

# Verbal cluster order and processing complexity

Jelke Bloem

Arjen Versloot

Fred Weerman

Amsterdam Center for Language and Communication  
University of Amsterdam  
{j.bloem, a.p.versloot, f.p.weerman}@uva.nl

## Abstract

We examine a case of word order variation where speakers choose between two near-synonymous constructions partly on the basis of the processing complexity of the construction and its context. When producing two-verb clusters in Dutch, a speaker can choose between two word orders. Previous corpus studies have shown that a wide range of factors are associated with this word order variation. We conducted a large-scale corpus study in order to discover what these factors have in common. The underlying generalization appears to be processing complexity: a variety of factors that are related to verbal cluster word order, can also be related to the processing complexity of the cluster's context. This implies that one of the word orders might be easier to process — when processing load is high, speakers will go for the easier option. Therefore, we also discuss which of the two word orders is easier to process. By testing for associations with factors indicating a higher or lower processing complexity of the verb and its context, we find evidence for the hypothesis that the more frequent of the two orders is easier to process.

## 1 Introduction

Verbal clusters in Dutch present an interesting example of grammatical variation. In clusters of two verbs, both possible orders are grammatical, leading to optionality:

- (1) *Ik denk dat ik het **begrepen** heb*  
I think that I it understood have  
'I think that I have understood it.'
- (2) *Ik denk dat ik het **heb** begrepen*  
I think that I it have understood  
'I think that I have understood it.'

Speakers may produce either order in similar contexts, and the choice between different variants does not appear to be determined by grammatical factors. A notable aspect of this optionality is that the difference in word order is generally assumed not to correspond to a meaning difference. However, when speakers choose between constructions in these situations, they do not do so randomly.

This variation cannot be explained on the basis of syntax alone. In the most widespread account of the syntactic mechanism behind verbal clusters, both constructions are of equal derivational complexity, making an explanation in terms of syntactic complexity unlikely. This account, first proposed by Evers (1975), states that verbal clusters are formed through the mechanism of verb raising. In verb raising, the main verb is generated as the complement of the head verb of the cluster, and then moves up to join the head verb. Figure 1 illustrates this process. The main verb *begrepen* is raised to attach to the governing verb *heb*, forming a complex head. It can attach on either side of the head verb, resulting in either a 2-1 (Figure 1b) or 1-2 (Figure 1c) ordered verbal cluster (1 is the head, 2 is the participle). The claim that verbal clusters are formed using this special mechanism is illustrated by the fact that verbal clusters are very rarely interrupted by non-verbal material. Furthermore, in multiple verb constructions with more than two

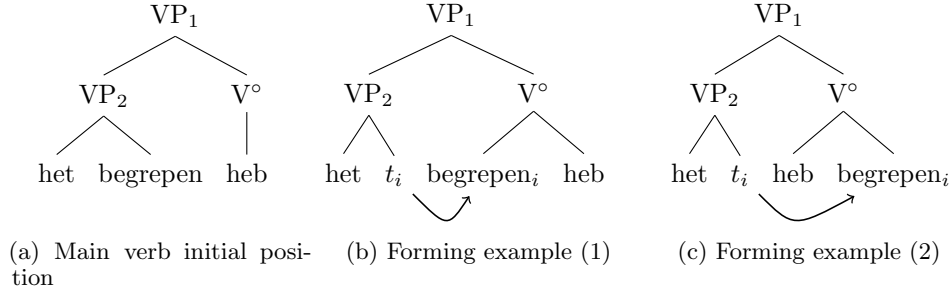


Figure 1: Verb raising to generate the verbal clusters of examples 1 and 2, following the analysis of Evers (1975).

verbs the *Infinitivus Pro Participio* (IPP) effect shows up, meaning that modal verbs that would normally be participles are infinitives instead (Wurmbrand, 2006).

Corpus studies of similar variation phenomena have shown that a large portion of the variation can be statistically accounted for using multifactorial statistical models that incorporate a variety of linguistic factors beyond syntax. De Sutter (2005) and Bloem et al. (2014) found that this is also the case for Dutch verbal clusters, quantifying the impact of morphosyntactic as well as semantic factors, and properties such as sentence length and word frequency. The multifactorial models employed in these studies reveal interesting patterns in the variation and associations with linguistic factors, but to explain why all of these different factors are involved when people choose to use one word order or the other, one needs to generalize over the factors that can be found by measuring things in a corpus.

Various generalizations have been suggested. Several of the observed effects regarding factors such as sentence length can be interpreted as effects of adhering to the rhythm of the Dutch language (Schutter, 1996), evenly distributing information weight throughout a sentence (De Sutter et al., 2007), or minimizing processing complexity (De Sutter, 2005; Bloem et al., 2014). Processing complexity refers to the amount of cognitive resources or cognitive effort required to produce or comprehend an utterance. Speakers prefer to minimize their use of cognitive resources, formulating sentences in a way that minimizes processing complexity when they have multiple grammatical ways to communicate something (i.e. Jaeger and Tily (2011) for an overview, Levy and Jaeger (2006) and Fedzechkina et al. (2012) for experimental work). When this happens, we call this a processing effect.

In this work, we will explore the suggestion that minimizing processing complexity is a general factor that can explain the effects observed in multifactorial studies so far. De Sutter (2005) originally formulated this idea for Dutch verb cluster order variation on the basis of two factors that were found to correlate with verb cluster order in his study — main verb frequency, and differences between verb clusters with arguments that are complements or adjuncts. He argues that the 1-2 word order is stylistically preferred, and is therefore more likely to be used when speakers have the spare cognitive capacity to do so, while the 2-1 order is otherwise more likely to be used. De Sutter ends his dissertation with the suggestion that in future work, other factors from his multifactorial study should be interpreted from the same perspective of processing complexity. Bloem et al. (2014) take this idea and suggest that it can be tested by studying whether factors indicating ease of processing correlate with one particular verb cluster word order, which would be the ‘default’ order. They also state that this default order is not necessarily the 2-1 order, and that there are arguments for considering the 1-2 order to be the default as well. That makes the processing complexity generalization empirically testable – is there a word order that correlates with processing complexity, and if so, which one? The present study aims to answer this question.

We address the issue by gathering empirical evidence for factors that affect verbal cluster order using a large amount of corpus data. Considering the wide range of factors that have been found to be associated with verbal cluster order variation in earlier studies, we expect that semantics, information structure and processing complexity all play a role in the choice between orders. However, in this study we focus on factors related to processing complexity, while controlling for other influences as much as possible. We hypothesize that any factor that cannot be attributed to semantics (i.e. it does not affect the meaning of the utterance), should be attributed to processing effects. We largely follow the methodology of Bloem et al. (2014), but with more factors relating to processing complexity, and analyzing the previously studied factors in terms of processing complexity. This approach can be viewed as analogous to that of De Sutter et al. (2007). They use a multifactorial model to study rhythm as a possible generalization over factors relating to verb cluster word order, but conclude that information weight is a more plausible generalization. Their suggestion is to create another multifactorial model that includes new and more specific factors relating to information weight.

In the next section, we will consider some relevant studies and insights, firstly on the topic word order variation, secondly on processing complexity, and lastly on Dutch two-verb clusters and the extent of the order variation there. Section 4 discusses our methodology and the corpus that we used to test our hypotheses, and we discuss factors that relate to both verbal cluster order and processing complexity in section 5. In section 6 we present our results, followed by a discussion of the findings and a conclusion in section 7 and 8 respectively.

## 2 Related work

### 2.1 Word order variation

Word order variation in Dutch verb clusters can be viewed as a case of a syntactic alternation, or near-synonymous constructions, where speakers have two alternative forms available that are both grammatical. Syntactic alternations have frequently been studied using multifactorial models based on corpus data, with a focus on the study of factors that can be measured in an objective way. This is the perspective from which we approach the phenomenon. While we discuss background from generative, construction grammar and functional linguistic theory to describe the observed factors, we do not intend to take a specific theoretical stance. The different theories predict that different factors shape grammar, but they do not place strong restrictions on how speakers decide between alternative near-synonymous constructions when multiple options are grammatical. To summarize some theoretical discussions on the nature of syntactic alternations, we will briefly compare Dutch verbal cluster order variation to the history of a better-known case of grammatical variation, the English dative alternation.

- (3) He gave his friend the ticket.
- (4) He gave the ticket to his friend.

This optionality, a choice between the double object construction shown in example (3) and the prepositional dative construction in (4), presents a similar puzzle as the Dutch verbal cluster order variation. The dative alternation was already discussed as a problem for transformational theories of language by Oehrle (1976). Moravcsik (2010) summarizes the problem that this alternation presents: it violates the principle of isomorphism, the one-to-one relation between meaning and form. This has been claimed to be a near-universal property of language (Haiman, 1980), and thus cannot easily be abandoned. Moravcsik (2010) discusses two kinds of solutions that linguists have pursued in resolving this conflict.

The first solution is to argue that the two forms actually have a single underlying structure, and are therefore actually the same thing, preserving the one-to-one relation. Larson (1988) worked out an implementation for this in terms of generative mechanisms, relating them transformationally. In this proposal, the two constructions are different surface representations of the same underlying form. This approach has been applied to Dutch verbal clusters as well (Evers, 1975).

The second solution is to argue that the two forms actually have two different meanings, which are influenced by pragmatic aspects that are context-dependent. Beck and Johnson (2004) take this approach, claiming that the two constructions have different underlying forms with different argument structures, giving rise to different meanings.

From the perspective of generative linguistics, optionality between multiple near-synonymous constructions is often considered to be a question of how objects generated by the grammar are used (Embick, 2008), and therefore outside of the scope of syntactic theory. Even in an approach where optionality is sensitive to meaning, such as in Beck and Johnson (2004), the meaning arises from the verb’s argument structure, which does not differ between the two word orders in the case of verbal clusters.

The multiple meaning approach is also common in construction grammar. Goldberg (1992) accounts for apparent usage differences between the two forms of the dative alternation by claiming that the meaning difference comes directly from the ditransitive argument structure, which carries its own semantics. Bresnan et al. (2007) later showed that information-structural properties can also account for a large portion of the variation between the two forms, using corpus evidence. In construction grammar, constructions are stored pairs of form and meaning, or form and function, in which at least some aspect of the construction’s function is not predictable from its component parts. Any difference in form therefore implies a difference in meaning or function (Goldberg, 2002). This assertion provides some additional testable predictions about the factors that may influence verbal cluster order variation. If the two verbal cluster orders in Dutch are two constructions, they must have differences in meaning or function. However, the notion of ‘allostructions’ has also been proposed — variant structural realizations of a partially underspecified construction (Cappelle, 2006). For two-verb clusters, this would mean a verb cluster construction that is underspecified for word order, with two possible allostructions in which word order is specified. This idea formalizes the relation between two forms of an alternation, but without one form being an underlying form of the other. In this view, there does not need to be a functional or meaning difference between two alternative word orders.

Word order variation can also be influenced by functional word order principles, which have been claimed to shape grammars, for example by Hawkins (2014). This idea can be extended to word order preferences, when an alternation is present and multiple options are grammatical. Some of the factors that have been discussed in the context of verb order variation can be interpreted as consequences of such principles. This perspective is discussed by Coussé (2008). She provides an overview of several functional word order principles, such as the adjacency principle, which states that conceptually related elements are placed together in the linear ordering of an utterance. The salience of certain positions (such as the first and last position in a sentence) is also discussed, as well as several principles of information structure, such as ‘topic-before-comment’, and complexity-related principles such as ‘light-before-heavy’. We will now discuss some theories that have been used to account for such complexity principles, and how they can be related to syntactic alternations.

## 2.2 Processing complexity

Processing complexity, more specifically facilitating processing by avoiding complex utterances, can be considered to be a functional motivation for variation as well, although this idea does not seem to be spelled out explicitly in generative or construction grammar literature. It has been discussed extensively in psycholinguistic literature, however. Two main approaches can be identified here: constraint-satisfaction models, and resource-limitation (or memory-based) models (Levy, 2008).

Constraint-based models consider various parallel alternative interpretations or parses of a sentence during processing, and relate processing difficulty to expectation. MacDonald (2013) presents a range of phenomena in which cognitive constraints and processing difficulty are shown to affect language production, including word order. Such processing effects have been modeled computationally by Hale (2001), using a measure of surprisal. Within this line of research, Levy (2008) has proposed a resource-allocation theory of processing difficulty, using probabilistic word models. The study mathematically proves that previous work on language processing can be linked to surprisal — Hale’s surprisal

measure is shown to be equivalent to Levy’s model. In this model, speakers would be able to choose between multiple possible parses when comprehending a sentence, preferring the interpretation with least surprisal and evenly distributed surprisal throughout the sentence, thereby minimizing processing difficulty (Jaeger, 2010). It has also been shown that structures that are dispreferred according to such probabilistic models can be produced, but are produced less fluently (Cook et al., 2009), showing that these models may be able to account for processing during production as well as comprehension. There are many different approaches to modeling surprisal though, with different representations (i.e. n-gram models or construction grammars) and different units (i.e. words or constructions) (Jaeger and Tily, 2011).

The other approach, resource-limitation models, focuses on the idea that there is some limited cognitive resource, such as memory, that imposes processing limitations. In this approach, processing difficulty is often defined in terms of dependency length — it is more difficult to process words that depend on each other when there is a larger distance between them. Gibson’s (1998)’s work is an example of this. In this theory, there is a memory cost for keeping track of obligatory syntactic requirements that need to be fulfilled with a dependency relation, as well as a cost for integrating an element with the current parse. This approach predicts higher processing costs for longer dependencies. This dependency length idea also corresponds to the first efficiently principle proposed by Hawkins (2014), Minimize Domains, who argues that these principles play an important role in shaping the grammar of a language. These principles are used to account for cases of language variation, including word order patterns. The basic idea of Minimize Domains can be applied to verb clusters fairly easily — the auxiliary verb is the head of the clusters and therefore dependency relations to e.g. the subject attach to it. If the auxiliary verb comes first, as in the 1-2 order, the dependency link can be made sooner. Therefore, the 1-2 order is easier to process following this approach, at least syntactically.

These two approaches to modeling language processing do not appear to be mutually exclusive, they just have a different focus. Furthermore, they do not deal with all domains of language, but mainly with syntax, and are therefore not complete accounts of how language processing works. As noted by De Sutter (2009, p. 226–227), it can be difficult to apply these models or principles of word ordering to the study of verb clusters, because most word order variation studies discuss interconstituent alternations, while verb clusters are an intraconstituent alternation (within the verb phrase).

