & humility’
 [colawafint,LawAfEl:49.117]

- *Verb Subject Object*: Due to the possibility of concurrent postposition of both S and O,²⁷ *V S O* is ambiguous with respect to VO/OV and INFL-medial/final base structure. (11) is an instance of *V S O* word order.

(11) *ðonne ofergæþ an fyr ealle ðas woruld*
 ðonne [ofergæþ]_V [an fyr]_S [ealle ðas woruld]_O
 Then cover a fire all this world
 ‘Then a fire will cover all this world’
 [coaelhom,+AHom_11:296.1640]

²⁷ We have found evidence which supports subjects and objects concurrently postposing (e.g., *Aux V S O*) we revisit concurrent postposition in §4.3.

- *Verb Object Subject*: The same analysis as in *Verb Subject Object* above. (12) is an instance of *V O S* word order.

(12) *đa gelicode Gode peos ben*
đa [gelicode]_V [Gode]_O [peos ben]_S
 then pleased God this prayer
 ‘then this prayer pleased God’
 [cocathom2,+ACHom_II,_45:336.30.7538]

Although all the word order patterns in this section (patterns without particles or auxiliaries) do not lead to definitive evidence for either VO/OV or INFL-medial/final base structure, the section allowed us to introduce the application of some of our assumptions, as well as the method of syntactic analysis and reasoning which we will employ to uncover unambiguous structural evidence in the more complicated word order patterns that follow.

4.2 Evidence from word orders with a particle and without an auxiliary verb

Certain word orders with particles and without auxiliary verbs provide unambiguous evidence for VO and INFL-medial underlying base phrase structures. As outlined in Assumptions (6a), (6b), (6c) and (6d) we assume that particles are base generated in the complement of V and cannot postpose because of a heaviness constraint on postposition movement and can only move out of the complement position through topicalization or in tandem with V by means of verb projection raising.

Therefore, the relative position of the particle is useful for identifying the underlying structure in certain word orders. We found thirteen word orders that were instantiated in the corpus out of the 24 (=4!) possible. They are listed in Table 2. For clarity of discussion, we present these word orders in three groups:

- *V after S or O, with the particle before the verb*: In these word orders, *S p V O* (13), *O S p V* (14), *S O p V* (15), *S p O V* (16), *O p V S* (17), since the particle is before the verb; either V and p exhibit canonical word order²⁸ (OV) or the verb has moved rightward over the particle (VO) to INFL-final.²⁹ Thus, the two structural analyses that are possible for

28 By ‘canonical word order’ we mean the linear order of labels assuming no movement and the underlying grammatical structure of the language. For example, the canonical word order in head final languages is *O V Aux*.

29 Unlike Pintzuk’s hypothesis where rightward verb movement implies the underlying word order to be OV — here we do not make that assumption and entertain the possibility that OE

these word orders are: OV in base (with INFL ambiguous) or INFL-final VO (with or without object postposition/scrambling or subject postposition or movement to Spec(IP) or topicalization). These ambiguous word orders do not offer definitive evidence for VO/OV or INFL-medial/final base structure.

- (13) *seo fullfremede lufe ut adrifð þone þeowlice*
 [seo fullfremede lufe]_S [ut]_P [adrifð]_V [þone þeowlice]
 the perfect love out drives the servile
ege
ege]_O
 fear
 ‘the perfect love drives out the servile fear’
 [coalcuin,Alc_[Warn_35]:484.377]
- (14) *þæt geswell heo of animeð*
 [þæt geswell]_O [heo]_S [of]_P [animeð]_V
 that tumor she off removed
 ‘she removed that tumor off’
 [coherbar,Lch_I_[Herb]:5.2.341]
- (15) *Issac me ut drifeð*
 [Issac]_S [me]_O [ut]_P [drifeð]_V
 Issac me out drives
 ‘Issac drives me out’
 [cogregdC,GDPref_and_3_[C]:14.200.19.2602]
- (16) *hwý he swa swyðe nyðer his setle gecure*
 hwý [he]_S swa swyðe [nyðer]_P [his setle]_O [gecure]_V
 why he so deep down his seat choose
 ‘why did he choose his seat so deep down?’
 [conicodC,Nic_[C]:336.323]
- (17) *him fore andredan ge Pirrus se reþa Creca*
 [him]_O [fore]_P [andredan]_V [ge Pirrus se reþa Creca]
 himself for fears also Pirrus the cruel Creca
cýning
cýning]_S
 king
 ‘Pirrus the cruel Creca king also fears himself’
 [coorosiu,OrHead:1.10.11]

exhibits INFL-final VO structure. However, recall that we restrict verb raising to INFL-final OV languages (Assumption 5).

- *V after S or O, with the particle after the verb*: In these word orders — *S V p O* (18), *S V O p* (19), *O V p S* (20), *S O V p* (21 and 22) and *O S V p* (23 and 24), since the particle is after the verb, the word orders are either VO in base or the verb has moved leftward over the particle.

(18) *se stream berð aweg Placidum*
 [se stream]_S [berð]_V [aweg]_P [Placidum]_O
 the stream carries away Placidus
 ‘the stream carries Placidus away’
 [cocathom2,+ACHom_II,_11:95.97.1943]

(19) *Moyses wearp hy upp*
 [Moyes]_S [wearp]_V [hy]_O [up]_P
 Moyses threw him up
 ‘Moyses threw him up’
 [cootest,Exod:9.10.2699]

(20) *raþe þæs þær com upp micel wæl deadra*
 raþe [þæs]_O þær [com]_V [upp]_P [micel wæl deadra
 quick there there came up many slain dead
fisca
 fisca]_S
 fish
 ‘Many dead fish came up there quickly’
 [coorosiu,Or_6:3.136.6.2865]

S O V p is a word order in which either O or p are out of canonical order. Since the particle cannot move rightward (unless it has been projection raised in tandem with V), the object has been scrambled (the subject is before the object, indicating that the object has not topicalized). We found 10 examples of this word order: 9 matrix clauses (such as Example 21) and one subordinate clause *hig hyne drifon ut* in a CP-recursion environment, i.e., the subordinate clause is governed by the verb *gehyrde* introduced by an overt complementizer *þæt*’ (see Example 22). In both types of clauses, the position of the subject could not be pinpointed, as the subject could have topicalized. If the object has undergone low scrambling and lies somewhere between INFL and VP, then the verb must be in situ in a VO base since its surface position is after the scrambled O and has not moved to INFL.³⁰ In case

³⁰ Given the order of *V p*, the verb and particle could not have been projection raised due to Assumption (5) concerning verb projection raising.

of high scrambling, the verb can be in INFL-medial — we take up discussion of verb movement to INFL in §5.3.

- (21) *he his agene sceap læt ut*
 [he]_S [his agene sceap]_O [læt]_V [ut]_P
 he his own sheep lets out
 ‘he lets his own sheep out’

[cowsgosp,Jn_[WSCp]: 10.4.6600]

- (22) *ða se hælend gehyrde þæt hig hÿne*
 ða se hælend gehyrde þæt [hig]_S [hÿne]_O
 when the savior heard that they him
drifon ut
 [drifon]_V [ut]_P
 pushed out
 ‘when the savior heard that they pushed him out’

[cowsgosp,Jn_[WSCp]:9.35.6581]

In *O S V p* the object has topicalized or scrambled out of the VP because it precedes S. We provide two examples of this word order from the corpus, (23) and (24). In particular, our example (23) is identical to Pintzuk & Haerberli (2008: Example 10) who make a case for a topicalized object in Spec(CP) with S occupying Spec(IP) and V in INFL-medial having moved leftward over the particle. They found only one example of this underlying structure (Pintzuk & Haerberli 2008: 376). They propose that this analysis can extend to subordinate clauses by virtue of CP-recursion (Pintzuk & Haerberli 2008: Endnote 10). However, it is unclear how they would analyze this word order in non-CP-recursion environments. We found eight examples of *O S V p*: five matrix clauses and three subordinate clauses. Of the three subordinate clauses, we found one clause (24) where there is no CP-recursion under our working definition and assumptions.³¹ In (24), the subordinate clause *oð þæt hi man gynde ongean eft to Iudan* functions as a temporal adjunct with the complementizer *oð* which translates to ‘until’, which according to Haegeman (2012: 160) can only introduce a central-adverbial phrase thus disallowing CP-recursion.³² In such an

31 Here *þæt* is serving as an accusative determiner as tagged in the corpus and not as the complementizer ‘that’ and is lost in translation.

