

A matter of omission

A study on the comprehension and acceptability of doubly center-embedded sentences in English

Joshua Koster

S4060717

31-07-2016

Bachelor dissertation

S. Frank

1.0	Introduction	1
2.0	Theoretical Framework	2
	2.1 <i>Disappearing syntactic nodes hypothesis</i>	2
	2.2 <i>The least recent nodes hypothesis</i>	3
	2.3 <i>The syntactic prediction locality theory (SPLT)</i>	4
3.	Method	8
	3.1 <i>Materials</i>	8
	3.3 <i>Participants</i>	9
	3.4 <i>Procedure</i>	10
	3.5 <i>Data analysis</i>	10
4.	Results	11
	4.1 <i>Comprehensibility</i>	11
	4.2 <i>Acceptability</i>	11
5.	Discussion	12
6.	Conclusion	15
	References	16
	Appendix A: an example test-stimulus	17
	Appendix B: All test-sentences per Condition	17

Various studies have shown that double-embedded sentences with a missing VP can be found more comprehensible than their grammatical counterparts, in English. Ernst (2015) among others has found that this does not hold true for Dutch, but used a different methodology. The goal is to find out if changing the methodology to match that of Ernst will change the finding of a missing VP₂ effect, either to match the original research by Gibson and Thomas (1998), or to match the finding by Ernst. Gibson and Thomas give three main theories that can predict which verb is preferably omitted. Frazier's *disappearing syntactic nodes hypothesis* (1985), Gibson's *least recent nodes hypothesis* (1991) and Gibson's *syntactic prediction locality theory* (1998). The results confirm that using an online questionnaire does not seem to change earlier found results by Gibson and Thomas; sentences with a missing VP₂ are rated as more comprehensible than sentences with a missing VP₁ or a missing VP₃. This study further backs Gibson's syntactic prediction locality theory, and rejects earlier studies in finding a missing VP effect through the *disappearing syntactic nodes hypothesis* and the *least recent nodes hypothesis*.

1.0 Introduction

Texts are not always clear. This can be due to a multitude of reasons, relating to the reader, the writer or the text. The clarity of texts is such a curious phenomenon that sometimes ungrammatical sentences are rated as more comprehensible than grammatical sentences. For example, this phenomenon can occur with sentences that have two (or more) sentences *center-embedded* within themselves, as in sentence (1).

(1) [_s The ancient manuscript [that [_s the paper shredder [that [_s the students in the library used]] effortlessly shredded]] was missing a page].

This sentence has two relative clauses (RC) 'the paper shredder ... effortlessly shredded' and 'the students in the library used' in the main clause 'the ancient manuscript was missing a page'. According to research conducted by Chomsky and Miller (1963) sentences like these are complex sentences that are difficult to understand, while not entirely unacceptable. Although the use of sentences like these is discouraged, they are not grammatically incorrect. However, various research has shown that when sentences have doubly embedded relative clauses and one of the verb phrases (VP's) was omitted, the sentence seemed to become more comprehensible (Gibson & Thomas, 1999; Christiansen & MacDonald, 2009; Frazier, 1985). This phenomenon is called an *grammaticality illusion*.

The study by Gibson and Thomas (1999) is part of research dedicated to the *structural forgetting* theory, which tries to explain why verb phrases could be forgotten in sentences with embedding. They try to explain why sentences like (2) are perceived grammatical, despite the fact that its second VP is missing.

(2) * The ancient manuscript that the paper shredder that the students (...)VP₂ effortlessly shredded was missing a page.

According to English syntactic rules these sentences are ungrammatical, though it seems as if our brains want shorter sentences that require less processing power, as research predicts that omitting a VP, such as the VP₂, makes sentences seem better than their grammatical counterparts (Frazier, 1985; Gibson, 1991; Gibson, 1998). The effect of ungrammatical sentences seeming grammatical, as in sentence (2) is called *the missing VP effect* (Gibson & Thomas, 1999). There are three main theories that try to predict which VP should be forgotten for better comprehension: Frazier's (1985) *disappearing syntactic nodes hypothesis*, Gibson's (1991) *least recent nodes hypothesis* and Gibson's (1998) *Syntactic prediction locality theory*.

2.0 Theoretical Framework

2.1 Disappearing syntactic nodes hypothesis

Frazier's disappearing syntactic nodes hypothesis is the first proposal of a process of structural forgetting, as a part of the missing VP effect. This hypothesis predicts that the VP₂ will be forgotten such as in sentence (2). For this effect to take place, though, relative clause pronouns have to be removed.

“At points of high complexity, syntactic nodes will tend to be forgotten if they dominate no lexical material or only dominate material that has already been semantically combined” (Frazier, 1985, pp. 178). According to Frazier, a noun phrase (NP) is semantically combined when it receives a thematic role, or when it shares the same thematic role as a non-lexical RC pronoun through coindexation. This means that as in Figure 1 ‘the famous soccer player’ is semantically combined through the verb “thought (selfish)” as it assigns the thematic role. But also “his team mates” receive a thematic role from ‘thought selfish’, this time because it coindexes with a non-lexical RC pronoun. According to the hypothesis when semantic combination has completed, this part of the sentence structure can be pruned away, which would leave only “the ball” and “had a puncture”.

