A quantitative exploration of the functions of auxiliary *do* in Middle English[1]

LORENZO MORETTI ●
University of Zurich

One of the questions that still surrounds the history of auxiliary *do* is what function it had
during the Middle English period (*c.*1100–1500). Scholars have put forward different
hypotheses, suggesting that it could serve, among others, as a perfective marker
(Denison 1985), agentive marker (Ecay 2015) and habitual marker (Garrett 1998). The
present article reports on a quantitative study that aims to shed further light on this
issue. By means of a collexeme analysis, this article investigates the semantic features
of the infinitives that occur with auxiliary *do* in several Middle English corpora. The
results show that auxiliary *do* was not connected to verbs with specific semantic
profiles, but it was employed in different contexts and had various functions.
Specifically, the data suggest that auxiliary *do* was used (i) as an accommodation tool
to facilitate the use of low-frequency verbs, particularly of French origin, and (ii) as an
aspectual particle to mark both perfectivity and habituality. It is argued that the
multifunctionality of auxiliary *do* in Middle English played a crucial role in the
preservation of the construction before it spread to the NICE (i.e. negation, inversion,
code and emphasis) environments.

Keywords: auxiliary *do*, collostructional analysis, corpus linguistics, Middle English,
language contact

1 Introduction

Present-day English features an operator, which is usually referred to as auxiliary *do* (other
terms used in the literature are periphrastic *do*, dummy *do* and *do*-support), which is
semantically empty but syntactically obligatory in negation, inversion, code and
emphasis environments (the so-called NICE properties; see for a summary Huddleston
& Pullum *et al.* 2002: 92–112) when no other auxiliary is present. Initially, when it
developed between the end of the twelfth and the beginning of the thirteenth century
(Ellegård 1953), auxiliary *do* had a different distribution, as it was restricted to

---

[1] I would like to thank Tine Breban for discussing various aspects of this paper with me and two anonymous reviewers
for their helpful comments.

declarative sentences only, where it was followed by an infinitive verb (*do* – INF), an example of which is provided in example (1).[2]

(1) and   also   that   ther   schold   be   founden weys   to   the   having
    and   also   that   there   should   be   found   ways   to   the   having
    awey   specialy   of the Duc of Orlians   and   also   of   the   king
    away   especially   of the Duke of Orleans   and   also   of   the   king
    as   welle   as   of   the   remanant   of   my   forsayd   prysoners
    as   well   as   of   the   remaining   of   my   aforesaid   prisoners
    that   god   **do**   defende.
    that   God   do   defend
    'and, also, that there should be found ways for the detention of especially the Duke of Orleans, the king and the remaining of my aforesaid prisoners that God defends.'
    (PCEEC: SIGNET,122.061.348)

The auxiliary *do* construction that we observe in the example above possesses some unique features. Like its modern descendant, it has no lexical content in declarative sentences, with the entire *do* – INF construction considered equivalent to a construction with a finite verb only (Ellegård 1953). At the same time, it is also syntactically optional, as auxiliary *do* never became mandatory in non-emphatic declarative sentences (Ellegård 1953; Denison 1993). The fact that auxiliary *do* was both semantically empty and syntactically optional in Middle English raises the question of why it remained part of the language system. Even if we understand language as a highly redundant system which comprises constructions that 'run the gamut from full generality to complete idiosyncrasy' (Langacker 1988: 113), the survival of auxiliary *do* still represents an oddity. Why would speakers maintain a construction in their linguistic inventory if this construction had no meaning and was not syntactically required? The obvious answer is to assume that auxiliary *do* had some kind of function. Scholars in the past have addressed this issue and proposed a number of possible functions (see section 2). However, methodological progress has provided us with new means to investigate this question. On the one hand, the advent of digitalised corpora has allowed us to rely on larger data sources and more accurate retrieval processes. On the other hand, recent years have seen the development of a variety of statistical methods which have proved useful to examine the synchronic behaviour and the diachronic developments of linguistic constructions (e.g. Hilpert 2013).

Thus, the purpose of this article is to address this issue and explore the function, or functions, that auxiliary *do* possessed in declarative sentences in Middle English from a quantitative perspective. This is a crucial period in the life of auxiliary *do*, as it

---

[2] Here and in the following examples, auxiliary *do* is not translated in the idiomatic translation in order to avoid any confusion with the modern emphatic *do*.

follows the coming into being of auxiliary *do* and precedes its spread to the NICE environments, and is therefore ideal to investigate the reasons why auxiliary *do* was preserved in the language system. In order to do so, a collexeme analysis will be used to investigate the semantic features of the infinitives that occur with auxiliary *do* in the period under investigation. The purpose of collexeme analyses is to determine the degree of attraction/repulsion between a construction and a lexical item; in the present study, this statistical method will be used to generate a ranked list of the infinitives that are most attracted (i.e. that most preferably occur) to auxiliary *do*. This will allow us to assess whether auxiliary *do* was used in combination with specific classes of verbs and, therefore, help us get a better understanding of the functions it had. The results of the statistical analysis indicate that auxiliary *do* was attracted to low-frequency verbs, particularly of French origin. On this basis, it is suggested that one function of auxiliary *do* was to be used as what will be called an ACCOMMODATION tool (in the spirit of De Smet & Shaw 2022) that aided the integration of infrequent verbal stems, be they either of French or native (i.e. Germanic) origin. Furthermore, it is shown that auxiliary *do* could also occur, although to a lesser extent, as a marker of aspectuality, as it is found in habitual as well as in perfective contexts. In light of these results, it is suggested that the multifunctionality of auxiliary *do* played a crucial role in the preservation of the construction before it acquired the NICE properties and joined the auxiliary system (see Budts 2021 for a recent contribution on this topic).

This article is structured as follows. Section 2 discusses previous accounts of the functions of auxiliary *do* in Middle English. Section 3 illustrates the methodological apparatus of this study. Section 4 presents the results of the statistical analysis, which are then discussed in section 5. Section 6 concludes the article.

2   Previous studies

Studies on the functions of auxiliary *do* in early English declarative sentences began to appear during the first half of the twentieth century (e.g. Engblom 1938; Ellegård 1953; Dahl 1956). What caught the attention of these early scholars was that all the very first occurrences of auxiliary *do* occur in poetical compositions, while the appearance in prose texts is dated only two centuries later, i.e. during the late fifteenth century. Ellegård (1953), in particular, noted that auxiliary *do* was frequent when the infinitive it occurs with was the last element of the line. This particular distribution led Ellegård to assume that auxiliary *do* was a metrical device used by poets to manipulate the word order of the verse and allow the placement of the infinitive verb at the end of the line in order to facilitate rhyme (see, for instance, Ellegård 1953: 208). This hypothesis has been tested in a recent quantitative study, in which it has been shown that auxiliary *do* was indeed used as a metrical device, but that it also had other functions, such as maintaining the correct number of beats in the verse and facilitating the use of verbs borrowed from French (see Moretti 2023).

In many accounts, the functions of auxiliary *do* are connected to the construction from which it is assumed to have originated.[3] The perfective hypothesis proposed by Denison (1985), for instance, stems from the idea that auxiliary *do* developed from the causative construction *do* – INF, a construction sporadically attested in Old English (see Fischer 1989; Denison 1993: 256; Timofeeva 2011) and still particularly infrequent in early Middle English (Ellegård 1953: 44), an example of which is provided in (2).