32 See Frey (2012) and Alshammari (2018) for a similar treatment in German and Arabic respectively.

instance, the object could not topicalize to Spec(CP), and hence the object has been scrambled or topicalized to Spec(IP), with the subject remaining in Spec(VP). Thus V is in situ, which leads to the conclusion that (24) has an underlying VO structure. We discuss the V in-situ analysis in more detail in §5.3.

- (23) *þæne se geatweard læt in*
 [þæne]_O [se geatweard]_S [læt]_V [in]_P
 that-one the gatekeeper lets in
 ‘the gatekeeper lets that-one in’

[cowsgosp,Jn_[WSCp]: 10.3.6596]

- (24) *and bicnode gehwanon mid blawunge him*
 and bicnode gehwanon mid blawunge him
 and summoned on-all-sides with blowing himself
fultum, oð þæt hi man gynde ongean
 fultum, oð þæt [hi]_O [man]_S [gynde]_V [ongean]_P
 assistance, until that them they drove back
eft to Iudan
 eft to Iudan
 again to Iudan
 ‘and summoned on all sides with blowing (of trumpets)
 assistance to himself, until they drove them back again to Iudan’

[coalive,+ALS_[Maccabees]:631.5251]³³

- *V before S and O and particle after S*: This group includes the word orders: *V S O p* (25), *V O S p* (26), *V S p O* (27). In these word orders, since the particle is after the subject and the verb starts the clause, V must have moved (to COMP-medial or INFL-medial).³⁴

³³ The phrase *eft to Iudan* here is a prepositional adjunct and is not relevant to our analysis.

³⁴ The focus of this section is identifying evidence for VO/OV or INFL-medial/final. We return to the discussion of V-to-I movement in §5.3.

Word order	Analysis	Count
S p O V	Ambiguous	5
S O p V	Ambiguous	80
S O V p	INFL medial or VO	10
O S p V	Ambiguous	12
O S V p	VO	1/8
S p V O	Ambiguous	8
S V p O	Ambiguous	26
S V O p	Ambiguous	26
O p V S	Ambiguous	1
O V p S	Ambiguous	2
V S O p	Ambiguous	23
V S p O	Ambiguous	23
V O S p	Ambiguous	2

Table 2 Word orders without an auxiliary but with a particle. Unambiguous word orders are in boldface. A ‘/’ in the rightmost column is used to differentiate the count of non-CP-recursion subordinate clauses from the total count of clauses. This breakdown is only given when the distinction is relevant to the discussion in the text.

(25) *ne tihþ he hyne hrædlice up on*
 ne [tihþ]_V [he]_S [hyne]_O hrædlice [up]_p on
 NEG pull he him promptly up on
restedæge
 restedæge
 sabbath-day
 ‘he does not promptly pull him up on the Sabbath day’

[cowsgosp,Lk_[WSCp]:14.5.4821]³⁵

³⁵ This is a conjoined clause from the entire phrase *þa cwæð he to him andswariende: hwylces eowres assa oððe oxa befealþ on anne pytt and ne tihþ he hyne hrædlice up on restedæge* which translates to ‘And answered them saying, which of you has an ass or an ox fallen into a pit, and you not immediately pull it out on sabbath day?’ Here, the word *hrædlice* which means ‘promptly’ and the phrase *on restedæge*, which translates as ‘on the sabbath day’, function as an adverb and adjunct, respectively, and are irrelevant to the discussion.

- (26) *Ne tio hie mon næfre of*
 ne [tio]_V [hie]_S [mon]_O næfre [of]_P
 NEG draw them one never off
 ‘one should never draw them out’

[cocura,CP:22.171.1.1159]

- (27) *þa ahof Paulus up his heafod*
 þa [ahof]_V [Paulus]_S [up]_P [his heafod]_O
 then lifted Paul up his head
 ‘Paul then lifted up his head’

[coblick,LS_32_[PeterandPaul[BiHom_15]]:187.303.2429]

All instances of *V S O p*, *V O S p* and *V S p O* occur in matrix clauses; thus, the position of the verb is ambiguous. In all these word orders, the position of the object is difficult to pinpoint as it could be post-posed/scrambled out of base position — these word orders are ambiguous as to VO/OV underlying base structure.

In this section, particle position was the key to identifying a word order pattern that provides unambiguous evidence of underlying VO structure (*O S V p*). If the structural position of the object could be identified, the pattern *S O V p* would be an unambiguous instance of either INFL-medial or VO in base. However, since the position of O can’t be determined it stands as an ambiguous word order (we revisit this point in §5.3). We now turn to word orders with an auxiliary and without a particle.

4.3 Evidence from word orders with an auxiliary verb and with no particle

Auxiliary verbs help to identify INFL-headedness, since the auxiliary along with the subject can mark the underlying structure. Also, movement to INFL is restricted — Aux and V cannot move in tandem to the left³⁶ which, in certain word order patterns, allows decisive identification of underlying INFL-medial headedness.

The word order *O Aux S V* is unambiguously *INFL-medial* in subordinate clauses without CP-recursion. Without CP-recursion, the Aux cannot move to COMP-medial. The position of Aux is marked by O, being topicalized to

³⁶ We assume that verb raising/projection raising implies INFL-final with the possibility of Aux landing in INFL-final position (recall that we do not assume that INFL has to be occupied) i.e., in cases of verb raising/projection raising, both the verb and the auxiliary can move to the right.

Spec(IP) or high scrambled, and S, which remains in the lower subject position of Spec(VP). Thus, *O Aux S V* has INFL-medial structure in non-CP-recursion subordinate clauses. We found 74 examples of *O Aux S V* where 28 are in subordinate clauses and 2 of those subordinate clauses were in non-CP-recursion environments — illustrated in (28) and (29).

In (28), the subordinate clause *geþenc nu hu weorðlic & hu foremærlíc þe wolde se man þincan* is in a *wh*-island and hence not in a CP-recursion environment. In (29), the subordinate clause *þa him wæron þas lytlingas acennede* is a central-adverbial temporal clause and is introduced by *þa* which translates to ‘when’. In this context ‘when’ relates to the time of the event of the main clause, thus the subordinate clause is a central-adverbial clause (see [Haege-man 2012: Table 4.3](#)), hence is not in a CP-recursion environment.

- (28) *geþenc nu hu weorðlic & hu foremærlíc þe*
geþenc nu hu weorðlic & hu foremærlíc [þe]_O
 consider now how honourable and how eminent you
wolde se man þincan.
 [wolde]_{Aux} [se man]_S [þincan]_V
 would the man seem
 ‘consider now, how honourable and how eminent would the man
 seem to you?’

[coboeth, Bo:41.142.25.2848]

- (29) *Isaac wæs lx wintre, þa him wæron þas*
Isaac wæs lx wintre, þa [him]_O [wæron]_{Aux} [þas
 Isaac was 60 years, when him were his
lytlingas acennede
lytlingas]_S [acennede]_V
 children born
 ‘Issac was 60 years old, when his children were born to him’

[cootest, Gen:25.26.1035]

The other word orders that exist in this category but are ambiguous for both INFL headedness and VO/OV are documented in Table 3. Their status as ambiguous word order patterns has been established through non-trivial analysis, which we explicate below. Indeed, in previous work, some of these word

orders were used as unambiguous evidence for some aspect of OE grammar.³⁷ Furthermore, some word orders also reveal other aspects of OE grammar, specifically that S and O can be concurrently postposed and that there is unambiguous evidence for V-to-I movement, albeit ambiguous as to whether the movement is to the left or right. For the purposes of presentation, we again divide the word orders into groups.