As we can see in figure 1, 'the ball' is left to dominate the RC pronoun, but as the RC pronoun is not lexical the hypothesis stipulates that the pronoun gets pruned as well, leaving just the latter part of the sentence intact.

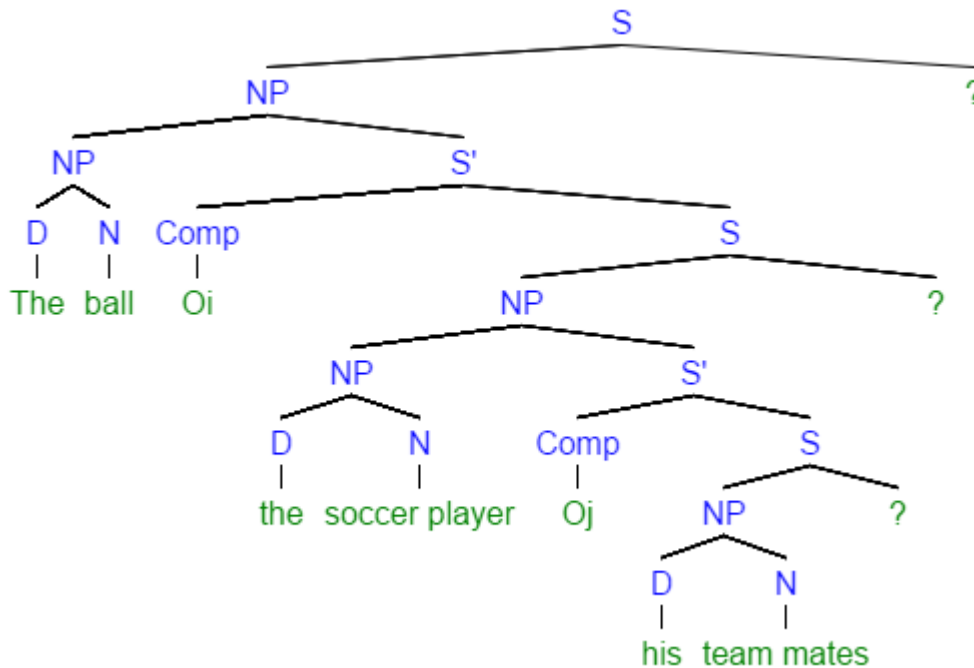


Figure 1: simplified version of syntactic tree sentence. The O's stand for the omitted RC pronouns and the small index for the referent above them in the tree.

Despite that it creates a situation in which there is an option for a missing verb, this hypothesis seems to be mainly hypothetical and specific. Not only does it require for a part of the sentence structure to disappear based on a theory of semantic combining, it also hypothesises that the sentence as used in figure 1 would be grammatical, despite the sentence's missing pronouns. Due to the prevalence of lexical RC pronouns in sentences like (1), this hypothesis predicts that there should be no missing VP effect in doubly embedded sentences.

2.2 The least recent nodes hypothesis

According to Gibson's (1991) least recent nodes hypothesis "if a structure requires more processing load than the available capacity, then selectively remove (forget) nodes directly dependent on the least recent words in the input string until the load associated with the structure is lowered below the desired threshold (p. 162)". This means that due to the sentence being doubly center-embedded, after having seen the second "that" the first NP and related RC

pronoun are pruned, as in sentence (3). The reader will proceed with ease until the word “was” is reached, at this point the reader will realise that something is wrong and rate the sentence as unacceptable.

- (3) [_s ~~The ancient manuscript~~ [~~that~~ [_s the paper shredder [that [_s the students in the library ~~used~~ _{VP1}]] effortlessly shredded _{VP2}]] was missing a page _{VP3}].

2.3 The syntactic prediction locality theory (SPLT)

This structural forgetting theory by Gibson (1998) predicts that when a point is reached where memory complexity is highest, the syntactic prediction with the highest memory load should be forgotten. The theory has two separate components that lead to complexity, both a memory cost and an integration cost. The complexity cost rises per discourse referent that is introduced after a sentence has started, though an exception is made for the _{VP3} as this is cost-free due to the importance of the predicate. Thus, according to the theory, the maximal complexity in double-embedded sentence processing occurs when reading the third NP, which in turn will result in the _{VP2}, that comes after, to be forgotten.

2.4 Further research

In a study by Gibson and Thomas (1999) double embedded-sentences with or without a missing VP were presented. The goal of the research was to find out how sentences with a missing VP were rated on comprehensibility, compared to sentences with all VP's. Every test sentence had four versions: one grammatical version (4), one with a missing _{VP1} (5), one with a missing _{VP2} (6) and one with a missing _{VP3} (7). The ‘*’ denotes that the sentence is deemed ungrammatical. (...) shows the omission of a VP.

- (4) [_s The ancient manuscript [that [_s the paper shredder [that [_s the students in the library used _{VP1}]] effortlessly shredded _{VP2}]] was missing a page _{VP3}].

