

The driving forces behind verbal cluster order variation from Early to Late Modern Dutch

Julie Van Ongeval (Vrije Universiteit Brussel, FWO Vlaanderen)

Rik Vosters (Vrije Universiteit Brussel)

AUTEURSMANUSCRIPT NA PEER REVIEW

Abstract: This paper investigates the effect of both structural and sociolinguistic variables on the development of the Dutch verbal cluster order variation. More precisely, the paper evaluates which specific set of these variables has the largest effect on the verb order variation from the sixteenth to the nineteenth century. To this end, a corpus linguistic study with a random forest analysis was conducted, using the Historical Corpus of Dutch (1550-1850), which contains administrative texts, pamphlets, and ego-documents from four different regions in the Northern (Holland and Zeeland) and Southern Netherlands (Brabant and Flanders). The results of this corpus study show that the sociolinguistic variables (viz. period, year, time, and region) are the most important predictors of the Dutch verb order variation, whereas the impact of the structural variables is overall limited (viz. the type of auxiliary verb, the morphological structure of the past participle, the distance between the conjunction and the cluster, and the absence or presence of the final clause position). Thus, the sociolinguistic variables seem to be the driving forces behind this syntactic change.

Keywords: history of Dutch, verb order variation, sociolinguistic and structural variables, syntax

1. Introduction

One of the most well-studied syntactic variables by Dutch linguists is the word order variation in verb clusters. In Dutch, the order for finite (auxiliary (AUX)) and non-finite verbs (infinitive (INF) or participle (PART)) in subordinate clauses is not fixed, resulting in free syntactic variation between a PART/INF+AUX or a AUX+PART/INF order (De Sutter, Speelman & Geeraerts, 2005, p. 97; Coussé & De Sutter, 2012, p. 73). Since the colored dialect maps of Pauwels (1953), the PART/INF+AUX variant (as in example 1¹) has traditionally been termed

¹ All the examples in the paper are taken from the Historical Corpus of Dutch (Van de Voorde, 2022; Van de Voorde et al., 2023).

the green order, and the AUX+PART/INF variant (as in example 2) the red order (Coussé & De Sutter, 2012, p. 74):

(1) Onnoodig ware het te zeggen dat de kerk op het luisterrykste
Unnecessary was it to say that the church at the most glorious
versierd.PART was.AUX.

decorated.PART was.AUX.

‘It went without saying that the church had been decorated in the most glorious manner.’

(EGO-1850-VL-3-RAK_Clementia_Hiers_Kortrijk_1849_1871)

(2) Het scheen dat ons vlaenderlant was.AUX verlaeten.PART.

It seemed that our Flanders was.AUX abandoned.PART.

‘It seemed like our Flemish homeland was abandoned.’

(EGO-1650-VL-3-RAK_Jacques_Inbona_Brugge_1645_1666)

Research on this verb order variation in Dutch has mainly focused on the impact of structural (viz. morphosyntactic, prosodic, discursive, and semantic) and sociolinguistic (viz. time, region, and text genre) variables, resulting in a clear overview of the most important factors driving this syntactic variation from a dialectological, usage-based, as well as formal perspective (see for instance the overview of Bennis & Coussé, 2012, pp. 1-4). Most studies, however, have concentrated on either structural or sociolinguistic variables, without considering the impact of both these two types of factors (De Sutter, 2007, p. 328). In addition, verb order variation has mainly been studied from a synchronic perspective. The few diachronic studies that are available (De Meersman, 1990; BurrIDGE, 1993; Coussé, 2008; Coupé, 2015) focus almost exclusively on literary and official or legal texts. Coussé (2008) does investigate the verb order variation in a wider range of narrative texts, also including two types of ego-documents (i.e. diaries and travel accounts), but only does so for Holland.

To fill these two research gaps, the goals of this contribution are twofold. First, we will investigate the impact of both sociolinguistic (viz. time, region, text genre) *and* structural (viz. the morphological structure of the past participle, the type of auxiliary verb, the distance between the subordinate conjunction and the verbal cluster, and the presence or absence of the final clause position) variables. Given this broad range of variables under investigation, we will make use of a random forest analysis, a machine learning technique which will help us determine which variables are most influential in predicting verb order variation. Secondly, we will examine the verbal cluster order variation in a diachronic multi-genre corpus, covering

administrative texts, pamphlets, and ego-documents from four different centuries (sixteenth to the nineteenth century) as well as from four different regions in the Northern (Holland and Zeeland) and Southern Netherlands (Brabant and Flanders).

To maximize comparability with most of the previous research and for reasons of practical feasibility, we decided to consider only two-verb clusters, consisting of one of the auxiliary verbs *hebben* ‘hebben’, *worden* ‘to become’ and *zijn* ‘to be’, alongside a past participle (in either PART-AUX or AUX-PART order). In other words, clusters containing an infinitive or clusters made up of more than two verbs were excluded.

The paper begins by providing a brief overview of the studies focusing on the role of the intra- and extra-linguistic factors governing the Dutch verb order variation (section 2). The next section (section 3) describes the corpus and methodology used. Finally, the results will be presented (section 4), followed by a conclusion of how sociolinguistic factors seem to be the driving motor of this diachronic change (section 5).

2. Word order variation in Dutch verb clusters

According to Dutch grammars and language advice books, the choice between the AUX-PART or the PART-AUX order in verb clusters is not guided by any fixed, absolute rules. The leading Dutch grammar *Algemene Nederlandse Spraakkunst* (1997, 2021), for instance, states that “the placement of the past and passive participle as an independent verb is in principle free, when used verbally” (my translation). However, various dialectological, usage-based, formal, and psycholinguistic studies have shown that this syntactic choice is driven by both extra- and intra-linguistic variation and is thus not entirely free (Arfs, 2007, pp. 225-226; Bennis & Coussé, 2012, pp. 1-4). The word order variation in Dutch verb clusters would not only be influenced by time, region, and register, but also by morpho-syntactic, semantic, prosodic, and discursive variables (Arfs, 2007, pp. 225-226; Coussé & De Sutter, 2012; De Sutter, 2007, p. 305), which we will discuss below.