- *Word orders which start with a verb or auxiliary: Aux S V O (30) and Aux S O V (31)* are word orders in which Aux is before the subject, indicating that Aux has moved leftward (either to INFL or COMP) over S; however given the possibility of object scrambling/postposition the position of the verb and object is unclear. Thus, both word orders are ambiguous.

(30) & *is nu hellegeat belocen rihtwisum*
 & [is]_{Aux} now [hellegeat]_S [belocen]_V [rihtwisum]
 & is now hell-gate closed-to righteous
mannum
 mannum]_O
 man

‘& the gate of hell is now closed to righteous men’

[cocathom1,+ACHom_I,_15:305.165.2895]

(31) *meahton hy þone here ahwær utan*
 [meahton]_{Aux} [hy]_S [þone here]_O ahwær utan
 could they that army anywhere outside
betreppan
 [betreppan]_V
 entrap

‘they could entrap that army anywhere outside’

[cochronC,ChronC_[Rositzke]:992.3.1234]

In the pattern *Aux O S V*, the position of the Aux (before the subject) indicates that the Aux has moved leftward to COMP-medial or INFL-medial. Our search of the corpus revealed 18 instances of *Aux O S V*, of which one is a subordinate clause introduced by the complementizer *þæt* and governed by the matrix verb, i.e., it is in a CP-recursion

³⁷ For example, see Pintzuk (1991) where word orders such as *S O Aux V* and *O S Aux V* starting with two heavy constituents are categorically analyzed as INFL-final whereas they are ambiguous in our analysis.

environment — it is instantiated in YCOE as (32). Here, the subordinate clause *þæt næfre ne sceal yfele mannen na god getidan* is the direct object of the verb *seggen* and has moved to the right.

- (32) *Get ic þe wylle an oðer seggen, þæt næfre ne*
 Get ic þe wylle an oðer seggen, þæt næfre ne
 still I one will an other say, that never NEG
sceal yfele mannen na god getidan
 [sceal]_{Aux} [yfele mannen]_O na [god]_S [getidan]_V
 shall evil men NEG good happen
 ‘still I will say to another one, that no good shall ever happen to
 evil men’

[coeluc1, Eluc_1_[Warn_45]:110.8]

Another example of ambiguity is the word order *Aux O V S* (33) where S has clearly been postposed. The object could have been scrambled with the Aux moving to INFL-medial/COMP-medial, or the object could be projection raised with the verb, leaving the Aux in base or moved to INFL-final. There is no evidence for INFL-medial/final or VO/OV base structure.

- (33) *þa wearð him æteowed wundorlic gesihð*
 þa [wearð]_{Aux} [him]_O [æteowed]_V [wundorlic gesihð]_S
 then was him shown wondrous vision
 ‘then a wondrous vision was shown to him’

[cochronC, ChronC_[Rositzke]:992.3.1234]

Aux V S O (34) is a word order in which the object has been clearly postposed. What is interesting about *Aux V S O* is that both Aux and V appear before the subject, indicating that the subject is not in base and has been postposed. This suggests that the postposition of subject and object can occur concurrently in OE (see Pintzuk 1991: 161-162). *Aux V S O*, however, is ambiguous in terms of underlying INFL or VP structure.

- (34) *nu todæg is geworden hæl þisum*
 nu todæg [is]_{Aux} [geworden]_V [hæl]_S [þisum
 Behold today is befalling salvation this
hirede
 hirede]_O
 family
 ‘Behold today salvation is befalling this family’

[coaelhom,+AHom_17:179.2451]

Aux V O S (35) is another instantiated word order pattern in which the fact that both S and O could have been postposed concurrently renders it ambiguous.³⁸

- (35) *Wearð þa geopenad his earman wife his*
 [Wearð]_{Aux} þa [geopenad]_V [his earman wife]_O [his
 was then revealed-to his poor wife his
manfullan behat þam hetolan deofle
 manfullan behat þam hetolan deofle]_S
 wicked promise that hateful devil
 ‘his wicked promise to the hateful devil was then revealed to
 his poor wife’

[coaelive,+ALS_[Basil]:405.726]

- *Word orders starting with the Subject*: Word orders that start with a subject and end with an object are ambiguous, as the subject may have topicalized and/or the object postposed, giving no clue as to the relative position of V and Aux.³⁹ The word orders *S Aux V O* (36) and *S V Aux O* (37) fall into this category. For *S V Aux O* there is clear evidence of postposition of the object since V, Aux, and O are not in any canonical order, and both V and Aux are before O, while in *S Aux V O*, the object could be in base VO order.

- (36) *we magon wundrian þæs mæran*
 [we]_S [magon]_{Aux} [wundrian]_V [þæs mæran
 we might wonder-at the distinguished
bisceopes
 bisceopes]_O
 Bishop
 ‘we might wonder-at the distinguished Bishop’

[coaelhom,+AHom_27:134.4008]

³⁸ Though we should note that we found only one occurrence of this word order in YCOE, but it serves as an example of how a corpus-driven search might spark interest in more targeted corpus-based studies which may reveal an increased number of relevant examples supporting a specific linguistic claim.

³⁹ For Pintzuk (1991) and similar work, the word order *V Aux* would unambiguously identify base INFL-final.

Word orders	Count
O S V Aux	93
S O V Aux	606
O S Aux V	104
S O Aux V	174
O V Aux S	2
S Aux O V	683
S Aux V O	273
S V Aux O	21
O Aux V S	21
Aux S V O	63
Aux O V S	4
Aux O S V	18
Aux V O S	1
Aux V S O	3
Aux S O V	177

Table 3 Ambiguous word orders with an auxiliary verb and no particles.

- (37) *man bletsian sceal brydguman & bryde*
 [man]_S [bletsian]_V [sceal]_{Aux} [brydguman & bryde]_O
 one bless shall bridegroom & bride
 ‘one shall bless the bridegroom & the bride’

[cochdrul,ChrodR_1:71.0.905]

Another word order that starts with a subject and is ambiguous is *S Aux O V* (38). For this word order, it is unclear whether the object is scrambled/projection-raised or in base with Aux moving leftward, thus it is ambiguous.

- (38) *He nolde hine gehælan*
 [he]_S [nolde]_{Aux} [hine]_O [gehælan]_V
 he wouldn’t him heal
 ‘he wouldn’t heal him’

[coaelhom,+AHom_6:332.1030))]

S O V Aux (39) is a word order which could be OV in base⁴⁰ with Aux in base, or it could be VO in base with the object scrambled and Aux movement to INFL-final, thus making it ambiguous for our purposes.

- (39) *þæt menn hit gehyran mihton*
 þæt [menn]_S [hit]_O [gehyran]_V [mihton]_O
 that men it hear might
 ‘that men might hear it’

[coaelive,+ALS_[Basil]:405.726]

Similarly, *S O Aux V* (40) could be an example of object scrambling with VO in base or verb raising with INFL-final OV.

- (40) *he cristene men wæs ehtende*
 [he]_S [cristene men]_O [wæs]_{Aux} [ehtende]_V
 he Christian men was persecuting
 ‘he was persecuting Christian men’

[cobede,BedeHead:1.6.16.7]

- *Word orders starting with an Object*: Word orders starting with an object followed by a subject indicate that the object has topicalized or scrambled and thus nothing exists to mark the position of Aux or V. Thus, *O S Aux V* (41) and *O S V Aux* (42) are ambiguous.

- (41) *his we magon wundrian*
 [his]_O [we]_S [magon]_{Aux} [wundrian]_V
 him we might wonder-at
 ‘we might wonder-at him’

[coaelive,+ALS_[Christmas]:72.58]

- (42) *hine mon byrgan sceolde*
 [hine]_O [mon]_S [byrgan]_V [sceolde]_{Aux}
 him one hide should
 ‘one should hide him’

[cobede,Bede_4:11.288.30.2917]

⁴⁰ Recall that we don’t assume mandatory V-to-I movement which, in our analysis here allows Aux to remain dominated by the VP, i.e., in base position.