More broadly, the connection between word order variation and processing difficulty has been made before in discussions of various optionality phenomena. Generally, a multifactorial corpus study is performed, an approach pioneered by Gries (2001). Those studies often show that a multitude of factors affect the choice between the two options leading to a hypothesis that generalizes over all of the factors under discussion. In this first multifactorial corpus study by Gries, on optional particle movement in English, this hypothesis is called the Processing Hypothesis (PH):

The multitude of variables (most of which are concerned with the direct object NP) that seems to be related to Particle Movement can all be related to the processing effort of the utterance. (Gries, 2001)

One example of a variation phenomenon where processing difficulty plays a clear role that cannot be confused with information structure, is the optionality of *that* in English non-subject relative clauses, as in (5):

- (5) They have all the water (that) they want.

Wasow et al. (2011) explain this optionality in terms of predictability and processing. A processing account has been suggested for another Dutch-language phenomenon as well. In a study on the optionality of the Dutch *om*-complementizer, Bouma (2013) suggests that *om*-insertion is related to linguistic processing. It serves to reduce ambiguity in the contexts in which it is used, which reduces the processing complexity of the clause.

These studies show that processing complexity has been cited as an explanation for other instances of grammatical variation, and the models of language processing discussed in this section can be used to infer what factors might affect the processing complexity of a sentence.

### 2.3 Verb clusters

Verb clusters are a widely studied phenomenon in Dutch syntactic literature, both in terms of the generative mechanism behind them, and in terms of their optionality (Evers, 1975; den Besten and Edmondson, 1983; Haegeman and van Riemsdijk, 1986; Zwart, 1996; Wurmbrand, 2004). Some of the first studies of Dutch verbal clusters were focused on regional variation (Pauwels, 1953). The mechanism of verb raising is generally used to explain the existence of verbal clusters, as illustrated in section 1. A broad overview of verb raising across Germanic languages, as well as different theoretical accounts of the phenomenon, is provided in Wurmbrand (2006).

While various terms have been used to describe the two orders, we will follow Stroop (1970) and call the construction in Example (1) the *descending* order, and the construction in Example (2) *ascending*. This is because the finite auxiliary is considered to be the verb that is highest in the syntactic tree, while the main verb is the lowest. This lets us number the verbs: example (1) shows a 2-1 order, because the head comes last, and the main verb comes before it. Larger clusters can thus be adequately described, e.g. a 4-3-2-1 cluster, where 4 is the main verb. In this work, we will be using this numbering to refer to the different orders.

In this section, we provide an overview of verbal cluster constructions with two verbs. We first discuss clusters with participial main verbs, which exhibit more order variation, and then clusters with infinitival main verbs, which exhibit less variation. Lastly, we discuss the ‘te’-infinitival cluster constructions. The discussion approximately follows the structure of the description of verb groups in the *Algemene Nederlandse Spraakkunst* (Haeseryn, 1997, p. 946–1076), but in far less detail. For a more complete description of possible Dutch verb clusters, we refer to this work.

The grammatical examples in this section come from the 145 million word Wikipedia part of the Dutch Lassy Large corpus (van Noord, 2009), which we have also used as our data source. Observations on which verbs occur in the different constructions also come from this corpus, unless otherwise noted. The identifiers displayed after the translation of each example are the identifiers of the sentence in the corpus. This corpus will be discussed in section 4.

There are various types of two-verb clusters to consider, all of them exhibiting free variation but with different probabilities for each order. Examples (1) and (2) from the introduction show two-verb clusters with a participial main verb, and auxiliary heads. When we discuss clusters with auxiliary verbs (or ‘auxiliary clusters’), we are referring to those auxiliaries that function as auxiliaries of time, *zijn* ‘to be’ and *hebben* ‘to have’, the passive auxiliary *worden* ‘to be’, and *zijn* ‘to be’ as a copular verb. These three verbs are the most frequent auxiliary verbs that take participial main verbs, and categorizing them in this way is in line with earlier work (e.g. De Sutter 2005). There are a few constructions in which modals or other types of verbs occur with a participial main verbs, but they are said to be headed by an elided infinitive (Haeseryn, 1997, p. 960), so we have grouped them with the infinitival clusters. This distinction between auxiliary and other clusters is made because these particular auxiliary verbs exhibit most of the order variation in Dutch two-verb clusters. We use this terminology for practical reasons, without meaning to imply any cognitively real categories or an overarching ‘auxiliary cluster’ parent construction. Most previous studies on two-verb cluster variation concern only these three verbs.

Besides the clusters with participial main verbs and auxiliary heads, there are other kinds of verbs that can head a verbal cluster, though they exhibit less variation and stronger preferences for the 1-2 order. These clusters generally have an infinitival main verb. The most frequent of these are the modal auxiliary verbs. It is possible to have a modal verb as a head, in which case the main verb is an infinitive. It is clear that the 2-1 order (example 6) and 1-2 order (example 7) are both grammatical:

- (6) ... *dat iedereen hem ongestraft doden mocht*  
 ... that anyone him unpenalized kill may  
 ‘... that anyone could kill him with impunity.’ [wik\_part0021/9811-21-2]
- (7) ... *zodat hij de Jedi kan uitroeien*  
 ... so he the Jedi can exterminate

‘... so that he can exterminate the Jedi.’ [wik\_part0013/5954-13-2]

Despite this, by counting constructions in a corpus, we can observe that the 1-2 order in example (7) is far more common in this construction, and some authors even say there is no optionality here (e.g. Zuckerman 2001). In the Wikipedia part of the Lassy Large corpus, modal clusters occur in the 2-1 order only 0.5% of the time. However, they are considered to be grammatical, and seem to be completely acceptable to Dutch speakers. For comparison, auxiliary clusters as defined above occur in the 2-1 order 35.2% of the time.

The modal verbs that are used in this construction are *kunnen* ‘can’, *moeten* ‘must’, *mogen* ‘may’, *willen* ‘want’, *zullen* ‘shall’, and *hoeven* ‘need’, as discussed in the ANS (Haeseryn, 1997, 984–1006), though *hoeven* is used with the ‘te’ infinitival marker. Modal clusters are generally treated as a different construction in the literature, as different grammatical rules may apply to it, particularly in other Germanic languages. For example, in West Flemish and Afrikaans, modal clusters occur in 1-2 order while auxiliary clusters occur in 2-1 order, with exceptions in a few contexts (Wurmbrand, 2006).

Another type of auxiliary verb used with infinitival main verbs is the causal auxiliary *laten* ‘let’, used in a 1-2 order in example (8):

- (8) ... *dat de luisteraar de muziek in 3-dimensionale stijl laat ervaren*  
 ... that the listener the music in 3-dimensional style let experience  
 ‘... that lets the listener experience the music in three-dimensional style.’  
 [wik\_part0018 /8288-34-2]

This construction might also be considered grammatical in the 2-1 order, but this appears to be quite rare. In the Lassy Large corpus, only 0.114% of the clusters with *laten* are in the 2-1 order, and some of those are annotation errors. The *laten* clusters behave similarly to the modal clusters in this, and also in the fact that they take infinitival main verbs.

In addition to this, there seems to be a restricted number of non-auxiliary verbs that can occur as heads of clusters. These are known as (other) grouping verbs. Some of these verbs are verbs of perception (*horen* ‘hear’, *zien* ‘see’), some are stative verbs (*blijven* ‘stay’, *gaan* ‘go’, *liggen* ‘lie’, *zitten* ‘sit’, , *staan* ‘stand’), and there are various others. These verbs are also grammatical in both orders, with 2-1 orders occurring 10.6% of the time when averaging over all of the verbs.

In the cluster types we discussed so far, the clusters with auxiliaries had participial main verbs, while the modal and other clusters had infinitival main verbs. However, there is another verbal cluster construction to consider, in which an auxiliary cluster can be made infinitival: the *te*-infinitival construction. These clusters are characterized by the fact that the main verb is marked by the infinitival marker *te*. These constructions have also been called pure-infinitival clusters, to avoid confusion with other clusters where the main verb is infinite (e.g. modal clusters). Furthermore, unlike other modal verbs, the modal and negative-polar verb *hoeven* (need) is used in this construction. Additionally, this construction also occurs with some grouping verbs, including some verbs that cannot group into a cluster in any other way. This is also known as the second infinitival construction. Lastly, there is the ‘third construction’, in which a grouping verb is used with a *te*-infinitival marker, yet the main verb is not infinitival. These constructions are not accepted by all speakers, and have been noted by den Besten and Broekhuis (1992) to have some interesting syntactic properties.

The examples below show typical *te*-infinitival constructions with auxiliary verbs:

- (9) ... *waaraan de soort is te herkennen*  
 ... by which the species is to recognize  
 ‘... by which the species can be recognized.’ [wik\_part0297/485044-4-1]
- (10) ... *die te maken hadden met muziek*  
 ... that to make had with music  
 ‘... that had to do with music.’ [wik\_part0568/1276744-4-1]

These clusters have not been the focus of any quantitative corpus study, but they appear to have limitations on their optionality with a few particular verbs. When *hoeven* or a grouping verb is used, only the 1-2 order is attested.

Now, we will summarize previous work on explaining the order variation.

## 2.4 Verb cluster order variation

There have been various studies trying to account for verbal cluster order variation in cases where multiple orders are grammatical. The rules and mechanisms discussed in generative literature allow for a lot of optionality, as was discussed earlier, and thus mainly outline the constructions in which order variation can occur. Coussé et al. (2008) provide a summary of recent work on verbal cluster variation, summarizing three dissertations on the topic (De Sutter, 2005; Coussé, 2008; Arfs, 2007). A diverse set of factors that may influence the use of 2-1 and 1-2 orders has been discussed in these works, and Coussé et al. (2008) group them into four broad categories.

**Contextual** factors include the regional background of the speaker and mode of communication. This is not intraspeaker variation, however. **Rhythmic** factors relate to the hypothesis that speakers may change the order of verbs to match the standard stress pattern of Dutch, though this may not be so important in written texts. De Sutter (2009) did not find a strong effect of stressed syllables near the cluster in his corpus study of written texts, except when two stressed syllables would have been directly adjacent. **Semantic** factors are described mostly in terms of lexical semantics. One finding in this area is that adjectival participles have a preference for the 2-1 order. Lastly, **discourse** factors are mentioned with syntactic priming as an example, although this could also be considered to be a processing factor. From all this, Coussé et al. (2008) conclude that the choice of verbal cluster order is influenced by a complex set of interacting factors. Therefore, any model representing this phenomenon would need to take many factors into account. Such a study has been conducted by De Sutter (2009), and expanded by Bloem et al. (2014), testing the association of 10 factors with various types of verbal clusters.

In a psycholinguistic study, Hartsuiker and Westenberg (2000) showed that verbal cluster orders can undergo structural priming, providing evidence that both verbal cluster orders can be considered to be distinct constructions. Under this assumption, variation between the orders may be motivated by semantic or functional differences, following Goldberg (2002).

In his dissertation, De Sutter (2005) provides some linguistic explanation for the order variation and its correlation with various factors that he observes. At the end of his dissertation, he suggests that there may be an underlying psycholinguistic effect of processing. This suggestion of a processing effect is based on two factors, the frequencies of the orders in various modes of language, and a priming effect. He assumes the 2-1 order to be easier to process, because it appears more often in modes of communication in which the ‘production pressure’ is greater — spoken rather than written, interaction rather than monologue, spontaneous rather than edited (De Sutter, 2005, p. 321). Based on this, he suggests that the 2-1 order is the default or most economical order, produced when not enough time or resources are available to produce the 1-2 order. Under this hypothesis, producing the 1-2 order would involve some additional process of sociostylistic encoding (De Sutter, 2005, p. 322-324), a hypothesis that has been discussed before, although de Sutter’s notion of style is fairly broad. One reason for assuming a sociostylistic process is the apparent preference for the 1-2 order in prescriptive grammars and editing guidelines. This idea is explored further in De Sutter (2007), where 5 factors that correlate with verb cluster word order, are interpreted in terms of cognitive complexity. In this article, he argues that factors that correlate with the 1-2 order have a lower processing cost.

The idea that there is a default order that is easier to process is an interesting one, which we will explore further in the next section. However, we will argue that it is not necessarily the 2-1 order. The issue discussed by De Sutter (2007) can be stated more broadly: Do either of the orders occur in contexts that are generally more difficult to process?



### 3 The processing hypothesis

De Sutter (2005) suggests that the stylistically preferred word order is the 1-2 order, shown in example 2. If this is true, one would expect to find the 1-2 order in contexts that are less difficult to process, and the 2-1 order in contexts that are more difficult to process, as is argued by De Sutter (2007). We will call this the **default 2-1 order hypothesis**. Additional evidence for a default 2-1 order is the observation that young children use the 2-1 order more frequently (Zuckerman, 2001).

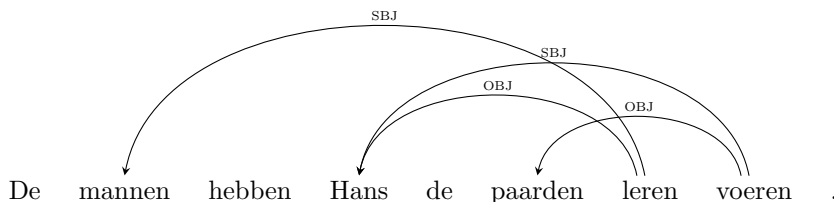
However, Meyer and Weerman (in press) argue that these 2-1 constructions in young children are instead a result of the base word order of Dutch (Object-Verb, OV). It has been observed that Dutch children produce verb-final utterances from the beginning (Wijnen, 1997), as well as in their first subordinate clauses (Van Kampen, 2010). Therefore, this seems to be a pattern or construction that is acquired early, and children use it before they produce groups of verbs or verb clusters (Wijnen, 1997).

