Not another book on Verb Raising
Dros, L.D.H.

Author: Dros, L.D.H.
Title: Not another book on Verb Raising
Issue Date: 2018-02-07

CHAPTER 4

Factoring verb clusters

4.1 Introduction[55]

The previous chapter discussed the results of a comparative judgement task, in which speakers judged grammatical, unrealized orders as more acceptable than ungrammatical orders. It was argued that speakers' knowledge of the grammatical system gives them the ability to distinguish orders that are not realized, but can be generated by Merge, from orders that cannot be generated by Merge. While this explanation accounts for a substantial portion of the results of the comparative judgement task, some results are left unexplained. For instance, there is no clear reason why speakers have ranked the 3-1-2 order much higher than the 1-3-2 order. These two orders co-occur in many Dutch varieties. The question that arises is what drives the choice between these two orders.

According to Chomsky (2005, 2007, 2008), three factors play a role in the development of human language.

I. First factor: genetic endowment, i.e. Universal Grammar (UG).

II. Second factor: experience, i.e. intake from the environment.

III. Third factor: principles not specific to the language system, which include principles of data analysis and principles of efficient computation.

Chomsky and his cohorts assume that Universal Grammar constitutes minimally (and perhaps only) the operation Merge (Chomsky 2000; Hauser et al.

[55] I would like to thank Markus Bader for his comments on an earlier version of this chapter.

2002). In the previous chapter many of the results of the comparative judgement task were linked to this First factor. The Second factor, experience, was demonstrated to play a limited role. While speakers ranked orders that were realized in their own language variety the highest, their experience with other language varieties did not affect their judgements. The chapter did not discuss a potential effect of Third Factor principles. In some approaches, grammars are assumed to be shaped completely by principles that Chomsky would describe as Second and Third Factor principles. Proponents of such an approach explain the acceptability of a construction functionally, by referring to general properties of language use, human behavior, and human cognition (see for instance Comrie 1989; Hawkins 1994; Croft 2003; Haspelmath 2008, among others). For instance, Hawkins (1994) argues that many word order preferences follow entirely from general properties of language processing. This makes language processing a potential source of the speakers' preferences in verb cluster orders. For this reason, this chapter explores whether the speakers' preferences are entirely or partially driven by processing preferences.

There are two ways to investigate the effect of properties of language processing on speakers' judgements. First, one could empirically measure processing difficulty by conducting a psycholinguistic experiment, such as measuring reading times. Orders that are less acceptable are then predicted to be associated with slower reading times. A problem with such an approach is that the precise cause of a potential change in reading times would remain unclear. As a consequence, such an experiment would not enlighten us on the precise factors that are causing a slowdown or speedup in reading times. Another possibility is to apply an established theoretical model of processing preferences and see what predictions that model makes for the acceptability of the different verb cluster orders. This is the approach taken here. Unfortunately, this is not without problems. There are very diverse models of language processing and some models downright contradict each other. Furthermore, most models are not worked out in enough detail to be straightforwardly applied to new data. In fact, many key properties are not made explicit by the designers of the models. This makes it very difficult, if not impossible, to make reliable predictions. In order to demonstrate these problems, the next section provides an evaluation of various established processing models.

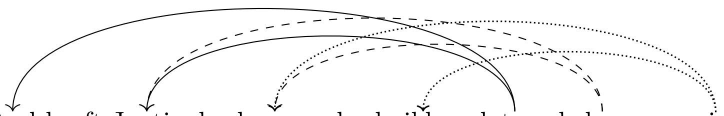
It will become evident during this chapter that none of the processing models can account for the results of the comparative judgement task by themselves. However, a combined model that takes both the grammar and language processing into account provides a better explanation for the results.

4.2 The diverse ideas on language processing

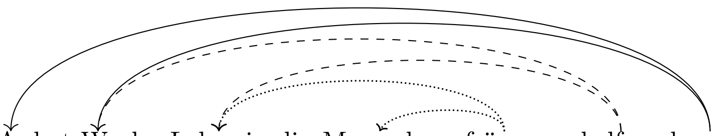
4.2.1 Intervening material makes processing harder

Crossed and nested dependencies

Bach et al. (1986) were among the first to discuss the difficulties involved in processing verb clusters. They focussed on the relation between arguments and verbs. An argument can sometimes have a dependency relation with more than one verb. For instance, in the sentence *Alfred helped the teacher collect the marbles*, the teacher is the one being helped, as well as the collector of the marbles. In languages with sentence-final verb clusters, new dependency relations can arise before others are closed off. In the Dutch head-initial order, this leads to multiple crossed relations of dependency, as illustrated in (77). In the German head-final order, this leads to multiple nested relations in, as illustrated in (78). According to Bach et al., such constructions are costly, because they require multiple arguments to be kept in memory before they are assigned to their respective verbs.



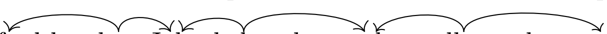
(77) Aad heeft Jantje de lerares de knikkers laten$_1$ helpen$_2$ opruimen$_3$.
Aad has Jan the teacher the marbles let help up.collect
'Aad let Jan help the teacher collect the marbles.'



(78) A. hat W. der Lehrerin die Murmeln aufräumen$_3$ helfen$_2$ lassen$_1$.
A. has W. the teacher the marbles up.collect help let

Bach et al. (1986) demonstrated that speakers of German have more difficulty comprehending sentences with multiple dependencies than speakers of Dutch. They concluded that nested orders are more difficult to process than crossed dependencies (see also Joshi (1990) and Rambow and Joshi (1994)). A crossed dependency is easier to parse, they argue, because it allows the language user to immediately integrate the verbs into the structure. In nested dependencies, on the other hand, the higher structure into which the lower verbs need to be integrated has not yet been encountered.

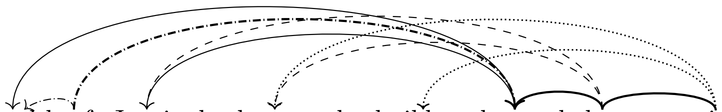
In English sentences, the arguments are all adjacent to the verb they belong to. This is illustrated in (79). Bach et al. argue that this explains why the English order is easier to comprehend and is found more acceptable.



(79) Alfred has let$_1$ John help$_2$ the teacher collect$_3$ the marbles.

However, it is generally assumed that the projection of the lower verb functions as the complement of the higher verb. There is hence a dependency relation between the verbs as well. It is not clear why Bach et al. (1986) do not take the relation between the verbs in this clause into account. In fact, those relations make differences in complexity between languages much less apparent. As illustrated in (80) and (81), in English, there are six relations that are nested within another dependency relation. In Dutch there are five relations that are nested, but some of these are nested within multiple other relations. As a consequence, it is no longer so obvious that the English order has fewer nested dependencies than the Dutch order.

(80) Alfred has let$_1$ John help$_2$ the teacher collect$_3$ the marbles.



(81) Aad heeft Jantje de lerares de knikkers laten$_1$ helpen$_2$ opruimen$_3$.
Aad has Jan the teacher the marbles let help up.collect
'Aad let Jan help the teacher collect the marbles.'

There are a number of assumptions one could make to ensure that the English order falls out as the easier order. But it is clear that this requires a more sophisticated calculation metric, with well-defined and supported variables, which is not provided by Bach et al. (1986). As will become clear below, this is a problematic issue for more recent theories of language processing as well.

The number of intervening words

Bach et al. (1986) established that the Dutch and German sentence-final verb clusters become increasingly more difficult to process when there are more dependency relations in the sentence. This finding is in line with a broader claim, according to which structures that have longer phrases embedded within them are generally harder to process (Hawkins 1994). Hawkins discusses the observation that in English, center-embeddings are ungrammatical when the embedded material is clausal, but not when the embedded material is an NP. This is illustrated in (82).

(82) a. *[Did [$_S$ that John failed his exam] [surprise Mary]]?
b. [Did [$_{NP}$ this fact] [surprise Mary]]?

According to Hawkins (1994), language processing occurs more rapidly and efficiently when constituents that belong together are closer to each other. Structures that contain complex center-embedded structures are more difficult to process because they make it harder to recognize and produce phrase structure groupings. Since noun phrases are typically smaller than clauses, they can center-embed more easily.[56]

The preference to reduce the number of words required to recognize phrasal combinations is an instantiation of a more general processing principle proposed by Hawkins, which he dubs 'Minimize Domains.' According to this principle, the efficiency and complexity of a sentence is affected by the number of words that intervene between (syntactically and semantically) related items (Hawkins 2004, 2014).