A third word order where the subject is not in canonical position, *O Aux V S* (43), is also ambiguous since the object could have topicalized or scrambled, and either the Aux or the V could have moved/raised. The subject has been clearly postposed.

- (43) *him bið oftogen his bigleofa*
 [him]_O [bið]_{Aux} [oftogen]_V [his bigleofa]_S
 him were taken-away his provisions
 ‘His provisions were taken away (from) him.’

[cocathom1,+ACHom_I,_19:329.113.3710]

For *O V Aux S* (44), the word order could be canonical OV in base or could be VO in base with the object topicalized/scrambled and Aux moved to INFL-final with the subject postposed in both cases.

- (44) *ne ure mod gedrefan mæge grim wea*
 ne [ure mod]_O [gedrefan]_V [mæge]_{Aux} [grim wea]_S
 NEG our mind disturb can cruel affliction
 ‘nor can cruel affliction disturb our mind’

[coverhom,HomS_1_[ScraggVerc_5]:191.965]

4.4 Evidence from word orders with both an auxiliary verb and a particle

The word orders with both auxiliary and particles are often more telling of INFL head position and VO/OV base structure, as particles do not move independently of the verb unless they are topicalized (see Assumption 6c) and the presence of Aux restricts the leftward movement of V. In fact, all the word orders discussed in this section unambiguously reveal either (or both) VO/OV or INFL headedness. The surface position of the main verb relative to the particle helps us pinpoint the base structure of VP, i.e., if the particle is before the verb, the word order is OV; otherwise, it is VO. The word orders in this category, enumerated and compiled in Table 4, yet again, are grouped below for discussion.

- *Word orders that start with a subject followed by an object:* If the object is sandwiched between the subject and both the auxiliary and the main verb, the object has not topicalized or postposed.

The word orders *S p O V Aux* (45) and *S O p V Aux* (46) are definitively OV with V in base (since it is to the left of Aux) and are both ambiguous regarding INFL (by our assumptions, Aux could still be in base — Assumption 3 §3.4).

- (45) *þæt he þonne aweg hine astyrian*
 þæt [he]_S þonne [aweg]_P [hine]_O [astyrian]_V
 that he then away him move
mæge
 [mæge]_{Aux}
 could
 ‘that he could then move him away’
 [coverhom,HomS_4_[ScraggVerc_9]]

- (46) *he hine forð healdan sceolde*
 [he]_S [hine]_O [forð]_P [healdan]_V [sceolde]_{Aux}
 he him forth/forward hold should
 ‘he should hold him forth/forward’
 [cocura,CP:58.443.28.3178]

The word orders *S O p Aux V* (47) and *S O Aux p V* (48) have O, V, Aux, and p, out of canonical word order. For *S O Aux p V*, since p is before V, it is OV in base. However, the position of the object is ambiguous — this word order could be an instance of verb projection raising or could have high object scrambling with Aux moving to INFL medial from OV base. *S O p Aux V*, however, is the result of verb raising because Aux has not moved to the left (Aux is after the particle), and thus this surface word order must be generated by verb raising to the right, indicating INFL-final OV base phrase structure.

- (47) *þæt hi hine ut sceoldon wurpan*
 þæt [hi]_S [hine]_O [ut]_P [sceoldon]_{Aux} [wurpan]_V
 that they him out should throw
 ‘that they should throw him out’
 [coeust,LS_8_[Eust]]

- (48) *he þonne þone ceap nelle fore gesellan*
 [he]_S þonne [þone ceap]_O [nelle]_{Aux} [fore]_P [gesellan]_V
 he then that price NEG for sell
 ‘he then does not sell for that price’
 [colawine,LawIne:74.1.183]

Word order	Analysis	Number
S p O V Aux	OV	1
S O p V Aux	OV	23
S O Aux p V	OV	3
S O p Aux V	INFL-final OV	1
O S p V Aux	OV	3
O S Aux p V	OV	4
S Aux V O p	VO	1
O Aux S p V	OV	4
S Aux O V p	VO	1
S Aux O p V	INFL-medial OV	2/11
Aux S O p V	OV	3
Aux S V O p	VO	1

Table 4 Word order patterns containing both an auxiliary and a particle.

- *Word orders that start with an object followed by a subject*: In these word orders, the object could have topicalized out of the VP base or could have scrambled with the subject in Spec(VP). The particle then acts as a marker of the base order in VP.

O S Aux p V (49) is a word order that has *p* before *V*, which makes it OV; however, the position of *Aux* is ambiguous, as it could have moved to INFL-medial, or *p V* could have been projection raised to the right over *Aux*. All examples of this word order are matrix clauses. Thus, the position of the subject is ambiguous, since the object could have topicalized to Spec(CP).

(49) & *hine se geatwerd ne wolde in*
 & [hine]_O [se geatwerd]_S ne [wolde]_{Aux} [in]_p
 & him the gatekeeper NEG would in
forlætan
 [forlætan]_V
 let
 ‘& the gatekeeper would not let him in’

[coverhom,HomS_24_[ScraggVerc_1]:22.22]

Finally, *O S p V Aux* (50) is a word order where the particle is before the verb, revealing OV base structure; however, the *Aux* could be in

base, thus making it ambiguous for INFL headedness. Similarly to *O S Aux p V* all examples of this word order are matrix clauses, thus the position of the subject is ambiguous.

- (50) *ond hit nænig mon ut cyþan ne*
 ond [hit]_O [nænig mon]_S [ut]_p [cyþan]_V ne
 and it not-any man out known NEG
moste
 [moste]_{Aux}
 must
 ‘and no man must make it known’

[coalex,Alex:41.1.525]

- *Word orders starting with a subject, where the verb is followed by the particle:* In word orders *S Aux V O p* and *S Aux O V p* the particle follows the main verb; hence, they are VO in base.

In *S Aux V O p* (51) Aux, V, and O are in canonical VO order. Aux could reside unmoved in base, thus rendering this word order ambiguous as to INFL position.

- (51) *and man ne mihte swaðeah macian hi*
 and [man]_S ne [mihte]_{Aux} swaðeah [macian]_V [hi]_O
 and one NEG might as-yet act himself
healfe up
 healfe [up]_p
 half up
 ‘and no one could put up half of them’

[coelive,+ALS_[Swithun]:431.4487]

We found one example of *S Aux O V p* (52) in a subordinate clause which is a peripheral-adverbial clause, since it is introduced by *þæt* which translates to ‘such that’ in this context and refers to the *result* of the verb of *becymð* which translates to ‘befall’ in the discourse structure (see Haegeman 2012: Table 4.3, for a similar analysis of the complementizer ‘so that’ which refers to a *result* discourse structure and is classified as a peripheral-adverbial clause). Hence, it is in a CP-recursion environment, making the position of S and Aux ambiguous, as the subject could be topicalized and Aux could be in COMP-medial. The word order, however, is VO in base as *p* is after V. The object *þæt drincgemett* must have scrambled from its canonical VO base — however, the exact position (high or low) remains elusive.

- (52) *þæt unwæstmbernys on eard becymð þæt ma ne
 þæt unwæstmbernys on eard becymð þæt [ma]_S ne
 that barrenness on earth befalls, that one NEG
 mæge þæt drincgemett bringan forð
 [mæge]_{Aux} [þæt drincgemett]_O [bringan]_V [forð]_P
 may that (measure-of-drink) bring forth
 ‘that barrenness befalls the earth, such that one may not bring
 forth that (measure of drink)’*

[cochdrul, ChrodR_1:6.28.167]

- *Word orders with Aux sandwiched between O and S with the particle before the verb: S Aux O p V and O Aux S p V* are word orders in this subcategory. In both cases, since the particle is before the verb, these word orders are OV in base.

For *S Aux O p V*, the subject could be in Spec(IP) or topicalized to Spec(CP) (see Assumptions 1 and 2). For subordinate clauses without CP-recursion, the subject must have moved to Spec(IP) with the Aux in INFL-medial position. We found eleven instances of this word order, of which ten were subordinate clauses and two of the ten were not in CP-recursion environments — one such example is instantiated in (53), where the subordinate clause *þa cristenan mihton hira ut ascufan* is a central-adverbial temporal clause introduced by the complementizer *ær* (translates to ‘before’ which can only introduce central-adverbial clauses, see Frey 2012, Alshammari 2018, Haegeman 2012 and others).

