

Judgements about double-embedded relative clauses differ between languages

Stefan L. Frank and Patty Ernst

Centre for Language Studies, Radboud University Nijmegen
P.O. Box 9103, 6500 HD Nijmegen, The Netherlands

ARTICLE HISTORY

Compiled April 21, 2017

ABSTRACT

When the middle verb phrase is removed from an English double-embedded sentence, the remainder of the sentence is read faster in spite of the ungrammaticality. It has been shown that this “missing-VP effect” is reversed in German and Dutch. The current study demonstrates that the same cross-linguistic difference holds for sentences judgments: Dutch double-embedded sentences are considered less comprehensible and acceptable when the middle verb phrase is removed whereas the same is not the case in English. The results, in combination with earlier findings, give rise to a hybrid account according to which the missing-VP effect is caused by properties of the language as well as properties of working memory.

KEYWORDS

Grammaticality illusion; missing-VP effect; cross-linguistic difference; acceptability; comprehensibility

1. Introduction

A *grammaticality illusion* occurs when an ungrammatical sentence is perceived as acceptable (Phillips, Wagers, & Lau, 2011). One well-known example of such an illusion is the so-called *missing-VP effect*: When a double center-embedded relative clause structure, as in (1a), is turned into an ungrammatical string by removing the middle verb phrase (VP), as in (1b), it does not become any harder to understand (Gibson & Thomas, 1999). Some studies have even found higher comprehensibility or acceptability when the second VP is missing (Christiansen & MacDonald, 2009; Gimenes, Rigalleau, & Gaonac’h, 2009). Gibson and Thomas (1999) ascribe this effect to working-memory limitations that cause the prediction of the second VP to be structurally forgotten when three consecutive noun phrases are processed.

- (1a) The ancient manuscript that the new paper shredder that the students in the library used effortlessly shredded was missing a page.
- (1b) The ancient manuscript that the new paper shredder that the students in the library used was missing a page.

The missing-VP effect has also been observed in word-reading times: Vasishth,

Suckow, Lewis, and Kern (2010) found that the final verb and post-verbal region are read faster in the ungrammatical condition than in correct double-embedded sentences, at least in English. Interestingly, this effect reversed when German native speakers read German double-embedded sentences. In this case, the grammar violation caused a *slowdown* in reading. Vasishth et al. (2010) argue that this is because German has verb-final relative clauses, which is to say that the verb is always located at the end of a relative clause. Consequently, speakers of German often encounter sentences where the verb appears late, which may increase their ability to keep verb predictions in working memory so that they are less prone to structural forgetting than English speakers.

In another reading-time experiment, Frank, Trompenaars, and Vasishth (2016) showed that the missing-VP effect does not occur in Dutch either; a language that, like German, has verb-final relative clauses. They also found the effect to reappear when Dutch or German native speakers read double-embedded structures in English, which shows that the cross-linguistic difference is not due to properties of the speakers (i.e., higher verbal working-memory capacity in Dutch and German speakers compared to English speakers) but is caused by properties of the languages. For example, Dutch and German word order makes consecutive VPs much more common in those languages than in English. Sensitivity to these statistics could speed up the processing of three consecutive VPs in Dutch/German compared to English.

These results do not imply that a missing-VP effect can never arise in verb-final languages. As a case in point, Häussler and Bader (2015) found longer reading times in grammatical double-embedded German sentences compared to versions without the second VP, if the entire structure was presented as a complement clause (as in: “I believe that the ancient manuscript that ...”). The present study investigates sentences without such a complementizer, that is, sentences like (1a) and (1b), for which the absence of the missing-VP effect in Dutch and German is well established, at least when reading time is the dependent variable.

The question remains whether the cross-linguistic difference also appears in a sentence rating task. Häussler and Bader (2015) claim that binary grammaticality judgments in German reveal a missing-VP effect, based on the fact that “in a substantial number of cases” (p. 10) the missing-VP sentences are judged as grammatical. However, the acceptance rate was even higher for grammatical sentences (81% versus 33%, for sentence without an initial complementizer). Previous studies on English and French have found comprehensibility or acceptability ratings to be equal or lower in the grammatical condition (Christiansen & MacDonald, 2009; Gibson & Thomas, 1999; Gimenes et al., 2009). Had these researchers measured *higher* scores for grammatical sentences instead (as Häussler and Bader do), they would have concluded that the missing-VP effect does not exist. Therefore, it seems premature to conclude there is a missing-VP effect in German grammaticality judgments.

The current study investigates whether the missing-VP effect occurs in comprehensibility/acceptability ratings of Dutch equivalents to sentences like (1a) and (1b). If it does, this would mean that subjective ratings and reading times diverge in their response to the grammaticality manipulation. Although this may seem unlikely a priori, it is certainly possible (e.g., see Gimenes et al., 2009), for example because reading times are sensitive to automatic cognitive processes whereas explicit judgments depend on conscious, deliberative processes. Furthermore, it is well known that word-reading times closely follow word probabilities (e.g., Smith & Levy, 2013), which depend on the statistics of a language’s word-order patterns. If the presence of the missing-VP effect indeed depends on the language’s statistical properties, as Frank et al. (2016) argue,

the cross-linguistic difference may be restricted to reading times and not appear in a sentence rating task. Although a recent study by Lau, Clark, and Lappin (in press) has shown that sentence acceptability ratings, too, correlate with probabilities, it stands to reason that ratings are less probability-sensitive than reading times because of the nature of the sentence judgment task.

