How to understand silence: Voice mismatches in ellipsis in English.

Benjamin Bruening, Anna Koppy, Bilge Palaz, Satoshi Tomioka, and Rebecca Tollan (University of Delaware)

draft, December 31, 2025

Abstract

Structure generation accounts of ellipsis propose that silent syntactic structure is generated but unpronounced in ellipsis. Reactivation accounts propose instead that the antecedent is reactivated at the ellipsis site, with no silent structure. Existing psycholinguistic findings have not distinguished these hypotheses (Phillips & Parker, 2014). We address this issue by investigating voice mismatches in VP ellipsis in English (e.g., *A new language was supposed to be* <u>*taught*</u> *in the fall semester, but no professors could _ because the schedule was too tight*). In Experiment 1, participants' performance at recognizing that a word did not appear in the sentence is poorer when a related word did (e.g., *teach* when *taught* appeared), versus a non-related word (e.g., *worst*). This effect is greater when the elided clause is in the active voice, as in the example, versus the passive (…*but no new language was*…). The contrast between active and passive conditions cannot be accounted for solely in terms of antecedent reactivation, since the antecedents are identical. Experiment 2, which compares sentences with VP ellipsis with sentences containing an overt pro-form (e.g., *do that*…), provides further evidence in support of this conclusion. The processing of VP ellipsis must include more than just antecedent reactivation. The findings of these two experiments are thus most consistent with the structure generation account of ellipsis.

1. Introduction.

An important task for arriving at the correct interpretation of a sentence is building an accurate mental representation of its grammatical structure. Research on language comprehension has focused on how we go about this task, with one domain of particular interest being how we arrive at an understanding of sentences containing VP ellipsis, like those in (1).

(1) **a.** Sometimes projects included affordable housing… but far too often, they **did not.** (*New York Times*, 2015).

b. My only concern was that he would pop up and panic, but he **did not**. (*USA Today*, 1996).

(accessed from Corpus of Contemporary American English; Davies, 2008*)*

Although both sentences in (1) have the same two-word sequence "did not," they are interpreted very differently: as something like "did not [include affordable housing]" in (1a) and “did not [pop up and panic]” in (1b). The important question is how the same two-word sequence can have very different interpretations in context. Specifically, what cognitive mechanism(s) do speakers and comprehenders deploy to achieve these different interpretations? To date, existing empirical observations do not distinguish between two broad theoretical approaches (Phillips & Parker, 2014).

In the first approach, the *structure generation* approach, comprehenders (and speakers) construct the full structural representation for an elided VP, it is just not pronounced. In (1b), for instance, there would be a silent VP “[pop up and panic]” in the structural representation of the sentence. Proponents of this approach include Ross (1969), Sag (1976), Fiengo & May (1994), Chung, Ladusaw, & McCloskey (1995), Lasnik (2001), and Merchant (2001, 2013). In the second approach, the *reactivation* approach[1], there is no syntactic structure at the ellipsis site, except perhaps for a silent pro-form. This pro-form, or the ellipsis itself, points back to its antecedent and reactivates it. In (1b), the antecedent VP “pop up and panic” would be reactivated and used in calculating the meaning of the sentence, but the structure corresponding to it would not be built at the ellipsis site. This is the stance taken by, among others, Hardt (1999), Ginzburg & Sag (2000), Culicover & Jackendoff (2005, 2019), Martin & McElree (2008), Kim & Runner (2022), Nykiel & Kim (2022).

Many psycholinguistic studies of the processing of VP ellipsis have shown that certain kinds of information are accessed at the ellipsis site. For example, Shapiro, Hestvik, Lesan, & Garcia (2003), using a cross-modal lexical priming paradigm, found priming for *tie* at the ellipsis site in an example like (2).

(2) The mailman bought a tie for Easter, and his brother, who * was playing volleyball, did * too, according to the salesclerk.

Priming is found at the ellipsis site (the second star), but not earlier (the first star), indicating that *tie* is being reactivated rather than its activation persisting from the antecedent VP (additionally, the subject of the antecedent, *mailman*, does not show any priming at the ellipsis site). Findings like this indicate that the antecedent VP is being reactivated, but they do not show whether any syntactic structure is being built or not. Similarly, Xiang, Grove, & Merchant (2019) showed evidence for a structural priming effect, where an antecedent VP with the argument structure NP PP followed by VP ellipsis increased production of NP PP frames over a condition where the same antecedent is followed by an intransitive clause. Once again, however, this only shows that the antecedent is being reactivated at the ellipsis site, it does not show that a full syntactic structure is built there.

---

[1] This approach is also referred to as the *non-structural approach* (e.g., Merchant, 2013) or the *direct interpretation* approach (e.g., Culicover & Jackendoff, 2019). We will refer to it as the *reactivation approach* to highlight how it works in processing.