Hawkins' theory can explain a number of cross-linguistic word order patterns. Consider for instance the difference between the two sentences in (83). While both sentences are fully grammatical, Hawkins states that sentence (83b) is easier to process than (83a), due to the fewer number of words that are minimally required to recognize and combine the subject with its predicate.

(83) a.

	$[_S$ [that	their	time	should	not	be	wasted]	$[_{VP}$ is	important.]]
S:	1	2	3	4	5	6	7	8
VP:								1	2

b.

	$[_S$ it	$[_{VP}$ is	important	[that	their time should not be wasted.]]]
S:	1	2			
VP:		1	2	3	

Hawkins argues that, in line with the Minimize Domains principle, sentences such as (83b) are indeed preferred over (83a), as evidenced by frequency patterns in corpora.

There are a few problems with Hawkins' approach, that become clear as soon as one attempts to apply his theory to new data. A single sentence contains many words that are, syntactically or semantically, related to each other. Theoretically, all of these relations might be subject to the Minimize Domains principle. It is unclear whether all these relations are relevant and, if so, how the processor deals with multiple preferences. Consider for instance the differences between the Dutch and German verb cluster orders in (77) and (78), that were observed by Bach et al. (1986). To be able to investigate whether these differences can be explained by Hawkins' model, it is very important to know which relations play a role. There are many relations that might affect the preferred order of verbs. The NP has a relation with one, maybe two verbs, and the verbs are also related to each other. Unfortunately, Hawkins (1994, 2004, 2014) is inexplicit on the question whether all these relations are as important for the processor. As a consequence, there are many ways in which one could apply his theory to these data. One way to go forward is to follow Bach et al.'s

[56] The word *typical* is crucial here. According to Hawkins, the functional disfavor for longer embeddings and the average length of sentential phrases has led to a "grammatical response that differentiates between [nominal and sentential embeddings] by blocking the structure that is regularly worse." (Hawkins 1994:24) This assumption is required to account for the fact that nominal embeddings are grammatical even when they are longer than sentential embeddings.

(1986) methodology and consider all and only the relations between the nouns and the verbs. This, however, does not lead to any differences between crossed and nested dependencies. The mean distances are identical between the two orders. This is illustrated in (84).[57]

(84) a. NP_1 NP_2 NP_3 NP_4 V_1 V_2 V_3.

(1)	(2)	3	4	(5)			dependencies V_1
	(1)	(2)	3	4	(5)		dependencies V_2
		(1)	(2)	3	4	(5)	dependencies V_3
							Mean: 5 words

b. NP_1 NP_2 NP_3 NP_4 V_3 V_2 V_1.

(1)	(2)	3	4	5	6	(7)	dependencies V_1
	(1)	(2)	3	4	(5)		dependencies V_2
		(1)	(2)	(3)			dependencies V_3
							Mean: 5 words

This implementation of Hawkins' theory hence cannot account for Bach et al.'s finding that crossed dependencies are easier to process than nested dependencies, as such structures do not differ in overall dependency lengths. There are a few ways to deal with this result. First, one might conclude that this particular implementation of Hawkins' model is incorrect. Another possibility is that this implementation of the model is correct, and the factors that make crossed dependencies easier to process follow from processing properties that are independent of the preference to minimize domains. Such an approach is discussed in section 4.3.1. A final possibility is that the model is incorrect. Potentially, the language users' preferences that were observed by Hawkins follow from different properties of language processing. One might then consider a different processing model, such as the model presented by Gibson (1998, 2000), which is discussed in the next subsection.

The number of intervening discourse-referring items

Gibson's (1998; 2000) view on language processing is comparable to that of Hawkins' in that language processing is affected by the distance between related items. However, he argues that intervening discourse referents in particular lead to additional processing costs; each discourse-referring element that intervenes between two related items increases the processing cost. This model

[57] In fact, if one applies Hawkins' calculation metrics, the overall costs even make the wrong prediction that nested dependencies are easier to process than crossed dependencies. This is not illustrated here, for reasons of simplicity.

can explain differences between the difficulty of nested structures with a first person pronoun, as in (85a), and nested structures with a definite noun phrase as in (85b), or a full noun phrase as in (85c).

(85) a. The reporter [who the senator [who **I** met] attacked] disliked the editor].
b. The reporter [who the senator [who **the professor** met] attacked] disliked the editor].
c. The reporter [who the senator [who **John** met] attacked] disliked the editor].

Warren and Gibson (1999) conducted an experiment that revealed that participants find the sentence with the intervening pronoun much easier to understand than the other two sentences. Hawkins' locality theory is unable to account for the contrast between (85a) and (85c), as these sentences show no differences in the number of words. Gibson (2000) argues that elements that are more accessible in the discourse are easier to process. According to him, substantial processing costs are associated with the processing of the head noun of an NP that refers to a new discourse object and the head verb of a VP that refers to a new discourse event. Pronouns typically refer to entities or individuals that are focussed and highly accessible in the discourse, which makes them easier to process.[58] A nice aspect of Gibson's model is that it can account for many of the locality effects observed by Hawkins (1994, 2004, 2014), such as the differences between the sentences in (83).

Crucially, Gibson mentions that it is not only discourse-referring nouns and verbs that can cause processing cost increments. In fact, Gibson mentions that adjectives and contentful prepositions in particular, but perhaps all types of words, can cause (some) processing cost increments. This assumption causes problems in the implementation of this theory, as it it unclear how such different costs should be weighed. This is discussed further in section 4.3.

Both Hawkins' (1994; 2004; 2014) and Gibson's (1998; 2000) models are based on the assumption that an increased distance between related elements is associated with processing difficulties. This assumption has been challenged by a number of experiments that have revealed *anti*locality effects, in which intervening material facilitated, rather than hindered, processing. This is discussed

[58] There are other conceivable explanations for the contrast in (85). For instance, the fact that pronouns in English are specified for case might make it easier to determine the structure of the clause. This might make interruptions by pronouns less complex. One way to investigate this option is by considering interruption by a third person pronoun. If case is the distinguishing variable, intervention of this pronoun should be as acceptable as interruption by a first person pronoun. Gibson (1998) reports an experiment performed by Gibson and Warren (1998) that demonstrated that doubly nested relative clauses are easier to process when a first or a second person pronoun is in the subject position of the most embedded clause, than similar structures in which a third person pronoun is in the subject position. This seems to suggest that discourse indeed affects language processing.

in the next section. It will become apparent that there are various incompatible ideas on the precise factors that influence language processing.

4.2.2 Intervening material can make processing easier

Consider the differences between the two sentences in (86). In these sentences, the relative clause has a relation with the head noun, and the verb has a relation with the object. The overall distances between these related items are larger in sentence (86a), in which the relative clause precedes the main verb, than in sentence (86b), in which a relative clause is extraposed. Locality-based processing models thus make the prediction that sentence (86b) will be easier to process than sentence (86a).

(86) a. Er hat die Rose, [$_{\text{RC}}$ die wunderschön war], hingelegt.
He has the rose, that beautiful was, laid.down.
'He has laid down the rose that was beautiful.'

b. Er hat die Rose hingelegt, [$_{\text{RC}}$ die wunderschön war].
He has the rose laid.down, that beautiful was.
'He has laid down the rose that was beautiful.'

Indeed, this prediction is borne out in a corpus study by Uszkoreit et al. (1998). However, other studies have revealed conflicting results. Konieczny (2000) reports on a reading experiment in which the matrix verb was consistently read faster in sentences with an interrupting relative clause, compared with sentences with an extraposed relative clause, even when locality-based principles predict an extraposed relative clause to be easier. This finding is further underlined by an experiment involving acceptability judgements. Sentences with an intervening relative clause were judged systematically better than sentences with an extraposed relative clause. On the basis of these results, Konieczny hypothesizes that intervening relative clauses can facilitate processing, because they provide more time to anticipate the upcoming matrix verb. He claims that processing of a word becomes easier when it is preceded by other material, in disagreement with locality-based theories.

However, the results of the acceptability judgements were not as clear-cut as one might expect on the basis of Konieczny's theory. While the acceptability judgements indeed revealed that intervening relative clauses are generally more acceptable, the interruption of a relative clause became somewhat less acceptable when the relative clause was longer. This result is more in line with locality-based models. The three types of experiments thus show contradicting results:

- The corpus data support locality-based theories.

- The reading times support antilocality theories. Intervening relative clauses lead to shorter reading times, even when locality-based theories predict extraposition to be better.
- The acceptability judgements support both theories. Relative clauses are generally preferred to show up in an intervening position, even when locality-based theories predict extraposition to be better. However, the degree of the acceptability is affected by the size of the elements involved.