In addition, we investigate if the difference between verb-initial and verb-final languages also appears when sentence comprehension is facilitated by semantic support. The sentence items of Vasishth et al. (2010) and Frank et al. (2016) contained nouns and verbs that allowed for several combinations of agent, action, and patient. For example, the sentence “The mother who the daughter who the sister found frightened greeted the grandmother” can only be understood through syntactic analysis because the meaning of the individual words do not provide any cue about who does what to whom. In sentence (1a), on the other hand, it stands to reason that the paper shredder did the shredding (and that it was the manuscript that was shredded) even if word order is ignored. Possibly, such semantic support could lead to more shallow parsing (Sanford & Sturt, 2002) or to prioritizing semantic over syntactic analysis (Townsend & Bever, 2001). This, in turn, could mean that the language’s word order (a purely syntactic parameter) is no longer relevant to the presence of the missing-VP effect.

In Experiment 1, native Dutch speaking participants rated Dutch double-embedded sentences with semantic support, similar to the English items from the Gibson and Thomas (1999) study. Results showed that there was no missing-VP effect: Grammatically correct sentences were rated as more comprehensible and more acceptable than the ungrammatical versions. Experiment 2 is identical to Experiment 1 except that it uses English stimuli. Consistent with Gibson and Thomas (1999), sentences with a missing second verb phrase were rated as more acceptable and comprehensible than grammatically correct sentences. Finally, Experiment 3 replicates Experiments 1 and 2 in a within-subjects design by presenting both Dutch and English items to native Dutch speakers of English as a second language. This experiment confirmed the findings from the first two experiments: the presence of the missing-VP effect is language dependent.

2. Experiment 1: Dutch

2.1. Method

2.1.1. Materials

Twelve target sentences were constructed, inspired by the English double center-embedded object relative (OR) structures from Gibson and Thomas (1999). The main difficulty with creating Dutch equivalents is that Dutch relative clauses are ambiguous between subject and object relative, with a strong preference for a subject relative (SR) reading in absence of other (e.g., semantic) cues (Mak, Vonk, & Schriefers, 2002, 2006). For this reason, all our sentences had an inanimate first noun phrase (NP), facilitating an OR reading. Further, only the third NP and its verb are plural, which syntactically disambiguates towards OR. In the General Discussion, we will return to the issue of Dutch SR/OR-ambiguity and discuss potential problems with our solutions.

The general structure of grammatical target sentences is depicted in (2) below. Note that the numbering of NPs and verb phrases (VPs) denotes their linear order rather than subject-verb dependencies. The form of the relative pronoun *die* or *dat* depends on the grammatical gender (common or neuter, respectively) of the modified noun.

Table 1. Example item from Experiment 1 in each condition, with English gloss for the grammatical (V0) condition.

Cond.	Sentence
V0	Het eeuwenoude manuscript dat de nieuwe papiervernietiger die de studenten in de bibliotheek gebruikten zonder enige moeite versnipperde miste een pagina. <i>The centuries-old manuscript that the new paper-shredder that the students in the library used without any difficulty shredded missed a page.</i> “The ancient manuscript that the new paper shredder that the students used in the library shredded without any difficulty was missing a page.”
V1	Het eeuwenoude manuscript dat de nieuwe papiervernietiger die de studenten zonder enige moeite versnipperde miste een pagina.
V2	Het eeuwenoude manuscript dat de nieuwe papiervernietiger die de studenten in de bibliotheek gebruikten miste een pagina.
V3	Het eeuwenoude manuscript dat de nieuwe papiervernietiger die de studenten in de bibliotheek gebruikten versnipperde zonder enige moeite.

(2) NP1_{sing,inanim} *die/dat* NP2_{sing} *die/dat* NP3_{plu} VP1_{plu} VP2_{sing} VP3_{sing}.

Each target sentence comes in four conditions: The complete and grammatically correct sentence (condition V0) and without the first, second, or third VP (conditions V1, V2, and V3, respectively). Table 1 shows one example in all four conditions. All twelve target sentences are listed in the Appendix.

In condition V0, NP1 is the unambiguous subject of VP3. The innermost relative clause “NP2_{sing} *die/dat* NP3_{plu} VP1_{plu}” is unambiguously OR because the plural verb must agree with the plural subject NP3. The outermost relative clause “NP1_{sing} *die/dat* NP2_{sing} VP2_{sing}” is syntactically ambiguous between SR and OR but the choice of nouns and verbs was such that only an OR reading made sense. Semantic constraints also prevent other combinations of subject nouns and verbs. Consequently, when one VP is removed, it is semantically evident which NP is left without a verb.

In condition V3, the final verb phrase (VP2) can be interpreted as an intransitive because its verb is always optionally transitive. As can be seen in Table 1, this condition involves a word-order change within VP2. This is necessary because word-order constraints in Dutch would make condition V3 highly unacceptable if the original order of VP2 were retained.