(2) The mayere of Norwich **dede** a-rest the baylly of Normandys,
     the   mayor   of Norwich did   arrest the bailie of Normandy
     'The mayor of Norwich had the bailie of Normandy arrested' (PCEEC:
     PASTON,I,331.107.327)

Denison claims that when the subject of the infinitive verb is left unexpressed, the pragmatic focus shifts from the agency of the action to the realisation of the action, which in turn makes *do* a marker of perfectivity. The evidence presented by Denison concerning the development of *do* as a perfective marker consists of the semantic features of the verbs that co-occur with auxiliary *do* at early stages. Denison found that the vast majority of predicates that combine with *do* in the thirteenth and fourteenth centuries express accomplishments and, to a lesser extent, achievements (in the sense of Vendler 1967), which, according to Denison (1985: 54), are the semantic classes of verbs that are more compatible with perfective particles. Then, once causative *do* was replaced by causative *make* towards the end of the Middle English period, perfective *do* became isolated, lost any perfective meaning, and joined the auxiliary system.

Aspectuality seems to play an important role in the history of auxiliary *do*. There are examples in modern southwestern and Irish dialects where auxiliary *do* occurs in non-emphatic declaratives and seems to express habituality (Ihalainen 1976). Habitual uses of *do* were noted by grammarians of the seventeenth and nineteenth centuries, who distinguish between habitual forms, which are conveyed by *do* and an infinitive verb, and actual forms, which are expressed by the present tense (Elworthy 1877). The same holds for the past tense, where habitual *did* – INF constructions can be replaced by *used to* – INF constructions. An example of *do* used as a habitual marker is provided in (3), in which *do* is used to describe generic, that is, habitual, actions (Ihalainen 1976).

(3) The surplus milk they did make into cheese and then the cheese **did** go to the different markets, that's how that did work. (from Ihalainen 1976: 615)

Building on examples such as (3), Garrett (1998) argues that instances of habitual *do* can be traced back to Middle English, when habitual instances are marked by the presence of adverbial adjuncts, verbal complements or more general contextual clues (see Garrett 1998: 297–300). Indeed, it is true that some examples lend themselves to a habitual

---

[3] Other proposals on the origin of auxiliary *do* that are not included in this section are the so-called Celtic hypothesis (Poussa 1990) and the elliptic origin (Visser 1963–73: 1489–91) (see Budts (2021: 16–23) for a comprehensive summary of these hypotheses).

interpretation. Look, for instance, at example (4), where the adverb *ilome* 'frequently' allows the interpretation of *dude him baþie i-lome* as habitual.

(4) A preost was wilene in one stude that **dude** baþie i-lome
    a priest was once in one place that did bath frequently
    'A priest was once in a place that frequently bathed him' (Southern English Legendary: 423:97, from Garrett 1998: 297)

However, habituality is not always clearly discernible in the examples provided by Garrett. For instance, in example (5) the construction involving *do* does not seem to contain any trace of habituality. Although Garrett argues that the context allows a habitual interpretation of *do*, the text in which this example appears is a poem and, given the presence of *die* at the end of the previous line, it seems likely that *do* is used as a metrical tool to place *lie* at the end of the verse in order to rhyme with *die*.[4]

(5) þerof gan he die; at Teukesbiri in toumbe his body **did** lie
    thereof PTC he die at Tewkesbury in tomb his body did lie
    'From it he died; his body lay in a tomb at Tewkesbury' (Mannyng Chron.Pt.(2) (Petyt) 213–14, from Garrett 1998: 299)

A further function of auxiliary *do* put forward by Fischer & van der Wurff (2006) is that the construction was used to support the integration of French verbal stems. Following the Norman Conquest (1066), the Germanic core of the vocabulary was substantially incremented by French loanwords (see, e.g., Dalton-Puffer 1996; Durkin 2014). These new verbs could be difficult to integrate in the native inflectional system, and a strategy to avoid hybrid forms, i.e. a French stem with an English past tense in *-ed* or a present in *-est*, would be to use auxiliary *do* with a foreign infinitive (Fischer & van der Wurff 2006: 155).

    Lastly, there is a recent contribution from Ecay (2015), who carried out a quantitative investigation of the behaviour of auxiliary *do* in Middle English. Ecay examined with what frequency auxiliary *do* occurred in clauses that feature verbs with different argument structures. Specifically, he divided his data set into unergative (e.g. *work*, *dance*), unaccusative (e.g. *die*, *arise*), experiencer (e.g. *care*, *fear*) and transitive (e.g. *ask*, *love*) verbs, and measured the incidence of auxiliary *do* in each context. The results of his analysis show that in affirmative declaratives, auxiliary *do* is robustly attested with unergative, transitive and experiencer-subject verbs, i.e. verbs that require the presence of an agentive external argument, while it is virtually absent with unaccusative verbs. On this basis, Ecay (2015: 80) argues that '*do*-support in affirmative declaratives is generated by a grammar which uses *do* to mark the presence of an (agentive?) external argument'.

---

[4] Note that the verb *gan* has been shown to be a metrical particle in Middle English poems, as it was used to place the infinitive at the end of the verse in order to facilitate rhyme (e.g. Funke 1922; Mustanoja 1960: 611–14; Smyser 1967).

3 Methodology

3.1 Corpus and data collection

The corpus data discussed in this article stem from three main sources: the *Parsed Corpus of Early English Correspondence* (PCEEC; see PCEEC 2006), the *Penn Parsed Corpus of Middle English*, 2nd edition (PPCME2, Kroch & Taylor 2000) and the *Helsinki Corpus* (HC, Rissanen *et al.* 1991). PCEEC and PPCME2 are corpora annotated for part-of-speech which include different genres and cover different time spans. PCEEC is made up exclusively of personal letters composed between 1350 and 1710, which altogether account for approximately 2.2 million words. However, since this study focuses on the functions of auxiliary *do* in Middle English, only a subcorpus containing the data for later periods, which are M3 (1350–1419) and M4 (1420–1499) following the periodisation introduced by the HC (Kytö 1996), was used in this article. The total word count of the letters in PCEEC from these two periods is 19,505 and 384,037 words, respectively. PPCME2 contains only prose texts dated between 1150 and 1500 and amounts to *c.* 1.2 million words.[5] Lastly, the HC is a more comprehensive corpus that contains English texts from early Old English (–850) to the end of the Early Modern English period (1500–1710). The Middle English section of the HC includes several texts that are also part of the other two corpora; therefore, I have only selected texts like dramas and mystery plays which are known to contain a conspicuous number of auxiliary *do* examples (Nurmi 1999: 236). The dramas consulted are the following: *Ludus Coventriae*, *Mankind*, *The Wakefield Pageants in Towneley Cycle*, *The York Plays* and *Digby Plays*, which account for 19,720 words.

The total word count of the data set amounts to 1,543,685, with the word count for each period included in this study given in table 1. As mentioned above, PCEEC and PPCME2 are tagged for parts-of-speech (e.g. pronouns, lexical nouns, proper nouns) and present a tag specific to the verb *do*, i.e. DO, which facilitated the extraction of the data relevant for this study.[6] The construction searched for was the verb *do* followed by an infinitive verbal form (pos tag VB). HC, on the other hand, contains only raw data, which means that every instance of *do* was analysed manually to determine whether it was followed by an infinitive verb. The software used to collect the data was *AntConc 3.5.9* (Anthony 2020). The collection procedure yielded an initial data set of 368 tokens. The following step consisted in the deletion of all the examples in which *do* was not an auxiliary or for which the interpretation was uncertain (see section 3.2). In this way, the original data set was narrowed down to 112 examples of auxiliary *do*, which were distributed across the different periods as illustrated in table 2.