- (53) *þa com þam Deniscan scypum þeah ær
 þa com þam Deniscan scypum þeah ær
 then came those Deniscan ships however before
 flod to ær þa cristenan mihton
 flod to, ær [þa cristenan]_S [mihton]_{Aux}
 flood-tide to, before those Christians could
 hira ut ascufan
 [hira]_O [ut]_P [ascufan]_V
 theirs out shove*

‘Then, however, the flood-tide came to the Danish ships, before the Christians could shove theirs out’

[cochronC, ChronC_[Rositzke]:897.38.997]

For *O Aux S p V*, we found four examples, all of them in matrix clauses (see one instantiated in 54). This word order is ambiguous with regard to INFL-headedness as the Aux can be in COMP-medial or INFL-medial with O scrambled to a high position or topicalized.

- (54) *hine nolde se geatweard in lætan*
 [hine]_O [nolde]_{Aux} [se geatweard]_S [in]_p [lætan]_V
 him wouldn't the gatekeeper in let
 'the gatekeeper wouldn't let him in'

[coverhomE,HomS_24.1_[Scragg]:148.112]

- *Word orders starting with Aux*: The word order *Aux S O p V* (55) has the particle before the main verb and the auxiliary at the beginning of the phrase. Since the particle is before the main verb, it is OV in base. However, the underlying position of INFL remains ambiguous; the three examples we found for this word order are matrix clause questions with the possibility of Aux moving to COMP-medial.

- (55) *hwī woldest ðu us ut alædan*
 hwī [woldest]_{Aux} [ðu]_S [us]_O [ut]_p [alædan]_V
 why would you us out lead
 'why would you lead us out'

[cootest,Exod:14.10.2934]

Aux S V O p (56) is a word order instantiated by a single matrix clause example in the corpus. The position of Aux is ambiguous, as it could be in COMP/INFL-medial. For *Aux S V O p*, since the particle follows the verb, the word order is VO in base.

- (56) *forþon sceal mon ær clæsnian þa*
 forþon [sceal]_{Aux} [mon]_S ær [clæsnian]_V þa
 therefore should man before cleanse the
yflan wætan aweg
 [yflan wætan]_O [aweg]_p
 harmful liquids away
 'therefore one should cleanse the harmful liquids away'

[colaece,Lch_II_[2]:30.1.16.2676]

Word order	Landing Site	Counts
S Aux O V p	Ambiguous	1
S O V p	Ambiguous	10

Table 5 Unambiguous evidence of scrambling.

5 MOVEMENT OPERATIONS IN OE

5.1 Object movement

Scrambling has been widely studied in OE. Several studies (see discussion in §2.4) ascribe object scrambling only in certain cases (negative, pronominal, and quantified). As per our Assumption (7), we posit two scrambling sites for objects: (i) high: between Spec(IP) and INFL and (ii) low: between INFL and Spec(VP). We found two word order patterns that unambiguously exhibit scrambling (see Table 5).

The word order *S Aux O V p* exhibits object scrambling as the object is out of canonical order with respect to V and p. In this word order, the object has been scrambled out of VP (see the analysis in §4.4). In the single instance of *S Aux O V p* (52) that we present in this study, the object *þæt drincgemett* is neither a pronoun nor a negative object.⁴¹

The word order *S O V p* follows a similar analysis. In three of the ten instances of *S O V p*, the object is not a negative object, pronoun, or quantified object. One example we present of *S O V p*, (21), was discussed in Elenbaas & Van Kemenade (2014: Example 24b) to argue for general object scrambling in VO contexts. We present an additional example (57) where the object in question ‘*ða dura*’ is not negative, quantified, or pronominal. Thus, the evidence presented in this study aligns with the proposal presented in Elenbaas & Van Kemenade (2014) — object scrambling is not restricted to negative objects, quantified objects, and pronouns.⁴²

⁴¹ This example was also presented in Pintzuk & Taylor (2006: 264) but was taken as an exception to their proposal of object scrambling in negative, quantified, or pronominal contexts. However, Elenbaas & Van Kemenade (2014: Example 24) use the same example to make a case for object scrambling without any constraints.

⁴² Pintzuk & Taylor (2006: 264) by reason of their finding only a single example of a positive object which preposes, exclude positive objects from scrambling; however, they do provide some leeway: ‘If this turns out to be the wrong assumption and positive objects do, in fact, prepose, the adjustment necessary to the model will be very small and will not affect the pattern presented by the data, which is our main interest.’

Word order	Count
S V Aux O	21
Aux V S O	4

Table 6 Unambiguous evidence of object postposition.

- (57) *þa oðre ða dura bræcon þær adune*
 [þa oðre]_S [ða dura]_O [bræcon]_V þær [adune]_P
 the others the doors broke there down
 ‘the others broke down the doors there’

[cochroNE,ChronE_[Plummer]:1083.23.2787]

It is important to note that all the instances of unambiguous object scrambling that we have uncovered in this study are ambiguous with respect to their landing site (high or low).

We also see unambiguous evidence of object postposition in two word orders: *S V Aux O* (37) and *Aux V S O* (30) (see Table 6) where O is the last element of the clause. In all instances of either word order, the object was never a pronoun, supporting the heaviness condition of postposition proposed in previous studies (Van Kemenade 1987, Pintzuk 1991). Furthermore, we also observe that none of the examples of postposition of objects involve negative objects, thus, supporting the findings of Pintzuk & Taylor (2006). This finding was also observed by Taylor & Pintzuk (2015) when they investigated post-verbal occurrence of negative objects in V Aux word order and found none — we can corroborate this with our only word order with V Aux formation with a post-verbal object, *S V Aux O*, which does not contain any negative objects.

5.2 Subject movement

In terms of the position of the subject Pintzuk & Haerberli (2008) claim that pronominal subjects occupy the high position, Spec(IP). We have identified two word order patterns, *O S V p* and *O Aux S V* (enumerated in Table 7), whose particular instances have the subject staying in its base-generated low position. In all the instances compiled in Table (7), the subject follows the object and the object has either been scrambled or topicalized in a non-CP-recursion subordinate clause environment; therefore, the subject could not

Word order	Subject position	Count
O S V p	low	1
O Aux S V	low	2
S Aux O p V	high	2

Table 7 Unambiguous evidence of the position of the subject.

have moved to the high position. We found an example of the word order *O S V p* (24), where the subject *man* is in Spec(VP) and is a pronoun.

We also have identified one word order *S Aux O p V* where for two instances of subordinate clauses without CP-recursion, the subject is in the high position Spec(IP) (without the possibility of being topicalized). Both instances have a DP subject. The two instances are very similar with minor spelling differences, but are from two different sources. We have discussed one instance (53) in §4.4, we present the other instance in (58).

- (58) *þa com þam Dæniscan scypum þeah ær*
 þa com þam Dæniscan scypum þeah ær
 then came those Deniscan ships however before
flod to ær þa cristenan meahton hira
 flod to, ær [þa cristenan]_S [meahton]_{Aux} [hira]_O
 flood-tide to, before those Christians could them
ut ascufan
 [ut]_p [ascufan]_V
 out drive

‘Then, however, the flood-tide came to the Danish ships before the Christians could shove theirs off’

[ChronD_[Classen-Harm]:897.46.919]

We also uncover five word order patterns: *Aux V S O* (34), *Aux V O S* (35), *O Aux V S* (28), *Aux O V S* (33), *O V Aux S* (44) where the subject, following V and Aux, is postposed. Additionally the examples presented for *Aux V S O* (34) and *Aux V O S* (35) support the observations about concurrent postposition made in Pintzuk (1991: 161-162). They are enumerated and compiled in Table 8. In all these instances, the subject was never a pronoun, thus satisfying the heaviness constraint imposed on postposition.

Word order	Count
Aux V S O	4
Aux V O S	1
O Aux V S	21
Aux O V S	4
O V Aux S	2

Table 8 Unambiguous evidence of subject postposition.