In this view, children put the head verb in last position, strictly following the base order. The other verb, which they might also analyze as an adjective, then comes before it (a 2-1 order). Meyer & Weerman theorize that children first learn about verb raising (which forms clusters) when they acquire the 1-2 order. Before verb raising, the head verb is in final position, following the base OV order of Dutch. Raising the verb to form a 2-1 cluster is therefore a vacuous movement. The surface order will be the same as the underlying structure (the head comes last), and provides no evidence of any sort of special verb raising mechanism to child learners of the language. On the other hand, raising the verb to create a 1-2 order violates the base word order of Dutch, and is therefore more straightforward to notice and learn. 2-1 orders can simply be interpreted as OV orders, until the learner figures out the mechanism of verb raising from the 1-2 order evidence. This idea is supported by Meyer and Weerman’s observation that Dutch children only overgeneralize the 1-2 order once they have acquired verbal cluster constructions, not the 2-1 order. More evidence can be found in observations used to account for dialectal variation in Dutch three-verb clusters (Barbiers et al., 2010, 2016). It is theorized that participles can be adjectives as well as verbs in these constructions even in adult language — and in particular, that they are adjectival when used in the initial position in a verb cluster. When we follow this idea of initial adjectival participles, 1-2 orders would be the earliest form of verb raising. This may mean that this word order is more entrenched in speakers’ grammars or lexicons, and therefore easier to process. If this is true, we would expect to find the 1-2 order in contexts that are more difficult to process, and the 2-1 order in contexts that are less difficult to process. We will call this the **default 1-2 order hypothesis**.

An additional argument for the default 1-2 order hypothesis can be found in the literature on a related phenomenon that has undergone much psycholinguistic investigation: cross-serial dependencies in Dutch. When a verb cluster consists of multiple verbs that have arguments, the verbs and arguments are ordered with crossing dependencies in Dutch as in (11), while in German, nested dependencies are preferred (12):

- (11) *De mannen hebben Hans de paarden leren voeren.*  
 The men have Hans the horses teach feed.
- (12) *Die Männer haben Hans die Pferde füttern lehren*  
 The men have Hans the horses feed teach.  
 ‘The men taught Hans to feed the horses’

The object and subject dependency relations in the Dutch example cross each other as shown in this simplified dependency tree:



In these verb cluster constructions with embedded arguments, the Dutch clusters are in the 1-2(-3) order and the German ones in the (3-)2-1 order. An experiment by Bach et al. (1986) showed that longer Dutch verb cluster constructions (3 or more verbs) with crossed dependencies are less difficult to process than German ones, as the subjects gave better comprehensibility judgements and made fewer errors in a question answering task. They hypothesize that the processing advantage comes from being able to build ‘higher’ elements of the structure first. This is also the case when the head verb comes first in a 1-2 order, and may facilitate processing of 1-2 orders. Even though Bach et al.’s (1986) study only showed such difficulties for larger clusters, it might be the case that outside of an experimental setting, with more factors contributing to processing load, there is a processing difference even for two-verb clusters, which can be mitigated by choosing the order that is easier to process. After all, there is no categorical difference, just a tendency or probability towards one of the two orders, which may be stronger when several factors that complicate processing are involved.

Under both De Sutter’s (2007) hypothesis and the one that follows from Meyer and Weerman (in press), speakers are expected to be more likely to use the more entrenched order in linguistic contexts that require more complex processing, as the more entrenched order is easier to process.

While both studies work from different theoretical backgrounds, both ideas of defaultness are compatible to some degree. Meyer and Weerman (in press) write “One question that remains central in verb cluster literature, and that is central in the present study, revolves around which of these orders is the default one (or, if the reader prefers, unmarked, basic or most economic – we take a neutral stand here), and which is derived”. While the notion of a derived form comes from generative grammar, it is clear that few assumptions on the nature of defaultness are made here. De Sutter (2007) does not explicitly define defaultness, though it is noted that the default order is the one that is used in situations do not have the time or the need to choose another order. Furthermore, De Sutter states that, on the basis of dialect research, one could conclude that the most frequently used word order (across dialects, presumably) in the most basic language variety in the most basic communicative situation, is the default order.

In this work, we would like to take a similar neutral stance as Meyer and Weerman (in press), taking economy or ease of processing as the main characteristic of a default word order. We will not commit to a particular theoretical framework’s notion on the nature of defaultness beyond what is mentioned here. This ‘ease of processing’ definition is compatible with De Sutter’s functional definition, though it is not necessarily compatible with his second point we paraphrased above: the most frequently used word order across dialects is not necessarily the most economical and therefore the default. For Dutch verb clusters, it appears to be the case that the 1-2 order spreads more easily in contact situations. It is more frequent in the western parts of the Netherlands which experienced more language contact. Furthermore, it has rapidly spread in the Frisian language area once Dutch-Frisian bilingualism became the norm there, even though Frisian is traditionally a language with the 2-1 order (Meyer et al., 2015; Hoekstra and Versloot, 2016). Therefore, it may as well be that the 1-2 order is more economical in Dutch. In a comparison between English, German and Dutch a similar trend is evident — English, the language with the most language contact, only allows 1-2 orders. German, with relatively little language contact, only allows 2-1 orders, while Dutch is in between, both in terms of contact and in terms of verb cluster word order (Meyer et al., 2015). We do agree that the most basic language variety is probably the most economical one, however, it is not clear whether the 2-1 order is still the most prevalent in spoken Dutch, because the data used in Stroop’s (2009) study of spoken Dutch did not take speaker age into account. In a recent study by Olthof et al. (2016), a change towards the 1-2 order in spoken language was observed in younger speakers in the Corpus of Spoken Dutch. It may be the case that only older speakers have a 2-1 order preference in informal spoken Dutch.

While the definitions of Meyer and Weerman (in press) and De Sutter (2007) are comparable, these two approaches lead to opposite predictions regarding the default cluster order. De Sutter’s hypothesis implies that a rather high-level process, sociostylistic encoding, is the main factor in deciding between these word order variations, a choice that speakers do not seem to be aware of. While this is possible, we believe that an

explanation in terms of more basic cognitive mechanisms would be more plausible, such as an entrenched 1-2 order suggested by the acquisition evidence and the work on cross-serial dependencies. Therefore, in addition to searching for factors of processing complexity, we will also compare our observations to these two default order hypotheses. There is of course a third possibility, that there is no default or most economical verb cluster word order at all. If our investigation results in insignificant evidence or conflicting evidence for both orders as the default, this would be a possible conclusion.

## 4 Method and data

To be able to discuss the effect that different factors have on verbal cluster order variation, we have created a multifactorial model from written corpus data. We model verbal cluster order as a binary variable, in which the orders can be 1-2 (ascending) or 2-1 (descending). This is the dependent variable. This variable is modeled in terms of the factors of the multivariate model, the independent variables, which will be discussed in the next section. In this, we follow the methodology of Bloem et al. (2014), but with some additional or differently operationalized variables, and an expanded data set.

Verbal cluster order variation is one of many language variation phenomena where it has been established that multiple factors contribute significantly to the observed variation. Such phenomena are best studied using multifactorial models, rather than testing each factor one by one. The statistical power of a multifactorial model is greater. Testing each factor while applying the necessary corrections for running multiple tests, also increases the chance of type II errors, a failure to reject a false null hypothesis. Starting with Gries (2001), such multifactorial models have been successfully used in the study of language variation. These models allow the researcher to examine how much a factor contributes to choosing either one or the other construction, while controlling for the other factors. In this way, the causes of a particular variation can be quantified in probabilistic terms, and the relative importance of each factor can be seen. Being corpus studies, these quantitative studies generally also emphasize evidence from larger samples of language and operationalize their factors in a precise way.

To obtain a large sample of verbal clusters, we used the Lassy Large corpus (van Noord, 2009) as our source of automatically annotated Dutch language data. This is a corpus of written Dutch that has been automatically annotated with full syntactic dependency trees by the Alpino parser for Dutch. This parser is currently the state of the art for Dutch, performing with an average concept accuracy (in terms of correct named dependencies) of 86.52% (van Noord, 2009). The use of this corpus limits the factors that can be investigated to those available in the corpus, which excludes accent and rhythm-related factors for example, as there is no prosodic annotation. From the corpus, we obtained the verbal cluster constructions of interest by defining them at the level of dependency syntax, extracting all matching constructions. This process avoids subjectivity, beyond the definition of the construction.

There are two main reasons why we chose to study this phenomenon using such a large corpus. Firstly, we are investigating probabilistic effects. In Dutch verbal cluster order variation, almost anything goes, but some combinations of factors may go together better than others. Probabilities and probabilistic associations are more accurately estimated when there is more data. Secondly, a large corpus allows us to control for information structure. Information-structural factors, such as givenness, are an important class of factors that are often invoked when explaining variation phenomena. However, they are not easily annotated, especially not automatically. Because they are not included in most corpora, the next best thing is to control for them. One way to do that is by having a very large corpus, in which many instances of the construction of interest will occur, in many different contexts with many different information-structural properties. This will balance out the effects of these properties.<sup>1</sup>

We extracted verbal clusters from the Wikipedia part of the Lassy Large corpus, which consists of the entirety of the Dutch version of the freely editable online encyclopedia Wikipedia on the 4th of August, 2011 (about 145 million words). From this data, 827.709

<sup>1</sup>A longer discussion on using automatically annotated corpora to study verbal cluster variation and an overview of some linguistic variation studies using such corpora is provided by Bloem et al. (2014).

two-verb verbal clusters were extracted, and 75.03% of these are in the 1-2 order. We chose to use this part of the corpus as it represents average standard Dutch. As discussed earlier, there is clear regional variation in verb cluster order preference, so this needs to be taken into account when choosing a corpus. While this corpus has been written by many speakers from different parts of the language area, non-standard language is likely to be edited out by other editors of Wikipedia. Furthermore, it has been argued that a corpus can still be representative of individual variation, when different language variants are included in a corpus. In a historical corpus study involving different variants of Frisian, Versloot (2008, p.64–66) showed that the number of variants of a particular word at a particular point in time correlates with the variance of that word within a single text at that time.

From this corpus, we extracted all two-verb verbal clusters from subordinate clauses, as defined in the introduction. This includes clusters with any sort of auxiliary, modal or grouping verb as the head, in contrast to earlier work using the same corpus (Bloem et al., 2014) in which a limited number of head verbs were investigated and in which main verb clusters were also included. This explains the difference in numbers.

Most of the factors we will discuss have also been tested for associations with verbal cluster order by De Sutter (2009) and validated by Bloem et al. (2014). Nevertheless, because we are testing a few additional factors that relate to processing complexity, we replicated the whole experiment and will report on the new results. Because the model is multifactorial, we cannot leave out factors that were already tested before. We also included more different cluster types in the data set. De Sutter (2009) used a limited set of auxiliary clusters from a smaller, manually annotated newspaper corpus. Bloem et al. (2014) included modal clusters and extracted data from the Lassy Large corpus, the same corpus that we are using here. This work includes the other types of two-verb verbal clusters described in section 2.3, resulting in a larger data set. Furthermore, we have defined some factors more precisely than previous studies did. Nevertheless, our focus in this work will not be on comparing these results to previous work, but rather on checking which factors that relate to processing complexity, are also associated with a particular order. We will also check whether these results are consistent: are many of the factors that are more difficult to process, associated with one particular order? The results described in section 6 are largely consistent with those of Bloem et al. (2014) in cases where comparison is possible, unless specified otherwise.

## 5 Measures of processing complexity

In this section, we will examine some factors that are empirically measurable for verbal cluster constructions, discuss how they relate to processing complexity, and how we operationalize them in our corpus study. Most of the factors have been discussed in previous literature as being relevant to verbal cluster order, or to other word order variation phenomena. Most of the factors were already tested for associations with verbal cluster order by De Sutter (2009) and validated by Bloem et al. (2014). Only factors for which the values could be automatically extracted from a written corpus were selected. We do not aim to provide a complete overview of studies linking these factors to verb cluster order variation — such background can be found in De Sutter (2005) for most of the factors. Rather, our focus here is on establishing connections between these factors and the notion of processing complexity by discussing empirical, mostly psycholinguistic work in which the factors, or similar factors, are discussed.

Processing complexity refers to the amount of cognitive resources or cognitive effort required to produce or comprehend an utterance. Processing complexity can be measured using methods from psycholinguistics, such as reaction time or reading time. Faster times indicate faster processing, which indicates lower complexity. Evidence for ease or complexity of processing ideally comes from such psycholinguistic studies. However, there isn't always direct evidence, so we also discuss theoretical work that may predict higher or lower processing effort for particular factors. Examples of such theoretical work are the resource-limitation and constraint-based models discussed in section 2.2. These models make predictions regarding the complexity of processing certain kinds of constructions.

Effects of cognitive effort can be found throughout different domains of language. In the domain of pragmatics and discourse, information structure seems to affect the

processing depth or processing efficiency of utterances. For example, reaction times to tasks involving words that are in focus, are faster. Wang et al. (2014) summarize various processing effects in this domain. In morphology, morphological complexity has been used as a basis for constraints on possible combinations of morphemes, and has also been related to processing cost (Plag and Baayen, 2009). And in the lexicon, lexical access speed can be seen as an indication of complexity, for example in the theory of Gibson (1998).

There are many models of processing complexity, focusing on different aspects of language processing. Our summary in section 2.2 discussed only a few. We will not define processing complexity in terms of any particular model of sentence processing, because as we discussed in the previous section, verbal cluster variation has already been shown to be influenced by factors related to various domains of linguistics. We are not aware of any single model of processing complexity that covers all of these domains, from word frequency to discourse integration, and there is also the aforementioned problem that most work has been done on interconstituent alternations, while we are dealing with an intraconstituent alternation. Instead, we will start from these factors which were previously identified, and some additional ones commonly used in grammatical variation studies that relate to complexity, and discuss studies that empirically measure processing complexity in their domain.

**Syntactic priming** Linguistic priming is the idea that units of language get activated when they are perceived or used, and some traces of this activation remain for a while, making easier and/or faster to re-activate them. The phenomenon was first considered at the level of words, but was later applied to other linguistic units as well. This effect is often measured in a priming task, where participants will have a faster reaction to a word or pattern that they have seen before, even if they did not consciously perceive it (in the case of masked priming).

A syntactic priming effect was first tested experimentally by Bock (1986). In a primed sentence production task, she found that particular syntactic forms were more likely to be used when that form had been used in the priming sentence, even though other connections between the prime and the sentence to be uttered were controlled for and minimized. The two alternations tested in the experiment were transitives in active and passive voice, and the dative alternation. Syntactic priming for the same constructions was later found in Dutch as well (Hartsuiker and Kolk, 1998).