2.1 Extra-linguistic variation

The diachronic corpus studies of Burridge (1993) (14th to 16th century), Coussé (2008) (13th to 20th century), Coupé (2015) (14th to 16th century), and De Meersman (1990) (13th to 16th century) have demonstrated that the choice for the AUX-PART or PART-AUX order varies greatly across time. While there was still no clear usage preference for any of the two linguistic variants in the thirteenth century, the use of PART-AUX gradually increased from the fourteenth century onwards, subsequently becoming the dominant variant in all regions of the

fifteenth-century Dutch language area (Coussé, Arfs & De Sutter, 2008, pp. 40-41; Coupé, 2015, pp. 95-96). However, from the second half of the sixteenth century onwards, the AUX-PART order was reintroduced in the centers of Brabant and Holland and, from there, was adopted in the peripheries of Zeeland and Flanders through a so-called diffusion process, giving renewed competition to the PART-AUX variant (Coussé, 2008, pp. 171-172; Coussé & De Sutter, 2012, p. 93; Coupé, 2015, p. 76).

After its late sixteenth-century revival, the AUX-PART variant continued to gain ground at the expense of PART-AUX. As such, the AUX-PART order is again used about as frequently as the PART-AUX order today (Coussé, Arfs & De Sutter, 2008, p. 41; Coussé & De Sutter, 2012, p. 94). Furthermore, the apparent-time study of Olthof et al. (2017, pp. 53-55) has shown that in contemporary Dutch the youngest groups of speakers use increasingly more AUX-PART clusters than older groups of speakers, indicating that the change towards more AUX-PART is still in progress. The modern verb order variation can therefore be described as an example of ‘synchronic layering’, since it is a snapshot in a language change that already started in the second half of the sixteenth century and is still ongoing today (Coussé & De Sutter, 2012, p. 74, p. 94).

In addition to these temporal differences, the Dutch verb order variation also differs by region, both from a macroscopic and a more microscopic perspective (De Sutter, Speelman & Geeraerts, 2005, pp. 100-101). From a macroscopic view, the PART-AUX order is observed more in the Southern than in the Northern Netherlands, which, according to De Sutter, Speelman & Geeraerts (2005, pp. 100-101, p. 112), could possibly be linked to the two regions’ separate historical standardization processes. In the seventeenth century, the AUX-PART order was incorporated into the developing Northern standard language (Coussé, 2008, pp. 182-184). From the nineteenth century onwards, the Southern Netherlands began to adapt more to the Northern standard, and thus to the AUX-PART order, which is why the current macroscopic regional variation is expected to further decrease in the future. However, this explanation based on the different Northern and Southern Dutch standardization processes remains a hypothesis that has yet to be tested empirically (De Sutter, Speelman & Geeraerts, 2005, p. 112, 125; for a recent investigation of the diffusion of orthographical change in the light of pluricentricity and standardization in the North versus the South, however, see Rutten, Van de Voorde & Vosters 2023).

Despite the clear regional difference between the North and the South, the two parts of the language area cannot be considered as homogeneous from a more microscopic perspective (De Sutter, Speelman & Geeraerts, 2005, p. 112). As for the Southern Netherlands, the AUX-

PART variant is more commonly used in Brabant than in Flanders (Haeseryn, 1990, p. 153; De Sutter, Speelman & Geeraerts, 2005, p. 112). Regarding the Northern Netherlands, the AUX-PART order is generally more preferred in Limburg, South-Holland, Utrecht, and Gelderland, whereas the PART-AUX order appears more in the Northern provinces of Friesland, Groningen and Drenthe (Pauwels, 1953, in Coussé, Arfs & De Sutter, 2008, p. 32). Nonetheless, De Sutter, Speelman & Geeraerts (2005, p. 112, 124) noted no statistically significant regional differences within the Northern Netherlands, except for a stronger preference for AUX-PART in Zeeland.

The final well studied type of extra-linguistic variation concerns stylistic differences. The PART-AUX order is considered as the most entrenched and informal variant, whereas AUX-PART is characterized as the more formal, stylistically marked variant, already from its introduction in the upcoming Northern standard onwards. As such, PART-AUX occurs more in informal, spoken language, while the AUX-PART order is more typical of written, formal language (Bloem, Versloot & Weerman, 2017, pp. 116-117; Coussé, 2008, pp. 182-184; De Sutter, Speelman & Geeraerts, 2005, pp. 116-117, 124-125). However, despite its association with formal standard language use, the AUX-PART variant was not prescribed as the most ‘correct’ form in metalinguistic works (e.g. grammars, dictionaries, spelling guides) until the twentieth century. Only in 1940 did the first explicit normative prescription appear in Charivarius’ (1998, p. 48) work:

“As a rule, place the auxiliary before the verb. So don’t write: *Het is nodig dat dit toegestaan wordt* [It is necessary that this is allowed; PART - AUX], but: *Het is nodig dat dit wordt toegestaan* [AUX - PART]. In case of doubt, let your sense of sound and meter decide (my translation and italics).”

Before this prescriptive injunction, Dutch verb order variation was treated without expressing a preference for one of the two variants. Some grammarians, such as Den Hertog (1892) and Brill (1852), even prescribed the PART-AUX order in specific syntactic contexts, such as in relative clauses (Van der Horst, 2002, pp. 547-549). Because of this lack of explicit normative prescriptions, Van der Horst (2002, p. 550) argues that the choice of the AUX-PART order as a formal and stylistically marked variant dates only from the second half of the twentieth century.

2.3 Intra-linguistic variation

Besides the temporal, regional, and stylistic differences described above, Dutch verb order variation is also governed by different types of intra-linguistic variables, viz. morpho-syntactic, prosodic, semantic, and discursive factors. In this contribution, our focus lies on four of the most well-studied morphosyntactic variables, namely the morphological structure of the past participle, the distance between the conjunction and the verb cluster, the presence or absence of the final clause position, and the type of auxiliary verb (De Sutter, 2007, pp. 305-306; Bloem, 2021, pp. 21-23; Bloem, Versloot & Weerman, 2017, p. 1980).