These results suggest that the nature of processing complexity is different for distinct modalities. This is very curious and this makes it very difficult to make predictions for different types of experiments. This hence requires a solid explanation.

There are a number of factors that have been proposed in the literature that might affect (anti)locality, namely: the difference in perception or production; the difference between local and global costs; the effect of normativity; the type of intervening items; and the complexity of the construction. Unfortunately, none of these factors can account for all results that were discussed so far.

Perception or production

Konieczny (2000) attributes the differences between reading times and acceptability judgements to a difference between language perception and language production. He hypothesizes that locality is a principle of language production and not of language perception. Other than reading times, acceptability judgements and corpus data concern language production and therefore exhibit locality effects. However, there is no theoretical reason why locality would only affect language production and not language perception. Linguistically, this is very unlikely. In fact, it is incompatible with the general idea that locality affects perception as well as production, for instance in long-distance wh-movement.

Additionally, the difference between language production and language perception does not explain the discrepancy between the corpus data and acceptability judgements. Konieczny mentions that corpus data might have been edited several times, but it is unclear how this can account for this difference. There is no clear reason why editing would result in more frequent extraposed structures.

Global or local costs, and normativity

Hawkins (2014) attributes the difference between corpus data and reading times to the fact that reading times reflect local points of processing load, in this case the reading times at the point of reading the main verb. Corpus data, on the other hand, reflect the structures selected based on global measures of processing ease. He argues that locality effects are not necessarily expected to be reflected in local reading times. Even if this is true, this does not account for the differences between the corpus data and the speakers' judgements. According to Hawkins, the judgements "suggest a possible confound resulting from a normative bias" (Hawkins 2014:56). However, there is no clear reason why normativity would lead to a preference for interrupting over extraposed relative clauses.

The types of elements involved

A third possible explanation for the difference between corpus data on the one hand and reading times and acceptability judgements on the other could come from the types of elements involved. This is an approach taken by Vasishth and Lewis (2006). They argue that items that are active in memory are easier to process. According to their theory, when an item α predicts the presence of a following item β, β becomes easier to process when it is preceded by α. In addition, intervening items that have a relation with either α or β reactivate the prediction of β, which facilitates processing of β; an antilocality effect. This effect is cancelled when the intervening item is similar to the predicted element. For instance, if a verb predicts the presence of a human-referring noun phrase, processing is not facilitated by an intervening noun that is also human-referring. Finally, if the intervening item does not reactivate the predicted element, or is very complex itself, processing becomes harder, because the activation of the predicted element begins to decay. This leads to a locality effect.

In this theory, intervening elements can thus facilitate processing, hinder processing, or neither facilitate nor hinder processing, depending on three key ingredients: predictions, similarity, and complexity. Unfortunately, Vasishth and Lewis (2006) do not explicate these notions. For instance, they assume items to be similar when they are "mutually similar along some dimension" (Vasishth and Lewis 2006:781). This notion is thus open to interpretation, which makes it very difficult to implement this theory to new data. This will become evident in section 4.3.2. Moreover, this theory cannot explain the discrepancy between the reading times and the acceptability judgements observed by Konieczny (2000). As these experiments included the same items, the similarity-based theory does not predict the results of these experiments to differ from each other.

Another problem for the similarity-based interference theory of Vasishth and Lewis (2006) comes from yet another processing experiment, conducted by Levy (2008). He demonstrated that antilocality effects can even show up when the intervening constituent does not reactivate the upcoming head. He reports on an experiment performed by Jaeger et al. (2005) that measured reading times at the matrix-clause verb in the sentences in (87). This experiment showed that reading times at the matrix verb are lower when there are more adverbs in the intervening relative clause. Crucially, there is no dependency relation between the adverbs in the relative clause and the elements in the matrix clause.

(87) a. The player [that the coach met at 8 o'clock] bought the house.

b. The player [that the coach met by the river at 8 o'clock] bought the house.
c. The player [that the coach met near the gym by the river at 8 o'clock] bought the house.

The complexity of the clause

Levy proposes that that the difficulty of a word is related to the probability of that word occurring in that context (see also Hale 2001). When a language user encounters a temporal phrase in the sentences in (87), it becomes less likely that another temporal phrase will be encountered, and more likely that the upcoming word will be the matrix verb. Processing the matrix verb is then easier when more material intervenes, which is reflected in the reading times.[59]

Levy (2008) thus argues that all preceding elements will limit the choice in the upcoming constituents. Constituents that are preceded by more material are easier to process. This theory would never predict locality effects to arise, contrary to fact. Levy discusses a finding by Grodner and Gibson (2005), who show that the embedded verb in an intervening relative clause is read more quickly in subject relative clauses than in object relative clauses, even though more elements have been encountered in an object relative clause. To account for this result, Levy hypothesizes that long-distance dependencies such as relativization are sensitive to locality, while local syntactic dependencies are only sensitive to expectations. The complexity of the clause would then affect whether locality plays a role. A problem for this explanation is that locality effects have been argued to occur in simplex constructions as well (reconsider Hawkins' examples in (83)). It thus remains a mystery if and when locality or antilocality effects do arise.

In addition to this problem, the difference between reading times and acceptability judgements that were observed by Konieczny's (2000) experiments cannot be attributed to the complexity of the construction, as these experiments used the same constructions.

4.2.3 Summary

This section looked at a variety of models of sentence processing. Two main types of processing preferences were distinguished: locality and antilocality. Where locality is often argued to be an effect of limitations on working memory, antilocality is argued to be helpful for language users' expectations. The various experiments discussed in this section exhibited contradicting results and none of the theories of language processing was able to account for all of

[59] In addition to antilocality effects, Levy's (2008) theory also predicts frequency effects. Since the expectations of language users are based on the experience they have with the words and types of constructions in a sentence, more common words or constructions will be less difficult to process (see also Keenan and Comrie (1977); MacDonald et al. (1994), among others). A problem with such frequency-based explanations is discussed in section 4.4.

the facts. Consequently, it remains unclear what variables play a role in language processing. If and when language processing is affected by locality and/or antilocality effects remains an open question. This makes it very hard to apply these models to the data under investigation here.

There is an additional problem for applying these models to three-verb clusters. In order to test the hypothesis that language processing affects the speakers' judgements in the comparative judgement task, the required notions need to be well-defined. Unfortunately, all of the processing models that were discussed in this section involve some abstract notions, such as 'relation', 'similarity', and 'complexity' that are not well-defined. This leads to problems in the implementation. Despite these problems, the next section investigates whether any interpretation of the processing models can account for the results of the comparative judgement task.

4.3 Processing verb clusters

The remainder of this chapter investigates if the word order preferences for verb clusters can be attributed – entirely or partially – to properties of language processing. If speakers' preferences are entirely driven by processing preferences, the assumptions about the grammatical structure of verb clusters that were proposed in chapter 3 should be set aside to the extent that they do not follow from functional factors. This means that none of the six potential orders of three-verb clusters should be excluded on the basis of grammatical properties. This would fit neatly in the idea that functional approaches "typically recognize no level more abstract than surface structure and have no real equivalent to abstract parameter settings" (Newmeyer 2005:43). By initially setting aside all assumptions about the grammatical structure of verb clusters that were proposed in chapter 3, it can be thoroughly investigated whether those assumptions are really necessary to account for the observed order preferences, or whether the preferences can be attributed entirely to language processing.

It is also conceivable that the informants' preferences are only partly driven by processing preferences, and the grammatical theory proposed in chapter 3 should be assumed. In fact, many current functional models are completely compatible with the presence of an abstract language system. Therefore, processing models that take the assumptions about grammatical structure into account should be investigated as well. Section 4.4 discusses such an approach.

The central hypothesis that is investigated in this section is depicted in (88).

(88) When speakers can choose between multiple verb orders and they prefer one order to another, the preferred order will be easier to process.

For reasons of simplicity, only one of the two types of verb clusters is discussed here, namely *moet kunnen zwemmen* 'must.MOD can.MOD swim.INF.' The test sentence for this cluster has the fewest arguments and no adverbs, which makes it the least complex.[60] The average results of the comparative judgement task for this cluster are repeated in figure 4.1.