The current study attempts to distinguish structure generation from reactivation by employing both a novel stimulus paradigm and a novel behavioral measure. We report on two experiments. In Experiment 1, we examine pairs of sentences like in (3), where the first sentence is in the passive voice (i.e., "be taught") and the second sentence contains elided VP material, shown with an asterisk in (3). In (3a), this second sentence is also in the passive voice with the verb "taught" elided ("but no new language was [taught]."). By contrast, the second sentence in (3b) is in the active voice, with "teach" elided instead ("but no professors could [teach]…"). (3b) therefore contains a **voice mismatch**: the voice of the elided clause differs from the voice of its antecedent.

(3) Passive-Passive versus Passive-Active.

a. A new language was supposed to be taught in the fall semester, but no new language was * because the schedule was too tight.

b. A new language was supposed to be taught in the fall semester, but no professors could * because the schedule was too tight.

Importantly, the verb "teach" is an irregular verb, meaning that the active stem "teach" has a different form from the passive stem "taught." In a novel word recall task, we ask speakers of English to read sentences like those in (3) and identify whether the active verb form, "teach," had been present in them. The correct response is NO: "taught" appears in both versions (in the first conjunct), but "teach" appears in neither. Our hypothesis is that, if processing ellipsis requires mentally constructing new material (as per the STRUCTURE GENERATION APPROACH), then English speakers should be more prone to misidentifying "teach" as having appeared in (3b), compared with in (3a). If, on the other hand, the earlier VP material (i.e., "taught") is only being reactivated at the ellipsis site (as per the REACTIVATION APPROACH), then a speaker's capacity to correctly reject "teach" should be no more burdensome in (3b) than in (3a). Importantly, the antecedent clauses in (3a) and (3b) are identical, so any difference in findings could not be attributed solely to reactivation of the antecedent.

Experiment 2 follows up by comparing VP ellipsis to a VP pro-form, *do that*. As mentioned earlier, some proponents of the reactivation approach treat VP ellipsis as containing a silent pro-form (e.g., Culicover & Jackendoff 2005). Those that do not treat them as similar in simply reactivating their antecedents. In most analyses, VP pro-forms do not contain any silent structure. The structure generation approach therefore predicts that VP ellipsis and VP pro-forms will be processed differently, while the reactivation approach predicts that they should be processed similarly.

Our findings, from both Experiment 1 and Experiment 2, favor the structure generation approach. The pattern of interference in word recall that we find is most consistent with silent structure being built at the ellipsis site.

Sections 2 and 3 describe Experiments 1 and 2, respectively. Section 4 is a general discussion and conclusion, where we discuss the implications of our findings and what proponents of the reactivation approach would have to do to account for them.

2. Experiment 1.

2.1. Method.

For Experiment 1, we created sentence pairs of the form in (3), using verbs that have different stem forms for active and passive (*teach-taught*, *break-broken, bring-brought, buy-bought, catch-caught, deal-dealt, do-done, drive-driven, freeze-frozen, sweep-swept, hide-hidden, keep-kept, ride-ridden, steal-stolen, take-taken, wear-worn)*. We selected stem pairs that are as different from each other as possible, in order to maximize the distinctiveness of the two mismatching forms. Our 2x2 factorial design (within subjects, within items) crossed VOICE (passive, active; as in 3) with RELATEDNESS of the target word: The target was either related to the verb form in the first of the two sentences (as "teach" is related to "taught") or an unrelated word (e.g., "worst"). The unrelated conditions were included to control for independent differences in word recall accuracy that might result from reading sentences with matched (passive/passive) or mismatched (passive/active) voice. See Table 1.

	Stimulus	
Voice = active	*A new language was supposed to be taught in the fall semester, but no professors could because the schedule was too tight.* ***Related target: TEACH***	*A new language was supposed to be taught in the fall semester, but no professors could because the schedule was too tight.* ***Unrelated target: WORST***
Voice = passive	*A new language was supposed to be taught in the fall semester, but no new language was, because the schedule was too tight.* ***Related target: TEACH***	*A new language was supposed to be taught in the fall semester, but no new language was, because the schedule was too tight.* ***Unrelated target: WORST***

Table 1. *An example item for Experiment 1 (4 conditions: active/related, active/unrelated, passive/related, passive/unrelated).*

The status of voice mismatches in VPE has been debated. One view argues that voice mismatch is grammatical in VP ellipsis (Hardt, 1993; Kim, Kobele, Runner, & Hale, 2011; Merchant, 2013), as in (4a) and (4b) (attested examples from Merchant, 2013 and Hardt, 1993). However, not every instance of voice-mismatched VPE is acceptable; see (5a) (from Arregui, Clifton, Frazier, & Moulton, 2006). Consequently, others argue that voice mismatch in VP ellipsis is ungrammatical (e.g., Arregui et al., 2006; Grant, Clifton, & Frazier, 2012; Clifton, Xiang, & Frazier, 2019).

(4) a. It engaged them in a way that I did not think they could be that early in the morning.
b. This information could have been released by Gorbachev, but he chose not to.