First, verb order variation is influenced by the morphological structure of the past participle in the verb cluster. If the participle is a separable particle verb (e.g. *aangenomen* ‘taken up’), AUX-PART is preferred. When the participle is not separable (e.g. *genomen* ‘taken’), PART-AUX is most often used (Bloem, Versloot & Weerman, 2017, p. 1979; De Sutter, 2007, p. 307). This is illustrated by the following two examples from the Historical Corpus of Dutch (Van de Voorde, 2022; Van de Voorde et al., 2023):

- (1) [...] waer door het veelyts gebeurt dat groot ende gewichtige resoltien
 [...] wherefore it often happens that large and important resolutions
 ter generaliteyt genomen.PART werden.AUX.
 at generalitat taken.PART were.AUX.
 ‘Which is why it often happens that large and important resolutions were taken
 by the States General’
 (ADM-1650-ZE-1-GAG_Notulen_Goes_1651_1653)
- (2) Het gerugt verspreid zich dat de Belgen de 18 Preliminairen
 The rumor spreads REFL. that the Belgians the 18 preliminaries
 hebben.AUX aangenomen.PART.
 have.AUX accepted.PART.
 ‘The rumor is spreading that the Belgians have accepted the 18 preliminary articles.’
 (EGO-1850-ZE-2-ZA_Elias_Tak_Middelburg_1831)

Second, the choice between the two variants also depends on the distance between the conjunction introducing the clause on the one hand and the verb cluster on the other hand: the more words there are in between the conjunction and the verb cluster, the higher the probability of AUX-PART instead of PART-AUX (Arfs, 2007, pp. 235-237; De Sutter, 2007, p. 308):

- (3) Gelooft gy dat ik dat afgezien.PART heb.AUX zonder afschrik?

Believe you that I that suffered.PART have.AUX without fear?

‘Do you believe that I suffered that without fear?’

(EGO-1850-VL-4-UBG_Francois_Auguste_Gevaert_Zingem_1850_1851)

- (4) [...] doch niet minder noodzakelijk komt het ons voor, dat, in the
[...] but niet less necessary comes it us for, that, in the
8^{ste} lid het beheer der kerkelijke goederen onder
8th section the management of the ecclesiastical goods under
het toezicht van den Grooten Kerkeraad worde.AUX geplaatst.PART.
the supervision of the Greater church council is.AUX placed.PART.
‘it appears to us no less necessary that, in the 8th section, the management of the
ecclesiastical goods is placed under the supervision of the Greater Church Council.’
(PAM-1850-ZE-2-BG_Pflt4625_Middelburg_1849b)

Third, verb order variation is determined by the presence or absence of the final clause position. When the verbal cluster itself occupies the final position in the subordinate clause, Dutch language users most frequently choose the PART-AUX order. In clauses where the verbal cluster is followed by other, non-verbal elements, the AUX-PART order prevails (De Sutter, 2007, pp. 307-308):

- (5) [...] terwijl het mij doelmatig voorkomt, dat de rapporten onmiddellijk
[...] while it me effective seems, that the reports immediately
aan den Kommandant gericht.PART worden.AUX.
to the commander addressed.PART are.AUX.
‘[...] while it seems to me expedient, that the reports be addressed immediately to the
commander.’
(PAM-1850-HO-4-KB_Pflt29152_Amsterdam_1850c)

- (6) Ick ben van seer goede handt geïnformeert, dat de continuatie van
I am of very good hand informed, that the continuation of
den Oorlogh in Vranckrijck is.AUX gheresolveert.PART noch voor vijf
the war in France is.PART resolved.PART still for five
Iaren.
years.
‘I have been very reliably informed that the continuation of the war in France has been
resolved for another five years.’

Finally, the type of auxiliary verb (viz. *worden* ‘to become’, *hebben* ‘to have’, *zijn* ‘to be’) also influences the choice for the PART-AUX or AUX-PART variant. According to De Sutter (2005, p. 256), the PART-AUX order occurs more often in combination with auxiliary ‘zijn’ than with ‘worden’ and ‘hebben’ respectively, whereas the AUX-PART order is more used in combination with ‘hebben’ than with ‘worden’ and ‘zijn’ respectively. This finding is in line with the more recent studies of Olthof et al. (2017) and Bloem (2021), who also found that AUX-PART occurs more often in clusters with ‘hebben’ than in those with ‘worden’ and ‘zijn’ respectively.

Bloem (2021) and De Sutter (2007) have shown that these four morphosyntactic variables all have a significant effect on the verb order variation in present-day Dutch. Thus, the variables all “have a unique effect on the word order choice that is not reducible to the effect of any of the other variables” (De Sutter, 2007, p. 318). Yet, they do not all exert an equally strong influence. According to De Sutter (2007, p. 319), the type of auxiliary verb and the morphological structure of the participle have a stronger effect than the absence or presence of the final clause position or the distance between the conjunction and the verbal cluster. Bloem (2021, p. 24) found similar relative effects as De Sutter (2007) for both the morphological structure of the participle and the distance between the conjunction and the verbal cluster, but not for the two other variables. For the type of auxiliary verb, he observed a smaller effect, whereas he observed a much stronger and reverse effect for the final clause position, although this may be explained by a different operationalization of these two variables (Bloem, 2021, pp. 24-26).

3. Empirical study

As discussed in the previous section, the existing literature clearly identifies a significant effect of a range of extra- and intra-linguistic variables governing the Dutch verbal cluster order variation. However, as most studies have treated these two types of variables separately, it remains largely undiscussed which specific set of these intra- and extra-linguistic variables has the largest impact on the verb order variation. Therefore, this study aims to investigate the combined effect of these two types of factors, from a diachronic perspective.

3.1 Historical Corpus of Dutch

To achieve the above-mentioned research objective, we conducted a corpus linguistic study using the Historical Corpus of Dutch (HCD). The Historical Corpus of Dutch (approximately 460,000 words) was compiled as part of the PhD project of Iris van de Voorde at Vrije Universiteit Brussel (VUB) and Leiden University. It is centered around three extra-linguistic dimensions, namely time, region, and text genre (Van de Voorde et al., 2023; Van de Voorde, 2022, p. 57).

First, the corpus covers the period from the sixteenth to the nineteenth century (ca. 1550-1850). These four centuries are all considered important phases in the history of the Dutch language (Van de Voorde, 2022, p. 61). From the Fall of Antwerp in 1585 and thus from the political division between the Northern and Southern Netherlands onwards, the socio-economic and linguistic center of gravity shifted from Brabant in the South to Holland in the North. As such, in the seventeenth century, the first Dutch standard language norms were codified and spread in and around the Northern provinces of Holland, often referred to as the “first phase in the Dutch standardization process” (Janssens & Marynissen, 2011, p. 137; Van der Wal & Van Bree, 2008, pp. 218-220). From the second half of the eighteenth century onwards, more and more meta-linguistic works also appeared in the Southern Netherlands, manifesting a distinct Southern normative tradition and identity (Vosters et al., 2012, pp. 139-140). However, from the nineteenth century onwards, the Southern Netherlands became more strongly oriented towards the Northern standard language model, resulting into a gradual convergence of Southern and Northern Dutch (Janssens & Marynissen, 2011, pp. 151-152; Willemyns, 2003, pp. 102-103).