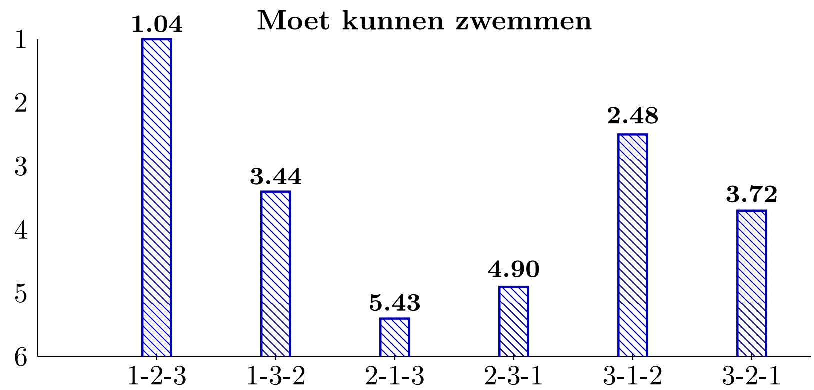


Figure 4.1: Results comparative judgement task for *moet kunnen zwemmen* 'must.MOD can.MOD swim.INF'

In the previous section, a number of preferences of language processing were discussed that might play a role in the speakers' judgements. Hawkins' (2004; 2014) and Gibson's (2000) locality-based theories would predict orders to be easier to process when related items are close to each other. Antilocality theories, such as Konieczny (2000) predict opposite effects, especially considering the fact that the test sentences were relatively uncomplicated. Vasishth and Lewis (2006) might predict either locality or antilocality effects, depending on the definition of 'similarity'. Finally, Levy (2008) might also predict an antilocality effect, however, his model also allows for frequency effects, which leads to different predictions. All of these possibilities are explored in this section.

4.3.1 Locality effects

Hawkins' minimal domains

According to Hawkins (1994, 2004, 2014), the human parser prefers linear orders that minimize the number of words in which syntactically and semantically related elements can be combined. As was mentioned in section 4.2.1, a single sentence can contain multiple relations, all of which might be affected by this processing principle. One type of relation discussed by Hawkins is that of phrasal combination; the human processor prefers to minimize the domains in which a mother phrase can be combined with its immediate constituents

[60] The theory in chapter 3 is inexplicit about the exact structural relations of verb clusters. For instance, the 3-1-2 order might involve a raised subject, or a PRO-subject. These assumptions might be relevant for language processing. However, since none of the processing models is explicit in how the syntactic derivation can affect language processing, one can only guess what effects these assumptions will have. Therefore, this issue is set aside here.

(ICs). Hawkins provides a metric that makes it possible to compare the distances of Phrasal Combination Domains (PCDs) in different sentences. For each sentence, the number of constituents that need to be combined are divided by the total number of words that are required to combine them. By aggregating the ratio scores for an entire construction, it is possible to compare competing constructions to each other.

This metric is illustrated for three-verb clusters in (89). In three-verb clusters, the immediate constituents of VP$_1$ can be recognized on the basis of two words: the head (V$_1$) and the head of its complement (V$_2$). The constituents of VP$_2$ can be recognized on the basis of V$_2$ and V$_3$. For VP$_3$ only V$_3$ needs to be processed. The IC-to-word ratios are illustrated for the 1-2-3 order in (89a) and for the 3-1-2 order in (89b).

(89)	a.	V_1	V_2	V_3.		
		1	**2**		PCD VP$_1$:	2/2 or 100%
			1	**2**	PCD VP$_2$:	2/2 or 100%
				1	PCD VP$_3$:	1/1 or 100%
						Mean: 100%
	b.	V_3	V_1	V_2.		
			1	**2**	PCD VP$_1$:	2/2 or 100%
		1	2	**3**	PCD VP$_2$:	2/3 or 66.7%
		1			PCD VP$_3$:	1/1 or 100%
						Mean: 88.9%

The domains of phrasal combination are minimal in the uniformly ascending 1-2-3 and descending 3-2-1 orders, and non-minimal in all other orders. There is no difference in the PCDs between these other orders. Consequently, the preference to Minimize Domains predicts the ranking: 1-2-3, 3-2-1 $>$ all other orders. This does not account for the results of the comparative judgement task, which for instance revealed a much higher ranking for the 3-1-2 than the 3-2-1 order.

Since Hawkins' Minimize Domains principle can affect various types of relations, other relations between elements in a sentence might also be affected by it. This leads to many conceivable implementations of this theory, all of which make different predictions for preferences in verb cluster orders. This becomes particularly clear when one compares two previous studies of verb clusters that apply Hawkins' theory. Both Culicover (2014) and Bloem et al. (2017) discuss the effect of Minimize Domains on structures with two-verb clusters, as in (90).

(90) a. S – O – $V_{1.\text{FIN}}$ – $V_{2.\text{MAIN}}$

b. S – O – $V_{2.\text{MAIN}}$ – $V_{1.\text{FIN}}$

According to Culicover (2014), Hawkins' theory makes the prediction that the 2-1 order is easier to process than the ascending 1-2 order, because the main verb immediately follows its arguments in the descending 2-1 order, while the auxiliary verb intervenes in the 1-2 order. On the other hand, Bloem et al. (2017) state that the 1-2 order should be easier to process, because the subject is in a dependency relation with the finite auxiliary verb. These two interpretations of Hawkins' theory thus make the exact opposite predictions. This contradiction shows very clearly that Hawkins' broad definition of a 'relation' is highly problematic. Since Hawkins is not explicit about the relevant relations, the orders of three-verb clusters could theoretically be affected by a large number of relations, in addition to the Phrasal Combination Domains: the θ-relation between the subject and the main verb, a relation between the auxiliary verbs and the main verb, and an agreement relation between the subject and the finite verb. Different combinations of these variables lead to different outcomes. The next graphs depict the results of three of these possible implementations.[61] Here, the results are plotted in a single graph with the results of the comparative judgement task. If any of these interpretations of the Minimize Domains principle were in line with the speakers' judgements, the graphs should be comparable to the results of the comparative judgement task. This is not the case. The 3-1-2 order, for instance, was found much more acceptable than the 3-2-1 order, but none of the implementations of the Minimize Domains principle led to a higher score for the 3-1-2 order than the 3-2-1 order. In addition, while the most unacceptable 2-1-3 and 2-3-1 orders often fall out as the most difficult orders, the difference with the other orders is not as large as one might expect based on the results of the comparative judgement task. Consequently, the results of the comparative judgement task cannot straightforwardly be attributed to the processing preference to Minimize Domains.

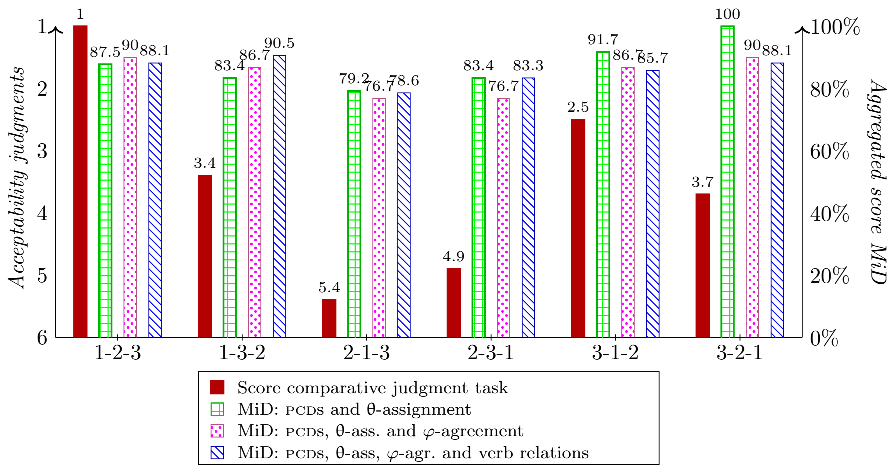


Figure 4.2: Comparing the Meertens Panel judgements with the predicted processing preferences, following Hawkins' locality-based principle.

Now, a problem presents itself. Since Hawkins is inexplicit about the relations

[61] The calculations can be found in appendix A.

that are relevant for the processor in addition to Phrasal Combination Domains, these implementations of the theory might be wrong. Potentially, a different set of relations could lead to the correct results. For this reason, many conceivable combinations of relations were applied to the data. So far, I have been unable to achieve the correct results. However, note that even if one of these selections had been successful, it would not be explanatory, as there is no independent evidence for selecting the winning set of relations.