(5) a. *This information was released, but Gorbachev didn't.
b. ?This information should have been released, but Gorbachev didn't.

Stockwell (2024) shows that voice mismatches are grammatical as long as a standard, focus-based condition is met, along with a condition stating that accommodated material in the antecedent (like the missing agent in the antecedent clause of 5a) cannot contradict the elided clause. Example (5a) violates the second condition. If the missing agent in the first conjunct is identified as *Gorbachev* via accommodation, it leads to the proposition that Gorbachev released this information, which contradicts the meaning of the second conjunct with the elided VP. On the other hand, modifying the sentence slightly as in (5b) does not violate the condition: There is no contradiction between "Gorbachev didn't release this information" and "Gorbachev should have released this information" (5b from Grant et al., 2012). To eliminate concerns about acceptability, we ensure that our stimuli abide by Stockwell's (2024) conditions (and all are judged acceptable by the native English-speaking authors).

2.2. Participants.

129 participants were recruited through Amazon's Mechanical Turk. All had an IP address from within the USA, and received payment of 4 USD, based on an estimated completion time of 20-25 minutes. Participants were asked about their English proficiency at the end of the experiment, but payment was not contingent upon how participants responded. Sample size was based on a similar recent experiment (REDACTED, to appear) where participants had to identify which of two words came last in a sentence.

2.3. Materials.

16 itemsets like in Table 1 were constructed and distributed across four presentation lists in Latin Square. We also included 28 filler items without any form of ellipsis, which served both

as distractor trials and as a measure of attention. An example appears in (6). In half of the fillers, the target word appeared in the sentence, while in the other half, it did not.

(6) The host asked if the tennis player Roger Federer won the Wimbledon final, hoping that someone would remember. (Target word: "tennis")

To encourage participants to read the sentences carefully, we also included a YES-NO comprehension question on four of the filler trials, which targeted the content of the prior sentence (e.g., "Did the host ask about the Wimbledon final?"). The correct answer was YES for two, and NO for the other two. The resulting 44 trials were presented in a pseudo-randomized order (no more than two critical trials were adjacent, and no two adjacent critical trials were of the same condition). The experiment was coded on the online platform *PC Ibex* (Zehr & Schwarz, 2018).

2.4. Procedure.

Participants gave informed consent by reading a disclaimer presented at the beginning of the experiment. They received the instructions shown in (7).

(7) Instructions

In this experiment, you will read some sentences followed by a single word [...]. Once you finish reading each sentence, a word will be displayed on the screen. You will be asked to identify whether the word appeared in the sentence or not. For example, in the sentence THE PROFESSOR TAUGHT THE STUDENT, the words STUDENT and TAUGHT both appeared, but the words GARDENER and TEACH did not appear.

Sentences were displayed in full on the center of the screen before disappearing at a key press (there was no time limit). After the sentence disappeared, a single word appeared on the next screen (e.g., "teach"). The participants' task was to identify whether this word had been present in the previous sentence. They could not return to any preceding screen. The experiment lasted 16 minutes on average.

2.5. Results.

3 participants were excluded from the analysis, because no data was recorded in PCIbex (n = 1), or they did not identify as L1 English speakers (n = 2). A further 8 participants were excluded because their accuracy rate on filler trials was less than 75% (i.e., they incorrectly identified whether the target word had or had not previously appeared), and/or they failed to answer at least 3 out of 4 comprehension questions correctly (i.e., above chance level). In total, data from 118 participants was included in the analysis. Their mean accuracy rate on

filler trials was 91.8%, indicating that they could generally identify the presence or absence of the target word accurately.

Our dependent measure for analysis of critical trials was whether participants correctly rejected the target word as having appeared (coded as 1) or failed to reject it (coded as 0). Participants answered accurately on 1585 out of 1888 total trials (84%). The results by condition are shown in Figure 1.