5.3 *Verb movement*

It has been widely accepted that OE is a V-to-I language with INFL overtly occupied (see the discussion in §2.3). That is, all surface word orders have compulsory V-to-I movement. In this study, we do not make such an assumption and entertain the possibility of the verb being dominated by the VP with tense hopping down from INFL. This assumption is corroborated in [Elenbaas & Van Kemenade \(2014: 158\)](#) where they argue for no V-to-I movement in their analysis of certain word orders. The two word order patterns uncovered by our study where the verb could be in base VP are analyzed below (enumerated in Table 9):

- i. *S O V p*: Ten instances of this word order were found in the corpus. The object is out of canonical order and follows S, and O has clearly been scrambled, as discussed above. If we were to assume a higher scrambling account, then V could have moved to INFL-medial, making it ambiguous for VO/OV. However, if we assume low scrambling, this word order can be argued as evidence for VO base, as V must be in base VP.
- ii. *O S V p*: Of the twelve occurrences of this word order, we see a single example (24) in a non-CP-recursion subordinate clause. The object in such an example is topicalized to Spec(IP) or scrambled with the subject remaining in Spec(VP) and the verb in base VP with VO word order.

Although the possibility of V staying in base has been noted before in [Elenbaas & Van Kemenade \(2014\)](#), it is important to put the implications of V staying in base VP in the context of diachronic change that we see as OE changes to Modern English with the rise of *do*-support in Middle English. In §2.3, we

Word order	Counts
S O V p (if low scrambling)	10
O S V p	1

Table 9 Evidence of no V-to-I movement.

discuss the diachronic evidence presented by previous research in terms of the placement of adverbs that leads many to advocate for a compulsory V-to-I analysis of OE. It could very well be argued that all examples of *S O V p* are instances of high scrambling and the singular example of *O S V p* that warrants a ‘*V in base*’ analysis is an error in writing/transcription. However, apart from diachronic evidence, there appears to be no unambiguous evidence for or against V remaining in base (see discussion in §2.3).⁴³

OE exhibits ample evidence of V-to-I movement. The word orders that unambiguously exhibit V-to-I movement (compiled in Table 10) are:

- *O Aux S V and O Aux S p V*: In both of these word orders, the object has been high-scrambled/topicalized with Aux moving to the left over the subject, which is in base.
- *V S p O, V S O p and V O S p*: The main verb has clearly moved out of VP (either to INFL-medial, or to COMP-medial through INFL-medial). In either case, they are instances of unambiguous evidence of V-to-I movement.
- *S Aux O V p*: Here, p is after V making the word order VO in base, and since the Aux is between the subject and the low-scrambled object it is clearly in INFL-medial.
- *Word orders having Aux before S and O without the possibility of both S and*

⁴³ As one reviewer pointed out, one could also entertain the possibility of OE having no mandatory V-to-I movement with affix hopping as an optional parameter. If affix-hopping was indeed present in OE, then it would be in competition with V-to-I movement. OE having optional/competing V-to-I/affix hopping is compatible with proposals by Kiparsky (1996), Haider (1993, 2000), Koopman (1995a), Vikner (2001) and others that INFL-final OV languages do not have V-to-I movement to INFL-final. Conceivably, this could lead to an INFL-medial grammar without V-to-I in OE, but if that is the case, then it would have to have been a short period before V-to-I movement outcompeted the affix hopping grammar before the Early Middle English period. However, with the small amount of evidence that we present toward a ‘*V in base*’ analysis, the claim of OE having affix-hopping needs further investigation with a corpus-based analysis involving adverbs in a diachronic setting.

Word order	Counts
O Aux S V	74
O Aux S p V	4
V S p O	23
V S O p	23
V O S p	2
S Aux O V p	1
Aux S V O	63
Aux O S V	18
Aux S O p V	3
Aux S O V	177
Aux S V O p	1

Table 10 Unambiguous evidence of V-to-I movement.

O being postposed: Aux S O V, Aux S V O, Aux O S V, Aux S O p V and Aux S V O p are word orders that exhibit V-to-I movement. They have Aux before S meaning that Aux has moved leftward over S.

We also see a singular unambiguous instance (47) of verb raising in our study in the word order *S O p Aux V*. Since we allow for the possibility of high and low scrambling, unambiguous evidence for verb raising is much more sparse than in certain other studies (e.g., [Haeberli & Pintzuk 2012](#)).

6 CONCLUSION AND FUTURE WORK

Previous research investigating Old English (OE) word order variation has been entirely corpus-based, i.e., where researchers search the corpus for specific patterns based on some linguistic insight or hypothesis. In this study, we investigate the variation in OE by adopting a largely corpus-driven approach to explore the York-Toronto-Helsinki Parsed Corpus of Old English Prose (YCOE). In the corpus-driven search of YCOE that is presented in the preceding sections, we computationally generate word order patterns using a restricted tagset of S(ubject), V(erb), O(bject), p(article) and Aux(iliary). After generating all word order patterns using the restricted tagset, we implemented queries to search for those word orders patterns minimally containing S, V, and O that are instantiated in YCOE.

Subsequently, the results of our queries were linguistically analyzed with a set of syntactic assumptions that hold cross-linguistically, to manually derive all underlying structures that license each and every word order pattern that resulted from our search of the corpus. Following a long tradition of historical linguistics, if there was a single linguistic analysis (e.g., INFL-final) consistent across all possible structural descriptions of a word order pattern, the word order pattern was considered as unambiguous evidence for that analysis.

To the best of our knowledge, this study is the first corpus-driven study of OE which documents and compiles a comprehensive set of word order patterns and their counts in YCOE.⁴⁴ In addition, we also re-analyze OE word order patterns that previous research has shown to be revealing about one or more aspects of OE grammar. Using this approach, we were able to corroborate many existing theories of OE grammar and challenge some.⁴⁵ We found:

- Evidence of VO and OV canonical word orders
- Evidence of INFL-medial and INFL-final word orders
- Evidence of V-to-I movement
- Scant evidence for no V-to-I movement
- Evidence of two structural positions for subjects, but no constraints on types of subjects that can occupy those those positions.
- Evidence for two scrambling positions for objects, but no constraints on types of objects that can scramble.
- Evidence of adjunction in postposition and constraints on what can or cannot postpose.

We also delve into embedded clauses which have rarely been investigated in OE — analyzing for CP-recursion or lack thereof (see [Salvesen & Walkden 2017](#), [Walkden & Booth 2020](#) for rare discussion and examination of V2 phenomena in embedded clauses). This study is also the first where the treatment of adverbial adjuncts in the framework proposed by [Haegeman \(2012\)](#) has been applied in the context of OE syntax for analyzing CP-recursion.

The documented distribution of patterns and frequencies presented in this paper can support future investigations of the syntax of OE grammar. More generally, we put forth a corpus-driven framework to investigate word

⁴⁴ Limited to the restricted tag set we employed.

⁴⁵ Citations and details can be found in previous sections.

order patterns in historical linguistic corpora. This framework could potentially guide future corpus-driven analyses of OE (and other languages) with a different set of linguistic assumptions and/or a different or extended tag-set.