Secondly, the Historical Corpus of Dutch contains texts from four different regions in the Northern and Southern Netherlands, namely from the most important linguistic centers Holland (North) and Brabant (South), and from more peripheral areas Zeeland (North) and Flanders (South) (Van de Voorde, 2022, p. 56). This distribution allows us to take North-South dynamics into account as well as to investigate to which extent this syntactic change occurred first in the more central or more peripheral provinces of the language area (Van de Voorde, 2022, p. 62).

Thirdly, the corpus consists of three different text genres, not only of hand-written administrative texts (e.g. registers of the bench of aldermen, minutes of the municipal council, and legislation for guilds and industries), but also of printed pamphlets (e.g. tidings, ordinances, and placards) and handwritten ego-documents (viz. diaries, travelogues, and (family) chronicles) (Van de Voorde, 2022, pp. 59-60, p. 66, p. 75, p. 81). The administrative texts and the pamphlets are more formal, were intended for a broader public, and were written by

professional scribes, whereas the ego-documents are more informal and incidental (Van de Voorde, 2022, p. 86). As such, this multi-genre corpus design enables us to compare more ‘literate’, formal, and standardized language use reflecting more top-down tendencies with more informal, orally colored, and everyday language use reflecting more bottom-up tendencies (Krogull, 2018, p. 50).

3.2 Methodology

Within the Historical Corpus of Dutch, we focused on two-verb clusters in subordinate *that*-clauses, consisting of one past participle and one of the auxiliary verbs *hebben* ‘to have’, *worden* ‘to become’ or *zijn* ‘to be’. Using a regular expression to search for all spelling variants of the subordinate conjunction *dat* ‘that’ (“\b(d|w){1,2}at{1,2}\b”) by means of the concordance tool *Antconc* (version 3.5.9), we first extracted all possible subordinate clauses introduced by *dat* (‘that’) from the corpus, and then filtered through them to check if one of the three above mentioned auxiliary verbs (including all conjugated forms and their spelling variants – see appendix 7.1 for a full list) appeared in the context to the right of the conjunction. After manual verification of each occurrence, we were able to extract all subordinate *that*-clauses containing one of the auxiliaries *hebben*, *worden* or *zijn* from the corpus.

The automated search yielded a total of 3,449 possible observations, and after manually removing all false positive results, including three-verb clusters (e.g. *soude afgelopen hebben* ‘would have ended’) or two-verb clusters with infinitives (e.g. *soude loopen* ‘would run’), 693 valid observations of the dependent variable remained. These observations were then annotated for the four independent intra-linguistic variables, namely for the type of auxiliary verb, the absence or presence of the final clause position, the morphological structure of the participle as well as the distance between the conjunction and the verbal cluster. More specifically, for each relevant verbal cluster, we determined manually what type of auxiliary verb (*hebben*, *worden* or *zijn*) it contained, whether the verbal cluster occupied the final clause position, and whether the participle of the cluster was a separable particle verb (e.g. *toe-gekomen v. gekomen* ‘arrived’). The distance between the conjunction ‘that’ and the participle was automatically calculated (in number of syllables) using a custom-written script in R. Subsequently, these annotated data were merged with the available extra-linguistic metadata for the corpus, and specifically the variables time period (viz. sixteenth, seventeenth, eighteenth, and nineteenth century), region (viz. Brabant, Holland, Flanders, and Zeeland) and text genre (administrative texts, pamphlets, and ego-documents).

Our analysis, which was carried out in R (R Core Team 2023), is based on a thorough visual and numeric exploration of the results using descriptive statistics, going over the different variables mentioned, as well as several patterns of interactions between these variables. In addition, we also ran a random forest analysis using the *randomForest* package (Liaw & Wiener, 2002), fitting a model which included the independent variables auxiliary verb, distance to the participle, final clause position, separable particle verb status, region, genre, century, as well as the more detailed numeric variable year of writing. We trained the model on a random partition of 70% of the original data, using the remaining 30% for model evaluation. This yielded an overall classification accuracy of 72% and an Area Under the Curve score (AUC) of 0.80, suggesting relatively strong discriminative power. For each variable in the random forest, we calculated the Mean Decrease in Accuracy (MDA) as an indication of variable importance, consisting of the “normalized difference of the classification accuracy for the out-of-bag data when the data for that variable is included as observed, and the classification accuracy for the out-of-bag data when the values of the variable in the out-of-bag data have been randomly permuted” (Cutler et al., 2007, p. 2786).

4. Results

4.1 Impact of structural and sociolinguistic variables

The above-described random forest analysis yielded a variable importance plot of all structural and sociolinguistic variables (see Figure 1). In descending order, this plot displays which of the variables are most important in terms of classification accuracy. Thus, the higher the MDA score of a variable in the plot, the more important it is in predicting the use of AUX-PART versus PART-AUX. MDA scores around zero – including negative scores – are typical for superfluous predictors, as is the case here for the choice of auxiliary verb and the final clause position (Levshina, 2015, pp. 291-299).

Figure 1 shows that the sociolinguistic variables (viz. genre, region, year, and period) by far have the largest impact on the verbal cluster order variation, with genre being the most important predictor, and period of writing (classified per century) scoring lower than its more fine-grained equivalent year of writing. The structural variables (viz. type of auxiliary verb, absence or presence of the final clause position, morphological structure of the past participle), which previous studies claimed all had a significant impact, play overall a less important role. The only structural variable that exerts a relatively strong influence is the distance between the conjunction and the verbal cluster, even though both Bloem (2021) and De Sutter (2007) found it to be the least significant of the four morphosyntactic variables. Do note, however, that we

used a more fine-grained operationalization of distance as syllable length, while this was counted in number of words in previous studies. In the present study, the least important predictors are the type of auxiliary verb and the final clause position, which play no role of importance in this syntactic change.