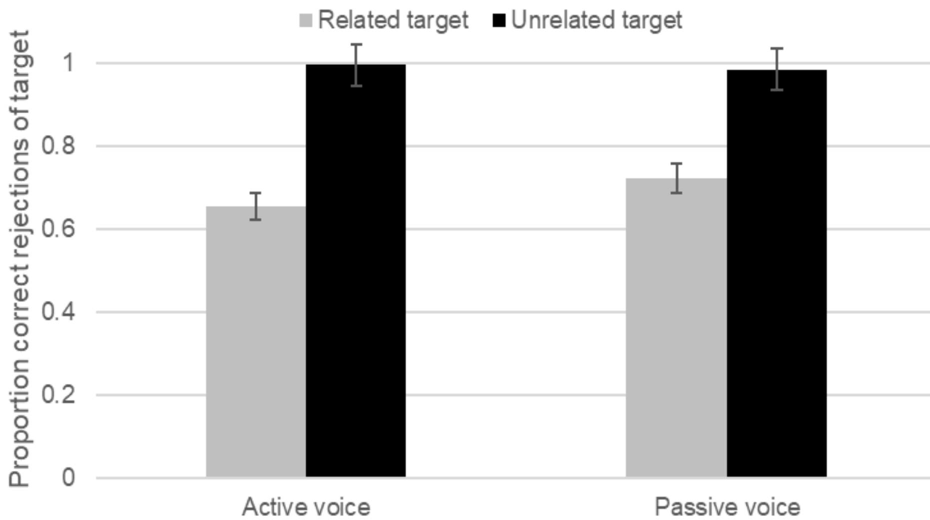

Figure 1. *Proportions of correct rejections of the target in active voice and passive voice conditions (related target in grey bars, unrelated target in black bars). Error bars indicate 95% C.I.s.*

We analyzed the data by fitting a 2x2 mixed-effects logistic regression with crossed random effects for participants and items (Baayen, Davidson & Bates, 2008), implemented in lme4 (R 4.3.2; Bates, Maechler, Bolker & Walker, 2015). The predictors VOICE and RELATEDNESS were each sum-coded. We report the output of the model with the maximal random effects structure that would allow for convergence without singular fit (Matuschek, Kliegl, Vasishth, Baayen & Bates, 2017). Descriptively, there was a higher accuracy rate in the passive voice conditions than the active voice conditions, but this effect was marginal (p = .093). However, we find a main effect of relatedness, with a higher proportion of correct rejections for unrelated targets than related ones (e.g., “worst” vs. “teach”; .99 vs. .69; β = -10.8, SE = 2.07, z = -5.22, p <.0001). We take this to indicate that a word like “teach” which has a morphologically related counterpart in the sentence (e.g., “taught”) is more likely to be misconstrued as having appeared in that sentence compared with a word that does not (e.g., “worst”), due to the forms having a shared verb lemma.

Importantly, there was a significant interaction of Relatedness*Voice (β = -2.86, SE = 1.08, z = -2.66, p = .0079): Participants were more accurate in the Related/passive condition than in the Related/active condition (.72 vs. .65; planned comparison β = -0.53, *SE* = 0.18, z = -2.95, p = .0032). We interpret this as participants having posited the verb "teach" at the ellipsis site in the active condition, leading them to be more likely to misconstrue the word as having appeared in the sentence proper. In contrast, they would have posited "taught" in the passive conditions, leading to no increased likelihood to misconstrue "teach" as having appeared, beyond the main effect of verb lemma-based interference.

This pattern is therefore predicted by the STRUCTURE GENERATION APPROACH to ellipsis, and is not expected by the REACTIVATION APPROACH. However, it is possible that English speakers *do* reactivate the antecedent verb, but then alter its morphological form so as to match the ellipsis environment, but without building new structure. This would presumably involve an additional step of processing. We examined target response-time data as well as sentence-reading time for the 1585 critical trials in which the target was correctly rejected, fitting a 2x2 mixed-effects linear regression to log-transformed RTs (following the same statistical protocol as described above). For target response, there was a main effect of Relatedness only, where participants responded slower to related targets than to unrelated targets (Related 3366ms vs. Unrelated 1890ms; log-transformed β = .28, *SE* = .026, t = 8.89, p <.0001). There was no effect of Voice (p = .33) and no interaction (p = .56). For sentence reading times[2], sentences in the Active conditions were read descriptively *faster* than those in the Passive conditions, but this difference was not significant (8099ms vs. 8806ms; p = .16). There was no significant effect of Relatedness or interaction (*p*s > .26), as expected (since the target word had not been presented by this point).[3] If participants *were* reactivating "taught" and converting it to "teach" in the active voice condition only, we might expect this to be reflected in longer sentence-reading and/or response times for the Active conditions versus the Passive conditions. We do not find the slightest hint of any such effect.

2.6. Discussion.

Experiment 1 found an effect of the voice of the elided clause on identifying whether a word appeared in the sentence or not. We found a main effect of relatedness, such that "teach" was incorrectly identified as appearing in the sentence more than an unrelated word like "worst." However, there was also an interaction, such that participants were significantly more likely to incorrectly identify "teach" as having appeared in the Active condition than in the Passive.

---

[2] We excluded a further nine trials from this analysis for which no sentence-reading time data was recorded in *PC Ibex*.

[3] For completeness, we also analyzed the data with the inclusion of incorrect response trials. Including these trials resulted in no meaningful differences (either qualitatively or quantitatively).

Importantly, unlike in all previous studies, the observed effect could not be due to reactivation of the antecedent of the elided clause, because the antecedents are identical in the two conditions. We reason that the parser must be doing *something* in addition to reactivating the antecedent VP. What could this additional something be? In the structure generation approach, it would be construction of the elided VP. In the Passive condition, a VP [taught] would be constructed, but in the Active condition, a VP [teach a new language] would be constructed instead. This representation would include the active form "teach," accounting for why participants incorrectly identified that word as having appeared in the sentence more than they did in the Passive condition, where only "taught" appeared. Thus, in terms of the representations of the sentences that are built, "teach" does indeed feature in the active variant (leading participants to identify it as having appeared in the sentence they had read), but not in the passive variant (where the representation includes the lemma-related variant "taught" instead).