REFERENCES

- Alshammari, Ahmad Radi H. 2018. *The syntax of temporal and conditional adverbial clauses in Najdi Arabic*. Newcastle, UK: Newcastle University dissertation.
- Authier, J.-Marc. 1992. Iterated CPs and Embedded Topicalization. *Linguistic Inquiry* 23(2). 329–336.
- Biber, Douglas. 2009. Corpus-Based and Corpus-driven Analyses of Language Variation and Use. In Bernd Heine & Heiko Narrog (eds.), *The Oxford Handbook of Linguistic Analysis*, 159–192. Oxford University Press.
- 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.
- Biberauer, Theresa & Ans Van Kemenade. 2011. Subject positions and information-structural diversification in the history of English. *Catalan Journal of Linguistics* 10. 1–53.
- Campbell, Lyle. 2013. *Historical linguistics: An introduction*. Edinburgh: Edinburgh University Press 3rd edn.
- Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht, Netherlands: Foris.
- Chomsky, Noam. 1982. *Some concepts and consequences of the theory of government and binding*. Cambridge, Mass: MIT Press.
- Den Besten, Hans & Jerold A. Edmondson. 1983. The verbal complex in Continental West Germanic. In Werner Abraham (ed.), *On the formal syntax of the Westgermania*, 155–216. Amsterdam: John Benjamins.
- Diesing, Molly. 1990. Verb movement and the subject position in Yiddish. *Natural Language and Linguistic Theory* 8. 41–79.
- Dryer, Matthew. 1992. The Greenbergian word order correlations. *Language* 68(1). 81–138.
- Elenbaas, Marion & Ans Van Kemenade. 2014. Verb particles and OV/VO in the history of English. *Studia Linguistica* 68. 140–167.
- Evers, Arnold. 1975. *The transformational cycle in Dutch and German*. Utrecht: Utrecht University dissertation.
- Evers, Arnold. 1981. Two functional principles for the rule 'Move V'. *GAGL: Groninger Arbeiten zur germanistischen Linguistik* 19. 96–110.
- Frey, Werner. 2012. On two types of adverbial clauses allowing root-phenomena. In Lobke Aelbrecht, Liliane Haegeman & Rachel Nye (eds.),

- Main clause phenomena: New horizons*, vol. 190, 405–430. Berlin: John Benjamins Publishing.
- Fuss, Eric & Carola Trips. 2001. Variation and change in Old and Middle English—on the validity of the Double Base Hypothesis. *The Journal of Comparative Germanic Linguistics* 4. 171–224.
- Gries, Stefan T. 1999. Particle movement: A cognitive and functional approach. *Cognitive Linguistics* 10. 105–145.
- de Haan, Germen & Fred Weerman. 1986. Finiteness and verb fronting in Frisian. In Hubert Haider and Martin Prinzhorn (ed.), *Verb Second Phenomena in Germanic Languages*, 77–110. De Gruyter Mouton.
- Haerberli, Eric. 1999. *Features, categories and the syntax of A-positions. Synchronic and diachronic variation in the Germanic languages*. Geneva: University of Geneva dissertation.
- Haerberli, Eric. 2002. Inflectional morphology and the loss of verb-second in English. In David Lightfoot (ed.), *Syntactic effects of morphological change*, 88–106. Oxford, UK: Oxford University Press.
- Haerberli, Eric. 2005. Clause type asymmetries in Old English and the syntax of verb movement. In Montserrat Batllori, Maria-Lluïsa Hernanz, Carme Picallo & Francesc Roca (eds.), *Grammaticalization and Parametric Variation*, 267–283. Oxford, UK: Oxford University Press.
- Haerberli, Eric & Tabea Ihsane. 2016. Revisiting the loss of verb movement in the history of English. *Natural Language & Linguistic Theory* 34. 497–542.
- Haerberli, Eric & Susan Pintzuk. 2012. Revisiting verb (projection) raising in Old English. In Dianne Jonas, John Whitman & Andrew Garrett (eds.), *Grammatical change: Origins, nature, outcomes*, 219–238. Oxford, UK: Oxford University Press.
- Haegeman, Liliane. 1988. Verb Projection Raising and the Multidimensional Analysis: Some Empirical Problems. *Linguistic Inquiry* 19(4). 671–683.
- Haegeman, Liliane. 2012. *Adverbial Clauses, Main Clause Phenomena, and Composition of the Left Periphery: The Cartography of Syntactic Structures, Volume 8*, vol. 8. Oxford University Press on Demand.
- Haegeman, Liliane & Henk Van Riemsdijk. 1986. Verb projection raising, scope, and the typology of rules affecting verbs. *Linguistic Inquiry* 17. 417–466.
- Haider, Hubert. 1993. *Deutsche Syntax, generativ: Vorstudien zur Theorie einer projektiven Grammatik*. Gunter Narr Verlag.
- Haider, Hubert. 2000. OV is more basic than VO. In Svenonius, Peter (ed.), *The Derivation of VO and OV*, 45–67. Amsterdam: John Benjamins.
- Haider, Hubert. 2005. How to turn German into Icelandic — and derive the OV—VO contrasts. *Journal of Comparative Germanic Linguistics* 8(1/2). 1–

53.

- Haider, Hubert. 2007. Encounters of the third kind—On the hybrid headedness property of Old Germanic language. Paper presented at *Information Structure and Word Order Variation in older Germanic*, Humboldt University, Berlin, 9–10 November 2007.
- Higgins, F. Roger. 1991. The fronting of non-finite verbs in Old English. Paper presented at the Linguistic Society of America Annual Meeting, Chicago, January.
- Hoey, Michael. 2009. Corpus-driven approaches to grammar: The search for common ground. In *Exploring the Lexis–Grammar Interface*, 33–47. John Benjamins.
- Hooper, Joan B. & Sandra A. Thompson. 1973. On the applicability of root transformations. *Linguistic Inquiry* 4(4). 465–497.
- Hunston, Susan. 2002. *Corpora in applied linguistics*. Cambridge University Press.
- Iatridou, Sabine. 1991. *Topics in conditionals*. Cambridge, MA: Massachusetts Institute of Technology dissertation.
- Iatridou, Sabine & Anthony Kroch. 1992. The licensing of CP-recursion and its relevance to the Germanic verb-second phenomenon. *Working papers in Scandinavian Syntax* 50. 1–24.
- Jenset, Gard B. 2013. Mapping meaning with distributional methods: A diachronic corpus-based study of existential *there*. *Journal of Historical Linguistics* 3(2). 272–306.
- Kayne, Richard. 1994. *The antisymmetry of syntax*. Cambridge, MA: MIT Press.
- Kiparsky, Paul. 1995. Indo-European origins of Germanic syntax. In Adrian Battye & Ian Roberts (eds.), *Clause structure and language change*, 140–169. Oxford, UK: Oxford University Press Oxford.
- Kiparsky, Paul. 1996. The Shift to Head-initial VP in Germanic. In Höskuldur Thráinsson, Samuel David Epstein & Steve Peter (eds.), *Studies in comparative Germanic syntax II*, 140–179. Dordrecht, Netherlands: Kluwer.
- Koopman, Hilda. 1995a. On verbs that fail to undergo V-second. *Linguistic Inquiry* 26(1). 137–163.
- Koopman, Willem. 1990. *Word order in Old English; with special reference to the verb phrase*. Amsterdam: University of Amsterdam dissertation.
- Koopman, Willem F. 1995b. Verb final main clauses in Old English prose. *Studia Neophilologica* 67(2). 129–144.
- Koopman, Willem F. 1997. Another look at clitics in Old English. *Transactions of the Philological Society* 95(1). 73–93.
- Krishnamurthy, Ramesh. 2008. Corpus-driven lexicography. *International Journal of Lexicography* 21(3). 231–242.