Hawkins' minimal domains + Haider's scope preference

Section 4.2.1 demonstrated that Hawkins' theory was insufficient to account for the observed preference for crossed over nested dependencies. It was hypothesized that Hawkins' theory has to be supplemented with another property of processing that can account for this preference. Culicover (2014) argues that the order variation patterns in verb clusters can be explained in terms of the interaction of two complexity biases. The first is based on Hawkins (1994, 2004): the preference for orders in which heads are positioned close to their dependents. The second is in the spirit of Haider (2003): a parser preference for orders in which scope-taking elements precede the material they scope over. For the scope-preference, the 1-2-3 verb cluster has the perfect order. This hence favors sentences with crossed dependencies over nested dependencies. For the dependency bias, Culicover argues that orders are preferred in which 3 is more to the front, closer to the arguments of the sentence. Culicover (2014) does not discuss the dependency relation between the subject and the finite verb in this sentence. Nor does he discuss the dependency relation between the various verbs. Hawkins' theory does not explicate which of the many possible relations are of relevance (see also the discussion page 106). For reasons of simplicity, Culicover's assumptions are adhered to here.

According to Culicover, all verb orderings are in principle available for all languages, and the interaction of the two complexity biases can lead to different effects. He argues that 1-2-3 and the 3-2-1 orders are cross-linguistically more frequent than the other orders, because they are both optimal for one of the complexity biases; 1-2-3 perfectly reflects the scope order, while the 3-2-1 order is perfect for dependency distances. However, the results of the comparative judgement task did not indicate that the 3-2-1 order is preferred over all other orders, such as the 1-3-2 order.

It is possible to draw different predictions from the two complexity biases. While the 1-2-3 and the 3-2-1 orders are indeed perfect for one of the two complexity biases, they also maximally violate the other complexity bias. The 1-3-2 and 2-3-1 orders, conversely, are neither perfect for both complexity biases, nor do they maximally violate both complexity biases. This is illustrated in table 4.1.

order	**complexity**	**perfect order**	**intermediate**	**maximal violation**
1-2-3	scope	✓		
	dependency			✓
1-3-2	scope		✓	
	dependency		✓	
2-1-3	scope		✓	
	dependency			✓
2-3-1	scope		✓	
	dependency		✓	
3-1-2	scope		✓	
	dependency	✓		
3-2-1	scope			✓
	dependency	✓		

Table 4.1: Complexity bias score for the verb orderings, following Culicover (2014)

Another potential ranking on the basis of these results might lead to 3-1-2 > all other orders > 2-1-3. This does not fully account for the ranking of the verb clusters provided by the Meertens Panel. For instance, the 1-3-2 order is not predicted to be more acceptable than the 2-3-1 order. Hence, the language users' preferences cannot be attributed solely to these preferences of language processing.

Bach et al.'s multiple dependencies

Bach et al.'s (1986) assumption that orders in which material that can immediately be integrated into the structure are easier to parse, cannot explain the results of the comparative judgement tasks either. Consider the relations in the 3-1-2 and the 1-3-2 orders. Regardless of whether the relations between the verbs are assumed to be relevant or not, these distances do not predict the 3-1-2 order to be more acceptable than the 1-3-2 order.

(91) a. ...dat iedereen moet$_1$ zwemmen$_3$ kunnen$_2$.
...that everyone must swim can
'...that everyone should be able to swim.'

b. ...dat iedereen zwemmen$_3$ moet$_1$ kunnen$_2$.
...that everyone swim must can

One might stipulate that the relation between the subject and the finite verb is not relevant. In that case, the 3-1-2 order is correctly predicted to be more acceptable, as it would involve one crossed dependency less. However, the 3-2-1 order is then wrongly predicted to be much more acceptable than both of these orders, as no relations are nested or crossed in the 3-2-1 order, as is illustrated in (92). This is not in line with the results of the comparative judgement task, which clearly indicated that the 3-1-2 order is more acceptable than the 3-2-1 order.

(92) ...dat iedereen zwemmen$_3$ kunnen$_2$ moet$_1$.

To conclude, the results of the comparative judgement task cannot be attributed to the number of nested and crossed dependencies.

Gibson's costs of integration and storage

Where Hawkins measures distance in terms of the number of words, Gibson (2000) argues that distance is affected by the number of discourse referents that intervene between two related items. His locality theory has two components: (i) the costs of structural integrations, which involves connecting a word into the structure built thus far, and (ii) keeping the structure in memory, which includes keeping track of incomplete dependencies. In addition to this, there are costs involved with discourse integrations. Gibson mentions that the exact relationship between integration and memory costs is still unclear. Therefore, both types of costs are considered here.

Gibson first focusses on the combined costs of discourse and structural integrations. Discourse integrations concern accessing or constructing the discourse structure of each discourse referent. Discourse referents are defined as words introducing new discourse elements, namely NPs and the main verbs of VPs that refer to events. Structural integrations concern connecting incoming words into the structure. Gibson argues that the processing costs for structural integrations are affected by discourse referents that intervene between items that need to be connected to each other. The distance-based integration cost profile for the 1-2-3 and 3-2-1 orders is presented in table 4.2. The amount of energy that is required to integrate a word in the structure is quantified in terms of energy units (EUs). The top row displays the costs associated with constructing the new discourse referents in the sentence: the subject and the verb that refers to a new discourse events, i.e. the lexical verb. The second row displays the costs of structural integration. In three-verb clusters, the head of VP_1 needs to be structurally connected with the subject and V_2, and the head of VP_2 needs to be structurally connected with V_3.

In the sentence with a 1-2-3 cluster, the first discourse referent that is encountered is 'Jan'. This integration thus consumes 1 energy unit. Subsequently, the highest verb of the cluster, 'moet', is integrated with the subject. No new discourse referents have been processed since the subject 'Jan' was processed, so the distance-based integration cost is 0 energy units at this point. The next point at which integration takes place is at the middle verb 'kunnen', which is integrated with V_1 without integration costs. Finally, the main verb is integrated into the structure. The cost of building the new discourse structure is 1 energy unit, because 'zwemmen' is a discourse referent. It costs 0 energy units to structurally integrate this phrase, since there are no intervening discourse referent heads between V_2 and 'zwemmen'.

	...dat ...that	Jan Jan	moet$_1$ must	kunnen$_2$ can	zwemmen$_3$. swim
discourse integration	0	1	0	0	1
structural integration	0	0	0_S	0_{V1}	0_{V2}
total costs	0	1	0	0	1
	...dat ...that	Jan Jan	zwemmen$_3$ swim	kunnen$_2$ can	moet$_1$. must
discourse integration	0	1	1	0	0
structural integration	0	0	0	0_{V3}	$1_S + 0_{V2}$
total costs	0	1	1	0	1

Table 4.2: Word-by-word integration costs for the 1-2-3 and 3-2-1 verb clusters.

In the 3-2-1 cluster, integration of the verbs occurs slightly differently. The lowest verb V_3 cannot be integrated into the structure until the middle verb is processed. At that point, V_2 is integrated as the head for V_3. No costs are associated with this integration, since the middle verb is not a discourse referent. At the point of processing the highest verb of the cluster, V_1, both the subject and the middle verb can be integrated into the structure. One discourse referent, 'zwemmen', was processed since the subject was input, so this integration consumes 1 energy unit.

Gibson assumes that "the overall intuitive complexity of a sentence depends to a large degree on the maximum complexity incurred at any processing state during the processing of a sentence" (Gibson 2000:105). So the difficulty of a sentence is not calculated by summing the different costs, but by looking at local difficulties. This assumption can be used to compare the complexity of the different verb clusters to each other. Since all verb orders consume maximally one energy unit at one point, there should be no difference in the complexity involved in integrating the different verb clusters. In other words, integration costs do not entail that different verb cluster orders are associated with different processing costs. Hence, this cannot be used to account for the results of the comparative judgement task.[62] A problem for this implementation is that Gibson mentions that not only discourse referents, but all words cause some processing cost increments. Since he does not explicate how to model different degrees of processing costs, it is very unclear how to weigh such effects. These

[62] If one were to assume that the 3-1-2 and the 1-3-2 orders involve a nominalized 3, it becomes even harder to account for the results of the comparative judgement task. These orders involve an additional discourse referent, as V_2 is now the main verb. As a consequence, the 3-1-2 order will have a higher maximal processing cost than the 3-2-1 order, even though the 3-1-2 order was found much more acceptable in the comparative judgement task.

potential costs are therefore set aside here.

The second component in Gibson's (2000) processing theory concerns storage costs; the costs involved with storing the words that are required by other words. Gibson argues that there is a storage cost associated with each syntactic head that is still required to complete the current input string as a grammatical sentence. For instance, in the sentences under investigation, at the point of processing the complementizer, at least two syntactic heads are still required to form a grammatical sentence: a noun and a verb. There is therefore a cost of 2 'memory units' at this point. Consider the storage costs associated with the different verb orders, depicted for 1-2-3 and 3-2-1 in table 4.3. In the 1-2-3 cluster, the first encountered verb is the finite modal verb, which requires an infinitival main verb. Conversely, in the 3-2-1 cluster, the first encountered verb is the main verb, which requires a finite verb.