The fact that the intra-linguistic variables generally appear to be less important than suggested by previous research may be related to the diachronic nature of our study, as studies such as Bloem (2021) and De Sutter (2007) only examined their role in explaining verb cluster variation in present-day Dutch. It is to be expected that morphosyntactic variables which today are significant factors driving this type of variation, have not had a stable effect over time (Olthof et al., 2017, p. 54), especially over the large time scale covered in the present study.

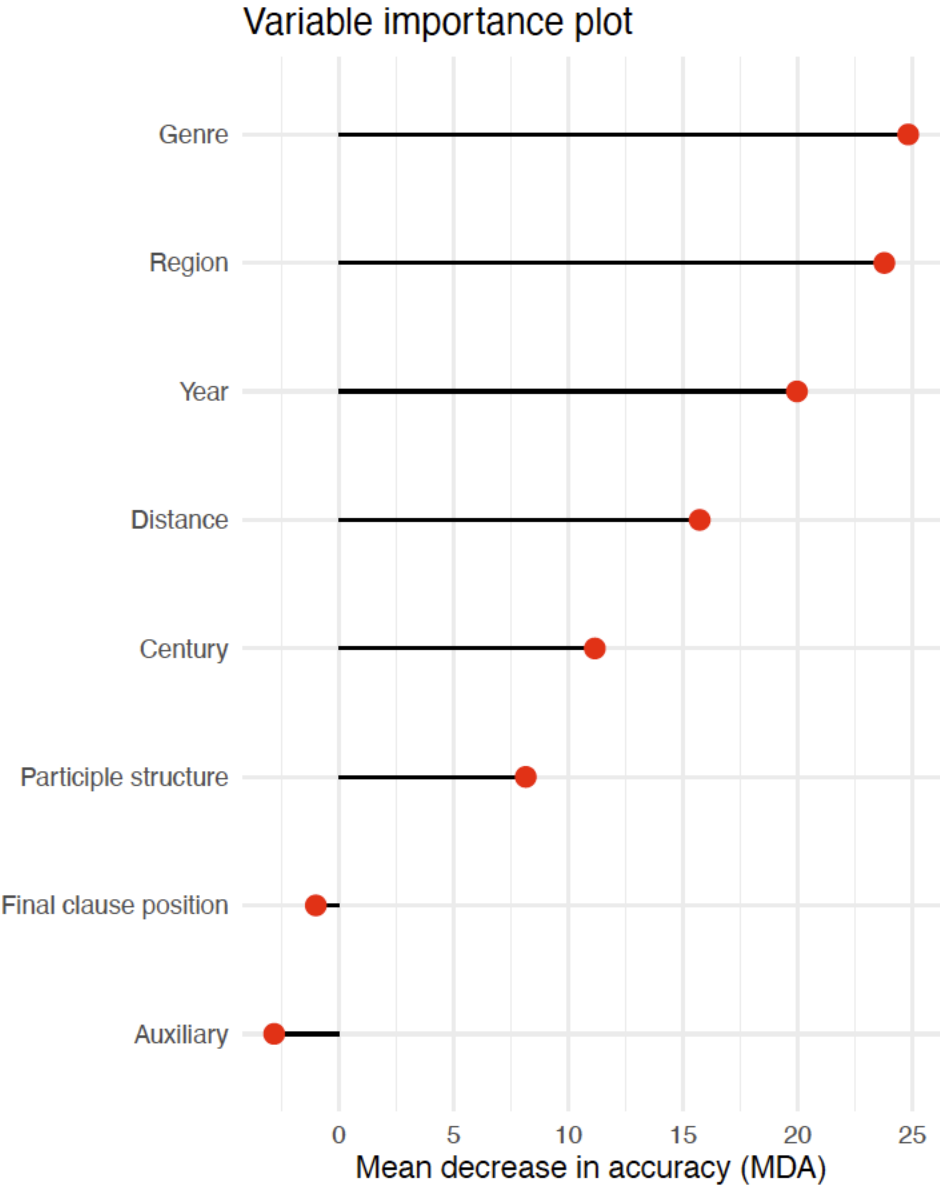


Figure 1: Impact of the extra- and intra-linguistic variables on the Dutch verbal cluster order variation

4.2 Extra-linguistic factors as the driving forces behind the Dutch word order variation

The random forest model in section 4.1 revealed that the sociolinguistic variables under investigation are the most important drivers of the Dutch verb order variation, whereas the impact of the structural variables is overall more limited. In this final empirical section, we will delve deeper into our results for this subset of variables, in order to describe in greater detail the direction of the extra-linguistics variables' impact, not only in isolation but also in interaction with each other. As these interaction patterns in combination with the relatively sparse data do not allow for a more elaborate statistical analysis – several regression models ended up either not converging or returning highly inflated confidence intervals –, we will limit ourselves to an exploration of the relative frequencies of use of both word order variants.

Figure 2 shows how the regional differences vary across the different centuries. From its reintroduction in the sixteenth century, the AUX-PART order is more prevalent in Brabant than in the other provinces, suggesting that the Southern center was leading in this syntactic change. This observation is also in line with the work of Coussé & De Sutter (2012). From the seventeenth century onwards, however, the use of the AUX-PART variant stagnates in Brabant, and strongly increases in the other provinces. After this period, a clear North-South-divide emerges with the North taking the lead, more or less in line with the traditional model of pluricentric development where Northern varieties take the lead in the standardization process (cf. section 2). From the seventeenth century onwards, the AUX-PART order became increasingly associated with the upcoming Northern Dutch standard, and only the Southern region of Brabant shows gradual convergence, while Flanders even moves away from AUX-PART, back towards the original PART-AUX order.