A large number of studies have been conducted to show that this priming effect is not lexical, but rather syntactic or constructional. It cannot be entirely attributed to other factors such as priming of thematic roles (see Pickering and Ferreira 2008 for an overview of evidence). Syntactic priming in verbal clusters was studied by Hartsuiker and Westenberg (2000), who showed that verbal cluster orders can be primed. They chose verbal cluster constructions to try to exclude any effects from semantics from their study, an effect they call ‘conceptual priming’. They make the assumption that the order variation is meaningless because the auxiliary verb of the cluster does not have intrinsic meaning. However, some semantic associations with verbal cluster order have been found (Bloem, 2016), and it can be argued that meaning arises from such associations. In addition, if phenomena like the past tense are viewed as a construction, auxiliary verbs express this meaning.

Nevertheless, their experiment showed that verbal cluster order is affected by a priming effect in spoken language. Later, De Sutter (2005) also observed such an effect in a corpus of Dutch newspaper text, though he discusses the possibility that it was a stylistic choice instead. He did not find a significant effect of priming distance, meaning how much text there was between the first instance of a cluster, the prime, and the second cluster. Such an effect might be expected due to gradual decay of the level of activation of the construction, though it has been observed in structural priming studies that activation from priming can last for a longer time. Bock and Griffin (2000) found no difference in the strength of priming between conditions where the prime and the target were separated by 0, 1, or 10 sentences, suggesting no decline within this timeframe. Nevertheless, De Sutter (2007) notes that this syntactic persistence effect he found in the newspaper corpus can be linked to processing ease, as re-using a recently activated pattern can reduce processing costs.

We consider priming as a factor related to complexity, because verbal clusters with a primed word order are easier to process. People react faster to primed stimuli and are more likely to reproduce them. If we observe sequences of the same verbal cluster order, this likely comes from a priming effect, and this indicates that processing difficulty does indeed affect verbal cluster order.

In operationalizing this factor, we distinguished three cases. When a verbal cluster is the first verbal cluster of the text (a text is a Wikipedia article from the corpus), we consider it unprimed, and it is categorized as ‘none’. When a verbal cluster is not the first verbal cluster of the text, we looked back and checked whether the previous verbal cluster was in 1-2 or 2-1 order. These are the two possible ‘primed’ states. Because Wikipedia articles may be edited by multiple people, there is not guaranteed to have been an actual priming effect — perhaps one cluster was produced, and then another cluster was inserted before it by a different author subsequently. In this case, it cannot be the case that the second cluster was primed by the first. Unfortunately that cannot be helped with this data set. It would not be helpful to exclude the variable either, because then the model may attribute the orders affected by priming to some other factor. Furthermore, there is independent evidence of verb cluster priming effects (Hartsuiker and Westenberg, 2000).

**Morphological structure of the main verb** The Dutch language has a variety of morphologically complex verbs. Some types of complex verbs can be affected by syntactic processes, thus making these verbs a complicated phenomenon on the boundary of morphology and syntax. We want to distinguish two types of complex verbs, which differ in complexity. Firstly, there are inseparable complex verbs such as *volbrengen* (to complete), where *vol* is a distinct morpheme meaning ‘full’. Secondly, there are also separable complex verbs, which contain a particle that may appear in a different place in the sentence in some contexts, such as *volhouden* (to endure, to keep up). One notable difference between them is that in the first type the stress is on the second syllable, but in the second type the stress is on the first syllable (when unseparated), the separable particle. This fact has been used to explain this factor as a rhythmic factor in previous work. However, there is a link with processing complexity as well.

The following examples illustrate the difference between the two types of complex verb:

- (13) *Ze vragen Anne of zij de wedstrijd volhoudt.*  
They ask Anne if she the match endures.  
‘They ask Anne if she can play out the match.’
- (14) \* *Anne volhoudt de wedstrijd.*
- (15) *Anne houdt de wedstrijd vol.*  
Anne endures the match PART.  
‘Anne endures the match.’
- (16) *Ze vragen Anne of zij de opdracht volbrengt.*  
They ask Anne if she the assignment completes.  
‘They ask Anne if she will complete the assignment.’
- (17) *Anne volbrengt de opdracht.*  
Anne completes the assignment.  
‘Anne completes the assignment.’
- (18) \* *Anne brengt de opdracht vol.*

In Dutch subordinate clauses, the finite verb appears at the end, with the particle directly in front of it (written together in Dutch orthography), as in (13). In main clauses, the finite verb has to appear in second position due to the V2 effect, while the particle still appears at the end, as in (15). Having the particle in second position as well is ungrammatical (14). Inseparable complex verbs show the opposite behaviour — leaving the morpheme in final position is ungrammatical (18). This factor has been associated with verb cluster ordering for a long time, i.e. by Meeussen and Vanacker (1951), who

claims that separable verbs only occur in the 2-1 order when there is no stress previous to it in the sentence.

According to Booij (2010), inseparable complex verbs should be considered as words, while separable complex verbs are neither words nor phrases. The principle of Lexical Integrity states that syntactic rules cannot operate on parts of words, therefore they cannot be words. Yet they also have word-like properties — in verbal clusters, they may occur in unseparated form and must have undergone some syntactic process as a syntactic unit. Their word stress patterns also follow those of single words (Booij, 2002). Booij considers them to be something in between words and phrases, and calls them ‘constructional idioms’. Similar claims have been made about particle verbs in German — Müller (2002) argues that they are syntactic combinations, not just morphological objects.

Particle verbs such as *afwassen* (to wash **up**) are one type of separable complex verb, which also can also be found in English. Other types of separable verbs include adjectives, e.g. *schoonmaken* (**clean**-make), nouns, e.g. *piano spelen* (**piano**-play), or even morphemes that do not occur independently, e.g. *teleurstellen* (disappoint) (Booij, 2002). We group all of these under the class of separable complex verbs. Separable complex verbs should be more difficult to process than other types of verbs, complex or non-complex, due to their phrase-like properties. Non-separable complex verbs behave like simplex verbs syntactically — besides not separating, they also cannot have inflectional markers intervening between the prefix and the verb root (Behrens, 1998, p. 704). While both separable and non-separable complex verbs may need to undergo morphological processing, only the phrase-like separable verbs need to undergo syntactic processing as well. Using evidence from priming studies, it has been argued that German complex verbs are morphologically decomposed in processing (Lüttmann et al., 2011; Smolka et al., 2015). Lüttmann et al. (2011) note that this is the case both for separable and non-separable complex verbs. However, these priming studies do not present the verbs in contexts in which they might be separated, presenting only single words or word pairs, and therefore these psycholinguistic studies do not tell us whether separable complex verbs are more difficult to process than nonseparable complex verbs. We are not aware of any study that tests this directly. However, it seems plausible for theoretical reasons, so we will assume that processing complexity is higher in the case where verbs are complex and separable.

In our study, this factor is operationalized such that any cluster where the main verb is annotated as a particle verb in the corpus, or any cluster with a separate particle in it, is considered a cluster with a separable verb, while others are non-separable.

**Length of the middle field** The linear distance between a syntactic head and its complement is considered to be a source of processing complexity, particularly in resource-limitation models of language processing. In these models, resolving a dependency between a head and a complement comes with an ‘integration cost’, and this cost depends on the length of the dependency (Gibson, 1998). We therefore consider the length of the middle field of the clause that the verbal cluster occurs in, to be a measure of processing complexity.

We follow the definition of De Sutter (2005, p. 153-154) in saying that the length of the middle field is the number of words between the connective of the start of the clause (or if there is none, just the start of the clause) and the verbal cluster, not counting elements such as particles or infinitival markers at the start of the cluster. De Sutter (2005) interprets this factor as word stress related: a longer middle field increases the chance of an accented morpheme appearing directly before the verbal cluster, causing it to clash with the accented syllable in the main verb. Speakers would then prefer putting the main verb in the final position to avoid this clash, causing 1-2 orders.

De Sutter (2007) discusses this factor in terms of processing, and interprets it in terms of a probabilistic model of processing: a longer middle field is more likely to provide specific cues on the nature of the upcoming main verb, leading to a more accurate prediction and a lower processing cost of the main verb. However, an opposite prediction can be made in terms of resource-limitation models of language processing: a longer middle field is more likely to lead to long dependency lengths and more dependencies between various parts of the middle field and the verb cluster, leading to higher integration costs when

those dependencies are resolved and attached to the verb cluster head (or main verb) when the verb cluster is processed. The larger number of words and constructions that need to be processed in a longer middle field should also add to the processing cost. It is also not necessarily the case that longer middle fields allow better prediction of the main verb — if there is just a single word in the middle field that has a high association strength with the main verb, this will be more informative than a large middle field that has a low association strength with the main verb. This is something that could be empirically tested with a statistical language model, but we are not aware of any such work.

Either way, we can conclude that the length of the middle field is a measurement of complexity, and we will assume that longer sentences cost more resources to process.

In our study, this factor is operationalized by taking the position in the sentence of the first verb in the cluster, and subtracting the position of the start of the subordinate clause.

**Syntactic depth of embedding** The processing complexity of sentences is often linked to their syntactic structure. The notion of center-embedding has been essential in discussions on the computational complexity of human language, starting with the claim that English is context-free and computationally a non-regular language, because it contains center-embedded structures that cannot be produced by a regular language (Chomsky, 1957). Limits on the complexity of these structures in corpus data were attributed to working memory limitations.

We define embedding in a very general sense, rather than just testing for center-embedding. We consider the depth of embedding to be the number of nodes in the syntactic tree of the sentence between the main verb of the verbal cluster and the head of the entire sentence. This is similar to the notion of ‘structural distance’ (O’Grady, 1997), except that structural distances are measured between two particular words or constituents, rather than to the root node.

In the processing literature, this factor is particularly important in constraint-based models of language processing. While resource-limitation models focus more on linear distance, constraint-based models focus more on expectation and prediction, often taking hierarchical structure into account to determine whether something is expected in a given context. It can therefore also be considered a measure of processing complexity.

Of course, the importance of this measure depends on how much of the apparent hierarchical syntactic structure of sentences is cognitively real. It has also been argued that language is not (or hardly ever) processed hierarchically, but rather sequentially. Frank et al. (2012) outline such a model, citing evidence from cognitive neuroscience, psycholinguistics and computational models. In this model, linguistic input is processed as a stream of components that have a linear order, rather than a tree of constituents. Embedding is therefore not a relevant notion in this model, and cannot be related to processing effects. In this view, embedding would not lead to a higher processing load. On the basis of generative theories of language, where embedding, caused by recursive processes, is a basic property of grammar, this point can also be argued. In this view, deeper embedding does not necessarily lead to a higher processing load, unless some kind of extrinsic, non-linguistic performance limit is hit and grammatical processing breaks down, as initially suggested by Miller and Chomsky (1963). From this, and from the observation that recursion is central to many human cognitive processes, it follows that more deeply embedded clauses would not be more difficult to process.

While it is a complex theoretical issue, we will for now assume that processing more deeply embedded structures is more complex.

In our study, this factor is operationalized by counting how far down in the syntactic tree of the entire sentence the head verb of the cluster is located. Each branching point of the tree is counted as one level. The trees of the Lassy Large corpus use a dependency tree format which is not binary-branching (van Noord, 2009). A visualized example of such a tree can be found in van Noord (2009, p. 119).

**Information value of the last preverbal word** This factor concerns the word class of the word that is used directly before the verbal cluster. The main reason why we chose to test only the last word before the verbal cluster is to allow comparisons to De Sutter



(2005), who did the same thing, but it may also be interesting for reasons of processing. Some late processing may take place on this word, affecting verbal cluster processing as well, particularly when function words are involved. In an ERP study, Hoen and Dominey (2000) found a left anterior negativity (indicating working memory load) between 400 and 700 ms after presentation of a function word. This was hypothesized to be an effect of predicting the next constituent, which is easier after a function word.

De Sutter (2005) found that the information value of the last preverbal word affects the order of the verbal cluster, linking it to prosody. De Sutter (2007) linked it to probabilistic processing, in the sense that a more informative word before the verb cluster will make prediction of the main verb, and therefore processing, easier. We do not see a particular reason why a highly informative word (such as a noun) before the cluster facilitates prediction more than a highly informative word earlier in the clause, i.e. an animate subject noun that might only occur with certain main verbs. De Sutter (2007) provides the following examples, where 19 shows a preverbal word with a high information value (a noun) and 20 shows a preverbal word with medium information value (an adverb):

- (19) ... *dat grote stukken landbouwgrond en infrastructuurwerken zijn verwoest*  
 ... that large tracts farmland and infrastructure were destroyed  
 ‘... that large tracts of farmland and infrastructure have been destroyed.’ De Sutter (2007, Ex. 29)
- (20) ... *dat ze al gebruikt zijn*  
 ... that they already used were  
 ‘... that they have already been used.’ De Sutter (2007, Ex. 30)

However, the following constructed combination of those sentences is also perfectly normal:

- (21) ... *dat grote stukken landbouwgrond en infrastructuurwerken al zijn verwoest*  
 ... that large tracts farmland and infrastructure already were  
 destroyed  
 ‘... that large tracts of farmland and infrastructure have already been destroyed.’

Even though the last preverbal word has medium information value here, we would argue that the clause is more informative than example (19), because it contains more information overall. Therefore, an observed effect of information value of the preverbal word does not necessarily mean the verb cluster was more predictable — it could have been predicted by something that came earlier in the clause, too.

Alternatively, this factor can be interpreted in terms of processing of the preverbal word itself. If that particular word requires more processing, or late processing because a functional projection needs to be created, this may affect the processing of the verb cluster. If the preverbal word is difficult to process, there may be fewer resources available to process the verb cluster. However, the literature is inconclusive on whether high-informative open-class words or low-informative function words are more difficult to process.

In psycholinguistics, differences have been observed in lexical access to open-class words (nouns, verbs) and closed-class words (function words). Bradley (1978) described a difference between the two categories: lexical decision response times to open-class words depend on their frequency, while those to closed-class words do not. This was taken as evidence for different processing mechanisms. More recently, Segalowitz and Lane (2000) argue against this model and the lexical decision methodology, and show that function words are accessed more quickly in general. This appears to be an unresolved issue, so while we can clearly state that this is a factor relating to processing complexity, it is not so clear which of the factor’s levels is actually the more complex one.

We operationalize this factor in the same way as De Sutter (2007) and Bloem et al. (2014), in terms of three classes: **highly** informational (nouns, verbs, numerals), **intermediately** informational (adjectives and adverbs) and **low** informational (pronouns, conjunctions and prepositions). The word class is taken from the annotation of the corpus.