	...dat ...that	Jan Jan	$moet_1$ must	$kunnen_2$ can	$zwemmen_3$. swim
requires	S + V_{FIN}	V_{FIN}	V_{MAIN}	V_{MAIN}	-
storage costs	2	1	1	1	0
	...dat ...that	Jan Jan	$zwemmen_3$ swim	$kunnen_2$ can	$moet_1$. must
requires	S + V_{FIN}	V_{FIN}	V_{FIN}	V_{FIN}	-
storage costs	2	1	1	1	0

Table 4.3: Word-by-word storage costs for the 1-2-3 and 3-2-1 verb clusters

No verb cluster order lead to maximum storage costs that exceed the 2 memory units arising at the point of processing the complementizer. As a result, this property of processing also does not make the prediction that different verb orders are associated with differences in processing costs. To conclude, no application of the various locality-based models is able to account for the results of the comparative judgement task.

4.3.2 Antilocality effects

Some have argued that intervening items can actually decrease, rather than increase, processing difficulty. However, simply reversing the predictions based on Hawkins' locality theory, does not lead to the correct results. For reasons of space, I shall leave it to the reader to verify this. Since Vasishth and Lewis (2006) argue that the types of elements in a clause affect whether locality or antilocality effects show up, this section focusses on predictions drawn from their theory.

According to Vasishth and Lewis (2006), items are easier to process when they are predicted to occur. So when an item α predicts the presence of a following item β, β becomes easier to process when it is preceded by α. In addition, intervening items that have a relation with either α or β reactivate the prediction of β, which facilitates processing of β. However, when these intervening items are similar to the predicted β, this facilitation is cancelled. The definitions of 'similar' and 'predicted' are hence crucial to the implementation of this theory. Consider the verb cluster under investigation.

(93) ...dat iedereen moet$_1$ kunnen$_2$ zwemmen$_3$.
...that everyone must can swim

First, when a language user processes the subject in this sentence, (s)he predicts there to be a finite verb and a main verb (or a finite main verb), as those items are still required to make a grammatical sentence. Potentially, a V_2 that precedes either of these verbs might reactivate these predictions and hence facilitate processing. However, depending on the definition of 'similarity', this reactivation might be cancelled. Unfortunately, Vasishth and Lewis do not properly define this notion, so it is unclear whether V_2 will facilitate processing or not. Both options are investigated here.

The first option is that V_2 reactivates the prediction of the finite and the lexical verb. In (93), the subject has activated the prediction of the finite verb and the main verb. In the orders 2-1-3 and 2-3-1, the presence of V_2 will reactivate this prediction, because it has a relation with both the finite and the main verb. This will therefore facilitate processing of V_1 and V_3. Orders in which V_2 precedes the other two verbs, 2-1-3 and 2-3-1, are then predicted to be the easiest to process. This is not in line with the results of the comparative judgement task, as these orders in fact received the worst scores.

The second option is that V_2 does not reactivate the prediction of the finite and the main verb, because it is considered 'similar' to the other verbs. In that case, an intervention of V_2 will neither hinder, nor facilitate processing. As a consequence, the different verb orders are not predicted to display differences in processing. This leads to the wrong results. The 2-1-3 order, for instance, would be predicted to be as acceptable as all other orders.

In addition, neither of these possible implementations of antilocality would predict a difference between the 1-3-2 and the 3-1-2 orders. Nevertheless, the 3-1-2 order was found much more acceptable than the 1-3-2 order. To conclude, antilocality cannot account for the results of the comparative judgement task.

4.3.3 The language users' expectations

As discussed in section 4.2.2, Levy (2008) argues that words are easier to comprehend in contexts where they are highly predictable. According to him, this is affected by the experience that language users have with the words and types of constructions in a sentence. He hypothesizes that constructions that are rare might be more difficult to process, because they are less likely to occur. Words are easier to process when they frequently occur in the position where they are observed, i.e. by statistical regularities. Following this hypothesis, the comparative judgements might be based on the probability of each verb occurring in its position in the different verb clusters.

There is a theoretical problem with the hypothesis that the lower frequency of a construction causes difficulties for the language user, as the reverse is also a theoretical possibility; the low frequency of this order might be the result of its difficulty. According to Hawkins (2004), structures that are preferred in performance are selected more frequently in usage. According to this view, the low frequency of a construction is a result of it being disfavored in language performance. This leads to a *chicken-and-egg* problem. Either a construction is more difficult to process because it is less frequent, or a construction is less frequent because it is harder to process. There is no clear-cut way to distinguish between these theoretical possibilities. For now, this problem is ignored and it is investigated whether frequency of occurrence could have affected the speakers' judgements.

A search was conducted using the PaQu search engine. This search engine enabled a search through 129.921 syntactically annotated sentences from the Spoken Dutch Corpus and 8.707.708 syntactically annotated sentences from LASSY large. The Spoken Dutch Corpus is a corpus of Dutch spoken by adults living in Flanders and the Netherlands, and LASSY Large is a corpus of sentences taken from Wikipedia articles.[63,64]

Unfortunately, this database did not contain enough data to determine the likelihood of the various words in the test sentence. For instance, the likelihood of encountering the modal verb 'moet' as the first verb in the test sentence could be determined by using trigrams. To that aim, one first needs to find all sentences that started with a complementizer followed by a noun or a pronoun and then any modal finite verb. This search only led to 162 instances in Lassy Large and 4 instances in the Spoken Dutch Corpus (SDC). These numbers are too small to make it possible to draw reliable conclusions. For that reason, the search was simplified and focussed on the occurrences of three-verb clusters. The results are depicted in table 4.4.

[63] I am grateful to Liesbeth Augustinus and Jelke Bloem for showing and explaining this search engine to me and to Erik Tjong Kim Sang and Gertjan van Noord for helping me perfect my search.

[64] The methodology for this search is described in appendix B.

	MOD.FIN-MOD.INF-V.INF		MOD.FIN-AUX.INF-PTCP	
	SDC	LASSY Large	SDC	LASSY Large
1-2-3	6	286	7	1315
1-3-2	0	0	5	115
2-1-3	0	0	0	0
2-3-1	0	0	0	0
3-1-2	0	0	7	969
3-2-1	0	0	0	0

Table 4.4: Frequencies in the two corpora using PaQu

The most striking results concern the MOD.FIN-MOD.INF-V.INF verb cluster. Since only one of those orders occur in standard Dutch, all others are equally improbable. These results hence do not account for the results of the comparative judgement task.

Although the frequencies of the orders for the MOD.FIN-AUX.INF-V.PTCP verb clusters closely match the ranking of the different verb orders, they are also not sufficient to account for the patterns. Especially since the 3-2-1 order is completely absent from standard Dutch, speakers are not expected to rank this order much higher than the 2-1-3 and the 2-3-1 orders.[65] In addition, these frequencies might make the prediction that the 1-2-3 order is more acceptable than the 3-1-2 order, contrary to fact. There is thus no evidence that shows that the speakers' judgements in the comparative ranking task are a result of experience.

It is conceivable that the processing models discussed so far do not make the correct predictions, because there are additional factors that play a role in the order of verbs in a cluster. Corpora of real-world texts have revealed many factors that play a role in the use of verb clusters within one language. The next subsection takes a short detour to discuss an approach that implements these factors in a processing model.

4.3.4 A usage-based processing model

Usage-based approaches often attribute variation to a combination of different factors. In overviews of corpus research by Coussé et al. (2008); Sapp (2011) and

[65] A potential explanation for the fact that speakers did find the 3-2-1 order in both verb clusters more acceptable than the non-occurring 2-1-3 and 2-3-1 orders comes from familiarity with what happens in varieties of Dutch. Perhaps speakers are familiar what happens in the varieties and their acceptability is (partly) based on that knowledge. Speakers may be familiar with the 3-2-1 order, since this order occurs frequently in the North. The experience-based theory might then make the prediction that speakers who have (more) contact with the varieties where this order occurs will find this order more acceptable. However, as was discussed extensively in chapter 3, this prediction is not borne out. The 3-2-1 order is not more acceptable for speakers from areas where that order is observed, than speakers from other areas.