We constructed 48 filler sentences in addition to the targets. Twelve of these were intentionally ungrammatical so that the ungrammaticalities are not restricted to target sentences. Gibson and Thomas (1999) did not include ungrammatical fillers, but Christiansen and MacDonald (2009) did.

Four lists were constructed such that each target item occurs in all four conditions across the lists but only once in each list, and the conditions occur equally often in a list. No two targets directly followed one another in a list, and the first and last two sentences of each list were fillers. To counter potential order effects, four more lists were created by reversing the order of the original lists, making a total of eight lists.

Table 2. Mean and by-subject standard errors of ratings in Experiment 1, on a scale from 1 (very incomprehensible/unacceptable) to 7 (very comprehensible/acceptable).

Rated aspect	Condition			
	V0 <i>M (SE)</i>	V1 <i>M (SE)</i>	V2 <i>M (SE)</i>	V3 <i>M (SE)</i>
Comprehensibility	4.07 (0.17)	3.63 (0.19)	3.43 (0.18)	3.03 (0.17)
Acceptability	2.83 (0.18)	2.38 (0.18)	2.29 (0.15)	2.02 (0.14)

2.1.2. Procedure

The experiment was conducted online as a Qualtrics questionnaire. Participants first provided basic demographic information and were asked about any reading difficulties they may have. After a brief explanation of the task, each participant was randomly assigned to one of the eight lists. They were presented with the 60 target and filler sentences but only one sentence was visible at a time. Each sentence had to be rated on two 7-point scales. The first scale was labeled “Zeer onbegrijpelijk / Zeer begrijpelijk” (“Very incomprehensible / Very comprehensible”) and the second was labeled “Zeer onacceptabel / Zeer acceptabel” (“Very unacceptable / Very acceptable”). The participants could not return to a previous item or skip a sentence but it was possible to close the questionnaire before completion, in which case the incomplete data were recorded. Participants who rated all 60 sentences were then asked to give their impression the experiment’s goal. None of the participants gave an answer that indicated any awareness of the grammaticality manipulation in the target sentences.

2.1.3. Participants

Adult, native speakers of Dutch were recruited via social media. Fifty-four people initiated the experiment but we only analyzed data from participants who rated at least two items in each of the four conditions. Data was also discarded from one participant who reported being dyslexic. This left 46 participants (37 females, age range 19–62 years, mean age 32.9) This should be sufficient to detect an effect of the same size as in Gibson and Thomas’s (1999) study, who tested 40 participants in a comprehensibility rating task using materials very similar to ours.

2.2. Results

Table 2 shows the mean comprehensibility and acceptability ratings for each condition. At first glance, there does not appear to be any grammaticality illusion: The scores in grammatical condition V0 are higher than in the three ungrammatical conditions. This was confirmed by an ordinal mixed-effects regression analysis, using the R package `ordinal`, which included by-subject and by-item random intercepts and random slopes of Condition. As is clear from the fitted regression model in Table 3, all three ungrammatical conditions resulted in significantly lower comprehensibility and acceptability scores than condition V0.

Table 3. Fixed-effect coefficients b (with condition V0 as reference level) and their z -scores and p -values from regression analysis of ratings in Experiment 1.

Condition	Rated aspect					
	Comprehensibility			Acceptability		
	b	z	p	b	z	p
V1	-0.79	-2.99	.003	-0.95	-2.53	.011
V2	-1.03	-3.12	.002	-0.92	-2.54	.011
V3	-1.53	-5.25	< .0001	-1.56	-4.78	< .0001

2.3. Discussion

Frank et al. (2016) found that the missing-VP effect does not appear in reading times on Dutch sentences. The current results are consistent with this finding, and extend it to ratings of sentence comprehensibility and acceptability. They contrast sharply with Gibson and Thomas’s (1999) results in an English sentence comprehensibility rating task, where scores in conditions V0 and V2 did not significantly differ but were higher than those in conditions V1 and V3. Likewise, Christiansen and MacDonald (2009) found that comprehensibility was higher in the ungrammatical condition V2 than in V0, and Gimenes et al. (2009) found the same in French.

It stands to reason that the difference between our results and those from earlier studies in English and French is caused by the difference in language of presentation. However, there are a number of confounds that prevent such a conclusion. In particular, our study was conducted online while the experiments by Gibson and Thomas (1999), Christiansen and MacDonald (2009), and Gimenes et al. (2009) took place under more controlled conditions, in the presence of the experimenter. To investigate if the English missing-VP effect also occurs under the precise conditions of Experiment 1 and with items that are as similar as possible, Experiment 2 repeats Experiment 1 but with English translations of the Dutch items.

3. Experiment 2: English

3.1. Method

3.1.1. Materials

The sentence items were (approximate) translations into English of the Dutch stimuli from Experiment 1 (see Appendix for the list of target sentences). The fillers were translated too, which required introducing different grammatical errors if the Dutch errors could not be translated. The item presentation order across eight lists was identical to that of Experiment 1.

3.1.2. Procedure

The experimental procedure was identical to that of Experiment 1 except that the entire questionnaire was presented in English.

Table 4. Mean and by-subject standard error of ratings in Experiment 2, on a scale from 1 (very incomprehensible/unacceptable) to 7 (very comprehensible/acceptable).