---

[5] PPCME2 includes one poem, the *Ormulum* (73,576 words), which has been excluded from this analysis to eliminate metrical uses of auxiliary *do*.

[6] The tags concerning the verb DO are the following: DAG (present participle), DAN (passive participle, both verbal and adjectival), DO (infinitive), DOD (past, including past subjunctive), DOI (imperative), DON (perfect participle), DOP (present, including present subjunctive).

Table 1. *Word count for each of the subperiods included in this study*

Period	Word count
M1 (1150–1249)	210,769
M2 (1250–1349)	146,575
M3 (1350–1419)	510,554
M4 (1420–1499)	671,828
Total	1,543,685

Table 2. *Observed frequency of auxiliary* do *by subperiods and corpora considered in this study*

Period	PPCME2	PCEEC	HC	Total
M1 (1150–1249)	0	0	0	0
M2 (1250–1349)	0	0	0	0
M3 (1350–1419)	11	5	0	16
M4 (1420–99)	30	44	22	96
Total	41	49	22	112

3.2 Data analysis

The identification of auxiliary *do* constructions in the Middle English period is not a straightforward exercise, since the string *do* – INF could also express a causative event. The presence of *do* in causative constructions is already attested in Old English texts, when it could take different types of complements, the more common being a *that*-clause (Royster 1922; Fischer 1989; Timofeeva 2011). Causative *do* with an infinitive verb began to occur with more frequency in early Middle English, although the newly emerged causative constructions involving *make* have been shown to be more productive already in the earliest Middle English periods (Moretti 2022). Structurally, there are no differences between causative and auxiliary *do* and their interpretation relies solely on contextual clues. There are examples where it can be safely assumed that the subject of the infinitive verb is not co-referential with the one of *do*. In such cases, the pattern *do* – INF expresses a causative event with the subject of the infinitive left unexpressed.[7] An example of causative *do* – INF is given in (2), and another one is provided in (6). In other instances, however, the subject of *do* and

---

[7] Except in some cases (see Goldberg 2001), causative constructions in which the subject of the infinitives is left unexpressed are no longer used in Present-day English. However, in Old and at least in early Middle English constructions like **he made build the castle* were rather common, particularly with some verbs (Denison 1993: 171).

the infinitive verb are co-referential, which means that *do* can be interpreted as an auxiliary, as shown in example (1) above and (7).

(6) And þis is cause þat I haue take no sewerte; wherfor   yf
    and this is cause that I have taken no certainty wherefore if
    it like yow to **do** write up th'endendure that I may haue
    it likes you to do write up the-indenture that I may have
    it that oon part vndir your seall,
    it that one part under your seal
    'And this is a cause that I have taken with no certainty; wherefore if you can have the indenture written up so that I can have that one part under your seal,' (PCEEC: PASTON, II,328.424.11000)

(7) and by the ryng of the churche doore as is aforsaide, that
    and by the ring of the church door as is aforesaid that
    then the saide attourneis **do** enter into the said parsonage
    then the said attorneys do enter into the said parsonage
    'and then by the ring of the church's door, as it is aforesaid, the said attorneys enter into the said parsonage' (PCEEC: CROMWEL,I,323.004.64)

However, there are some cases in which it is complicated for a modern researcher to determine whether *do* is a causative or an auxiliary verb, as both interpretations appear equally reasonable. This is illustrated in example (8), where it is unclear whether the subject of *do* and the subject of the infinitive are co-referential and is difficult, therefore, to assign an auxiliary (as in (a)) or a causative (as in (b)) interpretation.

(8) ye haue leuyd Thomas Grayngar to be your atornay at
    you have levied Thomas Graynger to be your attorney at
    Calles whyll ye go to the marte. I **do** send hym
    Calais while you go to the market I do send him
    a letter, and therein the schypys namys and the whette
    a letter and therein the ships' names and the weight
    of howr fathers
    of our fathers
    'you have levied Thomas Graynger to be your attorney at Calais while you go to the market. I am (a) sending him / (b) having him sent a letter with the names of the ships and the weight of our fathers' (PCEEC: CELY,81.064.1396)

These instances, which amount to 24 cases, have been tagged as ambiguous and excluded from the statistical analysis.

The first occurrence of auxiliary *do* in the data set dates to the last decade of the fourteenth century. There are earlier instances of the pattern *do* – INF, but they are either causative constructions, as in (6), or their interpretation is unclear between auxiliary and causative, as in (8). Clear examples of auxiliary *do* in early Middle

English, i.e. before the end of the fourteenth century, can be found in poetic texts, where the construction has different functions, as mentioned above (see Moretti 2023 for further details).

3.3   Statistical method

The method chosen to conduct the statistical analysis is a collexeme analysis (CA). CA is part of the family of methods known as Collostructional Analysis (Stefanowitsch & Gries 2003, 2005; Gries & Stefanowitsch 2004a, 2004b), which also includes distinctive collexeme analysis and diachronic distinctive collexeme analysis (see Hilpert 2008: 34–45 for an extensive introduction). The main purpose of CA is to carry out a semantic investigation of a given grammatical construction by identifying lexical elements (also called collexemes) that are typical of such construction. Specifically, CA measures the degree of attraction (or repulsion) between the items that fill two slots (A and B) in a pattern; this is calculated by assessing whether observed values deviate from what we would expect if the combination of A and B were free, and the corresponding statistical test is interpreted as a measure of attraction or repulsion between A and B.

In the present case, slot A is filled by auxiliary *do* and slot B by the infinitive verb. The extent to which an infinitive verb is attracted to auxiliary *do* is determined by the collostructional strength, which is measured through a comparison between the frequencies of the two elements of the construction under investigation both in conjunction and in isolation. For instance, if one is interested in investigating how strong the level of attraction between 'do' and 'understand' in 'do understand' is, the collostructional strength of 'understand' is calculated through the extraction of the frequency of the construction 'do understand' as a whole, and then of 'do' and 'understand' in isolation. The association measure used to calculate the collostructional strength is the Fisher–Yates Exact test, which according to Gries & Stefanowitsch (2004b: 101) is better suited to capturing rare collocations and overall provides better results than other association measures.[8] The final output of CA is a ranked list in descending order of the infinitives that are attracted to auxiliary *do*, which are referred to as collexemes. The software used for the statistical analysis is R (R Core Team 2017), while the package used to perform the CA is *collostructions* (Flach 2021).

A further piece of information that complements the results of the CA, and is particularly relevant when the data set is relatively small, concerns the analysis of how the collexemes of auxiliary *do* are distributed across the texts. It might be the case, in fact, that a given infinitive that ranks high in the list of the collexemes attracted to

---

[8] The use of the Fischer–Yates exact test was criticised by Schmid & Küchenhoff (2013). The main problem diagnosed by Schmid & Küchenhoff is that this test is not adequate as an association measure, and they have suggested using alternative tests like the odds ratio or the log odds ratio. However, Gries (2015) responded to Schmid & Küchenhoff and argued that the Fischer–Yates exact test presents several advantages besides the one cited in the main text (for more details see Gries 2015: 508). For these reasons, it has been preferred over other alternatives.

auxiliary *do* is attested only in a restricted number of texts. On the other hand, an infinitive that has a lower rank could be attested in a larger number of texts and be more equally distributed, which in turn means that it is more representative of any tendencies in the data. For this reason, I will calculate the degree of dispersion of every collexeme attracted to auxiliary *do*. The measure chosen is called *DP* (deviation of proportion) and was proposed by Gries (2008). *DP* is a value between 0 and 1, with values closer to 1 that indicate an uneven distribution, whilst values closer to 0 mean that the collexeme is spread out nicely across the texts and is, therefore, well dispersed.