In the reactivation approach, it is not clear what the *something* would be in addition to reactivation. We are aware of only one publication using this approach that addresses voice mismatches, Kim & Runner (2022). In Kim & Runner (2022: 490), both active *take to* and passive *be taken to* activate an abstract argument structure representation "[take.to(x,y,z)]" (see their example 108). Since this representation is abstract and shared by active and passive, the passive form can be the antecedent for the active and vice versa. However, we see no way for this approach to account for our findings. Both active and passive activate this abstract argument structure equally, so there is no reason to expect a difference when the elided clause is active versus passive. This is not to say that it is impossible to account for our findings without positing structure building. To do that, the reactivation approach would have to be augmented with something besides just reactivation of the antecedent. There would have to be some process that activates "teach" from "taught" in some mental representation, and this process would have to be tied to the syntactic voice of the clause containing the ellipsis, since it occurs more frequently in Active voice. Presumably this would be some process of integration, where the reactivated antecedent has to be integrated with the clause containing the elided VP. At this point, it is not clear what the necessary mental representation might be, or what the process of integration could be (see more on this in section 4).

3. Experiment 2.

In Experiment 2, we compare VP ellipsis with a VP pro-form, *do that*. In some versions of the reactivation approach (e.g., Culicover & Jackendoff 2005), VP ellipsis includes a silent pro-form that points back to the antecedent. In all versions of the reactivation approach, VP ellipsis should be processed similarly to VP pro-forms, which are generally taken not to have any silent structure. For all approaches, a VP pro-form would initiate a search for an antecedent and reactivate it once found, with no silent structure being built. The two approaches therefore make different predictions: The structure generation approach predicts that VP ellipsis and VP pro-forms should have different effects on a word recall task, while the reactivation approach predicts that they should behave the same.

The matter is not as simple as it initially appears, however, as one must consider what the processor has to do when it encounters a VP pro-form. Consider the sentence in (8), which contains an attested example of a voice mismatch with a VP pro-form.

(8) If the figures are moving, they <u>are perceived as what they are</u>. **To do this**, we must have in our heads a precise but flexible model of how people move. (Flambard 2018: 160, (209))

The VP pro-form *do this* must be reactivating the antecedent, as underlined in (8). This reactivated antecedent has to be integrated somehow with the clause that the VP pro-form is part of. It is not at all clear what the mental representation of the VP pro-form in this context is. The antecedent itself cannot be straightforwardly fit into the clause *to do this*, since it is in the wrong syntactic form. We are aware of two proposals for what the mental representation of the VP pro-form might be. As discussed above, Kim & Runner (2022) propose (for VP ellipsis) that the mental representation is an abstract argument structure like "[take.to(x,y,z)]." For (8), we might propose that the VP pro-form finds the underlined antecedent and activates an analogous abstract argument structure, namely, "[perceive.as(x,y)]," where x and y are filled in as *they* (the figures) and *what they are*. In the other proposal, Bruening (2019) analyzes the VP pro-form *do so* as copying the semantic representation of its antecedent. For instance, if the antecedent is *decorate the eggs*, a semantic function "*λe.decorate(e,the.eggs)*" (a predicate of events) is copied. For (8), a semantic function like "*λe.perceive.as(e,they,what.they.are)*" would be copied.

Note that both of these proposed representations are abstract. One is an abstract argument structure, the other is a semantic formula. It is not clear whether either of these would be expected to activate a given morphophonological form in processing. Kim & Runner (2022) and Bruening (2019) wrote "take.to" and "decorate," but we should not take that to mean that these specific lexical items will be activated at the pro-form site in processing. It is possible that the representation formed by the pro-form is a purely abstract, conceptual one. We know that such representations exist; for instance, there is an idiom *grasp/grab/clutch/seize at straws* that is regularly used with any of those four verbs. It is likely that this idiom is associated with a purely conceptual representation of the relevant action that links to all four verbs, all of which are distinct lexical items. A purely conceptual representation of this sort might be what is activated when a VP pro-form is encountered and its antecedent is found.

Turning back to our study, consider again the voice mismatch condition that yielded the greatest number of errors in judging whether the word *teach* was present in the sentence, like in (9).

(9) A new language was supposed to be taught in the fall semester, but no professors could <do that> because the schedule was too tight.