- Kroch, Anthony. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1(3). 199–244.
- Kroch, Anthony. 1994. Morphosyntactic variation. In Katherine Beals (ed.), *Proceedings of the 30th annual meeting of the Chicago Linguistic Society*, 180–201. Chicago: Chicago Linguistic Society.
- Kroch, Anthony & Beatrice Santorini. 1991. The derived constituent structure of the West Germanic verb-raising construction. In Robert Freidin (ed.), *Principles and parameters in comparative grammar*, 269–338. Cambridge, MA: The MIT Press.
- Kroch, Anthony & Ann Taylor. 1997. Verb movement in Old and Middle English: Dialect variation and language contact. In Ans Van Kemenade and Nigel Vincent (ed.), *Parameters of morphosyntactic change*, 297–325. Cambridge, UK: Cambridge University Press.
- Kroch, Anthony & Ann Taylor. 2000. Verb-object order in early Middle English. In Susan Pintzuk, George Tsoulas & Anthony Warner (eds.), *Diachronic syntax*, 132–163. Oxford, UK: Oxford University Press.
- Lasnik, Howard & Mamoru Saito. 1992. *Move α* . Cambridge, Mass.: MIT Press.
- Lightfoot, David. 1991. *How to set parameters: Arguments from language change*. Cambridge, MA: MIT Press.
- Lyubymova, Svitlana. 2019. Language corpora as data base and verification tool for cognitive linguistic research. *Rhetoric and Communications Journal* 38.
- McGillivray, Barbara. 2014. *Methods in Latin computational linguistics*. Leiden, Netherlands: Brill.
- Miyagawa, Shigeru. 2017. Topicalization. *Gengo Kenkyu (Journal of the Linguistic Society of Japan)* 152. 1–29.
- Morley, Barry & Patricia Sift. 2006. Towards automatic identification of directive speech acts. In Roberta Facchinetti & Matti Rissanen (eds.), *Corpus-based studies of diachronic English*, 95–112. Peter Lang.
- Pintzuk, Susan. 1991. *Phrase structures in competition: Variation and change in Old English word order*. Philadelphia, PA: University of Pennsylvania dissertation dissertation.
- Pintzuk, Susan. 1993. Verb seconding in Old English: verb movement to Infl. *The Linguistic Review* 10. 5–36.
- Pintzuk, Susan. 1996. Old English verb-complement word order and the change from OV to VO. *York Papers in Linguistics* 17. 241–264.
- Pintzuk, Susan & Eric Haeberli. 2008. Structural variation in Old English root clauses. *Language Variation and Change* 20(3). 367–407.
- 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*, 249–278. Oxford, UK: Blackwell.
- Randall, Beth. 2004. CorpusSearch 2. <https://corpussearch.sourceforge.net/>. [Online; accessed 19-July-2021].
- Ringe, Don & Joseph F Eska. 2013. *Historical linguistics: Toward a twenty-first century reintegration*. Cambridge, UK: Cambridge University Press.
- Roberts, Ian. 1997. Directionality and word order change in the history of English. In Ans Van Kemenade and Nigel Vincent (ed.), *Parameters of Morphosyntactic Change*, 397–426. Cambridge, UK: Cambridge University Press.
- Rögnvaldsson, Eiríkur & Höskuldur Thráinsson. 1990. On Icelandic word order once more. *Modern Icelandic syntax, Series: Syntax and semantics* 24. 3–40.
- Salvesen, Christine & George Walkden. 2017. Diagnosing embedded V2 in Old English and Old French. In Eric Mathieu & Rob Truswell (eds.), *Micro-change and macro-change in diachronic syntax*, 168–181. Oxford, UK: Oxford University Press.
- Salzmann, Martin. 2011. Resolving the movement paradox in verb projection raising: In favor of base-generation and covert predicate raising. *Empirical Issues in Syntax and Semantics* 8. 453–485.
- Santorini, Beatrice. 1992. Variation and change in Yiddish subordinate clause word order. *Natural Language & Linguistic Theory* 10. 595–640.
- Santorini, Beatrice. 1989. *The generalization of the verb-second constraint in the history of Yiddish*. Philadelphia, PA: University of Pennsylvania dissertation.
- Sheehan, Michelle, Theresa Biberauer, Ian Roberts & Anders Holmberg. 2017. *The Final-Over-Final Condition: A syntactic universal*. Cambridge, MA: MIT Press.
- Sinclair, John. 2004. *Trust the text: language, corpus and discourse*. London: Routledge.
- Steele, Susan. 1975. On some factors that affect and effect word order. In Charles N. Li (ed.), *Word order and word order change*, 197–268. Austin, TX: University of Texas Press.
- Struik, Tara & Ans Van Kemenade. 2020. On the givenness of OV word order: a (re) examination of OV/VO variation in Old English. *English Language & Linguistics* 24(1). 1–22.
- Struik, Tara & Ans Van Kemenade. 2022. Information structure and OV word order in Old and Middle English: a phase-based approach. *The Journal of Comparative Germanic Linguistics* 1–36.
- Stubbs, Michael. 2007. An example of frequent English phraseology: distri-

- butions, structures and functions. In Roberta Facchinetti (ed.), *Corpus linguistics 25 years on*, 87–105. Amsterdam, New York: Brill.
- Taylor, Ann. 2007. The York—Toronto—Helsinki parsed corpus of Old English prose. In Joan C. Beal, Karen P. Corrigan & Hermann L. Moisl (eds.), *Creating and Digitizing Language Corpora*, 196–227. London, UK: Palgrave Macmillan.
- Taylor, Ann & Susan Pintzuk. 2015. Verb order, object position, and information status in Old English. In Theresa Biberauer & George Walkden (eds.), *Syntax over time: Lexical, morphological, and information-structural interactions*, 318–335. Oxford, UK: Oxford University Press.
- Taylor, Ann, Anthony Warner, Susan Pintzuk & Frank Beths. 2003. The York-Toronto-Helsinki parsed corpus of Old English. Department of Language and Linguistic Science, University of York.
- Teubert, Wolfgang. 2005. My version of corpus linguistics. *International Journal of Corpus Linguistics* 10(1). 1–13.
- Tognini-Bonelli, Elena. 2001. *Corpus linguistics at work*. Amsterdam: John Benjamins.
- Tomaselli, Alessandra. 1995. Cases of verb-third in Old High German. In Adrian Battye & Ian G. Roberts (eds.), *Clause structure and language change*, 345–369. Oxford: Oxford University Press.
- Trips, Carola. 2002. *From OV to VO in early Middle English*. Amsterdam, Philadelphia, PA: John Benjamins Publishing.
- Trotzke, Andreas & Stefano Quaglia. 2016. Particle topicalization and German clause structure. *Journal of Comparative Germanic Linguistics* 19. 109–141.
- Van Kemenade, Ans. 1987. *Syntactic case and morphological case in the history of English*. Dordrecht, Netherlands: Foris.
- Van Kemenade, Ans. 1997. V2 and embedded topicalization in Old and Middle English. In *Parameters of morphosyntactic change*, 326–352. Cambridge University Press.
- Van Kemenade, Ans & Tanja Milicev. 2012. Syntax and discourse in Old and Middle English word order. In Dianne Jonas, John Whitman & Andrew Garrett (eds.), *Grammatical change: Origins, nature, outcomes*, 239–254. Oxford, UK: Oxford University Press.
- Vikner, Sten. 2001. Verb movement variation in Germanic and Optimality Theory. Ms., University of Tübingen.
- Vikner, Sten. 2017. CP-recursion and the derivation of verb second in Germanic main and embedded clauses. In *DGFS 2016 workshop V2 in grammar and processing: Its causes and its consequences*, 1–26.
- Walkden, George & Hannah Booth. 2020. Reassessing the historical evidence for embedded Verb Second. In Rebecca Woods & Sam Wolfe (eds.), *Re-*

- thinking verb second*, 536–554. Oxford, UK: Oxford University Press.
- Wallenberg, Joel C. 2009. *Antisymmetry and the conservation of c-command: Scrambling and phrase structure in synchronic and diachronic perspective*. Philadelphia, PA: University of Pennsylvania dissertation.
- Watanabe, Akira. 1992. Larsonian CP recursion, factive complements, and selection. In *North East Linguistics Society*, vol. 23 2, 18.
- Wurmbrand, Susi. 2000. *The Structure(s) of Particle Verbs*. Ms., McGill University and Universität Wien, Montreal and Vienna.
- Zimmermann, Richard. 2011. A diachronic survey of syntactic features in a compiled version of the Anglo-Saxon Chronicle. Paper presented at The York-Newcastle-Holland Symposium on the History of English Syntax 10 (SHES10).

Soumik Dey
2025 Hamilton Avenue
San Jose
CA-95125
sdey@gradcenter.cuny.edu
[https://www.linkedin.com/in/
dr-soumik-dey-ab8549a1/](https://www.linkedin.com/in/dr-soumik-dey-ab8549a1/)

William G. Sakas
Department of Computer Science,
Hunter College, CUNY
North Building, Room 1008
695 Park Avenue
New York NY 10065
sakas@hunter.cuny.edu
www.cs.hunter.cuny.edu/~sakas/