Rated aspect	Condition			
	V0 <i>M (SE)</i>	V1 <i>M (SE)</i>	V2 <i>M (SE)</i>	V3 <i>M (SE)</i>
Comprehensibility	3.44 (0.19)	3.35 (0.16)	3.98 (0.19)	3.18 (0.21)
Acceptability	2.50 (0.19)	2.65 (0.17)	2.86 (0.17)	2.34 (0.18)

Table 5. Fixed-effect coefficients b (with condition V0 as reference level) and their z -scores and p -values from regression analysis of ratings in Experiment 2.

Condition	Rated aspect					
	Comprehensibility			Acceptability		
	b	z	p	b	z	p
V1	-1.16	-0.62	> .5	0.42	1.25	> .2
V2	0.70	2.35	.019	0.69	2.02	.044
V3	-0.41	-1.52	> .1	-0.34	-1.00	> .3

3.1.3. Participants

Adult native English speakers were recruited via social media and other internet fora. After 100 people participated, we discovered that four of the eight lists were not presented correctly. The corresponding data were discarded and 41 additional participants were recruited, evening out the lists. From the two recruitment sessions combined, we only kept data from the 40 participants who completed at least two items in all four conditions. Data from two participants who reported being dyslectic were discarded, as was data from one participant who correctly guessed that the study was about how “sentence construction errors ... affect understanding”, leaving a total of 37 participants (22 females, age range 18–69 years, mean age 35.5).

3.2. Results

The mean comprehensibility and acceptability scores, presented in Table 4, suggest that participants experienced a grammaticality illusion when the second verb phrase was missing: The scores in condition V2 are slightly higher than in V0, whereas conditions V1 and V3 do not result in higher acceptability or comprehensibility ratings. This was indeed confirmed by an ordinal mixed-effects regression analysis (Table 5).

3.3. Discussion

Experiment 2 formed a conceptual replication of the paper-and-pencil study by Gibson and Thomas (1999). They found evidence for a missing-VP effect, in that removing the second VP from double-embedded sentences did not significantly affect comprehensibility, whereas removing either the first or third VP significantly resulted in reduced

ratings. Our results are somewhat different, but consistent, with this pattern. Instead of lower scores in the V1 and V3 conditions (compared to V0) we found higher comprehensibility and acceptability ratings in condition V2 compared to V0. This may be taken as evidence for a grammaticality illusion in *all three* ungrammatical conditions. However, the fact that most mean scores in the V1 and V3 conditions are numerically lower than for grammatical sentences is also consistent with a weak effect of the syntactic violation. Importantly, both ratings are significantly higher in condition V2 which confirms Gibson and Thomas’s (1999) claim that removing the second VP leads to an illusion of grammaticality.

Irrespective of the interpretation of these results (i.e., a grammaticality illusion in all ungrammatical conditions or only in V2), they are markedly different from those in Experiment 1. Consistent with the reading-time results from Frank et al. (2016), there appears to be a missing-VP effect for sentence judgments in English but not in Dutch. However, the difference between the outcomes of the two experiment could conceivably be due to differences between the two participant groups instead of the two languages. To directly test for an interaction between language and grammaticality in a within-subjects design, Experiment 3 presents the Dutch and English items to native Dutch speaking participants with high proficiency in English as a second language.

4. Experiment 3: Dutch and English

4.1. Method

4.1.1. Materials

Materials were the same as those of Experiments 1 and 2, except that the V1 and V3 conditions were not included. Four lists were created such that each item occurred in both languages (Dutch and English) and both forms (Grammatical: V0, and Ungrammatical: V2) across the lists but only once in each list. Languages were blocked such that each of these four lists first presented all Dutch items and then all English items. The first and last two sentences of a block were fillers, and two target items were always separated by at least one filler. A short text preceding each block stated the language of the upcoming sentences. This text was written in the language of that upcoming block. Four additional lists were constructed by reversing the presentation order of the four original lists, resulting in a total of eight lists (four starting with the Dutch block and four starting with the English block).

4.1.2. Procedure

Unlike Experiments 1 and 2, Experiment 3 was conducted in a controlled lab environment. Nevertheless, the rating task was administered using Qualtrics, as in the previous two experiments. All instructions were given in Dutch.

Following the rating task, participants filled out a short language background questionnaire, including self-ratings of English proficiency and amount of use on 7-point scales. As there was very little variability among participants, these ratings were not used for selecting participants or analyzing individual differences.

Next, participants performed two tests to ascertain their level of English proficiency. The first was the Vernon-Warden reading test (Hedderly, 1996), which consists of 42 sentences items with a missing word for which five options are provided. The task is to choose the correct word for as many items as possible within a 10-minute time

Table 6. Mean and by-subject standard error of ratings in Experiment 3, on a scale from 1 (very incomprehensible/unacceptable) to 7 (very comprehensible/acceptable).

Rated aspect	Language	Grammaticality condition	
		V0 <i>M (SE)</i>	V2 <i>M (SE)</i>
Comprehensibility	Dutch	4.18 (0.23)	3.33 (0.24)
	English	4.00 (0.25)	3.93 (0.21)
Acceptability	Dutch	3.22 (0.24)	2.32 (0.17)
	English	3.01 (0.22)	2.84 (0.17)

limit. The second was LexTALE (Lemhöfer & Broersma, 2012), a lexical decision task designed to measure English proficiency in non-natives.