**Definiteness of the last preverbal word** We will also discuss another property that these preverbal constituents can have: definiteness. This factor has been discussed in previous work, but has not been included in a multifactorial model before. De Sutter et al. (2007) viewed it as another measure of information weight, closely related to the aforementioned ‘information value’ factor. In our view, this property can also be related to processing rather than ‘information value’, as it indicates discourse-givenness of the referent. However, it only applies to noun phrases. In this work, we will not directly test givenness since this is very difficult to determine using automatic annotation methods. Givenness involves long-distance referencing in the discourse, across sentence or even paragraph boundaries, and we are not aware of any reliable automatic method for detecting it.

A definite noun phrase has usually been previously given in the discourse and is assumed to be already accessible, facilitating processing. Givenness also leads to phonological reduction, a further indication that given items are easier to process. However, it is often assumed that definite constituents have to be linked to a previously given discourse element, while indefinite constituents do not (e.g. by Heim 1982). This may be a complex process. Crain and Steedman (1985)’s discourse-based theory of NP attachment makes predictions along these lines, stating that indefinite NPs should be processed more easily because no additional sets of discourse objects are presupposed. This idea was later experimentally confirmed by Gibson et al. (1996), although this study only used a questionnaire, rather than a psycholinguistic measure.

Furthermore, studies of impaired populations of speakers suggest that definites are more difficult to process: Polite et al. (2011) tested English-speaking children with specific language impairment (SLI) and showed that they produce fewer definite articles than typically developing children in definite contexts, while there was no significant difference in indefinite article use. Testing Dutch children, Schaeffer et al. (2014) find a strong effect in this direction in children with High Functioning Autism (HFA), as well as those with SLI. Experimental results by Schaeffer et al. (2016) show that the difficulty that SLI and HFA children have, have different underlying causes: for HFA children, the issue appears to be related to a pragmatic impairment, while for the SLI children, who perform worse on tests of morphosyntax and working memory, it is a problem in maintaining discourse continuity. While the problems that HFA children have may not be indicative of general processing difficulties, but instead a pragmatic impairment, the findings for SLI children are relevant. These findings show that discourse integration, a higher-level cognitive process, takes place in the processing of definites, and that a population with fewer cognitive resources in the linguistic and working memory domains (children with SLI) has particular problems with it. We can therefore take this as evidence that definites are more difficult to process than indefinites.

In the present study, this factor is operationalized with three possible levels: a preverbal word is considered definite when it is a noun that is in a dependency relation with a definite determiner. When it is a noun in a dependency relation with an indefinite determiner, it is considered indefinite, and in any other case, it is considered ‘unknown’.

**Frequency of the main verb** Psycholinguistic theories generally assume that more frequent words are more easily activated, making them easier to process. This assumption can be found in the theory of Gibson (1998) mentioned in the section on syntactic depth: “high frequency lexical items require fewer energy resources to become activated than low frequency lexical items do”. Therefore, more frequent words are less complex to process. The frequency of the main verb is operationalized by counting the frequency of the verb lemma in the entire corpus (not only in verb clusters), and taking the logarithm of it.

**Extraposition** Extraposition of the prepositional object, also called PP-over-V, is another source of variation that can occur around verbal clusters, and one that may be linked to complexity as well. Even though Dutch embedded clauses are normally verb-final, when the main verb of the verbal cluster has a prepositional object, the object can optionally be positioned after the verbal cluster:

- (22) ... waar hij *als consul* aangesteld werd.  
 ... where he as consul appointed was.  
 ‘where he was appointed as consul.’

- (23) ... *wat werd veroverd* **door Hettieten**.  
 ... that was conquered by Hittites.  
 ‘that was conquered by Hittites.’

Example (23) shows the extraposition. Similar to verbal cluster order variation, there is no clear meaning difference between the two. This type of extraposition has been linked in previous work to various complexity-related factors that we also discussed in this study: inherence, morphological complexity, definiteness and heaviness (length), as well as habituality, information pattern, intonation pattern and register (Jansen, 1978; Canegem-Ardijns, 2006). Willems and De Sutter (2015) performed a multivariate corpus study of PP placement studying the effect of the weight of the middle field, the postfield and the extraposed PP itself. They found that the weight of the PP is the main factor, and that extraposition into heavy postfields is less likely. Weight was operationalized in terms of length and complexity (embedded structures), similar to our factors `LENGTH OF THE MIDDLE FIELD` and `DEPTH OF EMBEDDING`. This shows that complexity affects PP extraposition in some ways as well, and that the order variation caused by extraposition may be comparable to verbal cluster order variation. Willems and De Sutter (2015) state that their model requires more predictor variables because its performance leaves room for improvement, perhaps some of this remaining variation can be explained using other factors of processing complexity.

One interesting observation by Willems and De Sutter (2015) is that they find extraposition to be more common than non-extraposition. They find 58.4% extraposition based on 1,850 clauses from the Dutch Parallel Corpus of journalistic texts. If extraposition is such a common process, it might be expected that its processing difficulty is not very high — if it was difficult, the alternative of non-extraposition would be used instead. Furthermore, a commonly used construction is likely to be more entrenched, and according to constraint-based processing models, more predictable and therefore easier to process.

These studies show that extraposition is used to make a subordinate clause less difficult to process, but perhaps a direct link between extraposition and ease of processing can be found. We are not aware of any psycholinguistic study of PP-over-V extraposition, but a study by Levy and Keller (2013) on processing German verb-final structures may provide some insight. The German language is also verb-final, and also has verbal clusters that are theorized to have the same underlying mechanisms as Dutch verbal clusters. What Levy and Keller (2013) found is that the verb is more difficult to process when both an adjunct and a dative NP are present in a relative clause, compared to conditions where only one of them was present. Importantly, they only tested verb-final clauses in their relative clause experiment (Experiment 2), and controlled for sentence length. They did not test other subordinate clauses besides relative clauses.

In Dutch, PP adjuncts and dative NPs are elements that can be extraposed. If the processing difficulty observed by Levy and Keller (2013) is due to the presence of arguments before the verb, we would expect that subordinate clauses with extraposition are easier to process than those without. When there is extraposition, arguments that would have been before the verb come after the verb instead. Unfortunately Levy and Keller (2013) did not have a condition with arguments after the verb, as this is not possible in German. Nevertheless, if we assume that Levy and Keller’s (2013) findings also apply to Dutch, this would match what Van Haeringen (1956) already hypothesized. He claimed that extraposition may alleviate some of the ‘tension’ involved with not producing the verb until the end of the sentence, when the sentence is long. It would also help explain why extraposition was found to be associated with various complexity-related variables, e.g. by Willems and De Sutter (2015).

Levy and Keller’s (2013) study discussed PP adjuncts, but in Dutch, PP complements can also be extraposed. When De Sutter (2009) tested the effects on extraposition on verbal cluster order, he included a distinction between extraposition of adjuncts and extraposition of complements. De Sutter (2007) argues that complements are more costly to keep in memory than adjuncts due to their higher information weight, making a complement more difficult to process when extraposed — it has to be remembered for longer. One might also expect a tendency for the main verb and the complement to be closer together, following the adjacency principle. Complements are more strongly related

to the main verb semantically, which might cause additional processing difficulty if there is a great distance between them. For adjuncts, this is not the case. Example (23) shows an extraposed adjunct, example (24) contains an extraposed complement, as evidenced by the fact that the extraposed constituent is required to complete the meaning of the verb.

- (24) ... *waar hij aangesteld werd als consul.*  
 ... where he appointed was as consul.  
 ‘where he was appointed as consul.’

In summary, Levy and Keller’s (2013) findings are the most closely related psycholinguistic evidence that we could find, so we will take this to mean that processing verbal clusters with extraposed elements is easier than processing clusters without extraposition. However, it should be noted that the evidence comes from a somewhat different context.

The extraposition factor is operationalized as follows. When the verb cluster is the last element in the linear order of the sentence, we label this as ‘nothing’. When an adjunct follows the verb cluster in the clause, we consider this an extraposed adjunct. When a complement follows the verb cluster, this is considered an extraposed complement. When there is some other material after the verb cluster, such as another subordinate clause or part of the main clause, this is labeled as ‘Other continuation’. When there is punctuation after the verb cluster, followed by some other material, this is considered ‘Punctuation and continuation’. These distinctions are mainly made for technical reasons.

**Multi-word units** A multi-word unit (MWU) is a lexical unit, consisting of multiple words that together carry a different meaning than that of its constituent parts. They are also known as formulaic sequences. Collocations and idioms are types of multi-word units. Evidence from psycholinguistic studies, such as the self-paced reading experiment of Conklin and Schmitt (2008) and the eye-movement study of Underwood et al. (2004), suggests that MWUs are processed faster and remembered better, even when they span across syntactic boundaries (Tremblay et al., 2011). However, there are a few things to consider when multi-word units include (parts of) verbal clusters.

In his thesis on verbal cluster order, de Sutter discussed the similar concept of ‘inherence’, preverbal constituents that have a strong semantic connection with the main verb. In his 2005 study, he hypothesizes an association between this phenomenon and the 1-2 order, for reasons of prosody. An inherent preverbal constituent is assumed to have an accent, and the main verb has an accent too. The main verb would then be more likely to be uttered in the final position, to avoid disrupting the sentence rhythm. He finds that this is not true for regular verb phrases that fit the definition of inherence, but that there is an effect for fixed expressions, which he defines as any collocation between a verb and one or more non-verbs in the verb phrase. Collocations are defined using log-likelihood ratio tests, using statistical significance at  $\alpha = 0.05$  as a cutoff point for collocation-hood, and only collocations between two words (one verb and one non-verb) were tested. De Sutter (2007) also discusses this factor, again interpreting it in probabilistic processing terms — hearing the first part of a multi-word unit before a cluster will make it easier to predict what the main verb of the cluster will be.

We also interpret this finding as a complexity-related one, given the clear evidence for a relation between MWUs and processing difficulty discussed earlier. In this study, we will test several types of MWU constructions involving verbal clusters for associations with the 1-2 or 2-1 orders. Firstly, we can make a distinction between constructions with extraposable PP arguments and those without. MWU constructions without a prepositional object involve at least one verb of the cluster, and optionally something before the cluster:

- (25) ... *dat hij de revolutie de kop indrukken zou*  
 ... that he the revolution the head push-in would  
 ‘... that he would quell the revolution’

MWU constructions can also take a prepositional phrase as their object, which is not part of the multi-word unit:

- (26) ... dat hij **rekening** zou **houden met** de buren  
 ... that he account would hold with the neighbours  
 ‘... that he would take the neighbours into account’

In this example, the prepositional object *de buren* (the neighbours) is an open slot. It could be replaced with a wide variety of things that can be prepositional objects, such as *zijn moeder* (his mother), a conjunction like *zowel Nederlandstaligen als Franstaligen* (both the Dutch and French speaking) or even other multiword units such as *alles en iedereen* (everything and everyone).

Because these constructions have a prepositional object, extraposition as discussed in the previous section can also take place here. Example (26) is in fact an example of it. Both of these options are grammatical:

- (27) ... dat hij **rekening** zou **houden met** de buren  
 (28) ... dat hij **rekening met** de buren zou **houden**

We have included this distinction in the model as well for completeness’ sake, and to find possible effects of continuous and discontinuous sequences.

In summary, we assume that clusters with MWUs and transitive verbs, where there is no prepositional object to extrapose, are easier to process. When there is a prepositional object, we expect more processing difficulty since some syntactic processing will have to take place.

The LASSY Large corpus includes annotation of multi-word units, so we were able to extract this information without having to compute which words form multi-word units. The way multi-word units are annotated in the corpus is described by Odijk (2013). The factor is operationalized based on this information from the corpus. If one of the verbs was annotated as a multi-word part, it was categorized as a case of *INHERENCE*, with different values marking where the rest of the multi-word expression went (before the cluster, after the cluster, or there was no non-verbal component).

## 6 Results

In this section, we present two types of results. Firstly, table 1 lists all of the explanatory variables used in the model, along with their effect size. The model uses these variables to predict the binary dependent variable of verbal cluster word order. Secondly, table 2 presents the same variables, sorted by information gain. This allows us to see which of the variables contribute the most towards statistically explaining the observed variation. Unless noted otherwise, all of our factors account for differences significant at the  $\alpha = 0.001$  level. When working with huge, automatically-annotated corpora, sample sizes are so large that p-values become practically meaningless and everything becomes statistically significant. P-values depend on sample size. We therefore choose to focus on effect sizes and information gain instead.

Table 1 allows us to explore the effect of each factor on verbal cluster order. Most of the explanatory variables are categorical, meaning that they have several possible values which are not necessarily ordered. For example, the variable *TYPE OF AUXILIARY* has different (categories of) verbs as possible values, such as grouping verbs and modal verbs, but we cannot say that grouping verbs are ‘greater’ or ‘smaller’ than modal verbs. For these variables, one of the possible values was taken as the reference value, or baseline, to which all the other values are compared. For the variable *TYPE OF AUXILIARY*, the auxiliary verb *zijn* was taken as the baseline. The effect size listed for the verb *hebben* then indicates the odds of a 1-2 order in case there is the auxiliary verb *hebben* instead of *zijn*, all other conditions being the same. The effect size of each variable is given as an odds ratio. An odds ratio further from 1 in either direction indicates a stronger effect.

For example, these odds ratios show that in our data, *hebben*-clusters are two times more likely to be in the 1-2 order than *zijn*-clusters, while controlling for the other factors. Note that this also implies the reverse: *zijn*-clusters are two times more likely to be in the 2-1 order than *hebben*-clusters, because there are only two possible orders. It is simply a matter of which baseline we choose and which order (1-2 or 2-1) we decided to report on. The baselines were selected to be either the same as those in previous work, the most frequent value, or the most basic value (i.e. ‘none’). We will be reporting values for the

1-2 order (ascending or participle-final), because De Sutter (2009) also did this. As a result, odds ratios  $> 1$  indicate an effect of association with the 1-2 word order, odds ratios  $< 1$  indicate 2-1 order.

The variables `MAIN VERB LOG-FREQUENCY`, `SYNTACTIC DEPTH` and `LENGTH OF MIDDLE FIELD` are not categorical, but continuous. In these cases, an odds ratio cannot be interpreted in the same way. Instead, we show the  $\beta$  estimate, a different measure of effect size. It represents an effect on the dependent variable, cluster order, in terms of standard deviations.