4   Results

A CA was performed to investigate the degree of attraction between auxiliary *do* and the infinitives it co-occurs with in the period 1350–1499, which is when we have the first textual records of the auxiliary construction in the corpora consulted. The results of the CA are shown in table 3, which lists the top 20 most attracted infinitives of auxiliary *do* that are statistically significant at $p < 0.05$, along with their observed and expected frequencies, collostructional strength (CS) and their degree of dispersion (*DP*).

The first observation to be made concerns the poor dispersion of the collexemes listed in table 3. As can be seen, the *DP* values of every collexeme are close to 1, which indicates uneven distribution. If this result is expected given the low frequency of auxiliary *do* (see observed and expected frequencies in table 3), it becomes particularly relevant for the top two collexemes, namely *appear* and *understand*. These two verbs in fact appear to be strongly attracted to auxiliary *do* because they occur in formulaic expressions (Nurmi 1999: 236). That is, instances of *do – appear*, see example (9), occur only in two texts, of which one, the sermons of Robert Fitzjames, contains eight of them. Similarly, examples of *do – understand* occur exclusively in the Cely letters in the formulaic expression *I do well understand*, as shown in example (10).

(9) whom   our   lorde   had   with   his   grace   inspyred,   as   in   ye   fyrste
     whom   our   lord   had   with   his   grace   inspired   as   in   the first
     chapytre   of   the   fyrst   boke   of   Esdre   **doth**   appere
     chapter   of   the   first   book   of   Ezra   does   appear
     'whom our lord had inspired with his grace, as it appears in the first chapter of the book of Ezra'
     (PPCME2: CMFITZJA,A4V.62)

(10) Fordyrmor,   plesythe   yt   yow   to   vndyrstonde   I   hawe   resseywyd
     furthermore   pleases   it   you   to   understand   I   have   received
     an   letter   ffrom   yow,   the   whiche   I   hawe   rede   and   **do**   whell
     a   letter   from   you,   the   which   I   have   read   and   do   well
     vndyrstonde
     understand
     'Furthermore, it pleases you to know that I have received a letter from you, which I have read and well understand' (PCEEC: CELY,5.003.27)

Table 3. *Twenty most strongly associated collexemes of auxiliary* do *in the period 1350–1499*

VERB	GLOSS	*Obs*	*Exp*	CS	*DP*
apperen	appear	9	0	122.475	0.91
understanden	understand	12	0.1	108.815	0.89
specifien	specify	2	0	33.275	0.97
crien	cry, shout	3	0	30.825	0.96
bigilen	beguile	2	0	26.282	0.97
seisen	seize	2	0	25.613	0.97
declaren	declare	2	0	22.543	0.97
shouen	show	3	0	22.534	0.95
comen	come	4	0	21.093	0.94
appropren	appropriate	1	0	18.538	0.99
dedicaten	dedicate	1	0	18.538	0.99
overmacchen	overmatch	1	0	18.538	0.99
recchen	interpret	1	0	18.538	0.99
tracen	trace	1	0	18.538	0.99
wecchen	watch	1	0	18.538	0.99
asken	ask	2	0	16.705	0.97
braggen	brag	1	0	15.765	0.99
magnifien	magnify	1	0	15.765	0.99
reneuen	renew	1	0	15.765	0.99
satisfien	satisfy	1	0	15.765	0.99

Secondly, it can be noted that there is not a particular semantic profile to which auxiliary *do* is attracted. We find infinitives that express motion (e.g. *comen*), change of state (e.g. *seisen*), change of possession (e.g. *appropren*), as well as verbs of psychological state (e.g. *bigilen*). Moreover, there are some collexemes, see for instance *comen* and *seisen*, which usually occur in perfective contexts, and will be discussed more in detail in section 5.

An interesting result of the CA is that several collexemes that appear in table 3 occur rather sporadically in the data set. In fact, although it has to be borne in mind that the CA takes into account only infinitival forms, it is striking that the infinitive of verbs like *appropren*, *dedicaten*, *overmacchen*, *ricchen* and *tracen* is attested only once, the infinitive of *braggen*, *magnifien*, *satisfien* and *reneuen* occurs twice, while the infinitive of *specifien* appears three times in the data set used in this study. In addition, the infinitive forms of *seisen* and *bigilen* are found 14 and 12 times, respectively. Furthermore, it is worth noting that the majority of the verbs in table 3 are French borrowings. Specifically, we can identify 12 French-origin verbs, namely *apperen*, *specifien*, *crien*, *bigilen*, *seisen*, *appropren*, *declaren*, *magnifien*, *dedicaten*, *tracen*, *braggen* and *reneuen*. These collexemes do not share specific semantic features, as they belong to different semantic domains. We do in fact see verbs that express social interaction and verbal communication, like *declaren* and *crien*, verbs of psychological state, such as *bigilen*, and verbs that denote a change of state, like *seisen*. Instead, there

seems to be a correlation between low frequency and French borrowings. That is, several of the low-frequency verbs that are attracted to auxiliary *do* in table 3 are French borrowings. This correlation will be discussed more in detail in section 5.

5   Discussion

This study investigates which functions auxiliary *do* had during the Middle English period in declarative sentences by looking at the semantic features of the infinitives it occurs with. The first important finding of the CA reported in section 4 is that the infinitives that complement auxiliary *do* are not characterised in terms of a single, predominant meaning, but indicate that *do* was used in different contexts. One of the contexts identified by the statistical analysis is that auxiliary *do* often appeared in combination with verbs that are infrequently attested in the data set. Verbs following auxiliary *do* which fall into the infrequent class include: *specifien*, *bigilen*, *seisen*, *appropren*, *dedicaten*, *tracen*, *braggen*, *magnifien*, *reneuen*, *satisfien*, *overmacchen*, *recchen*, *wecchen*. Furthermore, as mentioned in section 4, the majority of these verbs are of French origin, since only *overmacchen*, *recchen* and *wecchen* are not borrowings. However, as mentioned above, CA does not provide a full picture of the frequency of these verbs, since it only takes into account infinitive forms. Thus, I collected every instance of these verbs and compared their frequency with the 10 most frequent Germanic verbs and the 10 most frequent French borrowings in the PPCME2.[9] The results are illustrated in figure 1.

   The plots in figure 1 show that the frequency of the verbs that occur with auxiliary *do*, regardless of their origin, is remarkably lower when compared with more frequent verbs. Specifically, we see that the frequency of the French borrowings that appear with auxiliary *do* is of 0.13232 per 10,000 words, while the frequency of the 10 most frequent French borrowings is 4.75628 per 10,000 words. An even greater disparity is found with verbs of Germanic origin, since those that occur with auxiliary *do* are attested with a frequency of 0.092127 per 10,000 words, while the frequency of the 10 most frequent Germanic verbs is 39.01762 per 10,000 words.