A complete experiment session would take at least 30 and at most 60 minutes.

4.1.3. Participants

Thirty-one native Dutch speaking participants (23 females, age range 20–41 years, mean age 25.5) who identified as speakers of English as a second language were recruited via the Radboud University participant registration system. They received a 10 euro gift voucher for their participation. None of the participants indicated any awareness of the grammaticality manipulation in the target sentences.

One participant with very low Vernon-Warden score was excluded from further analysis. All remaining participants scored in the top 10% to 50% for adult natives on the Vernon-Warden test, indicating that they have near-native English reading skills. LexTALE scores ranged between 61.25% and 100% (mean: 83.5%), which classifies the participants as upper intermediate or advanced/proficient second language users (see Lemhöfer & Broersma, 2012, Table 9).

4.2. Results

As can be seen in Table 6, Dutch grammatical sentences are rated as more comprehensible and more acceptable than the sentence with a missing second verb phrase. The same does not seem to hold in English, where the difference in ratings between the two sentence types is much smaller. This was confirmed by an ordinal mixed-effects regression analysis (Table 7) which revealed a significant interaction between Grammaticality and Language, such that the Grammaticality effect is larger in Dutch than in English. In fact, there was no significant Grammaticality effect for the English items.¹

4.3. Discussion

Experiment 3 replicated Experiments 1 and 2 with a within-subjects design and in a controlled lab setting. Again, the results showed that grammatical Dutch double-embedded sentences are considered more comprehensible and acceptable than ungram-

¹An analysis including the data from the single excluded participant resulted in nearly identical regression results.

Table 7. Fixed-effect coefficients b (with Grammaticality V0 and Language English as reference levels) and their z -scores and p -values from regression analysis of ratings in Experiment 3.

Factor	Rated aspect					
	Comprehensibility			Acceptability		
	b	z	p	b	z	p
Language Dutch	0.28	0.78	> .4	0.20	0.51	> .6
Grammaticality V2	-0.06	-0.21	> .8	-0.18	-0.51	> .6
Dutch \times V2	-1.22	-3.11	.002	-1.09	-2.72	.007

matical versions in which the second verb phrase has been removed, but that the same is not the case in English. In short, the missing-VP effect appears in English but not in Dutch.

A possible objection to this conclusion is that Experiment 3 contains a confound between language and nativeness because all participants had Dutch as their first language (L1) and English as a second language (L2). If double-embedded structures are particularly hard to process in L2, this could lead to lower ratings in the English V0 condition compared to English V2. However, there are several reasons why non-nativeness is unlikely to have caused the missing-VP effect in Experiment 3. First, if the sentences were harder to comprehend in L2, we would expect lower ratings for grammatical English compared to grammatical Dutch sentences. However, the data in Table 6 tell a different story: The cross-linguistic difference in grammaticality effect is mostly due to relatively high ratings in the English V2 condition. That is, the participants do not find English grammatical sentences (much) less acceptable and comprehensible than the Dutch equivalents, but they rate ungrammatical sentences more positively in English than in Dutch.

Second, the grammaticality effect for English sentences was much larger for the native speakers in Experiment 2 than the non-natives in Experiment 3. Although we should refrain from directly comparing the results of these two experiments because of the difference in setting (i.e., online versus in the lab), a larger missing-VP effect for natives than for non-natives is inconsistent with non-nativeness being the cause of the effect.

Third, insofar as the missing-VP effect is caused by working memory limitations, it is unlikely that non-nativeness plays a role, considering the evidence that working memory capacity does not differ between a bilingual's two languages (Keijzer, 2013; Lanfranchi & Swanson, 2005; Osaka & Osaka, 1992; Osaka, Osaka, & Groner, 1993).

Finally, a more fundamental reason to disprefer the non-nativeness account of the missing-VP effect is that this account is not parsimonious as it requires an alternative explanation for the same effect in native English speakers.

Ratings on English sentences were non-significantly lower in condition V2 than in V0. In contrast, we found significantly *higher* ratings for condition V2 in Experiment 2. An interesting possibility is that this is caused by language transfer; the phenomenon that native-language properties affect processing in a second language. Specifically, the native Dutch participants would apply their successful strategies for processing Dutch double-embedded structures when reading in English. However, a much simpler explanation for the difference between Experiments 2 and 3 is that it is caused by the experimental setting. Unlike the previous experiment, Experiment 3 took place in a

lab environment, which may have caused participants to be more attentive to the task and try harder to parse the sentences. Consistent with this explanation, Gibson and Thomas’s (1999) lab study also found no significant difference between conditions V0 and V2, using very similar materials.

5. General Discussion

Vasishth et al. (2010) and Frank et al. (2016) demonstrated that the missing-VP effect in sentence processing is language dependent, in that it appears in English but not in German or Dutch. The current results expand on this in two respects. First, we showed that the interaction between grammaticality (second VP present or missing) and language (Dutch or English) is not limited to reading times but also appears as a subjective illusion in sentence ratings: Dutch sentences with a missing VP were judged to be significantly less comprehensible and acceptable than their grammatical counterparts, whereas the same was not the case (or even reversed) for English sentences. Second, we found that the cross-linguistic difference is also present for sentences whose propositional content (the “who-does-what-to-whom”) is apparent from the semantic relations between agents, patients, and actions. In contrast, understanding the materials of Vasishth et al. (2010) and Frank et al. (2016) required a full syntactic analysis because the nouns and verbs used in these sentences made any agent-action-patient triplet semantically possible.