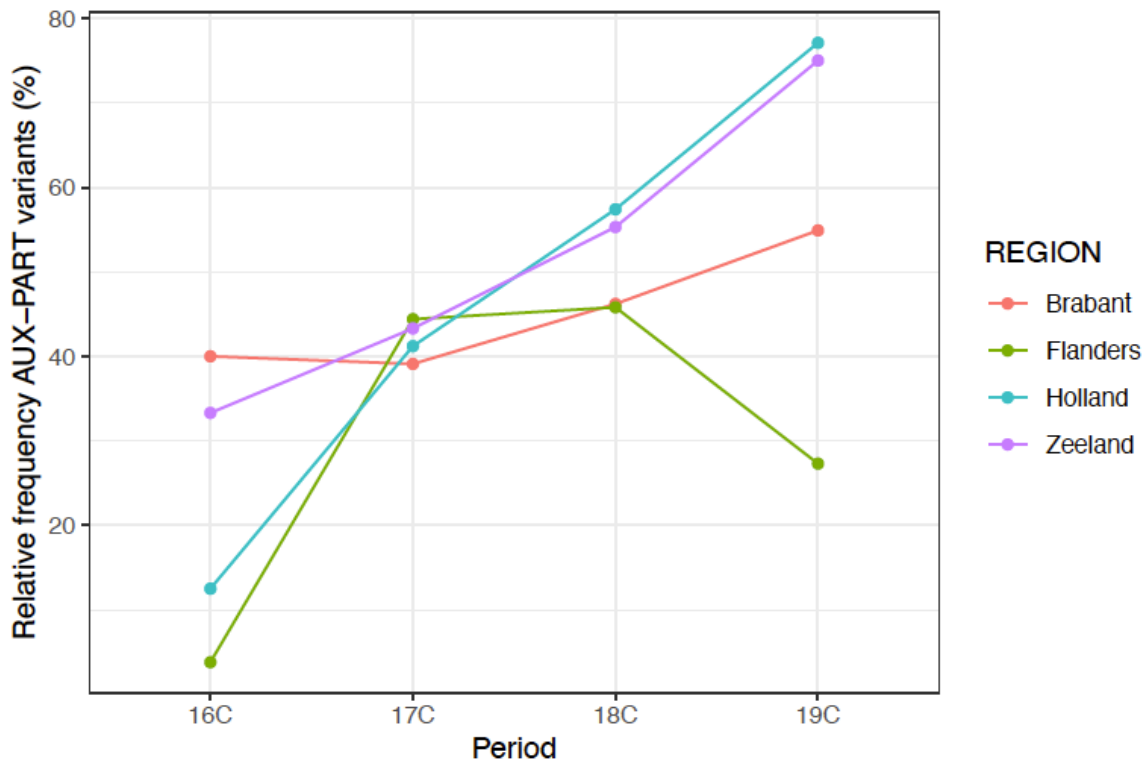


Figure 2: interaction between region and time period

Figure 3 reveals that the stylistic variation also changes through time. What stands out here is the continual and rapid increase of the AUX-PART order in administrative texts, representing more literate, formal, and standardized language use (cf. Krogull, 2018, p. 50). Although the AUX-PART order initially also spreads rather rapidly in the pamphlets and ego-documents, the number of AUX-PART variants in these two text genres remains nevertheless relatively low throughout the four centuries, compared to the administrative texts. In the ego-documents, representing relatively more informal, oral, and everyday language use (cf. Krogull, 2018, p. 50), the change trickles in the slowest: until the nineteenth century, the PART-AUX variant is used most. At the same time, however, AUX-PART is as common as PART-AUX in the pamphlets and has already become the dominant variant in the administrative texts.

Consequently, the AUX-PART variant seems to spread gradually from the more formal text types and higher social domains (cf. the administrative texts) to the more informal text types and private spheres (cf. the ego-documents), suggesting that its sixteenth-century revival should be considered as a top-down rather than as a bottom-up change. Even though the AUX-PART order was only explicitly prescribed as a formal-stylistically marked variant from the twentieth century onwards (Van der Horst, 2002), it was thus already implicitly associated with formality and prestige before that period, confirming Coussé's (2008) hypothesis.

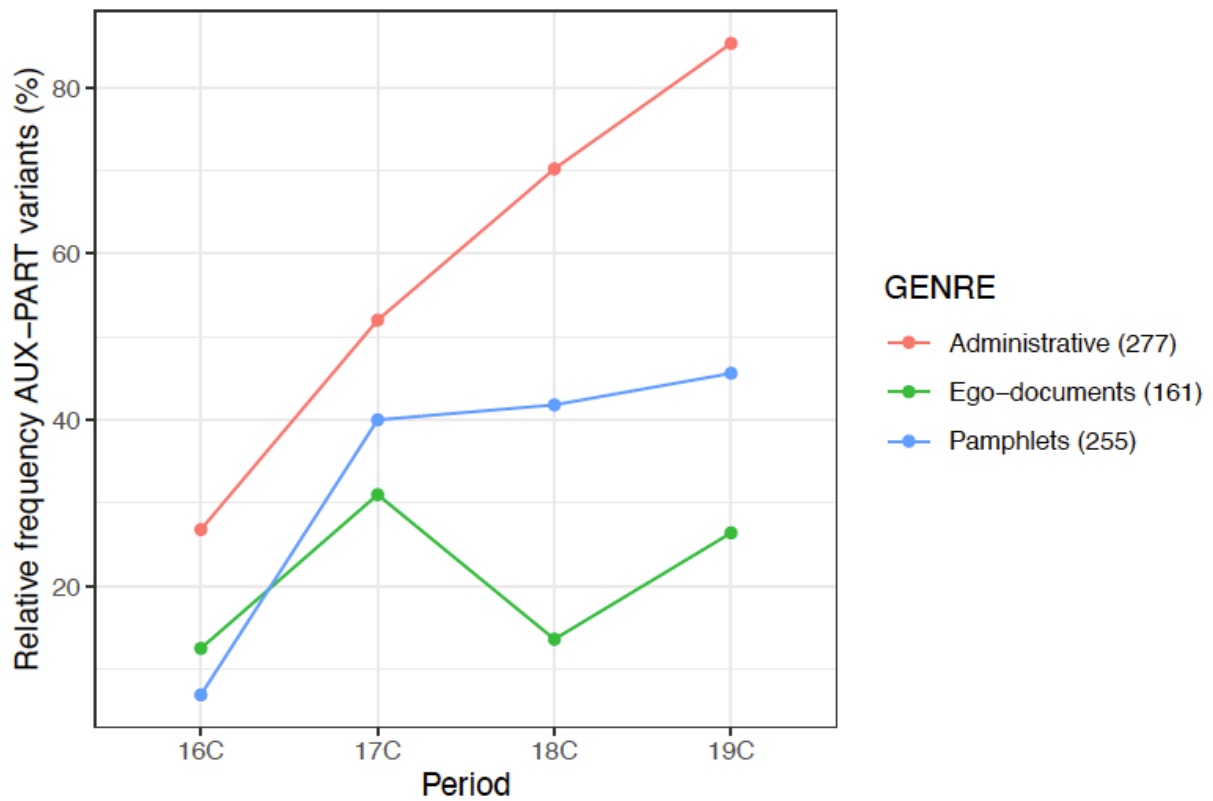


Figure 3: interaction between text genre and time period

Finally, Figure 4 illustrates the interaction between text genre and region. In Brabant, Flanders, and Holland, the AUX-PART order is used the most in the administrative texts and the least in the ego-documents. In the pamphlets, a clearer regional difference does occur: the AUX-PART order is more common in the Northern (Holland and Zeeland) pamphlets than in the Southern (Brabant and Flanders) ones, corresponding to the current macroscopic regional variation in the Dutch language area.