- (5) * [_s The ancient manuscript [that [_s the paper shredder [that [_s the students in the library (...) _{VP1}]] effortlessly shredded _{VP2}]] was missing a page _{VP3}].

- (6) * [_s The ancient manuscript [that [_s the paper shredder [that [_s the students in the library used _{VP1}]] (...) _{VP2}]] was missing a page _{VP3}].

(7) * [_s The ancient manuscript [that [_s the paper shredder [that [_s the students in the library used _{VP1}]] effortlessly shredded _{VP2}]] (...)_{VP3}].

Native English speaking participants were asked to rate 60 sentences on a 5-point Likert-scale, out of which twelve sentences were actual test stimuli. Participants were only shown one version of each sentence. Sentences were also carefully constructed to ensure that each verb in the item had a strong semantic/pragmatic selectional restriction. This meant that only one of the preceding noun phrases could be a plausible subject.

The results of the research by Gibson and Thomas (1999) showed that sentences with a missing VP₂ could be rated as just as easy to understand as the grammatical version of the sentence. The results seemed to disprove the *disappearing syntactic node* hypothesis; as it was hypothesised that relative clause structures would only disappear when the RC pronouns were non-lexical. The results were also not in line with the *least recent nodes hypothesis*. This second hypothesis predicted that the first VP in the sentence would be forgotten, not the second. The SPLT was not disconfirmed; it predicted that the second VP would be forgotten and this seems to be the case.

In a follow-up experiment by Christiansen and MacDonald (2009) they changed up the methodology to see when a sentence would be rejected based on a *self-paced reading* task. This means the sentence was presented word-for-word, as opposed to a whole sentence at a time and at each sentence node participants could decide whether the sentence still seemed grammatical or not. They used half of the experimental items from the Gibson and Thomas (1999) study. The results seemed to be slightly different from what was found in the original study: sentences with a missing VP₂ were rated significantly better than their grammatical counterparts. The study by Gibson and Thomas did find that sentences with a missing VP₂ were acceptable, but not more acceptable than the grammatical ones. Despite using different stimuli, Christiansen and MacDonald ran another experiment to control for the length of sentences, where adverbs were used to replace the missing verbs, and found the exact same results.

Vasishth, Suckow, Lewis and Kern (2010) also conducted a self-paced reading experiment, but they not only researched the missing VP₂ effect in English, but also in German. The difference between these two languages is interesting due to the fact they have the same word order for object RC's; both SOV. However, these languages have a different amount of occurrences of sentences in this order. In English SOV sentences are exceedingly rare, though in German these sentences are very common. Also, in German, RC's are marked at the beginning and end with

a comma, which potentially makes them easier to read or understand. In total they did six experiments to research the structural forgetting hypothesis in a language other than English.

The first experiment was a self-paced reading experiment and the second an eye-tracking experiment, these were both experiments in English. After that a German self-based reading experiment and a German eye-tracking experiment follow. In experiment one, they confirmed the VP₂ effect in English, as the ungrammatical Condition was significantly easier to read than the grammatical sentences. To make sure that this effect wasn't being duplicated by the mere fact that three VP's were presented in a row there was a replication of the experiment. In this experiment adverbs and PP's got added in between sentences to try and control for comprehensibility and reading times, but the results were the same as from experiment 1. An eye-tracking experiment was also conducted which also yielded the same results: a confirmation of the missing VP₂ effect. In German the opposite was found, in this case the grammatical sentences were read quicker at V1 and post V-1 than the ungrammatical sentences. The reading times were longer for ungrammatical sentences in German than in English, thus also disconfirming the missing VP effect in German. However, Varsishth et al. theorised, because German has obligatory commas at the start and end of their RC's this effect might be found in English too when commas are added. When adding commas to the English sentences, the same results were still found as in experiment 1 and 2. This meant that the missing VP effect still held true for English. A possible explanation for these results, as per the authors, is that the habit of getting RC's in between commas helps German speakers process the sentences better. German natives could, unconsciously, train themselves while reading sentences with RC's which would in turn make them easier to read after having practiced reading them a lot. An alternative explanation is that German word-order is the reason that this difference occurs. German is a so called *head-final* language, which means that the head always comes last. What this means is that verbs will also end up at the end of sentences, so the discourse referent must be remembered for longer than in English, as the verb will occur earlier in English. The missing VP effect is mostly based on memory overload, but maybe German speakers' overload comes later due to the nature of the word-order of the language.