5.1. *Ambiguity of Dutch relative clauses*

As mentioned in Section 2.1.1, Dutch relative clauses are ambiguous between subject-relative (SR) and object-relative (OR) readings. That is, the Dutch sentence fragment “Het manuscript dat de papiervernietiger versnipperde” (lit.: “The manuscript that the paper-destroyer shredded”) can in principle be interpreted to mean that the manuscript shredded the paper shredder. This because Dutch, unlike German, does not have case marking. We reduced the amount of ambiguity by having each sentence start with an inanimate noun, which leads to an initial preference for the (intended) OR reading (Mak et al., 2002). Furthermore, only the third noun was plural so that the innermost embedding is unambiguously OR as soon as the plural verb is encountered. Semantically, too, the two relative clauses could only be interpreted as OR. Nevertheless, the syntactic ambiguity of the first relative clause remains, which could lead to increased comprehension difficulty compared to the English sentences in which any ambiguity is already resolved at the first word following the relative pronoun. However, this is unlikely to have caused the cross-linguistic difference because, if anything, increased difficulty for Dutch should make Dutch double embeddings even harder to process, resulting in a *stronger* grammaticality illusion in Dutch than in English, as opposed to its absence.

In Frank et al.’s (2016) Dutch sentences, no useful semantic information was present and all nouns were animate. Consequently, these structures could not easily be interpreted as ORs. For this reason, an adjective was inserted after each relative pronoun, which syntactically disambiguates towards an SR reading. Frank et al. (2016) argue that this could not have caused the reversal of the missing-VP effect compared to English, and our current results indeed confirm that the illusion is also absent in Dutch double-embedded OR clauses.

5.2. *Explaining the missing-VP effect*

Three explanations of the missing-VP effect have been proposed in the literature. First, the structural forgetting account (Gibson & Thomas, 1999) claims that encountering a double-embedded structure results in working memory overload which leads to one of VP predictions to be forgotten. According to Gibson’s (1998) Syntactic Prediction Locality Theory, most memory is freed up if it is the prediction of the second VP that is dropped, which explains why only condition V2 of Experiment 2 results in relatively high comprehensibility and acceptability ratings. As explained in the Introduction, speakers of a verb-final language are more accustomed to keeping verb predictions in working memory and, according to Vasishth et al. (2010), this explain why the missing-VP effect does not occur in such languages.

Second, the interference account (Häussler & Bader, 2015) claims that the missing-VP effect is not caused by forgetting but by memory interference. More specifically, upon encountering the second verb phrase, there are two possible attachment sites (the first and second noun phrase) and a grammaticality illusion can occur when the first noun phrase instead of the second is incorrectly retrieved from memory. The first VP does not lead to such confusion because, at that point, the final noun phrase, to which it is to be attached, is still active in working memory. According to Häussler and Bader (2015), the reason why no grammaticality illusion arises when the third VP is deleted is that the first items of a list are more strongly represented in working memory than later items (except for the most recent ones) — this is the primacy effect that is well known from the working-memory literature (e.g., Page & Norris, 1998). If the second VP is correctly attached to the second noun phrase, the absence of a third noun phrase is detected relatively easily because of the first noun phrase’s primacy advantage. The advantage of the interference account over Gibson and Thomas’s (1999) structural forgetting account is that the former follows naturally from general properties of working memory rather than from a particular idea about syntactic complexity, such as the Syntactic Prediction Locality Theory.

Third, according to the language statistics account (Frank et al., 2016), the language’s statistical word-order patterns are central to the occurrence of the missing-VP effect. This can explain why English behaves differently from Dutch and German, irrespective of the participant’s native language. As explained in the Introduction, the occurrence of consecutive VPs are more probable in verb-final than verb-initial languages. Sensitivity to such statistics can cause the third VP to be highly unexpected in English while it is much less surprising in Dutch/German. Computational simulations using neural networks and other statistical models have shown that this is indeed a viable explanation of the cross-linguistic difference (Engelmann & Vasishth, 2009; Frank et al., 2016; Futrell & Levy, 2017).

None of the three accounts on its own can straightforwardly explain all the available empirical data. For structural forgetting to explain why native Dutch and German speakers display the illusion in English, it needs to assume that working memory capacity is language specific, something that is highly unlikely (see Section 4.3 and references therein). The interference account is unable to explain why the same sentence structures yield opposite effects in verb-initial (English and French) versus verb-final (German and Dutch) languages. The language statistics account is hard pressed to explain why dropping the first or third VP does not result in a grammaticality illusion in English.

The simplest model that can explain all results may be a hybrid statistical and working-memory account: Primacy and recency effects lead to stronger working mem-

ory activation of the first and last noun phrase, so the absence of their verb phrases is easily detected. For grammatical English double-embedded sentences, the low occurrence probability of three consecutive VPs leads to a sense of unacceptability (and slowdown in reading) on the third VP, even if it is syntactically required. In verb-final languages, in contrast, the third VP is less unexpected and therefore more acceptable.