   Let us start by discussing verbs that are borrowed from French, which represent the vast majority of these infrequent collexemes. One plausible reason for the presence of several French-origin verbs among the collexemes most attracted to auxiliary *do* is that it served as an 'accommodation tool' to facilitate their use. By accommodation tool is meant a strategy whereby a light verb, i.e. a verb with broad referential scope (as defined by Jespersen 1954: VI, 117–18), is inserted to create a two-verb construction in which the light verb, in this case *do*, is the element bearing the inflection and, more generally, the grammatical information, while the semantic content is mostly provided by the borrowed verb. An example is given in (11), where *dothe* is the inflected element,

---

[9] The reason why only the PPCME2 was consulted is that this process was automated, as I used the lemmatiser developed by Percilier & Trips (2020), available at http://basics-toolkit.spdns.org/app/lemmatizer/, and the part-of-speech tags present in the corpus.
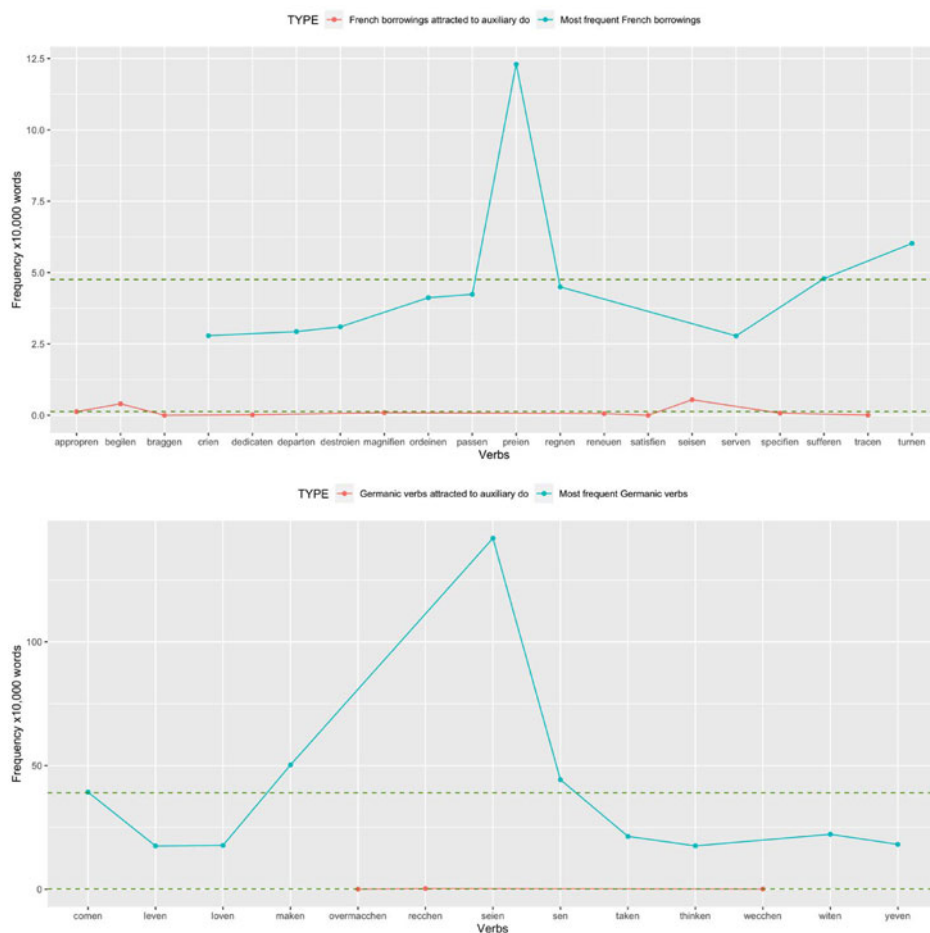
Figure 1. Normalised frequency of the verbs occurring with auxiliary *do* and of the 10 most frequent verbs borrowed from French (top plot) and of Germanic origin (bottom plot)

while the borrowing *spesyffide* (< French *specifier*) in the infinitive form is the element that provides the semantic information.

(11) Fordermore, Syr,   ye   schall understonde that I resayved a
     furthermore Sir   you   shall   understand   that I received a
     letter ffrom your mayster by mayster Bryane, the whyche I
     letter from   your master   by master   Brian   the which   I
     schall do according as your wryting **dothe** spesyffide
     shall   do according as your writing does   specify
     'Furthermore, Sir, you should understand that I received a letter from your master by master
     Brian that I should do as your writing specifies' (PCEEC: STONOR,II,90.083.1522)

The function of auxiliary *do* as an accommodation tool in Middle English is supported by typological and cognitive studies. Firstly, as Matras (2009: 175) claims, there is a

'near-consensus view' that, due to their morphological complexity, the process of borrowing verbs presents more difficulties than, for instance, nouns. This means that verbs are more likely to undergo processes of adaptation when integrated in the receiving language (Winford 2010: 173–5). Secondly, scholars have drawn attention to the fact that, from a cross-linguistic perspective, inserting a semantically light verb to facilitate the use of foreign verbal stems is a well-attested practice. Specifically, Wohlgemuth (2009: 102) claims that the introduction of a light verb is the second most common strategy behind what he calls 'direct insertion' (i.e. the replica of borrowed verbs in the recipient language without morphological or syntactic adaptation) from a typological perspective. Interestingly, it is also argued by Wohlgemuth that *do*-verbs are the most frequent type of light verbs used cross-linguistically. In addition, Jäger (2006: 160) has observed that accommodating foreign verbs is one of the main functions of *do*-periphrasis in languages across the world. The examples in (12a–c) illustrate the use of *do*-verbs to integrate foreign verbs in Bengali, Uzbek and Moroccan Arabic, respectively. In each case, the borrowed verb is left uninflected, while the grammatical content is carried by the *do*-verb (e.g. *qilmoq*, *kɔra*, *dar*).

(12) (a) Perevesti qilmoq 'to translate' (from Schlyter 2003: 162, as cited in Wohlgemuth 2009: 107)
　　　(b) Magnify kɔra 'to magnify' (from Bhattacharya 2001: 70, as cited in Wohlgemuth 2009: 107)
　　　(c) Dert-hum ontmoeten 'I met them' (from Versteegh 2010: 647)

A related perspective on cross-linguistic accommodation in situations of codeswitching is provided by Myers-Scotton & Jake (2014: 515), in which it is argued that borrowed verbs occur more frequently in non-finite forms: they bring along their semantic content, while argument and morphological integration is granted by different strategies of the receiving language, one of them being the insertion of a *do*-verb.

　　However, there are also three collexemes of Germanic origin, specifically *overmacchen*, *wecchen* and *recchen*, which are infrequently attested in the data set, as they occur only 3, 9 and 33 times in PPCME2, respectively. This suggests that auxiliary *do* was used to facilitate not only the use of infrequent foreign verbs, but of native ones as well. This implies a scenario in which highly frequent verbs were more morphologically integrated, or entrenched (on frequency and entrenchment see, among others, Langacker 1987; Bybee 2007; Schmid 2016a; Hilpert & Diessel 2017; Divjak & Caldwell-Harris 2019) within the Middle English language system and, as it appears, did not need an accommodation tool like auxiliary *do* to be used. Low-frequency items, on the other hand, are less entrenched and more difficult to retrieve mentally and, as neurolinguistic and psycholinguistic studies have shown, infrequent constructions come with a higher processing cost (Blumenthal-Dramé 2016). The possibility of inserting a semantically light verb like auxiliary *do* served to facilitate their use, as speakers could split grammatical and lexical functions, since *do* bears the grammatical load and the infinitive verb carries lexical information, reducing in turn the processing load and making the use of low-frequency verbs easier.