To research the hypothesis further that a head-final language could be of influence Frank, Trompenaars and Vashith (2015) tested the hypothesis whether the absence of a grammaticality illusion is due to the language being SOV or if it is due to a difference in the robustness of memory load in speakers. Each sentence had two Conditions, one grammatical version and one ungrammatical version with a missing VP₂. This study comprised three

experiments; in the first they mirror the German study by Vasishth et al. (2010) for Dutch, to test if the results from the earlier study are caused by word-order. The second experiment was to test if Dutch or German speakers, when compared to English speakers, have a more robust working memory due to their language having a SOV structure. To do this both these groups of speakers were tested for their reading times of English sentences. The results from experiment one show that the Dutch speakers indeed read ungrammatical sentences slower than grammatical ones. This mirrors the finding with German participants in the previous study. Thus there seems to be no grammaticality illusion in Dutch. German participants that read English sentences were found to read grammatical sentences slower than ungrammatical ones, thus mirroring the native English speakers. The same result was found for Dutch participants, they also read ungrammatical sentences faster than grammatical sentences. These results seem to point towards linking the grammaticality illusion to individual languages, as opposed to individual speakers, as there didn't seem to be a memory constraint involved in their L2. A native speaker should be aware of statistics of certain constructions appearing and will base their expectation on previous experiences (Vasishth et al. (2010)). It is then a question of seeing which language has a higher prevalence of three verbs appearing next to each other, and that is more likely in German or Dutch than in English

Ernst (2015) also conducted research using Dutch participants. In a replication of Gibson and Thomas (1999), where there were four different Conditions, but now with an added acceptability rating; the goal was to find out if it is possible to find a grammaticality illusion in Dutch. Ernst's methodology was also different as compared to Gibson and Thomas' study due to it being a questionnaire as opposed to a self-paced reading study. The reason acceptability was asked as well as comprehensibility is because Philips, Wagers and Lau (2011) define a grammaticality illusion as "when a sentence [...] is judged acceptable, we refer to it as a grammaticality illusion (p. 157)". This leaves a wider range for a grammaticality illusion to be found. In English both comprehensibility and acceptability have been used to find a grammaticality illusion (Frazier, 1985; Gibson & Thomas, 1999), but in Dutch acceptability had not yet been used to try and find a grammaticality illusion. The results show that sentences from Condition 0, the grammatical sentence with all three VP's, were rated significantly more comprehensible than sentences from Condition 2, with an omitted VP₂, for comprehensibility by subjects ($p = <.02$) and by items ($p = <.03$). For acceptability a significant difference was found between Condition 1 and Condition 3, but no others Conditions were significantly

different. This means that there seems to be no missing VP effect for Dutch with this methodology.

The aim of the current study is to replicate the results of the experiment by Gibson and Thomas using the methodology of Ernst (2015). The current study will be the first to contain questions for participants about both comprehensibility and acceptability. Also, possibly an online questionnaire will yield different results than a self-paced reading task or an eye-tracking experiment, as both different outcome variables are used and due to the fundamental difference of the task at hand. This experiment will test if omitting the second VP in a doubly center-embedded sentence in English will lead to a higher average comprehensibility and acceptability relative to Condition 1 and 3. Also, it is expected that Condition 2 is rated as not significantly harder to understand as compared to the grammatical Condition 0.

3. Method

3.1 Materials

The online questionnaire was made in Qualtrics. The questionnaire consisted of sixty total items, twelve target sentences and forty-eight fillers. All sentences were constructed based on the English translations of Ernst's (2015) Dutch items. For every sentence four different versions were made, a grammatical one with all verbs, and ungrammatical ones with either a missing VP₁, VP₂, or VP₃ (all test sentences can be found in Appendix B; an example sentence can be seen in Appendix A). The sentences were all constructed with the sentences of Gibson and Thomas (1999) in mind; doubly center-embedded sentences with a subject first and verb phrase last. An example sentence of all four Conditions will follow below. Ernst also accounted for making the sentences more acceptable when the VP₂ was omitted by making the V₂ one that could be used transitively and intransitively. This means that this verb could be used with another noun (transitive) "the man conducted a music piece" or without a noun (intransitive) "the man conducted".

Condition 0: "The old music piece that the composer who the musicians have abhorred for some time passionately conducted consisted of three parts."

Condition 1: "The old music piece that the composer who the musicians passionately conducted consisted of three parts."

Condition 2: "The old music piece that the composer who the musicians have abhorred for some time consisted of three parts."

Condition 3: “The old music piece that the composer who the musicians have abhorred for some time passionately conducted.”

The fillers were long and complex sentences, that contain a lot of relative clauses or subordination: “*Jean saw that Peter's brother bought a book in Sophie's neighbor's uncle's store, but that he simply walked out of the store without paying.* “. There were also 11 ungrammatical sentences, with incorrect congruence “*The use of euro coins, that are still being called francs by a lot of people, seem to still be difficult for a lot of people.*“, a wrongly inserted relative clause “*The opening of the company, that after a restructure has taken on an entire new team of employees, was visited well and why the board hopes the profit forecast will come true.*” The original study by Gibson and Thomas did not use ungrammatical fillers, but the study by Christianson and MacDonald (2009) did. As both acceptability and comprehensibility are being tested, the ungrammatical fillers serve as a distractor for the aim of the study as there are extra unacceptable sentences. It must be noted that a duplicate filler can be found in Ernst's questionnaire , “*The speaker, who tried to divide his attention equally over all speakers and had taken his time speaking about the jubilee, now turned his attention to a few recent achievements.*”. This was a small mistake, but for the sake of reproduction this was added to the current study's questionnaire too.