It may appear unrealistic to claim that syntactically correct (three-VP) English sentences receive a lower probability than ungrammatical two-VP structures, but the subjective (and implicit) probabilities assigned by the language comprehension system need not be based on complete and correct syntactic parses. As a case in point, Lau et al. (in press) found that the probabilities assigned by a context-free grammar display weaker correlation with acceptability ratings than probabilities from a recurrent neural network. The same has been reported for eye-tracking and EEG data (Frank & Bod, 2011; Frank, Otten, Galli, & Vigliocco, 2015).

6. Conclusion

In a series of three sentence-rating experiments, we have shown that the classical missing-VP effect reported by Gibson and Thomas (1999) holds up in English but not in Dutch. English double-embedded object-relative clauses from which the second VP is removed are considered to be more (or at least no less) acceptable and comprehensible than their grammatical counterparts, but this grammaticality illusion does not occur in similarly structured Dutch sentences. This cross-linguistic difference complements the earlier effects on reading times found by Vasishth et al. (2010) and Frank et al. (2016). Taken together, the reading-time and rating results form converging evidence for how word-order differences between languages can have subtle and perhaps unexpected consequences for sentence comprehension.

Acknowledgments

We are thankful to Joshua Koster for his help in constructing the English stimuli. The research presented here was funded by the Netherlands Organisation for Scientific Research (NWO) Gravitation Grant 024.001.006 to the Language in Interaction Consortium.

References

- Christiansen, M. H., & MacDonald, M. C. (2009). A usage-based approach to recursion in sentence processing. *Language Learning*, *59*, 126–161. doi: 10.1111/j.1467-9922.2009.00538.x
- Engelmann, F., & Vasishth, S. (2009). Processing grammatical and ungrammatical center embeddings in English and German: A computational model. In A. Howes, D. Peebles, & R. Cooper (Eds.), *Proceedings of 9th International Conference on Cognitive Modeling* (pp. 240–245).
- Frank, S. L., & Bod, R. (2011). Insensitivity of the human sentence-processing system to hierarchical structure. *Psychological Science*, *22*, 829–834. doi: 10.1177/0956797611409589

- Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, *140*, 1–11. doi: 10.1016/j.bandl.2014.10.006
- Frank, S. L., Trompenaars, T., & Vasishth, S. (2016). Cross-linguistic differences in processing double-embedded relative clauses: Working-memory constraints or language statistics? *Cognitive Science*, *40*, 554–578. doi: 10.1111/cogs.12247
- Futrell, R., & Levy, R. (2017). Noisy-context surprisal as a human sentence processing cost model. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 688–698). Association for Computational Linguistics.
- Gibson, E. (1998). Syntactic complexity: Locality of syntactic dependencies. *Cognition*, *68*, 1–76. doi: 10.1016/S0010-0277(98)00034-1
- Gibson, E., & Thomas, J. (1999). Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes*, *14*, 225–248. doi: 10.1080/016909699386293
- Gimenes, M., Rigalleau, F., & Gaonac’h, D. (2009). When a missing verb makes a French sentence more acceptable. *Language and Cognitive Processes*, *24*, 440–449. doi: 10.1080/01690960802193670
- Häussler, J., & Bader, M. (2015). An interference account of the missing-VP effect. *Frontiers in Psychology*, *6*, 766. doi: 10.3389/fpsyg.2015.00766
- Hedderly, R. (1996). Vernon-warden reading test, restandardised 1993 and 1994. *Dyslexia Review*, *7*, 11–16.
- Keijzer, M. (2013). Working memory capacity, inhibitory control and the role of L2 proficiency in aging L1 Dutch Speakers of near-native L2 English. *Brain Sciences*, *3*, 1261–1281. doi: 10.3390/brainsci3031261
- Lanfranchi, S., & Swanson, H. L. (2005). Short-term memory and working memory in children as a function of language-specific knowledge in English and Spanish. *Learning and Individual Differences*, *15*, 299–319. doi: 10.1016/j.lindif.2005.05.003
- Lau, J., Clark, A., & Lappin, S. (in press). Grammaticality, acceptability, and probability: A probabilistic view of linguistic knowledge. *Cognitive Science*. doi: 10.1111/cogs.12414
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*, 325–343. doi: 10.3758/s13428-011-0146-0
- Mak, W. M., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language*, *47*, 50–68. doi: 10.1006/jmla.2001.2837
- Mak, W. M., Vonk, W., & Schriefers, H. (2006). Animacy in processing relative clauses: The hikers that rocks crush. *Journal of Memory and Language*, *54*, 466–490. doi: 10.1016/j.jml.2006.01.001
- Osaka, M., & Osaka, N. (1992). Language-independent working memory as measured by Japanese and English reading span tests. *Bulletin of the Psychonomic Society*, *30*, 287–289. doi: 10.3758/BF03330466
- Osaka, M., Osaka, N., & Groner, R. (1993). Language-independent working memory: Evidence from German and French reading span tests. *Bulletin of the Psychonomic Society*, *31*, 117–118. doi: 10.3758/BF03334156
- Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, *105*, 761–781. doi: 10.1037/0033-295X.105.4.761-781