Bloem et al. (2014) the following factors are argued to affect order variation in verb clusters: *the type of clause, whether or not the sentence has an extraposed phrase, the finiteness of the verbs, the frequency of the main verb, the genre of the text, the information value of preceding word, the length of the middle field, whether the main verb is a particle verb or not, whether or not a participle has an adjectival status, the depth of embedding, priming, the region of the language user, rhythmic factors, the social class of the language user, the time period,* and *the type of auxiliaries involved.*

As there are so many, partly unrelated, factors involved, corpus linguists often make use of so-called multifactorial models to unravel the relative effect of these factors (see Lötscher 1978; De Sutter 2005; Sapp 2006; Arfs 2007; Coussé et al. 2008; De Sutter 2009; Dubenion-Smith 2010; Bloem et al. 2014, among others). Since multifactorial models can quantify over different variables, the effect of a factor can be measured while controlling for other factors. In this manner, Bloem et al. (2017), for instance, have revealed a ranking of various factors that affect the order of two-verb clusters across constructions. This is displayed in table 4.5, in which the highest ranked variable has the largest effect on the order of verbs in a cluster.

Rank	**Factor**
1	type of auxiliary
2	priming
3	te-infinitive
4	extraposition
5	length of the middle field
6	frequency of the main verb
7	information value of preceding word
8	morphological structure of the main verb
9	multi-word units
10	syntactic depth
11	definiteness of the preceding word

Table 4.5: Results Bloem et al. (2017)

The question that arises is why these different factors affect verb clusters. Bloem et al. hypothesize that many of these factors could be attributed to processing complexity. According to them, there is a relation between the complexity of these factors and the observed verb orders. They argue that the variables that are more difficult to process are often associated with the ascending 1-2 order and conclude from this that the 1-2 order must be easier to process. This conclusion follows from the assumption that more difficult contexts will be more likely to have verb orders that are easier to process. In this way, speakers can reduce the total processing complexity of a sentence. For instance, Bloem et al. (2017) argue that particle verbs are more complex than regular main verbs and they demonstrate that such particle verbs are more likely to occur with the 1-2 order. They conclude that the 1-2 order must be easier to process. They provide similar arguments for the factors ‘extraposition’, ‘length of the middle field’, ‘multi-word units’ and ‘definiteness’.[66]

The conclusion that the 1-2 order is always easier to process is questionable. First, it is not obvious that more difficult constructions require easier orders. In addition, there is no apparent reason for assuming that one order is easier to process to start with. It makes one wonder why the other order would ever be used. Furthermore, if one order is indeed easier to process, it is not obvious that this is the 1-2 order. The argumentation provided by Bloem et al. is not fully valid, as some results make the opposite prediction. For instance, Bloem et al. demonstrate that sentences with deeper structural embeddings are less likely to occur in the 1-2 order, even though these are presumably harder to process. There are more such problematic results. For instance, Bloem et al. (2017) argue that longer middle fields are harder to process, because they are more likely to lead both to additional dependencies and to longer dependency lengths (in line with Gibson 1998). Since they find that the 1-2 order is associated with longer middle fields, they argue that this illustrates that this order is easier to process. However, Bloem et al. also discuss an opposite (anti-locality) view suggested by De Sutter (2007), who argues that longer middle fields make it easier to accurately predict the properties of the upcoming verb. This would undermine their conclusion that the 1-2 order is easier to process.

To conclude, there is no validity to the assumption that language processing can explain the verb order patterns.

4.3.5 Intermediate conclusion

This section aimed at investigating whether properties of language processing can account for the results of the comparative judgement task. Many, partially contradicting, theories of language processing were discussed. To be able to apply those models to this type of data, it is essential for them to be well-defined. It became very apparent that this is not the case. As a matter of fact, the models are very imprecise, even on crucial aspects of the theory. For instance, the broad definition of a ‘relation’ provided by Hawkins (1994, 2004, 2014) can be applied to a lot of dependency relations, leading to an explosion of possible implementations of this model. The most salient interpretations of the various models have been applied to the data. Crucially, none of these implementations led to the correct results. There is no apparent way in which processing models can be applied to the data that leads to results that are in line with the results of the comparative judgement task. This is especially striking in light of the fact that the previous chapter has demonstrated that a grammatical analysis can account for a substantial part of the results of the comparative judgement task.

[66] Note that this model does not explain the relative weight of these factors. It remains a mystery why, for instance, extraposition has a larger effect on the order of verbs in a cluster than a longer middle field.

The next section aims to investigate if a processing model that takes the assumptions about grammatical structure into consideration is better able to account for the results.

4.4 Language processing as a Third Factor

The previous sections have demonstrated that there is no processing account that successfully predicts the results of the comparative judgement task. However, if one were to follow Chomsky's (2005; 2007; 2008) Third Factor model, processing preferences do not need to account for all order variation. In such a model, processing preferences can only select from structures made available by the grammar system (see also Altmann and Steedman (1988), a.o.). This limits the amount of variation that processing models need to explain.

The first thought that comes to mind in combining a processing approach and a grammatical approach is using a grammar-based model of complexity, such as the Derivational Theory of Complexity (DTC) (Miller and Chomsky 1963). According to this theory, the complexity of a structure is dependent on the number of transformations that are required to generate that structure. The relevance of issues of language processing for generative theory was re-established by Marantz (2005). According to him, generative theory is psycholinguistic by nature, as it addresses representations and computations in the minds and brains of speakers. The complexity of the derivation should therefore in principle affect language processing. As was argued in chapter 3, the various observed orders of verb clusters do not require any transformations – they can all be base-generated. These orders are hence not be affected by the DTC. Consequently, this theory cannot account for the results of the comparative judgement task. For instance, it remains unclear why the 1-3-2 and the 3-1-2 orders, which can both be base-generated, are not equally acceptable.

Another possibility is to combine the processing models discussed in section 4.3 with the grammatical model that was proposed in chapter 3. In this section it is investigated whether such a combined model can account for the results of the comparative judgement task. The question that immediately arises is how such a combined model would work. If one takes the assumption seriously that the human processor can only select from structures that have been made available by the grammar system, at least three scenarios are conceivable:

1. All orders that can be derived by the grammar are considered by the human processor, regardless of whether these orders are a part of the language variety. This means that only the 2-1-3 and 2-3-1 orders are excluded and the human processor is able to rank the other orders.

2. Only the orders that are grammatical in the language variety are considered. For standard Dutch, this means that the 3-2-1 and the 1-3.N-2 orders are never considered.

3. Only orders that are grammatical in the language variety and that are syntactically and categorically equivalent are considered by the processor. This means that the processor has nothing to say about the choice between the 1-2-3 and 3-1-2 orders.

Here, predictions are drawn from all these scenarios.

Scenario 1: All derivable orders are considered by the language processor

It is likely that speakers made use of the knowledge of their own language variety in the comparative judgement task. However, it is theoretically possible that the grammar only played a small role in the comparative judgement task and properties of the informants' own language varieties have not affected their ranking. This only excludes two verb orders from being assessed by the language processor; the 2-1-3 order can never be merged and the 2-3-1 order can never be derived with the verb clusters under investigation.

Let's consider this theoretical possibility. The processing model then needs to account for the ranking between 1-2-3, 3-1-2, 1-3-2 and 3-2-1. For the *moet kunnen zwemmen* 'must.MOD can.MOD swim.INF' cluster, this ranking is: 1-2-3 > 3-1-2 > 1-3-2 > 3-2-1. This does not improve the results. In section 4.3, it has been explicitly mentioned that the processing difficulties for these orders do not match their relative ranking, as the reader can verify for him- or herself.

Scenario 2: Only orders that are possible in the language variety are considered by the language processor

Potentially, the human processor only considers orders that can be derived in the language variety. Following this hypothesis, there are two orders that will not even be considered by the informants, because they are not a part of the language used in the comparative judgement task. First, the 3-2-1 order requires a different linearization, as was discussed in chapter 3. Secondly, the 1-3.INF-2 order requires nominalization of the infinitival 3 as well as interruption of this non-verbal element. Since no language variety has both these properties, this order should be ungrammatical. Consequently, language processing can only affect the 1-2-3, 3-1-2 and 1-3-2 orders for the *moet hebben gemaakt* 'must.MOD have.AUX made.PTCP' cluster, and the 1-2-3 and 3-1-2 orders for the *moet kunnen zwemmen* 'must.MOD can.MOD swim.INF' cluster.