3.2 Design

There was a within subjects design, with two independent variables, acceptability and comprehensibility. The independent variable was the Condition of the sentence. Either a sentence had no missing VP's (Condition 0), a missing VP₁ (Condition 1), a missing VP₂ (Condition 2) or a missing VP₃ (Condition 3). When taking the online survey participants were randomly assigned one of eight surveys. There were four different versions, but to control for order in which the sentences were shown the remaining four surveys had a reversed item order. The order of the questions themselves was not random, to make sure a participant didn't get to see multiple test sentences in a row. Every questionnaire had three sentences from each Condition, and through counterbalancing it was determined that every participant only saw one version of each sentence.

3.3. Participants

A total of 123 participants started the online survey, but only 31 participants managed to finish the study. Participants were recruited by links spread out over social media and internet forums such as Reddit, asking if they wanted to participate in a study about sentence comprehension.

As all other participants did not give a response to all the test sentences it was decided to leave these out of the analysis. The participants received no form of credit or payment, merely thanks at the end of the questionnaire. From the 31 participants, 10 were male and 21 were female. The mean age was 32 years (SD 11.8), with the youngest being 18 and the eldest 56. When asked about their highest level of current or past education 1 answered elementary school, 5 answered college, 6 people answered postgraduate school and 19 answered graduate or undergraduate school. All participants answered that they had no problems reading.

3.4 Procedure

Native English participants were shown an on-screen instruction, after which a set of meta-data questions were asked (age, sex, mother tongue, a.o.). If participants were under 18, or not native speakers of English, they were denied the opportunity to participate. Participants were then presented with 60 questions, that all looked like the example sentence as found in Appendix A. Participants were asked to rate sentences on a scale of 1 (very incomprehensible) to 7 (very comprehensible) for comprehensibility and 1 (very unacceptable) to 7 (very acceptable) for acceptability. The questions were shown one at a time, so participants did not have the opportunity to change their answers. Only when both answers were filled in a participant could click an arrow button to proceed to the next question. This was to ensure people couldn't re-read sentences after having seen them and to ensure answers that came from their intuition weren't edited.

3.5. Data analysis

The results were analysed in the statistics program SPSS version 22. As was replicated from Ernst's (2015) study, the results were analysed using a one-way ANOVA by subjects and items. According to the Shapiro-Wilk test of normality all Conditions in both categories were non-significant, which means an ANOVA analysis can be run. As this is a replication study to confirm or disconfirm the results found by Thomas and Gibson (1999) and Ernst, a paired-samples t-test was conducted to compare comprehensibility in Condition 1 to Condition 2 and Condition 2 to Condition 3. As the result of the differences between Condition 0 and Condition 2 are unimportant – to confirm the results of Thomas and Gibson Condition 2 is only necessary to be higher than Condition 0 - no additional t-tests were run. Due to the assumption of homogeneity being broken by the ANOVA by subjects analysis for acceptability the non-parametric Kruskal-Willis was run.

4. Results

Table 1 shows the average score that participants rated sentences per Condition for both comprehension and acceptability. The average comprehension is highest in Condition 2, with a 3.8 score out of 7. The average acceptability score was also highest in this Condition with a score of 2.9 out of 7. This means the ungrammatical Condition 2's scores are slightly higher than the grammatical Condition 0's scores.

Table 1: the mean of all participants' rating of comprehensibility and acceptability of the test sentences. The scale ran from 1, very incomprehensible/unacceptable, to 7, very acceptable/very comprehensible. The standard-deviation is displayed in brackets.

	Condition 0	Condition 1	Condition 2	Condition 3
Comprehensibility	3.67 (1.12)	3.45 (.69)	3.8 (.94)	3.41 (1.00)
Acceptability	2.76 (.83)	2.70 (.74)	2.90 (.86)	2.55 (.85)

4.1 Comprehensibility

The ANOVA by subjects and items revealed that there was no significant effect for the omission of VP's on the overall comprehensibility of sentences ($F_1(3,44) = .847, p = .476, F_2(3,120) = 1.481, p = .223$). A paired-samples t-test was run to compare the comprehension score in sentences with a missing VP₁ (Condition 1) and missing VP₃ (Condition 3) to a missing VP₂ (Condition 2). Sentences with a missing VP₁ ($M = 3.45, SD = .69$) as compared to a missing VP₂ scored significantly lower ($M = 3.8, SD = .94$); $t(30) = 1.83, p = .039$ (one-tailed). Sentences with a missing VP₃ when compared to a sentence with a missing VP₂ also scored significantly lower ($M = 3.8, SD = .94$); $t(30) = 1.82, p = .040$ (one-tailed). This means that the missing VP₂ sentences were rated significantly better than both sentences with a missing VP₁ or VP₃.