Table 2 gives us a ranking of the variables, in order of their contribution to the model. This is expressed as an Akaike Information Criterion (AIC) value, which is a measure of information loss. The model with the lowest AIC loses the least information, compared to the data it models.

The ranking was produced through stepwise regression. In this procedure, we start with a model with no independent variables (explaining nothing). Such a model has the maximum possible information loss (AIC). We then calculate which of the variables, when added to the model, would cause the greatest decrease in AIC. This variable will be ranked 1, and added to the model. Then we repeat the procedure. Out of all the remaining variables, we calculate which one would cause the greatest decrease in information loss, and add it. This variable will be ranked 2. We repeat the procedure until we ran out of variables, or the remaining variables do not decrease the information loss any more. In our case, we ran out of variables — all of the variables contribute to the model.

We will start with some general observations. The 1-2 order is more frequent overall. Even if we take only the clusters with the auxiliary verb ‘zijn’ (to be), which are relatively more likely to be in the 2-1 order, the 1-2 order is still the most common in terms of absolute frequency, with 57.39% of such clusters occurring in the 1-2 order for this construction. The reason for this relative preference is that the verb ‘zijn’ commonly functions as a copula verb, in which case the 2-1 order is more likely (according to De Sutter 2009) or even required (according to prescriptive grammars).

In the remainder of this section, we discuss the effect sizes of all the factors from section 5, listed in table 1.

Firstly, there are two factors included in the model that were included as controls, and we will not focus on them here: `TYPE OF AUXILIARY` and `‘TE’-INFINITIVE`. While these factors are associated with verbal cluster order variation, and therefore should be included to have an accurate model of verbal cluster variation, they have a clear meaning component, and can be considered to be different constructions on the basis of that. They control for the different verbal cluster constructions discussed in section 2.3 — `TYPE OF AUXILIARY` refers to the type of auxiliary verb used in the construction, and `‘TE’-INFINITIVE` encodes whether the construction was a *te*-infinitive construction. Following Bloem et al. (2014, p. 1979), auxiliary types with similar word order preferences and similar function are grouped. As we discussed in the introduction, it is not controversial that constructions, which are pairs of form and function, can be distinguished on the basis of their meaning. Such constructions can have different word order preferences simply for historical reasons or due to frequency effects, such as the modal verb clusters which are known to occur almost exclusively in the 1-2 order (and this is what we also found). Besides a control variable for the type of auxiliary, one might also expect a control variable for the main verb, or type of main verb. While this would probably improve the model, there were too many different main verbs in the data for the model to handle. However, effects related to the main verb have been observed (Bloem, 2016), and therefore it would be good to find a method of incorporating these effects in future work.

In table 2, we can observe that the two control variables are quite important in explaining the observed variation. The factor `TYPE OF AUXILIARY` is ranked first, with a large decrease in AIC. This follows from the fact that this factor distinguishes between different constructions with different meanings, which sometimes have clear word order preferences. A large part of this can probably be attributed to the difference between modal clusters and other clusters — as noted, modal clusters occur in the 1-2 order 99.5% of the time. This is a large deviation from the average, which is 75.03%. The factor `TYPE OF AUXILIARY` accounts for this, and is therefore an integral part of the

| Variable                                 | Categories                        | Odds ratio  | Frequency |
|--|-----------------------------------|-------------|-----------|
| TYPE OF AUXILIARY                        | <i>zijn</i>                       | <b>1.00</b> | 191.811   |
|  | <i>worden</i>                     | 1.47        | 277.531   |
|  | <i>hebben</i>                     | 2.00        | 97.969    |
|  | Grouping verb                     | 9.65        | 55.157    |
|  | Modal verb                        | 159.27      | 189.470   |
|  | Causal verb                       | 1135.14     | 14.855    |
| ‘TE’-INFINITIVE                          | No                                | <b>1.00</b> | 764.312   |
|  | Yes                               | 0.29        | 63.397    |
| PRIMING                                  | None                              | <b>1.00</b> | 228.227   |
|  | Primed for 1-2                    | 1.24        | 444.950   |
|  | Primed for 2-1                    | 0.59        | 154.482   |
| MORPHOLOGICAL STRUCTURE OF THE MAIN VERB | Non-separable                     | <b>1.00</b> | 810.401   |
|  | Separable                         | 7.06        | 17.308    |
| LENGTH OF MIDDLE FIELD                   | Per additional word               | 1.0390      | —         |
| SYNTACTIC DEPTH                          | Per additional level of embedding | 0.95        | —         |
| INFORMATION VALUE                        | Low                               | <b>1.00</b> | 416.663   |
|  | Intermediate                      | 1.00        | 139.715   |
|  | High                              | 0.79        | 240.170   |
| DEFINITENESS                             | Indefinite                        | <b>1.00</b> | 22.657    |
|  | Definite                          | 1.04        | 70.231    |
|  | Unknown/NA                        | 1.27        | 734.821   |
| FREQUENCY OF MAIN VERB (LOG)             | Per additional order of magnitude | 1.06        | —         |
| EXTRAPOSITION                            | Nothing                           | <b>1.00</b> | 329.211   |
|  | Extrapolated adjunct              | 0.75        | 144.673   |
|  | Extrapolated complement           | 0.65        | 75.807    |
|  | Other continuation                | 0.98        | 139.610   |
|  | Punctuation and continuation      | 0.91        | 138.408   |
| MULTI-WORD UNITS                         | None                              | <b>1.00</b> | 811.588   |
|  | PP object, before                 | 1.88        | 8.373     |
|  | PP object, after                  | 2.79        | 1.225     |
|  | Other/no object                   | 0.37        | 6.523     |

Table 1: Effect of different variables on the likelihood of 1-2 verbal cluster orders

model. Bloem et al. (2014, p. 1981) note that there are no great differences between a model with and without the modal clusters, which means that the factor is controlled for adequately.

The factor FINITENESS is ranked third. While it also has word order associations, *te*-infinitive clusters are quite rare compared to other constructions, so the overall impact of this factor is smaller.

What we are interested in here is the fact that even when we control for relevant factors, there is still a lot of variation remaining. For example, among verbal clusters with the auxiliary ‘have’ and a finite main verb, there are still clusters in both the 1-2 and 2-1 orders, and this variation can be linked to the processing complexity-related factors that we discuss below. We are more interested in these factors that cannot be explained away as semantic or lexical differences. Word order differences are less likely to be stored preferences in this case and more likely to hint at processing effects.

When using such a large multifactorial model, there is the risk that the different factors are not independent, and partially explain the same portion of the variation. For

| Rank | Factor                                   | AIC    | Decrease |
|------|--|--------|----------|
| 0    | (none)                                   | 463279 | —        |
| 1    | Type of auxiliary                        | 382538 | 80741    |
| 2    | Priming                                  | 378185 | 4353     |
| 3    | ‘te’-infinitive                          | 374378 | 3807     |
| 4    | Extraposition                            | 371413 | 2965     |
| 5    | Length of middle field                   | 369817 | 1596     |
| 6    | Frequency of the main verb               | 368744 | 1073     |
| 7    | Information value                        | 367806 | 938      |
| 8    | Morphological structure of the main verb | 366870 | 936      |
| 9    | Multi-word units                         | 366162 | 708      |
| 10   | Syntactic depth                          | 365674 | 488      |
| 11   | Definiteness                             | 365461 | 213      |

Table 2: Decrease in information loss when adding factors to the model in a stepwise regression procedure

example, it might be the case that complex particle verbs are generally also less frequent. Statistically, we can test for this by computing the variance inflation factor (VIF) of the variables in the model. This is a measure of multicollinearity — correlation between predictor variables. In our model, all of the VIFs are very low ( $< 2$ ). The factors therefore do not appear to correlate with each other, are independent, and do not statistically account for the same parts of the variation.

**Syntactic priming** The results in table 1 show that 1-2 clusters are more likely to be preceded by other 1-2 clusters (1.24) and less likely to be preceded by 2-1 clusters (0.59), compared to clusters that are unprimed. Table 2 tells us that the role of syntactic priming in explaining the observed variation is large — the factor is ranked second. This confirms the importance of processing effects in the model, as syntactic priming is a clear processing effect.

**Morphological structure of the main verb** In the data, separable verbs are 7.06 times more likely to occur in the 1-2 order, shown in table 1. De Sutter (2009) got an effect size of 3.87 here, and Bloem et al. (2014) observed 4.92. It is a robust observation that separable verbs are more likely to occur in the 1-2 order. We have also created lists of verbs that are most strongly associated with the 1-2 order, and many of the verbs at the top of this list are separable verbs.

While the effect size is large, the importance of this factor for the overall model is not as large. It is ranked eighth in terms of information gain, as shown in table 2. This is because most verbs are not separable complex verbs, so the factor is relevant to a limited portion of the data.

**Length of the middle field** According to table 1, clusters preceded by longer middle fields are more likely to occur in the 1-2 order. This matches the result of De Sutter (2007). This is 1.039 (3.9%) times more likely per additional word in the middle field. This effect is fairly important in the overall model, ranking fifth in table 2.

**Syntactic depth of embedding** In the data, clusters that are more deeply embedded are less likely to occur in the 1-2 order. From table 1 we can see that the effect size is 0.95 (-5.3%) per additional level of embedding. This effect is fairly small, and as a result, this is one of the factors of lesser importance to the model. It is ranked tenth in table 2.

It is interesting that the effect direction here is the opposite from that of the length of the middle field. Intuitively, one might expect that clause length and depth of embedding correlate with each other: longer sentences have bigger trees. Looking into this more closely, we found that the two factors only correlate moderately, with  $r = .32$ , too little to consider one to have a confounding effect on the other. Sentences with clusters and



long middle fields are not nearly always deeply embedded. This demonstrates that the two factors need separate explanations.

**Definiteness of the last preverbal constituent** Table 1 tells us that clusters preceded by a definite constituent are 1.04 (4%) times more likely to occur in the 1-2 order. It is our only finding that is not significant at the  $\alpha = 0.001$  level — for definite constituents compared to indefinite,  $p = 0.042$ . It then comes at no surprise that it is also ranked last in table 2. From the frequencies we can see that only 11% of verbal clusters were directly preceded by a noun marked for definiteness in our study, perhaps this is not enough to draw conclusions from. This result also contrasts with De Sutter et al.’s (2007) finding, who found that indefinites are more associated with the 1-2 order, rather than definites. However, this was not tested in a multifactorial model.

**Information value of the last preverbal constituent** In the data of table 1, clusters that are preceded by a word of high information value (or an open-class word) are 0.79 times more likely to occur in the 1-2 order, which means they are 1.27 ( $0.79^{-1}$ ) times less likely to occur in the 1-2 order, or 27% less. Interestingly, this result is different from those of De Sutter (2009) and Bloem et al. (2014), even though those studies defined the factor in the same way. The factor is ranked fairly low in table 2 (7th) even though it affects all clusters; perhaps the effect is not important enough to measure accurately across different models.

**Frequency of the main verb** We estimated verb frequencies by counting the number of occurrences of the verb’s root in the entire Lassy Large corpus. According to the model, clusters that contain main verbs of higher frequency are more likely to occur in the 1-2 order. The effect size is 6% per order of magnitude of frequency, so for each magnitude (e.g. 10.000 or 100.000 occurrences), the 1-2 order is 1.06 times more likely in our data, as shown in table 1. It is not a large effect, but it is robust, and a similar effect was found by De Sutter (2009). In table 2, this factor is ranked 6th, so it is one of the more important processing-related factors. One might think that more frequent main verbs occur in clusters with modal verbs more frequently, causing a difference, because modal verbs occur in the 1-2 order relatively more often. The observation is true: the average token frequency of main verbs in modal clusters is 340.199, while the average token frequency of main verbs in other clusters is 104.379. However, because we are using a multifactorial model, this factor is already taken into account, and should not influence the result much. We created a model of only the non-modal clusters to verify this, and the effect size was also 1.06.

**Extraposition** In the data, clusters with extraposed constituents are generally less likely to be in the 1-2 order, 0.75 for adjuncts and 0.65 for complements (Table 1). In table 2 this factor is ranked fourth, and therefore explains a relatively large portion of the variation.

**Multi-word units** In the results for this factor (Table 1), we can see that clusters that involve a fixed expression are more likely to occur in the 1-2 order when they have a prepositional object, but less likely to occur in the 1-2 order when they do not. The information gain of the factor MWU is fairly low, ranked 9 in table 2. This can be attributed in part to the frequency of multi-word units.

## 7 Discussion

In the preceding section, we have seen that all of the factors that we have investigated contribute in some way to a model that statistically accounts for the verbal cluster order variation that we found in the data. While some factors account for more of the variation than others, it is clear that all of these factors account for some of the order variation in verbal clusters. Furthermore, in section 5 we have learned that all of these factors, apart from the control variables, can be related to processing complexity. All of this allows us to conclude that processing complexity clearly plays a role in this order variation.

Now that we know this, we can also start to address the question of what the processing difference between the two constructions might be. Does either of the orders consistently occur in contexts that are more difficult to process? In section 5, we discussed for

each factor which value is the most difficult to process. For the factor INFORMATION VALUE, we could not learn from the literature whether open-class or closed-class words are more difficult to process. In the remaining cases, it was fairly clear which value was more difficult to process (e.g. longer sentences rather than shorter sentences, separable verbs rather than non-separable ones). There are three possible answers to the question posed here:

**Default 1-2 order hypothesis** This is the right answer when we find the 2-1 order in contexts that are less difficult to process, and the 1-2 order in contexts that are more difficult to process. This hypothesis is supported by ideas from Meyer and Weerman’s (in press) work on acquisition of verbal clusters.

**Default 2-1 order hypothesis** This is the right answer when we find the 1-2 order in contexts that are less difficult to process, and the 2-1 order in contexts that are more difficult to process. This hypothesis is supported by ideas from De Sutter’s (2005) thesis, where he proposes that sociostylistic processing may play a large role.