If we were to include a pro-form *do that* in place of ellipsis, as shown in angled brackets, what would we expect? The pro-form might activate a mental representation of an abstract argument structure, "[teach(x)]," or a semantic formula, "*λe.teach(e,a.new.language)*." These may or may not activate the morphophonological form *teach*. If it is activated, then we expect interference in the recall task, exactly as we found with VP ellipsis. (Note that in Experiment 1, if VP ellipsis were processed like a pro-form and pro-forms activate the stem/active form, then we should have found equal activation of *teach* across the two conditions, which is not what we found.)

Although it is not clear whether the mental representation created by a VP pro-form actually includes or activates a morphophonological form, we reason that if it does, it is likely to be the stem form of the verb, which in English is identical to the active form (i.e., *teach* in the above example). In comparing VP ellipsis to a VP pro-form, then, we reverse the direction of the voice mismatch, as in (10).

(10) **VP ellipsis and VP pro-form.**

a. Someone was supposed to teach organic chemistry in the fall semester, but **that wasn't done** because it was dropped from the curriculum.
b. Someone was supposed to teach organic chemistry in the fall semester, but **it wasn't** because it was dropped from the curriculum.

In (10a), the VP pro-form *that wasn't done* either activates a purely abstract, conceptual representation, or it activates the stem form *teach*. We have no reason to expect it to activate the passive participle form *taught* (beyond relatedness activation, since they are closely related). Thus, if we ask whether *taught* appeared in the sentence, we do not expect additional interference beyond that from relatedness. In contrast, in (10b), on the structural account, the processor must build silent structure at the ellipsis site that includes *taught*, and so we expect more interference and worse performance on the recall task compared to (10a). The reactivation approach expects no extra difficulty, as (10b) should be processed in basically the same way as (10a).

3.1. Method.

For Experiment 2, we created sentence pairs of the form in (10), again using verbs that have different stem forms for active and passive (*teach-taught*, *break-broken, bring-brought, buy-*

bought, catch-caught, sell-sold, choose-chosen, tell-told, freeze-frozen, wake-woken, speak-spoken, fly-flown, leave-left, steal-stolen, tear-torn, wear-worn). Due to budgetary constraints, we ran a smaller scale study, with 64 participants. In an effort to compensate for having a reduced sample size, we altered the nature of the baseline target conditions, such that participants would be less likely to achieve ceiling-level performance as they did in Experiment 1. This time, our 2x2 factorial design (again, within subjects, within items) crossed ELLIPSIS (ellipsis, pro-form; as in 10) with RELATEDNESS of the target word: The target was either related to the verb form in the first of the two sentences (as "taught" is related to "teach") or to a word in an adjunct to the antecedent clause (e.g., "spring," related to "fall"). Importantly, if the adjunct is reactivated in the second clause, this reactivation should not differ across the ellipsis versus pro-form conditions. An example itemset is given in Table 2.

	Stimulus	
Ellipsis = ellipsis	*Someone was supposed to teach organic chemistry in the fall semester, but it wasn't because it was dropped from the curriculum.* ***Verb-Related target: TAUGHT***	*Someone was supposed to teach organic chemistry in the fall semester, but it wasn't because it was dropped from the curriculum.* ***Adjunct-Related target: SPRING***
Ellipsis = pro-form	Someone was supposed to teach organic chemistry in the fall semester, but that wasn't done because it was dropped from the curriculum. ***Verb-Related target: TAUGHT***	Someone was supposed to teach organic chemistry in the fall semester, but that wasn't done because it was dropped from the curriculum. ***Adjunct-Related target: SPRING***

Table 2. *An example item for Experiment 2 (4 conditions: ellipsis/verb-related, ellipsis/adjunct-related, pro-form/verb-related, pro-form/adjunct-related).*

3.2. Participants.

Sixty-four participants, all self-reported L1 speakers of English, were recruited via *Prolific.com*. They received payment of 4 USD, based on an estimated completion time of 20-25 minutes.

3.3. Materials.

16 item sets like the one in Table 2 were constructed and distributed across four presentation lists in Latin Square. We also included the same 28 filler items as in Experiment 1 and corresponding comprehension questions as in Experiment 1. Again, the resulting 44 trials were presented in a pseudo-randomized order (no more than two critical trials were adjacent, and no two adjacent critical trials were of the same condition), and the experiment was coded on the online platform *PC Ibex* (Zehr & Schwarz, 2018).

3.4. Procedure.

The consent process, instructions, and procedure were the same as in Experiment 1.

3.5. Results.

4 participants were excluded from analysis because their accuracy rate on filler trials was less than 75% (i.e., they incorrectly identified whether the target word had or had not previously appeared), and/or they failed to answer at least 3 out of 4 comprehension questions correctly (i.e., above chance level). In total, data from 60 participants was included in the analysis. Their mean accuracy rate on filler trials was 90.8%, similar to that of Experiment 1, and again indicating that participants could generally identify the presence or absence of the target word accurately.

Like with Experiment 1, our dependent measure for analysis of critical trials was whether participants correctly rejected the target word as having appeared (coded as 1) or failed to reject it (coded as 0). Participants answered accurately on 737 out of 960 total trials (76.8%). The results by condition are shown in Figure 2. We analyzed the data following the same statistical procedure as for Experiment 1.