(94) a. Ranking for *moet hebben gemaakt* 'must.MOD have.AUX made.PTCP': 3-1-2 > 1-2-3 > 1-3-2

b. Ranking for *moet kunnen zwemmen* 'must.MOD can.MOD swim.INF': 1-2-3 > 3-1-2

The implementation of Hawkins (1994, 2004, 2014) theory massively improves if one follows these assumptions. As the results in figure 4.2 on page 107 demonstrated, there are two possible implementations of Hawkins' theory that lead to a higher score for the 1-2-3 than the 3-1-2 order for the *moet kunnen zwemmen* 'must.MOD can.MOD swim.INF' cluster. The guiding idea in these implementations is that the distance between 3 and its dependent V_2 is larger in the 3-1-2 order than in the 1-2-3 order. As a result, the 3-1-2 order is predicted to be worse than the 1-2-3 order.

To see how well Hawkins' theory handles the *moet hebben gemaakt* 'must.MOD have.AUX made.PTCP' cluster, one first needs to determine the domain distances. As this cluster has an additional argument and an adverb, the mean distances for this verb cluster are different than in the *moet kunnen zwemmen* 'must.MOD can.MOD swim.INF' cluster. Appendix A provides the calculations for this verb cluster. The implementation in which all conceivable relations are considered, comes very close to the ranking of the comparative judgement task.

Domains	1-2-3	1-3-2	3-1-2
PCDs + θ-ass	87.5 %	81.7 %	85.4%
PCDs + θ-ass + φ-agr.	80 %	75.3 %	76.3 %
PCDs + θ-ass + φ-agr. + verb relations	85.7 %	77.6 %	78.3 %
PCDs + θ-ass + φ-agr. + verb relations + relation adv. & V_3	81.3 %	76.3 %	81.1 %

Table 4.6: Predicted processing preferences for must.MOD have.AUX made.PTCP, following Hawkins' locality-based principle

(95) Results comparative judgement task for *moet hebben gemaakt* 'must.MOD have.AUX made.PTCP': 3-1-2 (mean 1.65) > 1-2-3 (mean 1.56) > 1-3-2 (mean 3.36)

If one follows these assumptions, it is possible to attribute a large part of the speakers' ranking to preferences of language processing, although it is not explained why the 3.PTCP-1-2 order receives a slightly better score than the 1-2-3.PTCP order.

The other processing models cannot be used to account for this part of the results of the comparative judgement task. None of the salient interpretations of those models provides the correct ranking for both verb clusters. This is not discussed in detail here, but is supported by the following conclusions that can be drawn from the discussion in section 4.3.

- As table 4.1 on page 109 indicates, Culicover's (2014) theory does not make the prediction that the 1-2-3 order is easier to process than the 3-1-2 order for the *moet kunnen zwemmen* 'must.MOD can.MOD swim.INF' cluster.

- Depending on the relations that are assumed to be relevant, Bach et al.'s (1986) theory either makes the prediction that the 1-2-3 order is more acceptable than the 3-1-2 order, or vice versa. This ranking is not predicted to depend on the types of verbs involved, contrary to fact.
- Gibson's (2000) theory does not make the prediction that any verb order is more acceptable (see the discussion on page 111-112).
- Vasishth and Lewis' (2006) antilocality theory would not predict a processing difference between the 3-1-2 and 1-3-2 orders (see the discussion on page 113).
- For an antilocality theory that reverses the predictions drawn from Hawkins (2014), there is no implementation that provides the correct results for both verb clusters.
- An account based on statistical regularities would wrongly predict the 1-2-3.PTCP order to be more acceptable than the 3.PTCP-1-2 order, as the 1-2-3 order is more frequent (see table 4.4 on page 115).

Scenario 3: Only orders that are possible in the language variety and that are syntactically and categorically equivalent are considered by the language processor

It is conceivable that the language processor can only choose between orders that are syntactically and categorically equivalent. In light of this, the language processor would never have to choose between the 1-2-3 and the 3-1-2 orders, as these orders are syntactically very different from each other. As was discussed in chapter 3, the 1-2-3 order is a three-verb cluster, while the 3-1-2 order is a two-verb cluster with a non-verbal 3. This entails that for the *moet hebben gemaakt* 'must.MOD have.AUX made.PTCP' cluster, only the choice between the 3-1-2 and 1-3-2 orders can be affected by language processing. The 2-1-3 and the 2-3-1 orders cannot be derived, the 3-2-1 order requires a different linearization, and the 1-2-3 order is syntactically different. For the *moet kunnen zwemmen* 'must.MOD can.MOD swim.INF' cluster, the 1-3-2 is not considered either, as this order is not part of the language variety.

This massively limits the part of the variation that can be explained by language processing. Only 3.PTCP-1-2 and the 1-3.PTCP-2 orders are true competitors. Indeed, the fact that the 3-1-2 order was found much more acceptable than the 1-3-2 order for this cluster can be attributed to properties of language processing. For instance, multiple interpretations of Hawkins' (2004; 2014) processing model indicate that the 3-1-2 order is easier to process, as is illustrated in table 4.6 above.

This combined approach can thus account for a large part of the comparative judgement task. This is illustrated in (96). Some rankings still remain unexplained, however. For instance, if processing can only guide a choice between orders that are syntactically and categorically equivalent, it remains unclear how the informants choose between the 3-1-2 and 1-2-3 orders. Potentially, the semantics of the participle has affected their choice here. The fact that the 1-2-3.PTCP order was ranked higher than 3.PTCP-1-2 suggests that the verbal interpretation of the participle was preferred over a stative interpretation.[67] Additionally, if processing does not assess orders that are not part of the language variety, it is unclear how informants chose between unrealized orders. These issues remain open for future research.

(96) a. MOD-AUX-PTCP: 3-1-2 > 1-2-3 > 1-3-2 > 3-2-1 > 2-1-3 > 2-3-1

? (not equivalent) | processing | +/−realized | +/−derivable | ?

b. MOD-MOD-INF: 1-2-3 > 3-1-2 > 1-3-2 > 3-2-1 > 2-3-1 > 2-1-3

? (not equivalent) | +/−realized | ? | +/−derivable | ?

4.5 Summary and discussion

The results of the comparative judgement task clearly showed that speakers have intuitions on orders that can and cannot occur in varieties of Dutch. Speakers distinguish between what does occur, what can occur, and what cannot occur. This chapter further investigated speakers' potential strategies in this task. It was hypothesized that language processing partly drives their intuitions.

The chapter started with an evaluation of a number of processing models proposed in the literature. A problem that immediately appeared is that there is no consensus on the type of preferences that affect language processing. In addition to this, it became apparent that the current models of language processing are inexplicit on a number of factors that are crucial for their theories, such as the definitions of 'related' and 'similar' items.

These issues are very problematic for any approach that aims to draw predictions from these models. A very illustrative example came from two contradicting interpretations of Hawkins' (2004; 2014) theory. For this reason, a variety of possible applications of the processing models were investigated.

[67] This may be due to properties of V_2. In the test sentence V_2 was *hebben* 'have'. It would be interesting to see if the results change with verb clusters in which *zijn* 'be' is the middle verb.

None of predictions drawn from those models could account for the results of the comparative judgement task. It can be concluded that it is not possible to attribute the informants' ranking to properties of language processing alone.

Subsequently, it was hypothesized that a combined model that integrates processing theories in the grammatical theory outlined in chapter 3 could successfully account for the observed data. The results indicated that this approach is successful, especially if one assumes that the language processor only considers orders that are (i) possible in the language variety and (ii) syntactically and categorically equivalent. One can conclude from this that preferences of language processing play a limited role. This result is especially striking considering that in many functionalist approaches the explanatory role of a formal theory is often downplayed in sake of a functional theory.

In light of this, an important additional observation that was not discussed so far, is that none of these processing models can account for the geographic co-occurrence patterns that were discussed in section 2.10 of chapter 2. Chapter 3 has demonstrated that the grammatical theory is able to account for these patterns. However, none of the processing models can explain these facts. For instance, processing models have nothing to say about why the 1-2-3.PTCP, 1-3.PTCP-2 and the 3.PTCP-1-2 orders co-occur, with the exclusion of the 3-2-1 order. Neither can the processing models explain why the 1-3.INF-2 order only occurs in transitional areas. These facts further support the grammatical theory outlined in chapter 3.

Chapter 3 argued that the grammatical model allows for a number of choices, for instance the choice between different orders of Merge. This chapter presented a possible solution for the question how the language user makes a choice between different options. The next chapter of this dissertation delves deeper into the assumption that the timing of Merge, to a certain extent, is free, providing a choice between the 3-1-2 and 1-3-2 orders. It focusses on the parallel between these constructions, particle-verb constructions and so-called *verb cluster interruption*. It is demonstrated that the extent to which verb clusters can be interrupted by non-verbal material can further enlighten us on the underlying structure of verb clusters.