4.2 Acceptability

The ANOVA by items revealed that there was no significant effect found for the omission of VP's on the overall acceptability of sentences ($F_1(3,44) = 1.03, p = .387$). An ANOVA by subjects is not possible, as the assumption of homogeneity was violated ($F = 3.02, p = .033$). As an alternative the non-parametric Kruskal-Willis test, which is robust for the assumption of homogeneity, was used to test significance of acceptability by subject. No significant effect was

found for acceptability of sentences based on Condition ($H_1(3) = 2.90, p = .407$). A paired-samples t-test was run to compare the amount of acceptability of sentences with a missing VP₁ (Condition 1) and missing VP₃ (Condition 3) to a missing VP₂ (Condition 2). There was no significant difference in the scores for sentences with a missing VP₁ ($M = 2.7, SD = .74$) compared to a missing VP₂ ($M = 2.90, SD = .86$); $t(30) = -1.16, p = .127$ (one-tailed). There was also no significant difference between the scores for sentences with a missing VP₃ when compared to a sentence with a missing VP₂ ($M = 2.55, SD = .85$); $t = -1.68(30), p = .52$ (one-tailed).

5. Discussion

The goal of this paper was to replicate the study by Ernst (2015) in English, to test if there was a significant effect between Conditions with omitted VP's. If looking at the results from Gibson and Thomas (1999) we might expect that the grammatical sentence, as in Condition 0 might be rated as just as grammatical as sentences with a missing VP₂. Also Conditions 1 and 3 might both be rated as less comprehensible and acceptable than Condition 2 with a missing VP₂. Though if the effect would be a result of the methodology of Gibson and Thomas, or if the results found by Ernst are not only confined to Dutch we might find a significantly higher score for the comprehensibility rating of the grammatical Condition 0 sentences when compared to the ungrammatical Condition 2 sentences.

The results of this study show the mean comprehensibility score is slightly higher in Condition 2 than in Condition 0, meaning that the ungrammatical sentences were rated slightly more comprehensible than the grammatical sentences. The difference in mean score between Conditions 1 and 2, and Conditions 2 and 3 show that a missing VP₂ is significantly more comprehensible than a missing VP₁ or missing VP₃. The current study therefore replicates the results found in Gibson and Thomas's study. Interestingly there was no significant effect for acceptability between Conditions 2 and 1, or between Conditions 2 and 3. This seems to mean that despite the fact that participants don't find the sentences any less comprehensible, they rate all ungrammatical sentences as unacceptable. This is made even more interesting due to the fact that the average rating of the grammatical sentences was lower (3.57) than the average of the sentences from Condition 2 (3.66). A correlation was also run to test if both acceptability and comprehensibility had any kind of linear relationship, the results show a weak to moderate correlation. This means that the more a sentence a sentence is found comprehensible the more it is found acceptable and vice versa.

The hypothesis of this study is confirmed; there was indeed no significantly higher rating of comprehensibility of a grammatical double-embedded sentence than a sentence with a missing VP₂. The second hypothesis, that sentences with a missing VP₁ and VP₃ would be rated significantly less comprehensible than a sentence with a missing VP₂ was also confirmed. Despite using a different methodology than the original experiment by Gibson and Thomas, this shows that the missing VP effect appears even when asked via an online survey.

In line with Gibson and Thomas (1999), this study finds more evidence for a missing VP effect while rejecting the disappearing syntactic nodes theory (Frazier, 1985) and the least recent nodes hypothesis (Gibson, 1991). The former expects there to be a missing VP effect if relative pronouns are excluded in sentences. However, if double center-embedded sentences with their relative clauses are in the sentence, it expects that all ungrammatical sentences are rated worse on comprehensibility than the grammatical sentences. The latter expects that the first NP is omitted, which also leads to omission of VP₁. One theory that does fit the results is the Syntactic prediction locality theory (SPLT) which theorises that the VP is omitted that leads to the highest memory cost, this being the VP₂.

The follow-up study by Christiansen and MacDonald (2009) also found a missing VP₂ effect and even found that a sentence with a missing VP₂ was rated significantly better than its grammatical counterpart. Despite the fact that the mean comprehensibility was rated higher in Condition 2 than in Condition 0 there was no significant effect found in this study. However the effect in Christiansen and MacDonald's study could be due to using only half of the experimental items from Gibson and Thomas's study, as well as being a self-paced reading task. This means that in the current study participants were still able to go back and read the entire sentence, instead of clicking past every word and then rate the comprehensibility and acceptability afterwards.

A self-paced reading task experiment was also conducted by Vasishth et al. (2010) hypothesising that there was no missing VP effect in German as opposed to English. The study did not only confirm that there was no grammaticality illusion in German, it also further confirmed a missing VP₂ effect for English. Moreover, there were also separate experiments to control for eye-movement and length of the sentence, but the same VP₂ effect was found. As there was no missing VP effect for German, a possible effect for not finding these same results in English as compared to German could be due to the sentences being head-final in the last two languages, meaning that due to the frequency of having to wait for a verb for a long time there is more robustness for long sentences.

In the follow-up study by Frank, Trompenaars and Vasishth (2016) this effect was replicated for Dutch; the participants in the study also read the ungrammatical sentences slower. Interestingly, when both Dutch and German speakers were tested in English, i.e. were given ungrammatical and grammatical sentences in English as a second language, both Dutch and German participants read the ungrammatical sentences faster. This seems to point towards a missing VP effect for a specific language, as opposed to all languages, as the SPLT theory by Gibson (1998) suggests.