**Neither** This is the right answer when we do not find a clear association between processing difficulty either of the word orders. There are still processing effects, but they do not clearly point to one order occurring in contexts that are easier to process than the other.

| Rank | Factor                                   | Supports  |
|------|--|-----------|
| 1    | Type of auxiliary                        | (control) |
| 2    | Priming                                  | (both)    |
| 3    | ‘te’-infinitive                          | (control) |
| 4    | Extraposition                            | 1-2       |
| 5    | Length of middle field                   | 1-2       |
| 6    | Frequency of the main verb               | 2-1       |
| 7    | Information value                        | (unclear) |
| 8    | Morphological structure of the main verb | 1-2       |
| 9    | Multi-word units                         | 1-2       |
| 10   | Syntactic depth                          | 2-1       |
| 11   | Definiteness                             | 1-2       |
| -    | Overall word order frequencies           | 2-1       |

Table 3: Summary of the factors that were tested, and which default order hypothesis they support

In table 3, we have summarized our results from the previous section with respect to the default word order hypotheses. The table lists the word order association for the more difficult to process condition of each factor (i.e. for length of the middle field, a longer middle field). For most factors, we observed that the value that is more difficult to process, is associated with the 1-2 order, thereby supporting the default 1-2 order hypothesis. However, there are a few exceptions. We will now briefly discuss each factor with respect to these hypotheses, starting with the overall frequency of the orders.

The 1-2 order is more frequent for all of the verbal cluster constructions we tested in our data, which might be seen as support for the default 1-2 order hypothesis. However, it should be noted that this corpus only contains written data. Previous work has found that spoken Dutch contains more 2-1 orders. For example, Stroop (2009) finds 63% 2-1 orders in the Corpus of Spoken Dutch (CGN) among clusters with a participle. This goes against the default 1-2 order hypothesis, since it is generally assumed that producing spoken language involves more processing difficulty, because it is more immediate. De Sutter (2005) also notes that text genres with more production pressure include relatively more 2-1 orders. These findings from spoken data rather support the default 2-1 order hypothesis, although it can be argued that there is more to entrenchment than just raw token frequency.

Next, we have some factors that do not really affect the hypotheses. We did find the PRIMING factor as operationalized in our study and that of De Sutter (2007), but as

noted by De Sutter, it applies equally to both word orders, so it has no bearing on the default word order hypothesis. The factor `INFORMATION VALUE` is not associated with a hypothesis in the table, because it was not clear to us which value is more difficult to process. However, De Sutter (2007) considered clusters preceded by a word of high information value to be easier to process. If this is true, our results would support the default 1-2 order hypothesis, as the more difficult condition of lower information values is more associated with the 1-2 order.

It is notable that medium-informative words (adverbs and adjectives) show the same effect as function words, while it seems unlikely that they are processed in similar ways. However, this is compatible with the views of Cinque (1998, 2010), who takes adverbs and adjectives to be specifiers of functional projections, thus making them quite parallel to functional elements. In this theory, the low and medium-informative words would all be specifiers, while the high-informative words are heads. But it could also be the case that this finding is simply not robust across different models or data sets. De Sutter (2009) also discussed this factor and in his study it was found to be the weakest effect. Furthermore, in De Sutter’s study the effect was in the other direction: words of high information value were 1.92 times more likely to occur in the 1-2 order, and medium information value words were 1.41 times more likely. The cause of this is either a difference between the corpora of the two studies (the Wikipedia domain versus the newspaper domain), or methodological differences. Furthermore, Bloem et al. (2014) found that words of high information value were 1.11 times more likely to occur in the 1-2 order, and medium information value words were 1.21 times more likely, again a different result, indicating that this finding may not be very reliable.

Several factors clearly support the default 1-2 order hypothesis. This is the case for `MORPHOLOGICAL STRUCTURE OF THE MAIN VERB` (separable participles are more difficult to process and associated with the 1-2 order), `LENGTH OF THE MIDDLE FIELD` (clusters with longer middle fields are more difficult to process and associated with the 1-2 order) and `DEFINITENESS` (definites are more difficult to process and associated with the 1-2 order).

The factor `MULTI-WORD UNITS` also matches the default 1-2 order hypothesis. Multi-word unit verbal clusters are easier to process and associated with the 2-1 order, except when they have an open slot for a PP. In that case, the association is the opposite, presumably because they will have to undergo some sort of syntactic processing. One part of the construction undergoes cluster formation and the other part optionally undergoes extraposition, even though both parts are stored as a unit. This may be a demanding task, though this has not been psycholinguistically tested, as far as we are aware. It is interesting to note that De Sutter (2007) finds an opposite effect for this factor: in his study, clusters in fixed expressions are 2.26 times more likely to be in the 1-2 order. Therefore, he interprets this as supporting a default 1-2 order. However, this study does not distinguish between fixed expressions with and without PPs, and the fixed expressions were not annotated in the same way, which may have led to different results.

For the factor `EXTRAPOSITION`, we can say that the overall result matches the default 1-2 order hypothesis, since extraposition eases processing, and extraposition is associated with the 2-1 order. Non-extraposition would be the more difficult condition, and that is associated with the 1-2 order.

Since we included adjunct and complement extraposition separately in the model, we can now also address the idea of De Sutter (2005), who suggested that complements might be more likely to be positioned closer to the main verb, due to their semantic relatedness. In the case of extraposition, that would predict 1-2 orders. However, extraposed complements are the condition that is the most likely to be in the 2-1 order in our results, so an adjacency effect seems unlikely. Therefore, it is more likely that both complement and adjunct extraposition depend on the processing complexity of the context. De Sutter (2007) suggested that complements might actually be more difficult to process when extraposed, rather than less difficult. Under that assumption, one could expect that the default order is more likely to be used in the case of complement extraposition. In our data, the ‘extraposed complement’ is associated more with the 2-1 order than the ‘extraposed adjunct’ condition. Under De Sutter’s (2007) assumption, this actually matches the default 2-1 order hypothesis, though the effect is small. Furthermore, we are

not aware of any psycholinguistic evidence suggesting that complement extraposition is more difficult to process, nor is it clear that complement extraposition would be more difficult than no extraposition.

Another link that could be investigated is the one predicted by Van Haeringen (1956), between extraposition and clause length. We did not include the length of the extraposition in our model, and there was no clear multicollinearity involving extraposition and length, but nevertheless this idea would be interesting to include in a multifactorial study. The idea was previously tested by De Sutter et al. (2008), though not in a multifactorial model. They did find a trend of longer extraposed elements correlating with more 1-2 orders. If long extraposed elements make the cluster more difficult to process, this supports the default 1-2 order hypothesis. It could be that the length of the extraposition and the length of the middle field, when taken together, show a clearer processing effect, or that the results are different when extraposition is measured in a way that controls for the total clause length. It has already been shown by Willems and De Sutter (2015) that the length of the PP affects whether it is extraposed, perhaps this is a factor of processing complexity that also relates to the verbal cluster the PP is extraposed over. This would be an interesting factor to include in future work.

One particularly interesting consideration here is that the PP-over-V extraposition in itself is also a grammatical phenomenon showing variation that may be related to processing difficulty, in the same way as the verbal cluster order variation. As we discussed, it has been linked to some factors related to processing in previous work. Willems and De Sutter (2015) explained some of the variation in PP placement in terms of complexity, and they state that their model requires more predictor variables because its performance leaves room for improvement. Perhaps some of this remaining variation can be explained using other factors of processing complexity that we just discussed. There could also be an interaction between the two phenomena, involving a balance of processing difficulty. For example, if a verbal cluster is produced in a ‘more difficult’ ordering, then the prepositional object may be produced in an ‘easier’ position, or the other way around. However, answering this question would require a larger corpus study of extraposition constructions and verbal cluster constructions at the same time, with a methodology similar to the current study.

Lastly, two of our factors appear to match the default 2-1 order hypothesis, rather than the 1-2 one. This was the case for the factor `SYNTACTIC DEPTH OF EMBEDDING`, a factor that was not studied by De Sutter (2007). It is unexpected that this factor and `LENGTH OF THE MIDDLE FIELD` show opposite results — deep embeddings are associated with the 2-1 order, but long middle fields are associated with the 1-2 order. One might expect that longer sentences also have more complicated structures with deeper embeddings, but interestingly, we found that `SYNTACTIC DEPTH OF EMBEDDING` and `LENGTH OF THE MIDDLE FIELD` only correlate weakly. Therefore, the two factors appear to have distinct effects on verb cluster order. An explanation such as the one proposed by Baumann (2016) might account for this — in this experimental reading time study, a distinction is made between ‘structural distance’, which is like depth of embedding, and ‘linear distance’, which is like length. Baumann shows that ‘structural distance’ better predicts reading times, and has a clearer processing cost. In this view, the factor `LENGTH OF THE MIDDLE FIELD` is not important for processing and the default 2-1 order hypothesis is supported. However, there is also a lot of evidence in support of resource-limitation models of sentence processing, in which ‘linear distance’ plays a larger role. Another possible explanation might be found in the theories that we mentioned in section 5, i.e. on non-hierarchical processing (Frank et al., 2012), which claim that embedding isn’t very relevant for processing effort in one way or another. Since this factor also ranked second-to-last in terms of information gain in the model (Table 2), its importance as a processing effect may be limited.

The other factor that we found to be associated with the default 2-1 order hypothesis is `FREQUENCY`, which showed more frequent words being associated with the 1-2 order. This factor was also studied by De Sutter (2005) and his results are consistent with ours. This appears to be a counter-example to the default 1-2 order hypothesis. There are some possible issues with using this factor in a regression model. When including frequency in the model, there are likely to be more instances of the frequent verbs in

the dataset of verb clusters, which may amplify the effect of the frequency factor in the regression model. Alternatively, it might be the case that some verbs are relatively more frequent in verbal cluster contexts, though this seems unlikely. As far as we are aware, any verb can be used in a verbal cluster construction, but further study is needed to verify this hypothesis. The consequence of this would be that these verbs are more easily activated in verbal cluster contexts only, resulting in a construction-specific frequency effect rather than a general one. Further testing for construction-dependent frequency effects may be helpful here to exclude this possibility.

Nevertheless, it appears that most of the processing-related factors that we investigated are compatible with the default 1-2 order hypothesis, as summarized in table 3. These findings lead to an explanation based on Meyer and Weerman’s (in press) acquisition study — the 1-2 word order is more entrenched in standard Dutch, as it is acquired as a construction earlier than the 2-1 order. Because the 1-2 order is more entrenched and more easily activated, it is used relatively more often in contexts that are more difficult to process.

## 8 Conclusion

In this work, we have aimed to provide a linguistic explanation for the observation that many different factors appear to play a role in verbal cluster order variation. We have shown that for this particular case of grammatical optionality, Dutch verbal cluster orders, it is necessary to go beyond semantic and syntactic factors in search of an explanation. Various factors that are associated with this word order variation were linked to processing complexity in this study. We claim that processing complexity is another motivation for variation, in addition to the more commonly discussed factors of semantics and information structure, which might apply to other alternations of near-synonymous constructions as well. Minimizing processing complexity is a plausible explanation for various facts that we observed from the corpus data.

All of the factors we discussed have a significant association with either of the word orders. The direction of the effect is fairly consistent. For most factors, it is the case that the more difficult condition is associated with the 1-2 order. We view this as evidence for a hypothesis we proposed based on Meyer and Weerman’s (in press) acquisition study — these factors are evidence that the more entrenched and more frequent 1-2 word order is used in contexts that are more difficult to process. However, not all of the factors we investigated support this idea, and the study was limited to written corpus data. Earlier work has shown a greater proportion of 2-1 orders in spoken Dutch, a fact that is difficult to reconcile with this hypothesis.

It is notable that one of the factors we found to affect verbal cluster word order, the extraposition factor, is in itself also a grammatical phenomenon showing variation that could be investigated using the same methodology that was used in this study. More generally, it would be interesting to invoke processing complexity as a factor to explain other instances of word order variation among near-synonymous constructions. Using large amounts of annotated linguistic data and multifactorial models, it is possible to test this empirically.

Furthermore, it would be interesting to investigate additional factors related to processing complexity to obtain more evidence for or against the default 1-2 word order hypothesis. One possible factor, which was not included in the present study, is the number of arguments of the verb. It has been found that verbs that are more associated with the 1-2 order are more often ditransitive, while verbs associated with the 2-1 order are more often intransitive (Bloem, 2016). This could be related to processing complexity as well. Using our methodology we are limited to factors that can be extracted from this automatically annotated corpus however, unless we analyze a part of the data manually. This would need to be done to look at most discourse or information-structural factors, such as ‘givenness’ or ‘focus’. The corpus also does not allow the inclusion of rhythm-related factors, which have been shown to affect verb cluster order in previous work, although these are unlikely to be related to processing complexity.

It might also be interesting to implement a model of language processing more directly, and test whether complexity in this model predicts verb cluster word order on corpus data. For example, a constraint-based model of language processing would predict that a

probabilistic language model can measure processing complexity. However, this is difficult because there are many possible ways in which such a model might be implemented, e.g. at the level of words, constructions, word classes, or a combination thereof. A further test would of course be a psycholinguistic experiment, in which the complexity of verb cluster contexts could be manipulated more directly, and processing difficulty could be indirectly measured.

Besides processing effects, there may be semantic or constructional factors that play a role in the choice between word orders that our model did not control for. Particular verbs or semantic classes of verbs can show a preference for one order or the other (Bloem, 2016). It would be interesting to incorporate this into a multifactorial model in future work, although this is difficult due to the large number of possible main verbs.

## References

- Arfs, M. (2007). *Rood of groen? De interne woordvolgorde in tweeledige werkwoordelijke eindgroepen met een voltooid deelwoord en een hulpwerkwoord in bijzinnen*. PhD thesis, Göteborg University.
- Bach, E., Brown, C., and Marslen-Wilson, W. (1986). Crossed and nested dependencies in German and Dutch: A psycholinguistic study. *Language and Cognitive Processes*, 1(4):249–262.
- Barbiers, S., Bennis, H., De Caluwe, J., and Van Keymeulen, J. (2010). De plaats van het werkwoord in zuid en noord. *Voor Magda. Artikelen voor Magda Devos bij haar afscheid van de Universiteit Gent*, pages 25–42.
- Barbiers, S., Bennis, H., and Hendriks, L. (2016). Merging verb cluster variation. *Manuscript submitted*.
- Baumann, P. (2016). Dependencies and hierarchical structure in sentence processing. *Unpublished manuscript*.
- Beck, S. and Johnson, K. (2004). Double objects again. *Linguistic Inquiry*, 35(1):97–123.
- Behrens, H. (1998). How difficult are complex verbs? Evidence from German, Dutch and English. *Linguistics*, 36(4):679–712.
- Bloem, J. (2016). Lexical preferences in Dutch verbal cluster ordering. In Bellamy, K., Karvovskaya, E., Kohlberger, M., and Saad, G., editors, *ConSOLE XXIII: Proceedings of the 23rd Conference of the Student Organization of Linguistics in Europe*, pages 70–93, Leiden. Leiden University Centre for Linguistics.
- Bloem, J., Versloot, A., and Weerman, F. (2014). Applying automatically parsed corpora to the study of language variation. In Hajic, J. and Tsujii, J., editors, *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 1974–1984, Dublin. Dublin City University and Association for Computational Linguistics.
- Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive psychology*, 18(3):355–387.
- Bock, K. and Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General*, 129(2):177.
- Booij, G. (2002). Separable complex verbs in Dutch: A case of periphrastic word formation. In Dehé, N., Jackendoff, R., Macintyre, A., and Urban, S., editors, *Verb-Particle Explorations*, pages 21–42. Mouton de Gruyter.
- Booij, G. (2010). Construction morphology. *Language and Linguistics Compass*, 4(7):543–555.
- Bouma, G. (2013). Om-omission in Dutch verbal complements. *Manuscript submitted*.
- Bradley, D. C. (1978). *Computational distinctions of vocabulary type*. PhD thesis, Massachusetts Institute of Technology.
- Bresnan, J., Cueni, A., Nikitina, T., and Baayen, R. H. (2007). Predicting the dative alternation. In Bouma, G., Kraemer, I., and Zwarts, J., editors, *Cognitive foundations of interpretation*, pages 69–94. KNAW, Amsterdam.