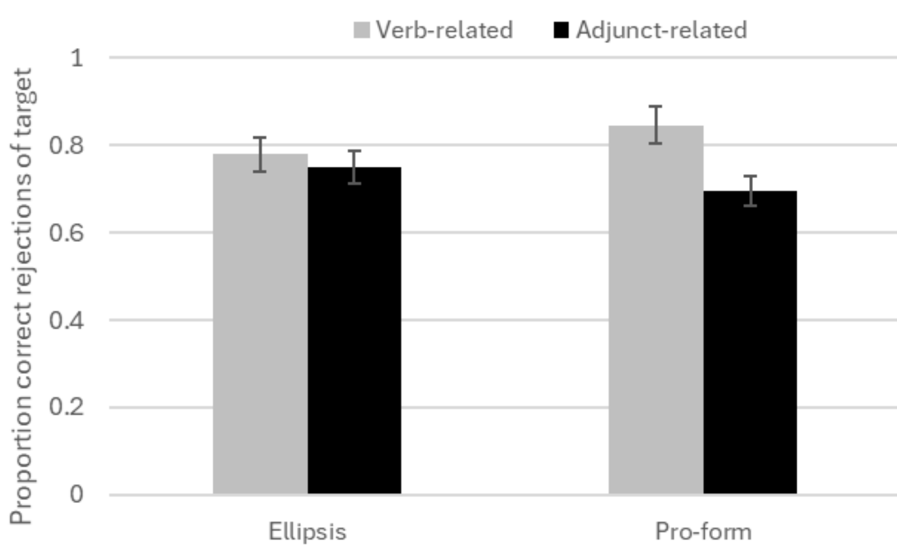

Figure 2. *Proportions of correct rejections of the target in ellipsis and pro-form conditions (target related to the verb in grey bars, target related to the adjunct in black bars). Error bars indicate 95% C.I.s.*

Importantly, we find a significant interaction of relatedness and ellipsis in the same direction as the relatedness*voice interaction in Experiment 1 (β = -1.09, *SE* = 0.39, *z* = -2.77, *p* = .0056): Nested planned comparisons revealed that participants demonstrated lower accuracy in correctly rejecting the verb-related target in the ellipsis conditions compared to the pro-form condition (.78 vs. .85; β = -0.66, *SE* = 0.3, *z* = -2.21, *p* = .027). Like with Experiment 1, we interpret this as participants having posited the verb "taught" at the ellipsis site in the ellipsis condition, leading them to be more likely to misconstrue the word as having appeared in the sentence. In contrast, if they are activating any morphophonological form in the pro-form condition, it would be the stem form "teach", leading to a lower likelihood of misconstruing "taught" as having appeared. As for the pair of adjunct-related conditions, the relevant difference trended in the opposite direction: Participants were descriptively more accurate in the ellipsis conditions compared with the pro-form conditions, but this difference was marginal (β = 0.44, *SE* = 0.26, *z* = 1.68, *p* = .093). We think that this difference most likely reflects some orthogonal performance differences involved in processing a sentence with VP ellipsis versus a sentence with a VP pro-form. Neither the main effect of ellipsis nor the main effect of relatedness was significant (both *p*s > .27).

We also examined target response-time data for the 737 critical trials in which the target was correctly rejected, fitting a 2x2 mixed-effects linear regression to log-transformed RTs (following the same statistical protocol as described above). The mean RTs are shown in Figure 3. Here, there was a main effect of Relatedness, where participants responded slower to verb-related targets than to adjunct-related targets (2780ms vs. 2489ms; log-transformed β = .14, *SE* = .054, *t* = 2.59, *p* = .019). There was also a main effect of ellipsis, where participants responded faster in the ellipsis conditions compared with the pro-form condition, overall (2475ms vs. 2800ms; log-transformed β = -.09, *SE* = .024, *t* = -3.8, *p* <.001). Importantly, the interaction of relatedness and ellipsis was also significant (β = .1, *SE* = .049, *t* = 2, *p* = .048): In both conditions, responses to the verb-related target are slower than those to the adjunct-related target, but this slow-down effect is more pronounced in the ellipsis condition than in the pro-form condition. This suggests that correctly rejecting a verb-related target was more burdensome in the ellipsis condition than in the pro-form condition, which is consistent with the view that participants were building VP structure in the ellipsis condition, structure that contained the target word.