Finally, Ernst (2015) tried to find a grammaticality illusion in Dutch through a new methodology; an online questionnaire. The results by Frank et al. (2016) were replicated; comprehension was significantly higher in the sentence with all three verbs, both by subjects as by items, as opposed to the sentence with a missing VP₂. Moreover, Condition 2 was not rated significantly better than Condition 1 or Condition 3. The acceptability by subjects also was highest for the grammatical sentence, though only the difference with Condition 3, with the VP₃ omitted was rated significant. The analysis by items showed that there was a significant difference between acceptability ratings per Condition 0 and Condition 2, though Condition 0 was rated higher on average acceptability. It is perhaps interesting to note that post-hoc the acceptability of both the grammatical and the ungrammatical sentences were calculated, and the mean score for the grammatical fillers was 4.64, whereas the mean score for ungrammatical fillers was 3.77. This means that all the fillers were more acceptable than the double embedded sentences, and that the grammatical fillers were rated higher on acceptability than the ungrammatical sentences. The reason this was tested was to try and find a way to measure to ensure participants were properly answered the questionnaire. It is important to note that finding a significant difference in Conditions in the current study shows that the missing VP effect is relatively robust for differences in methodology. When comparing this study to that of Gibson and Thomas some important differences should be noted. Gibson and Thomas's study only asked one question on a scale from one to five, with the scale going from 1, 'easy to understand' to 5, 'hard to understand'. This study used a scale from 1, 'very incomprehensible/acceptable', to 7, 'very comprehensible/acceptable', for two questions. As there is a larger scale it seems likely to be less difficult to find a difference between grammatical and ungrammatical sentences, as there is a larger scale to have a difference on. Additionally, the wording used to describe the scales was made more difficult; maybe the question asking is something is difficult to understand is different to the question asking about comprehensibility. Finally, Gibson and Thomas only used one question, whereas this study had two. Gibson and Thomas' study shows

that participants usually finished the task within 10-15 minutes. The average time a participant took to finish the questionnaire was 58 minutes, however, this time contained three times of more than 2 hours, one time of more than 3 hours and one time of more than 9 hours. If the extreme times are taken out of the analysis the average time is almost 23 minutes per participant. A common complaint at the end of the questionnaire was that it was too long and wondering if the goal of the study was testing participants' patience. Perhaps if going for such a long experiment a small reward would be suitable for participants that finish within a certain, acceptable, time – or an idea might be to cut some of the questions out. Gibson and Thomas chose to give participants \$4.00 for completion of the experiment. Though the fact that the only participants of the study was perhaps easier as they were recruited on campus and were in a controlled environment, where they were stimulated to finish the experiment. It is probable that this combination of reason explains so many drop-outs, as only 31 out of 123 participants decided to finish the study. Also, as opposed to Gibson and Thomas, this study had ungrammatical fillers, and not only grammatical fillers. In Ernst (2015) this was done to draw the attention away from the double-embedded sentences with mistakes, as they might be more easily recognised in Dutch. As this study aimed to fully replicate the study of Ernst, it was chosen to readopt this.

6. Conclusion

The current study confirms findings by Thomas and Gibson (1999) by finding that sentences with a missing VP₂ are rated as more comprehensible than sentences with a missing VP₁ and sentences with a missing VP₃. This study also strengthens Ernst's (2015) study by proving the missing VP₂ effect for English is not constrained to a certain methodology, but perhaps due to Dutch being a head-final language. The fact that the second verb phrase was preferentially omitted also strengthens the SPLT theory as described by Gibson (1998).

References

- Chomsky, N., & Miller, G.A. (1963). "Introduction to the Formal Analysis of Natural Languages." In: R.D. Luce, R.R. Bush, and E. Galanter (Eds.). *Handbook of Mathematical Psychology* (pp. 269-321), Volume 2, New York: Wiley.
- Christiansen, M. H., & MacDonald, M. C. (2009). A usage-based approach to recursion in sentence processing. *Language processing*, 59, 126-161.
- Ernst, P. (2015) *De grammaticaleitsillusie. Een studie naar het weglaten van werkwoordelijke gezegdes in complexe, ingebedde zinnen in het Nederlands*. B. thesis, Radboud University Nijmegen, Nijmegen.
- 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.
- Gibson, E. (1991). *A computational theory of human linguistic processing: Memory limitations and processing breakdown*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh. PA.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1-76.
- Gibson, E. (2000). Dependency locality theory: A distance-based theory of linguistic complexity. In: A. Marantz, Y Miyashita, & W. O'Neil (Eds.), *Image, language, brain: Papers from the first mind articulation project symposium* (pp. 95-126). Cambridge, MA: MIT Press.
- Gibson, E., & Thomas, J. (1999). Memory limitations and structural forgetting: the perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes*, 14(3), 225-248. DOI: 10.1080/016909699386293.
- Phillips, C., Wagers, M., & Lau, E. F. (2011). Grammatical illusions and selective fallibility in real-time language comprehension. *Experiments at the Interfaces*, 37, 147-180.
- 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.

Appendix A: an example test-stimulus

After the math professor had lost all her chalks, she had absolutely no idea how to explain the Pythagorean theorem to a ninth grade class.