A further observation to be made concerns the possibility that auxiliary *do* was used as a marker of aspectuality. It has been argued at length in the literature that aspect is to a large extent compositional (see, e.g., Verkuyl 1972; Mourelatos 1978: 196–7; Brinton 1988: 26; Verkuyl 1993: 17–23; Depraetere 1995: 4; Croft 2020: 31), since semantic distinctions are attributed to a wide range of complex expressions, which range from lexical verbs to verb phrases and full sentences. A method like CA, which offers a quantitative perspective on the interaction of auxiliary *do* and the infinitive verbs with which it co-occurs, does not take into account other elements in the surrounding context and is, therefore, not well suited to determine aspectual features. This, however, can be carried out with a detailed analysis of each example, which suggests that, in some cases, auxiliary *do* was used as an aspectual marker. Firstly, there are six instances, all attested in late fifteenth-century data, in which auxiliary *do* is clearly used as a marker of habituality. In such cases, the context in which auxiliary *do* occurs indicates that the situation expressed by the infinitive verb is distributed over a given interval of time. This is illustrated in examples (13)–(14), where the presence of the temporal adverb *dayly* invites a habitual interpretation of the entire situation, as it is implied that the actions of uttering and shipping are repeated over time or occur regularly.

(13) But   in   faith   I   knowe   wele   the   iuge,   W.   Wayte   his   mawment,
     but   in   faith   I   know   well   the   judge,   W.   Wayte   his   idol,
     hise   boy   Yimmis,   with   here   hevedy   and   fumows   langage   haue
     his   boy   Jimmy   with   their   impetuous   and   irascible   language   have
     and   dayly   **do** vttyr   lewd   and   screwd   dalyauns
     and   daily   do   utter   ignorant   and   evil   chatting
     'But in faith I know well the judge, his idol W. Wayte, and his boy Jimmy, with their impetuous and irascible language daily uttering ignorant and evil chatting' (PCEEC: PASTON,II,334.425.11063)

(14) At   thys   day   I   haue   schepyt   x   sarplerys   woll   and   daily   **do**   schepe
     at   this   day   I   have   shipped   ten   bales   wool   and   daily   do   ship
     'On this day I have shipped ten bales of wool and ship do so daily' (PCEEC: CELY,25.020.401)

Secondly, there are also thirteen instances in which auxiliary *do* appears in perfective contexts, where the situation described by the infinitive verb is portrayed as complete. In these cases, auxiliary *do* serves to indicate temporal sequence in the discourse, as it marks the completion of an event with respect to a subsequent event. Look, for instance, at example (15). Here, auxiliary *do* indicates that the action of seizing the land is complete and, furthermore, that it occurs before the following event, which is to warn John of Beston. In this context, auxiliary *do* is used to mark anteriority, a function which has been associated with perfective markers in previous studies (see Givón 1982; Hopper 1982a; Bybee *et al.* 1994). Note that a similar function has been

described by Ziegeler (2004) with respect to the use of auxiliary *do* in Early Modern English data.

(15) Item, the seid   Gonnore manased and thret   John of Beston
      also   the said   Gonnore menaced and threatened John of Beston
      for he wuld   not warn here-of and he **dede** sease all
      for he would not warn hereof and he did   seize all
      his lond in Routon   and warned hym that he shuld not
      his land in Roughton and warned him that he should not
      occupy hys lyme kyll nere no lond that he had in Routon.
      occupy his lime kiln near no land that he had in Roughton
      'Also, the said Gonnore menaced and threatened John of Beston that he would not warn him hereof and he seized all his land in Roughton and warned him that he should not occupy his lime kiln near any land that he had in Roughton.' (PCEEC: PASTON,II,65.268.7816)

The picture that emerges from this discussion is that auxiliary *do* does not appear to have just one function in late Middle English, but it seems to be associated with at least two. The fact that auxiliary *do* took up different functions depending on the context in which it appeared is supported by several studies (e.g. van der Auwera 1999; Jäger 2006; Schultze-Berndt 2008) in which it has been shown that *do*-periphrasis can be used to fulfil a variety of functions both cross-linguistically and within single languages. Such functions range from grammatical (i.e. expression of negation, aspectual marker, code) to discourse, style and avoidance of complex verb paradigms (see Jäger 2006 for more details). In the present case, the results of the CA and qualitative analyses have shown that late Middle English auxiliary *do* could be inserted to facilitate the use of low-frequency verbs – particularly, but not exclusively, foreign ones – and, to a lesser extent, as a marker of aspectuality to express habituality and perfectivity. This heterogeneity is due to the fact that a lexically empty or light verb like auxiliary *do* is a rather versatile construction that can potentially take up different functions across different contexts, as has been shown, for instance, for semantically light verbs like *take* and *have* (Brugman 2001). Moreover, these functions do not clash with one another. That is, the possibility of marking habituality and perfectivity did not prevent *do* from being used to support the use of low-frequency foreign verbal stems.

The hypothesis of auxiliary *do* as a multifunctional element contrasts with the studies reviewed in section 2, as in each of them the authors suggested that *do* had only one function in Middle English. More important, however, is the fact that multifunctionality may be the answer to the question posed in section 1, i.e. why was auxiliary *do* preserved in the language if it had no semantic content and was not syntactically obligatory? That is, the semantic emptiness and syntactic optionality that characterise auxiliary *do* in Middle English that, in theory, could have pushed it out of the language system, gave it enough flexibility to be employed in a variety of contexts and take up different functions. This development is interesting since, typically,

linguistic specialisation and the creation of what are called 'functional niches' (Traugott & Trousdale 2013: 18; De Smet *et al.* 2018; Traugott 2020) are crucial in order for infrequent constructions to be preserved in the language system. Specifically, constructions can become associated with a specific functional domain and find their place in the language system, allowing them in turn to be maintained (for a recent study involving the dative alternation in English, see Zehentner 2022). In the case of auxiliary *do*, however, it seems that the opposite, that is, the ability to be used in an array of contexts, facilitated its survival before *do* became associated with the modal system and gradually developed into the operator that it is now in Present-day English. In addition, the fact that auxiliary *do* was associated with a number of functions fits in well with the view that relationships between form and function are rarely one-to-one, but are typically organised many-to-many (Van de Velde 2014).