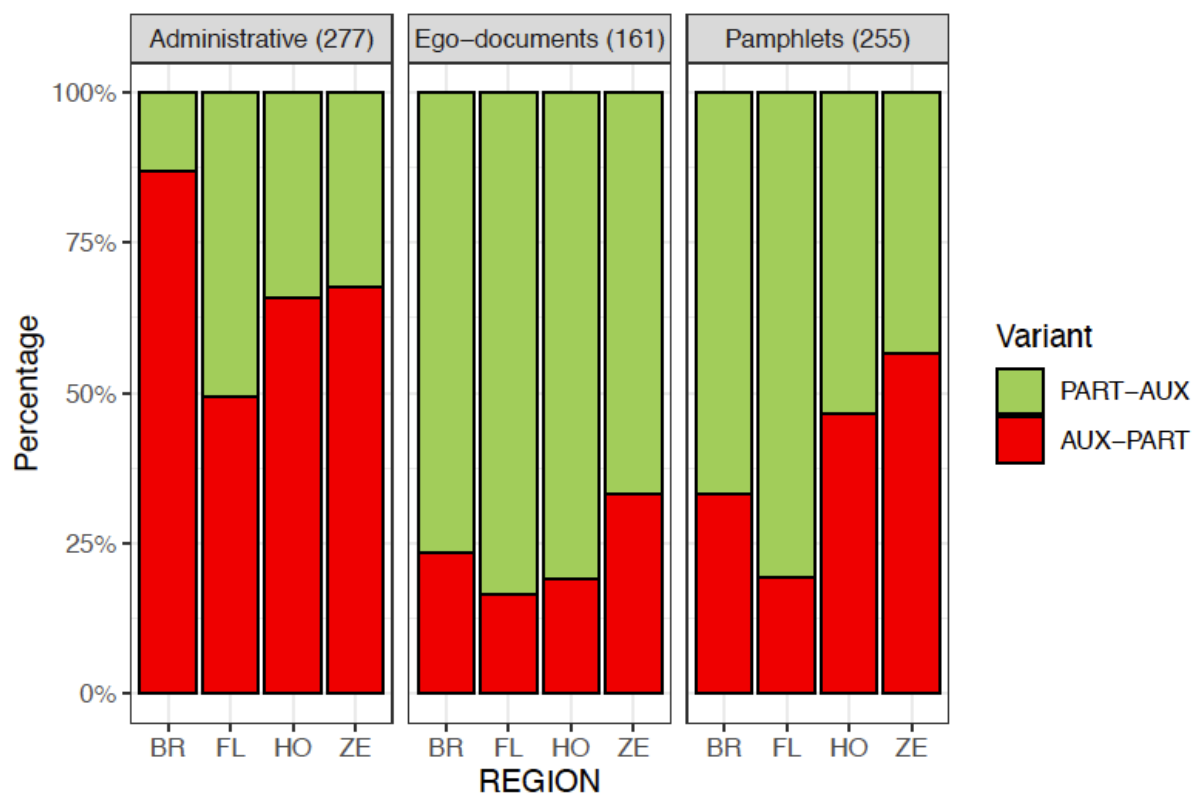


Figure 4: interaction between region and text genre

5. Concluding remarks

The present study aimed at examining the effects of both structural and sociolinguistic factors in the history of the Dutch verbal cluster order variation, determining which of these factors can be seen as the most important predictors for either an AUX-PART or a PART-AUX order as the preferred variant. Based on a random forest analysis, we can conclude that the sociolinguistic variables (viz. time, region, and genre) are the real motor driving this syntactic change, whereas the structural variables (viz. the type of auxiliary verb, the absence or presence of the final clause position, and the morphological structure of the participle) have overall a more limited impact, with the exception of the distance between the conjunction and the verbal cluster. One possible hypothesis is that the effect of the structural variables has not been stable during the period from the sixteenth to the nineteenth century, which could explain why their overall effect is smaller than previous synchronic studies on present-day Dutch suggested. Overall, by studying the impact of both structural and sociolinguistic variables in a broader diachronic, multi-genre corpus, we were able to draw a more nuanced picture of the driving forces behind this well-studied syntactic change in Dutch, adding to our understanding of how linguistic changes spread both through the language itself as well as through the communities of its users.