Please indicate the comprehensibility and acceptability of the above sentence.

	1	2	3	4	5	6	7	
Very Incomprehensible	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very Comprehensible
Very Unacceptable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very Acceptable

Appendix B: All test-sentences per Condition.

Sentence 1:

Condition 0: The ancient manuscript that the new paper shredder that the students in the library used effortlessly shredded was missing a page.

Condition 1: The ancient manuscript that the new paper shredder that the students in the library effortlessly shredded was missing a page.

Condition 2: The ancient manuscript that the new paper shredder that the students in the library used was missing a page.

Condition 3: The ancient manuscript that the new paper shredder that the students in the library used effortlessly shredded.

Sentence 2:

Condition 0: The song that the mother who the children scared in the living room thoroughly enjoyed singing was about a lost love.

Condition 1: The song that the mother who the children in the living room thoroughly enjoyed singing was about a lost love.

Condition 2: The song that the mother who the children scared in the living room was about a lost love.

Condition 3: The song that the mother who the children scared in the living room thoroughly enjoyed singing.

Sentence 3:

Condition 0: The game that the company that the parents hired for a lot of money organized seamlessly lasted the whole afternoon.

Condition 1: The game that the company that the parents organized seamlessly lasted the whole afternoon.

Condition 2: The game that the company that the parents hired for a lot of money lasted the whole afternoon.

Condition 3: The game that the company that the parents hired for a lot of money organized seamlessly.

Sentence 4:

Condition 0: The ball that the famous soccer player who his team mates thought selfish dribbled across the entire field had a puncture.

Condition 1: The ball that the famous soccer player who his team mates dribbled across the entire field had a puncture

Condition 2: The ball that the famous soccer player who his team mates thought selfish had a puncture

Condition 3: The ball that the famous soccer player who his team mates thought selfish dribbled.

Sentence 5:

Condition 0: The detached house that the real-estate agent who the enthusiastic house hunters called a lot would rather sell was for sale.

Condition 1: The detached house that the real-estate agent who the enthusiastic house hunters would rather sell was for sale

Condition 2: The detached house that the real-estate agent who the enthusiastic house hunters called a lot was for sale.

Condition 3: The detached house that the real-estate agent who the enthusiastic house hunters called a lot would rather sell.

Sentence 6:

Condition 0: The article that the editor who the journalists thought was unsuitable for the function enthusiastically corrected contained a number of spelling mistakes.

Condition 1: The article that the editor who the journalists enthusiastically corrected contained a number of spelling mistakes.

Condition 2: The article that the editor who the journalists thought was unsuitable for the function contained a number of spelling mistakes.

Condition 3: The article that the editor who the journalists thought was unsuitable for the function enthusiastically corrected.

Sentence 7:

Condition 0: The exciting book that the popular author who the reviewers meticulously criticized very confidently published was missing a number of pages.

Condition 1: The exciting book that the popular author who the reviewers very confidently published was missing a number of pages.

Condition 2: The exciting book that the popular author who the reviewers meticulously criticized was missing a number of pages.

Condition 3: The exciting book that the popular author who the reviewers meticulously criticized very confidently published.

Sentence 8:

Condition 0: The iron sculpture that the artist who the pupils helped lovingly forged was suddenly stolen on Friday night.

Condition 1: The iron sculpture that the artist who the pupils lovingly forged was suddenly stolen on Friday night.

Condition 2: The iron sculpture that the artist who the pupils helped was suddenly stolen on Friday night.

Condition 3: The iron sculpture that the artist who the pupils helped lovingly forged.

Sentence 9:

Condition 0: The old music piece that the composer who the musicians have abhorred for some time passionately conducted consisted of three parts.

Condition 1: The old music piece that the composer who the musicians passionately conducted consisted of three parts.

Condition 2: The old music piece that the composer who the musicians have abhorred for some time consisted of three parts.

Condition : The old music piece that the composer who the musicians have abhorred for some time passionately conducted.

Sentence 10:

Condition 0: The film that the strict director who the actors obeyed well vividly visualized was downloaded illegally.

Condition 1: The film that the strict director who the actors vividly visualized was downloaded illegally.

Condition 2: The film that the strict director who the actors obeyed well was downloaded illegally.

Condition 3: The film that the strict director who the actors obeyed well vividly visualized.

Sentence 11:

Condition 0: The luxury room that the cleaner who the guests didn't like mopped twice a day was not rented out often.

Condition 1: The luxury room that the cleaner who the guests mopped twice a day was not rented out often.

Condition 2: The luxury room that the cleaner who the guests didn't like was not rented out often.

Condition 3: The luxury room that the cleaner who the guests didn't like mopped twice a day.

Sentence 12:

Condition 0: The chic restaurant that the customer who the waiters excellently served had given a great review won a prize for its hospitality.

Condition 1: The chic restaurant that the customer who the waiters had given a great review won a prize for its hospitality.

Condition 2: The chic restaurant that the customer who the waiters excellently served won a prize for its hospitality.

Condition 3: The chic restaurant that the customer who the waiters excellently served had given a great review.