- Canegem-Ardijns, I. V. (2006). The extraposition of prepositional objects of adjectives in Dutch. *Linguistics*, 44(3):425–457.
- Cappelle, B. (2006). Particle placement and the case for "allostructions". *Constructions*, pages 1–28.
- Chomsky, N. (1957). *Syntactic structures*. Mouton.
- Cinque, G. (1998). *Adverbs and functional heads: A cross-linguistic perspective*. Oxford University Press.
- Cinque, G. (2010). *The syntax of adjectives: A comparative study*, volume 57 of *Linguistic Inquiry Monographs*. MIT Press.
- Conklin, K. and Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, 29(1):72–89.
- Cook, S. W., Jaeger, T. F., and Tanenhaus, M. K. (2009). Producing less preferred structures: More gestures, less fluency. In *Proceedings of the 31st Annual Meeting of the Cognitive Science Society*, pages 62–67.
- Coussé, E. (2008). *Motivaties voor volgordevariatie. Een diachrone studie van werkwoordvolgorde in het Nederlands*. PhD thesis, Ghent University.
- Coussé, E., Arfs, M., and De Sutter, G. (2008). Variabele werkwoordsvolgorde in de Nederlandse werkwoordelijke eindgroep: een taalgebruiksgebaseerd perspectief op de synchronie en diachronie van de zgn. rode en groene woordvolgorde. In Rawoens, G., editor, *Taal aan den lijve : het gebruik van corpora in taalkundig onderzoek en taalonderwijs*, pages 29–47. Academia Press.
- Crain, S. and Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological parser. *Natural language parsing*, pages 320–358.
- De Sutter, G. (2005). *Rood, groen, corpus! Een taalgebruiksgebaseerde analyse van woordvolgordevariatie in tweeledige werkwoordelijke eindgroepen*. PhD thesis, University of Leuven.
- 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. (2009). Towards a multivariate model of grammar: The case of word order variation in Dutch clause final verb clusters. In Dufter, A., Fleischer, J., and Seiler, G., editors, *Describing and Modeling Variation in Grammar*, pages 225–255. Walter De Gruyter.
- De Sutter, G., Speelman, D., and Geeraerts, D. (2007). Luisteren schrijvers naar hun innerlijke stem? De invloed van ritmische factoren op de woordvolgorde in geschreven werkwoordelijke eindgroepen. *Neerlandistiek*, 2007.
- De Sutter, G., Speelman, D., and Geeraerts, D. (2008). Prosodic and syntactic-pragmatic mechanisms of grammatical variation: The impact of a postverbal constituent on the word order in dutch clause final verb clusters. *International Journal of Corpus Linguistics*, 13(2):194–224.
- den Besten, H. and Broekhuis, H. (1992). Woordvolgorde in de werkwoordelijke eindreeks. In *GLOT 12*, pages 79–137.
- den Besten, H. and Edmondson, J. A. (1983). The verbal complex in continental West Germanic. In Abraham, W., editor, *On the formal syntax of the Westgermanica. Papers from the 3rd Groningen Grammar Talks.*, volume 3, pages 155–216. John Benjamins Publishing Company.
- Embick, D. (2008). Variation and morphosyntactic theory: Competition fractionated. *Language and Linguistics Compass*, 2(1):59–78.
- Evers, A. (1975). *The transformational cycle in Dutch and German*. Indiana University Linguistics Club, Bloomington.

- Fedzechkina, M., Jaeger, T. F., and Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences*, 109(44):17897–17902.
- Frank, S. L., Bod, R., and Christiansen, M. H. (2012). How hierarchical is language use? *Proceedings of the Royal Society B: Biological Sciences*, 279(1747):4522–4531.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68(1):1–76.
- Gibson, E., Schütze, C. T., and Salomon, A. (1996). The relationship between the frequency and the processing complexity of linguistic structure. *Journal of Psycholinguistic Research*, 25(1):59–92.
- Goldberg, A. E. (1992). The inherent semantics of argument structure: The case of the English ditransitive construction. *Cognitive Linguistics*, 3(1):37–74.
- Goldberg, A. E. (2002). Construction grammar. In *Encyclopedia of Cognitive Science*. Macmillan Reference Limited Nature Publishing Group.
- Gries, S. T. (2001). A multifactorial analysis of syntactic variation: Particle movement revisited. *Journal of quantitative linguistics*, 8(1):33–50.
- Haegeman, L. and van Riemsdijk, H. (1986). Verb projection raising, scope, and the typology of rules affecting verbs. *Linguistic inquiry*, 17(3):417–466.
- Haeseryn, W. (1997). *Algemene Nederlandse spraakkunst*, volume 2. M. Nijhoff.
- Haiman, J. (1980). The iconicity of grammar: isomorphism and motivation. *Language*, 56(3):515–540.
- Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies*, NAACL '01, pages 1–8. Association for Computational Linguistics.
- Hartsuiker, R. J. and Kolk, H. H. (1998). Syntactic persistence in Dutch. *Language and Speech*, 41(2):143–184.
- Hartsuiker, R. J. and Westenberg, C. (2000). Word order priming in written and spoken sentence production. *Cognition*, 75(2):B27–B39.
- Hawkins, J. A. (2014). *Cross-linguistic variation and efficiency*. OUP Oxford.
- Heim, I. (1982). *The semantics of definite and indefinite noun phrases*. PhD thesis, University of Massachusetts Amherst.
- Hoekstra, E. and Versloot, A. (2016). Three-verb clusters in interference frisian: A stochastic model over sequential syntactic input. *Language and Speech*, 59(1):43–58.
- Hoen, M. and Dominey, P. F. (2000). ERP analysis of cognitive sequencing: a left anterior negativity related to structural transformation processing. *Neuroreport*, 11(14):3187–3191.
- Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive psychology*, 61(1):23–62.
- Jaeger, T. F. and Tily, H. (2011). On language ‘utility’: Processing complexity and communicative efficiency. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(3):323–335.
- Jansen, F. (1978). Hoe krijgt een spreker zijn woorden op een rijtje? Taalgebruikaspecten van de PP over V konstruktie. In Kooij, J., editor, *Aspekten van woordvolgorde in het Nederlands*, volume 8 of *Publikaties van de Vakgroep Nederlandse Taal- en Letterkunde*, pages 70–104. Leiden.
- Larson, R. K. (1988). On the double object construction. *Linguistic inquiry*, 19(3):335–391.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3):1126–1177.



- Levy, R. P. and Jaeger, T. (2006). Speakers optimize information density through syntactic reduction. In *Advances in neural information processing systems*, pages 849–856.
- Levy, R. P. and Keller, F. (2013). Expectation and locality effects in German verb-final structures. *Journal of Memory and Language*, 68(2):199–222.
- Lüttmann, H., Zwitserlood, P., and Bölte, J. (2011). Sharing morphemes without sharing meaning: production and comprehension of German verbs in the context of morphological relatives. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 65(3):173.
- MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, 4:226.
- Meeussen, A. and Vanacker, V. (1951). De dubbele werkwoordgroep. *De nieuwe taalgids*, 44:36–50.
- Meyer, C., de Haan, D., Faber, M., and Weerman, F. (2015). Embracing ascending orders: Verb clusters in bilingual acquisition. Presented at Generative Approaches to Language Acquisition, Université de Nantes.
- Meyer, C. and Weerman, F. (in press). Cracking the cluster: The acquisition of verb raising in Dutch. *Nederlandse Taalkunde*.
- Miller, G. A. and Chomsky, N. (1963). Finitary models of language users. In Luce, R. D., Bush, R. R., and Galanter, E., editors, *Handbook of mathematical psychology*, volume 2, pages 2–419. Wiley.
- Moravcsik, E. A. (2010). Conflict resolution in syntactic theory. *Studies in Language*, 34(3):636–669.
- Müller, S. (2002). Syntax or morphology: German particle verbs. *Verb-particle explorations*, 1:119.
- Odiijk, J. (2013). Identification and lexical representation of multiword expressions. In *Essential Speech and Language Technology for Dutch*, Theory and Applications of Natural Language Processing, pages 201–217. Springer.
- Oehrle, R. T. (1976). *The grammatical status of the English dative alternation*. PhD thesis, Massachusetts Institute of Technology.
- O’Grady, W. (1997). *Syntactic development*. University of Chicago Press.
- Olthof, M., Westendorp, M., Bloem, J., and Weerman, F. (2016). Synchronic variation and diachronic change in dutch two-verb clusters. *Manuscript submitted*.
- Pauwels, A. (1953). *De plaats van hulpwerkwoord, verleden deelwoord en infinitief in de Nederlandse bijzin*. M. & L. Symons, Leuven.
- Pickering, M. J. and Ferreira, V. S. (2008). Structural priming: a critical review. *Psychological Bulletin*, 134(3):427.
- Plag, I. and Baayen, H. (2009). Suffix ordering and morphological processing. *Language*, 85(1):109–152.
- Polite, E. J., Leonard, L. B., and Roberts, F. D. (2011). The use of definite and indefinite articles by children with specific language impairment. *International Journal of Speech-Language Pathology*, 13(4):291–300.
- Schaeffer, J., van Witteloostuijn, M., and Creemers, A. (2016). Pragmatic article choice as compared to morphosyntax and working memory in children with High-Functioning Autism and children with Specific Language Impairment. *Manuscript submitted*.
- Schaeffer, J. C., Witteloostuijn, M. v., and Haan, D. d. (2014). Article choice in children with High Functioning Autism (HFA) and in children with Specific Language Impairment (SLI). *Linguistics in the Netherlands*, 31(1):107–128.
- Schutter, G. d. (1996). De volgorde in tweeledige werkwoordelijke eindgroepen met voltooid deelwoord in spreek-en schrijftaal. *Nederlandse taalkunde*, 1:207–220.
- Segalowitz, S. J. and Lane, K. C. (2000). Lexical access of function versus content words. *Brain and Language*, 75(3):376–389.

- Smolka, E., Gondan, M., and Rösler, F. (2015). Take a stand on understanding: electrophysiological evidence for stem access in german complex verbs. *Frontiers in human neuroscience*, 9:62.
- Stroop, J. (1970). Systeem in gesproken werkwoordsgroepen. *Taal en Tongval*, 22:128–147.
- Stroop, J. (2009). Twee- en meerledige werkwoordsgroepen in gesproken Nederlands. In e.a., E. B., editor, *Fons Verborum. Feestbundel Fons Moerdijk*, pages 459–469. Instituut voor Nederlandse Lexicologie, Leiden.
- Tremblay, A., Derwing, B., Libben, G., and Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, 61(2):569–613.
- Underwood, G., Schmitt, N., and Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In Schmitt, N., editor, *Formulaic Sequences: Acquisition, processing and use*, volume 9 of *Language Learning & Language Teaching*, pages 153–172. John Benjamins.
- Van Haeringen, C. B. (1956). *Nederlands tussen Duits en Engels*. Servire Den Haag.
- Van Kampen, J. (2010). Typological guidance in the acquisition of V2 Dutch. *Lingua*, 120(2):264–283.
- van Noord, G. (2009). Huge parsed corpora in LASSY. In Van Eynde, F., Frank, A., De Smedt, K., and van Noord, G., editors, *Proceedings of the Seventh International Workshop on Treebanks and Linguistic Theories (TLT 7)*, volume 12, pages 115–126. LOT.
- Versloot, A. P. (2008). *Mechanisms of language change: vowel reduction in 15th century West Frisian*. PhD thesis, University of Groningen.
- Wang, L., Li, X., and Yang, Y. (2014). A review on the cognitive function of information structure during language comprehension. *Cognitive neurodynamics*, 8(5):353–361.
- Wasow, T., Jaeger, T. F., and Orr, D. (2011). Lexical variation in relativizer frequency. In Simon, H. J. and Wiese, H., editors, *Expecting the unexpected: Exceptions in grammar*, pages 175–195. Walter de Gruyter, Berlin.
- Wijnen, F. (1997). Functionele categorieën in taalverwerving en taalgebruik. *Nederlandse taalkunde*, 2:178–198.
- Willems, A. and De Sutter, G. (2015). Reassessing the effect of the complexity principle on PP Placement in Dutch. *Nederlandse Taalkunde*, 20(3):339–367.
- Wurmbrand, S. (2004). West Germanic verb clusters: The empirical domain. In Kiss, K. É. and van Riemsdijk, H., editors, *Verb clusters: A study of Hungarian, German, and Dutch*, pages 43–85. John Benjamins Publishing Company.
- Wurmbrand, S. (2006). Verb clusters, verb raising, and restructuring. In Everaert, M. and van Riemsdijk, H., editors, *The Blackwell Companion to Syntax*, pages 229–343. Blackwell Publishing.
- Zuckerman, S. (2001). *The acquisition of "optional" movement*. PhD thesis, University of Groningen.
- Zwart, J.-W. (1996). Verb clusters in continental West Germanic dialects. In Black, J. R. and Motapanyane, V., editors, *Microparametric syntax and dialect variation*, volume 139 of *Current Issues in Linguistic Theory*, pages 229–258. John Benjamins Publishing.