6. References

- Arfs, M. (2007). Rood of groen? De interne woordvolgorde in tweeledige werkwoordelijke eindgroepen in Nederlandse bijzinnen. *Colloquium Neerlandicum*, 16, 223-245.
- Bennis, H. & E. Coussé (2012). Werkwoordsvolgorde in de rechterperiferie van de Nederlandse zin: inleiding. *Taal en Tongval*, 64(1), 1-9.
- Bloem, J. (2021). Processing verb clusters [Doctoral dissertation]. Universiteit van Amsterdam.
- Bloem, J., Versloot, A. & F. Weerman (2017). Applying automatically parsed corpora to the study of language variation. In J. Hajic & J. Tsuji (eds.), *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers* (pp. 1974-1984). Dublin City University and Association for Computational Linguistics.
- Bloem, J., A. Versloot & F. Weerman (2014). Verbal cluster order and processing complexity. *Language Sciences*, 60, 94-119.
- Burridge, K. (1993). *Syntactic change in Germanic. Aspects of language change in Germanic with particular reference to Middle Dutch*. John Benjamins.
- Charivarius (1998). *Is dat goed Nederlands?* Sdu Uitgevers.
- Coussé, E. (2008). Motivaties voor volgordevariatie. Een diachrone studie van werkwoordsvolgorde in het Nederlands [Doctoral dissertation]. Universiteit Gent.
- Coussé, E., Arfs, M. & G. De Sutter (2008). Variabele werkwoordsvolgorde in het Nederlands. Een taalgebruiksgebaseerd perspectief op de synchronie en diachronie van de rode en groene woordvolgorde. In G. Rawoens (ed.), *Taal aan den lijve. Het gebruik van corpora in taalkundig onderzoek en taalonderwijs* (pp. 29-47). Academia Press.
- Coussé, E. & G. De Sutter (2012). De historische wortels van de rode en groene volgorde in het Nederlands. *Taal en Tongval*, 64(1), 73-101.
- Coupé, G. (2015). Syntactic extension. The historical development of Dutch verb clusters [Doctoral dissertation]. Radboud Universiteit Nijmegen.
- Cutler, D.R., Edwards Jr, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., & J.J. Lawler (2007). Random forests for classification in ecology. *Ecology*, 88(11), 2783-2792.
- De Meersman, A. (1990). Bijzinsvolgorde in tweeledige ww-groepen. Een verkennende historische schets. *Taal en Tongval*, 3, 152-185.
- De Sutter, G. (2007). Naar een corpusgebaseerde, cognitief-functionele verklaring van de woordvolgordevariatie in tweeledige werkwoordelijke eindgroepen. *Nederlandse Taalkunde*, 12(4), 302-330.

- De Sutter, G., Speelman, D. & D. Geeraerts (2005). Regionale en stilistische effecten op de woordvolgorde in werkwoordelijke eindgroepen. *Nederlandse Taalkunde*, 10(2), 97-128.
- Haeseryn, W. (1990). Syntactische normen in het Nederlands. Een empirisch onderzoek naar volgordevariatie in de werkwoordelijke eindgroep [Doctoral dissertation]. KU Nijmegen.
- Haeseryn, W., Romijn, K., Geerts, G., De Rooij, J. & M.C. van den Toorn (1997, 2021). *Algemene Nederlandse Spraakkunst*. Martinus Nijhoff-Plantyn.
- Janssens, G. & A. Marynissen (2011). *Het Nederlands vroeger en nu* (4^{de} druk). Acco.
- Levshina, N. (2015). *How to do linguistics with R. Data exploration and statistical analysis*. John Benjamins.
- Liaw, A. & M. Wiener (2002). Classification and Regression by randomForest. *R News* 2(3), 18-22.
- Olthof, M., Westendorp, M., Bloem, J. & F. Weerman (2017). Synchronic variation and diachronic change in Dutch two-verb clusters. *Tijdschrift voor Nederlandse Taal- en Letterkunde*, 1(1), 34-60.
- Pauwels, A. (1953). *De plaats van hulpwerkwoord, verleden deelwoord en infinitief in de Nederlandse bijzin*. Drukkerij M. & L. Symons.
- R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <<https://www.R-project.org/>>.
- Rutten, G., Van de Voorde, I., & R. Vosters (2023). Transmission and diffusion in diachronic orthography. In M. Condorelli & H. Rutkowska (eds.), *The Cambridge Handbook of Historical Orthography* (pp. 596–616). Cambridge University Press.
- Van der Horst, J.M. (2002). Geschiedenis van een taalnorm: de doorbroken werkwoordelijke eindgroep (deel I). *Verslagen en Mededelingen van de Koninklijke Academie voor Nederlandse Taal- en Letterkunde*, 112(3), 529-556.
- Van de Voorde, I., Rutten, G., Vosters, R., Van der Wal, M.J. & W. Vandenbussche (2023). Historical Corpus of Dutch: A new multi-genre corpus of Early and Late Modern Dutch. *Taal en Tongval: Language Variation in the Low Countries*, 75(1), 114–132..
- Van de Voorde, I. (2022). *Pluricentriciteit in de taalgeschiedenis. Bouwstenen voor een geïntegreerde geschiedenis van het Nederlands (16^{de}-19^{de} eeuw)* [Doctoral dissertation]. Vrije Universiteit Brussel & Universiteit Leiden.
- Van der Wal, M.J. & C. Van Bree (2008). *Geschiedenis van het Nederlands* (5^{de} herziene druk). Spectrum.

Vosters, R., Rutten, G., Van der Wal, M.J. & W. Vandebussche (2012). The sociolinguistics of spelling. A corpus-based case study of orthographical variation in nineteenth-century Dutch in Flanders. In A. Van Kemenade & N. De Haas (eds.), *Historical linguistics 2009. Selected papers from the 19th International Conference on Historical Linguistics* (pp. 253-274). John Benjamins. Willemys, R. (2003). Dutch. In A. Deumert & W. Vandebussche (eds.), *Germanic standardizations. Past to present* (pp. 93-125). John Benjamins.

7. Appendix

7.1 List with possible spelling variants of the auxiliary verbs *hebben*, *worden*, and *zijn*

Hebben	Worden	Zijn
Hebben	Worden	Zijn
Hebbe	Worde	Sijn
Hebe	Wordn	Zyn
Hebbn	Worddn	Syn
Heb	Wordden	Ben
Hep	Wordde	Bent
Hebt	Wordt	Bendt
Hept	Word	Bend
Hebd	Wort	Bin
Hebdt	Wor	Bist
Hepd	Werden	Benne
Hepdt	Wert	Is
Heeft	Wer	His
Heefd	Werdt	Ist
Heefdt	Werd	Ys
Heef	Wierdt	Hys
Heft	Wiert	Yst
Hefd	Wierd	Es
Hefdt	Wier	Hes
Hef	Wird	Est
Hefd	Wirdt	Zijt
Hefdt	Wirt	Sijt
Hef	Wir	Zyt
Had	Werdden	Syt
Hat	Werde	Was
Hadt	Werde	Waz
Hadden	Werdn	Waren
Haden	Werddn	Waeren
Hadde	Woer	Waaren
Haddn	Woerdt	Waerren
Hade	Woert	Waarren

Hebban	Woerd	Warren
Heban	Wirde	
	Wirdden	
	Wirde	
	Wirdde	
	Wirddn	
	Wirdn	
	Wierden	
	Wierdden	
	Wierdde	
	Wierde	
	Wierdn	
	Woerden	
	Woerde	
	Woerdn	
	Woerdde	
	Woerdden	
	Woerddn	