- Phillips, C., Wagers, M., & Lau, E. F. (2011). Grammatical illusions and selective fallibility in real-time language comprehension. In J. Runner (Ed.), *Experiments at the interfaces* (pp. 147–180). Bingley, UK: Emerald Group Publishing.
- Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: not noticing the evidence. *Trends in Cognitive Sciences*, *6*, 382–386. doi: 10.1016/S1364-6613(02)01958-7
- Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*, 302–319. doi: 10.1016/j.cognition.2013.02.013
- Townsend, D. J., & Bever, T. G. (2001). *Sentence comprehension: The integration of habits and rules*. Cambridge, MA: MIT Press.
- Vasishth, S., Suckow, K., Lewis, R. L., & Kern, S. (2010). Short-term forgetting in sentence comprehension: Crosslinguistic evidence from verb-final structures. *Language and Cognitive Processes*, *25*, 533–567. doi: 10.1080/01690960903310587

Appendix A. Target sentences

The twelve Dutch and English target sentences are listed below. Brackets indicate the words that were removed in the V1, V2, and V3 conditions.

- 1 Het eeuwenoude manuscript dat de nieuwe papiervernietiger die de studenten [in de bibliotheek gebruikten]_{V1} [zonder enige moeite versnipperde]_{V2} [miste een pagina]_{V3}.
The ancient manuscript that the new paper shredder that the students [in the library used]_{V1} [effortlessly shredded]_{V2} [was missing a page]_{V3}.
- 2 Het liedje dat de moeder die de kinderen [lieten schrikken in de woonkamer]_{V1} [heel erg graag zong]_{V2} [ging over een verloren liefde]_{V3}.
The song that the mother who the children [scared in the living room]_{V1} [thoroughly enjoyed singing]_{V2} [was about a lost love]_{V3}.
- 3 Het spel dat het bedrijf dat de ouders [voor veel geld inhuurden]_{V1} [heel goed organiseerde]_{V2} [duurde heel de middag]_{V3}.
The game that the company that the parents [hired for a lot of money]_{V1} [organized seamlessly]_{V2} [lasted the whole afternoon]_{V3}.
- 4 De bal die de bekende voetballer die zijn medespelers [egostisch vonden]_{V1} [helemaal over het veld pingelde]_{V2} [was lek]_{V3}.
The ball that the famous soccer player who his team mates [thought selfish]_{V1} [dribbled across the entire field]_{V2} [had a puncture]_{V3}.
- 5 Het vrijstaande huis dat makelaar die de enthousiaste genteresseerden [heel vaak belden]_{V1} [het liefst zou verkopen]_{V2} [stond te koop]_{V3}.
The detached house that the real-estate agent who the enthusiastic house hunters [called a lot]_{V1} [would rather sell]_{V2} [was for sale]_{V3}.
- 6 Het artikel dat de redacteur die de journalisten [niet geschikt vonden voor zijn functie]_{V1} [enthousiast corrigeerde]_{V2} [bevatte een aantal spellingsfouten]_{V3}.
The article that the editor who the journalists [thought was unsuitable for the function]_{V1} [enthusiastically corrected]_{V2} [contained a number of spelling mistakes]_{V3}.
- 7 Het spannende boek dat de populaire schrijver die de recensenten [nauwlettend bekritiseerden]_{V1} [met veel vertrouwen publiceerde]_{V2} [miste een aantal pagina's]_{V3}.
The exciting book that the popular author who the reviewers [meticulously criticized]_{V1} [very confidently published]_{V2} [was missing a number of pages]_{V3}.
- 8 Het ijzeren standbeeld dat de kunstenaar die de leerlingen [hielpen]_{V1} [met veel liefde smeedde]_{V2} [was vrijdagochtend plots gestolen]_{V3}.
The iron sculpture that the artist who the pupils [helped]_{V1} [lovingly forged]_{V2} [was suddenly stolen on Friday night]_{V3}.
- 9 Het oude muziekstuk dat de dirigent die de muzikanten [al een tijd verafschuwden]_{V1} [vol overgave dirigeerde]_{V2} [bestond uit drie delen]_{V3}.
The old music piece that the composer who the musicians [have abhorred for some time]_{V1} [passionately conducted]_{V2} [consisted of three parts]_{V3}.
- 10 De film die de strenge regisseur die de acteurs [goed gehoorzaamden]_{V1} [levendig visualiseerde]_{V2} [werd illegaal gedownload]_{V3}.
The film that the strict director who the actors [obeyed well]_{V1} [vividly visualized]_{V2} [was downloaded illegally]_{V3}.
- 11 De luxe kamer die de schoonmaakster die de gasten [niet aardig vonden]_{V1} [twee maal per dag dweilde]_{V2} [werd niet zo vaak verhuurd]_{V3}.
The luxury room that the cleaner who the guests [didn't like]_{V1} [mopped twice a day]_{V2} [was not rented out often]_{V3}.
- 12 Het chique restaurant dat de klant die de obers [uitstekend bedienden]_{V1} [een goede beoordeling had gegeven]_{V2} [won een prijs voor klantvriendelijkheid]_{V3}.
The chic restaurant that the customer who the waiters [excellently served]_{V1} [had given a great review]_{V2} [won a prize for its hospitality]_{V3}.