6 Conclusion

This article has investigated the functions of auxiliary *do* in the Middle English period. Particular focus has been given to the semantic features of the infinitives that auxiliary *do* combined with in the period under investigation, which were examined by means of a collexeme analysis. The results of the statistical analysis have shown that auxiliary *do* could be used in a variety of contexts. Firstly, auxiliary *do* occurred with low-frequency verbs, of which the majority are borrowings from French. It has been suggested that auxiliary *do* functioned in this context as an accommodation tool to facilitate the use of infrequent verbs. This fits in nicely with previous typological and psycholinguistic studies which have shown that the insertion of semantically light verbs, like *do*, with borrowed and infrequent predicates helps their integration in the recipient language and their use. Secondly, although less extensively, auxiliary *do* was used in contexts that invite habitual and perfective interpretations. Habitual marker uses were characterised by the presence of temporal adverbs that indicate repetition, while perfective marker uses occurred in contexts where the situation is portrayed as complete. The fact that auxiliary *do* could appear in diverse contexts suggests that it could serve different functions, and challenges previous studies which have put forward hypotheses whereby auxiliary *do* only had a single function. Importantly, it has been argued that the multifunctionality of auxiliary *do* was a key factor in the preservation of the construction before it joined the auxiliary system and acquired the NICE properties. This is an interesting development, as infrequent constructions either decline or are assigned to a particular functional niche. In the case of auxiliary *do*, on the other hand, the opposite process, i.e. the possibility to be associated with a variety of functions, proved to be a successful strategy for the preservation of the construction in the language system. This is, however, not surprising, as language is a redundant system in which constructions entertain many-to-many relationships between form and function.

Author's address:

English Department
University of Zurich
Plattenstrasse 47
CH-8032 Zürich
Switzerland
lorenzo.moretti@es.uzh.ch

References

Anthony, Lawrence. 2020. *AntConc* (3.5.9). www.laurenceanthony.net/software. Tokyo: Waseda University.

Bhattacharya, Tanmoy. 2001. Bangla. In Jane Garry & Carl Rubino (eds.), *Facts about the world's languages: An encyclopedia of the world's major languages, past and present*, 65–71. New York: Wilson.

Blumenthal-Dramé, Alice. 2016. Entrenchment from a psycholinguistic and neurolinguistics perspective. In Schmid (ed.), 129–52.

Brinton, Laurel J. 1988. *The development of aspectual systems: Aspectualizers and post-verbal particles*. Cambridge: Cambridge University Press.

Brugman, Claudia. 2001. Light verbs and polysemy. *Language Sciences* 23, 551–78.

Budts, Sara. 2021. On periphrastic do and the modal auxiliaries: A connectionist approach to language change. PhD dissertation, Universiteit Antwerpen.

Bybee, Joan L. 2007. *Frequency of use and the organization of language*. Oxford: Oxford University Press.

Bybee, Joan L., William Pagliuca & Revere D. Perkins. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: Chicago University Press.

Croft, William. 2020. *Verbs: Aspect and causal structure*. Oxford: Oxford University Press.

Dahl, Torsten. 1956. *Linguistic studies in some Elizabethan writings: The auxiliary* do. Aarhus: Universitetsforlaget.

Dalton-Puffer, Christiane. 1996. *The French influence on Middle English morphology: A corpus-based study of derivation*. Berlin: De Gruyter Mouton.

Denison, David. 1985. The origins of periphrastic *do*: Ellegård and Visser reconsidered. In Roger Eaton, Olga Fischer, Willem F. Koopman & Frederike van der Leek (eds.), *Papers from the 4th International Conference on English Historical Linguistics*, 45–60. Amsterdam: John Benjamins.

Denison, David. 1993. *English historical syntax: Verbal constructions*. London and New York: Longman.

Depraetere, Ilse. 1995. On the necessity of distinguishing between (un)boundedness and (a)telicity. *Linguistics and Philosophy* 18, 1–19.

De Smet, Hendrik & Marlieke Shaw. 2022. Loan word accommodation biases: Markedness and finiteness. *Transactions of the Philological Society* 120(2), 201–17.

De Smet, Hendrik, Frauke D'hoedt, Lauren Fonteyn & Kristel Van Goethem. 2018. The changing functions of competing forms: Attraction and differentiation. *Cognitive Linguistics* 29(2), 197–234.

Divjak, Dagmar & Catherine Caldwell-Harris. 2019. Frequency and entrenchment. In Ewa Dąbrowska & Dagmar Divjak (eds.), *Handbook of Cognitive Linguistics*, 53–75. Berlin: De Gruyter Mouton.

Ecay, Aaron. 2015. A multi-step analysis of the evolution of English *do*-support. PhD dissertation, University of Pennsylvania.

Ellegård, Alvar. 1953. *The auxiliary* do: *The establishment and regulation of its use in English*. Gothenburg: Almqvist and Wiksell.

Elworthy, Frederick T. 1877. *An outline of the grammar of the dialect of West Somerset*. London: Trübner.

Engblom, Victor. 1938. *On the origin and early development of the auxiliary* do. Lund: CWK Gleerup.

Fischer, Olga. 1989. The origin and spread of the accusative and infinitive construction in English. *Folia Linguistica Historica* 8(1–2), 143–217.

Fischer, Olga & Wim van der Wurff. 2006. Syntax. In Richard Hogg & David Denison (eds.), *A history of the English language*, 109–98. Cambridge: Cambridge University Press.

Flach, Susanne. 2021. *Collostructions: An R implementation for the family of collostructional methods*. Package version v.0.2.0, https://sfla.ch/collostructions/

Funke, Otto. 1922. Die Fügung *ginnen* mit dem Infinitiv im Mittelenglischen. *Englische Studien* 56, 1–27.

Garrett, Andrew. 1998. On the origin of auxiliary *do*. *English Language and Linguistics* 2, 283–330.

Givón, Talmy. 1982. Tense-aspect-modality: The creole prototype and beyond. In Hopper (ed.), 135–63.

Goldberg, Adele E. 2001. Patient arguments of causative verbs can be omitted: The role of information structure in argument distribution. *Language Sciences* 23, 503–24.

Gries, Stefan Th. 2008. Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics* 13(4), 403–37.

Gries, Stefan Th. 2015. More (old and new) misunderstandings of collostructional analysis: On Schmid & Küchenhoff (2013). *Cognitive Linguistics* 26(3), 505–36.

Gries, Stefan Th. & Anatol Stefanowitsch. 2004a. Covarying collexemes in the *into*-causative. In Michel Achard & Suzanne Kemmer (eds.), *Language, culture, and mind*, 225–36. Stanford, CA: CSLI.

Gries, Stefan Th. & Anatol Stefanowitsch. 2004b. Extending collostructional analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics* (9)1, 97–129.

Hickey, Raymond (ed.). 2010. *The handbook of language contact*. Oxford: Wiley Blackwell.

Hilpert, Martin. 2008. *Germanic future constructions: A usage-based approach to language change*. Amsterdam: John Benjamins.

Hilpert, Martin. 2013. *Constructional change in English: Developments in allomorphy, word formation, and syntax*. Cambridge: Cambridge University Press.

Hilpert, Martin & Holger Diessel. 2017. Entrenchment in Construction Grammar. In Schmid (ed.), 57–74.

Hopper, Paul. 1982a. Aspect between discourse and grammar: An introductory essay for the volume. In Hopper (ed.), 3–18.

Hopper, Paul (ed.). 1982b. *Tense-aspect: Between semantics and pragmatics*. Amsterdam: John Benjamins

Huddleston, Rodney & Geoffrey K. Pullum *et al.* 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

Ihalainen, Ossi. 1976. Periphrastic 'do' in affirmative sentences in the dialect of East Somerset. *Neuphilologische Mitteilungen* 77, 608–22.

Jäger, Andreas. 2006. *Typology of periphrastic 'do'-constructions*. Bochum: Universitätsverlag Dr. N. Brockmeyer.

Jespersen, Otto. 1954. *A modern English grammar on historical principles*. London: Allen & Unwin.

Kroch, Anthony, Ann Taylor & Beatrice Santorini. 2000–. *The Penn–Helsinki Parsed Corpus of Middle English* (PPCME2). Department of Linguistics, University of Pennsylvania. www.ling. upenn.edu/ppche/ppche-release-2016/PPCME2-RELEASE-4 (accessed 9 August 2008).

Kytö, Merja. 1996. *Manual to the diachronic part of the Helsinki Corpus of English texts: Coding conventions and lists of source texts*. Helsinki: University of Helsinki.

Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar*, vol. 1: *Theoretical prerequisites*. Stanford, CA: Stanford University Press.

Langacker, Ronald W. 1988. An overview of Cognitive Grammar. In Brygida Rudzka-Ostyn (ed.), *Topics in Cognitive Linguistics*, 3–48. Amsterdam: John Benjamins.

Matras, Yaron. 2009. *Language contact*. Cambridge: Cambridge University Press.

Moretti, Lorenzo. 2022. A multivariate analysis of causative *do* and causative *make* in Middle English. *Linguistics Vanguard* 8(1), 165–76.

Moretti, Lorenzo. 2023. The functions of auxiliary *do* in Middle English poetry: A quantitative study. *Journal of English Linguistics* 51(1), 3–29.

Mourelatos, Alexander P. D. 1978. Events, processes, and states. *Linguistics and Philosophy* 2(3), 415–34.

Mustanoja, Tauno F. 1960. *A Middle English syntax*, part 1: *Parts of speech*. Helsinki: Société Néophilologique.

Myers-Scotton, Carol & Janice L. Jake. 2014. Nonfinite verbs and negotiating bilingualism in codeswitching: Implications for a language production model. *Bilingualism: Language and Cognition* 17(3), 511–25.

Nurmi, Arja. 1999. *A social history of periphrastic DO* (Mémoires de la Société Néophilologique de Helsinki 56). Helsinki: Société Néophilologique.

Parsed Corpus of Early English Correspondence, PCEEC tagged version. 2006. Annotated by Ann Taylor, Arja Nurmi, Anthony Warner, Susan Pintzuk & Terttu Nevalainen. Compiled by the CEEC Project Team. York: University of York and Helsinki: University of Helsinki. Distributed through the Oxford Text Archive. www-users.york.ac.uk/~lang22/PCEEC-manual/ corpus_description/index.htm (accessed 8 August 2008).

Percillier, Michael & Carola Trips. 2020. Lemmatising verbs in Middle English corpora: The benefit of enriching the *Penn–Helsinki Parsed Corpus of Middle English 2 (PPCME2)*, the *Parsed Corpus of Middle English Poetry (PCMEP)* and *A Parsed Linguistic Atlas of Early Middle English (PLAEME)*. In Nicoletta Calzolari *et al.* (eds.), *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*, 7170–8. Marseille: European Language Resources Association. https://aclanthology.org/2020.lrec-1.886.pdf

Poussa, Patricia. 1990. A contact-universals origin for periphrastic do, with special consideration of OE–Celtic contact. In Sylvia M. Adamson, Vivien A. Law, Nigel Vincent & Susan Wright (eds.), *Papers from the 5th International Conference on English Historical Linguistics*, 407–34. Amsterdam: John Benjamins.

R Development Core Team. 2017. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. www.r-project.org (October 2020).

Rissanen, Matti, Merja Kytö, Leena Kahlas-Tarkka, Matti Kilpiö, Saara Nevanlinna, Irma Taavitsainen, Terttu Nevalainen & Helena Raumolin-Brunberg. 1991. *The Helsinki Corpus of English Texts*. Department of Modern Languages, University of Helsinki. https:// helsinkicorpus.arts.gla.ac.uk/display.py?what=index (accessed 10 August 2022).

Royster, James Finch. 1922. Old English causative verbs. *Studies in Philology* 19(3), 328–56.

Schlyter, Birgit N. 2003 Sociolinguistic changes in Central Asian societies. In Jacques Maurais & Michael A. Morris (eds.), *Languages in a globalising world*, 157–87. Cambridge: Cambridge University Press.

Schmid, Hans-Jörg. 2016a. A framework for understanding linguistics entrenchment and its psychological foundations. In Schmid (ed.), 9–38.

Schmid, Hans-Jörg (ed.). 2016b. *Entrenchment and the psychology of language learning.* Berlin: De Gruyter Mouton.

Schmid, Hans-Jörg & Helmut Küchenhoff. 2013. Collostructional analysis and other ways of measuring lexicogrammatical attraction: Theoretical premises, practical problems and cognitive underpinnings. *Cognitive Linguistics* 24, 531–77.

Schultze-Berndt, Eva. 2008. What do 'do' verbs do? The semantic diversity of generalised action verbs. In Elisabeth Verhoeven, Stavros Skopeteas, Yong-Min Shin, Yoko Nishina & Johannes Helmbrecht (eds.), *Studies on grammaticalization*, 185–208. Berlin: De Mouton Gruyter.

Smyser, H. M. 1967. Chaucer's use of *gin* and *do*. *Speculum* 42, 68–83.

Stefanowitsch, Anatol & Stefan Th. Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2), 209–43.

Stefanowitsch, Anatol & Stefan Th. Gries. 2005. Covarying collexemes. *Corpus Linguistics and Linguistic Theory* 1(1), 1–43.

Timofeeva, Olga. 2011. Infinitival complements with the verb (*ge*)*don* in Old English: Latin influence revisited. *Leeds Studies in English* 42, 93–108.

Traugott, Elizabeth C. 2020. The intertwining of differentiation and attraction as exemplified by the history of recipient transfer and benefactive alternations. *Cognitive Linguistics* 31(4), 549–78.

Traugott, Elizabeth C. & Graeme Trousdale. 2013. *Constructionalization and constructional changes*. Oxford: Oxford University Press.

Van de Velde, Freek. 2014. Degeneracy: The maintenance of constructional networks. In Ronny Boogaart, Timothy Colleman & Gijsbert Rutten (eds.), *Extending the scope of construction grammar*, 141–79. Berlin: De Gruyter.

van der Auwera, Johan. 1999. Periphrastic 'do': typological prolegomena. In G. A. J. Tops (ed.), *Thinking English grammar: To honour Xavier Dekeyser, professor emeritus*, 457–70. Leuven: Peeters.

Vendler, Zeno. 1967. *Linguistics in philosophy*. Ithaca, NY: Cornell University Press.

Verkuyl, Henk J. 1972. *On the compositional nature of the aspects*. Dordrecht: D. Reidel.

Verkuyl, Henk J. 1993. *A theory of aspectuality: The interaction between temporal and atemporal structure*. Cambridge: Cambridge University Press.

Versteegh, Kees. 2010. Contact and the development of Arabic. In Hickey (ed.), 634–51.

Visser, F. Th. 1963–73. *An historical syntax of the English language*, vol. 3. Leiden: Brill Archive.

Winford, Donald. 2010. Contact and borrowing. In Hickey (ed.), 170–87.

Wohlgemuth, Jan. 2009. *A typology of verbal borrowings*. Berlin and New York. Mouton de Gruyter.

Zehentner, Eva. 2022. Competing constructions construct complementary niches: A diachronic view on the English dative alternation. *Language Dynamics and Change* 13(1), 34–73.

Ziegeler, Debra. 2004. Reanalysis in the history of *do*: A view from construction grammar. *Cognitive Linguistics* 15(4), 529–74.