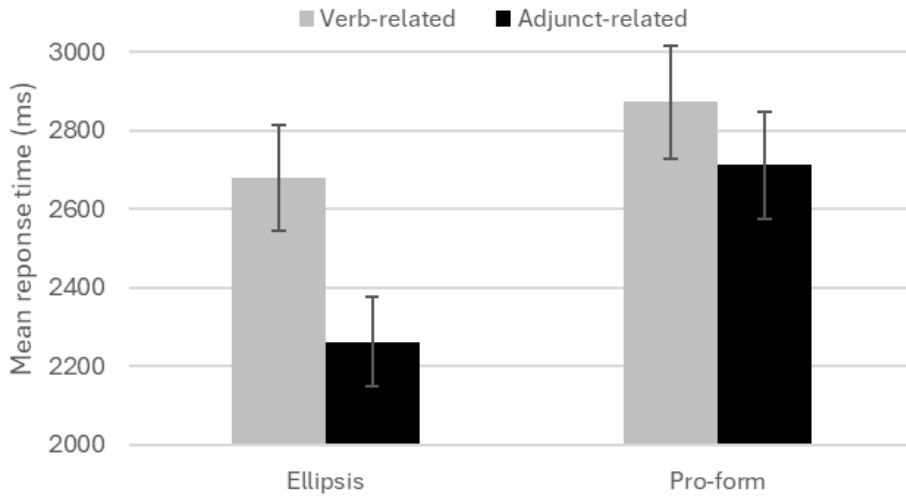

Figure 3. *Mean response times for correct rejections of the target in ellipsis and pro-form conditions (target related to the verb in grey bars, target related to the adjunct in black bars). Error bars indicate 95% C.I.s.*

3.6. Discussion.

The outcome of Experiment 2 provides further support for an approach in which processing VP ellipsis involves an additional step beyond reactivating the antecedent: Participants were tasked with reactivating this antecedent in both the pro-form conditions and the ellipsis conditions, yet we found that performance in the verb-related ellipsis condition was poorer compared with the verb-related pro-form condition. This outcome is consistent with the findings of Experiment 1, in which participants showed poorer performance in active-related conditions compared with passive-related conditions. Therefore, we arrive at the conclusion that interpreting VP ellipsis requires something extra besides antecedent reactivation.

It should be noted that previous experiments have compared the processing of VP ellipsis with the processing of VP pro-forms. In fact, many of these studies have looked at voice mismatches, just as our experiment did (although many started from the incorrect assumption that voice mismatches are ungrammatical with VP ellipsis). Most of these studies have found differences in the processing of VP ellipsis and VP pro-forms. For instance, Tanenhaus & Carlson (1990) and Mauner, Tanenhaus, & Carlson (1995) found differences between VP ellipsis and VP pro-forms in a timed “makes-sense” judgment task when the antecedent did not match in voice or syntactic category. Duffield, Matsuo, & Roberts (2009) and Duffield & Matsuo (2009), using materials similar to those of Tanenhaus & Carlson (1990), found an effect of mismatch on VP pro-forms as well as on VP ellipsis, but the effect was smaller with VP pro-forms. Bélanger (2004) found differences between VP ellipsis and VP pro-forms in reading times for a verification sentence. Roberts, Matsuo, & Duffield (2013) conducted an eye-tracking study and found differences in the processing of VP ellipsis and VP anaphors. Specifically, they found VP ellipsis to be more computationally costly at the earliest stages of interpretation, and the form of the antecedent had a greater effect on VP ellipsis than on VP pro-forms. Roberts et al. (2013) argue that these findings are best explained by structure generation taking place in VP ellipsis but not in VP pro-forms, consistent with our findings here. On the other hand, at least two studies found no differences between VP ellipsis and VP pro-forms: Murphy (1985) and Murphy (1990) compared the total reading times of sentences containing VP ellipsis with those containing VP anaphors with a variety of antecedents (e.g., shorter vs. longer, closer vs. further away, matching vs. mismatching voice), and found no difference between the two anaphor types. It is not clear at all what predictions the structure generation and reactivation approaches would make for total sentence reading times, and so we do not find these results particularly telling. We conclude that most of the available evidence accords with our findings in indicating that VP ellipsis and VP pro-forms are processed differently, and this is most consistent with the structure generation approach.

4. General discussion and conclusion.

In two experiments on VP ellipsis, we found differences in behavior that cannot be attributed solely to reactivation of the antecedent. In both experiments, the antecedents were identical in the contrasting conditions. Yet behavioral differences were observed which point to the syntactic properties of the elided clause having an effect on an elided lexical item. Specifically, the voice of the elided clause leads to greater activation of the form of the elided verb associated with that voice. In Experiment 2, this happened to a greater extent with VP ellipsis than it did with a VP pro-form, as predicted by the structural approach to VP ellipsis. The reactivation approach incorrectly predicts that VP ellipsis and VP pro-forms should behave identically. This is not what we found.

Our findings are therefore most consistent with the structure generation approach. In this approach, full syntactic structure is built at the site of VP ellipsis. This syntactic structure includes active or passive voice, such that the correct form of the elided verb is included at the ellipsis site (e.g., *taught*, when the antecedent VP had *teach*).

It should be noted that not all versions of the structure generation approach include morphophonological items in an elided clause. In some versions based on Distributed Morphology (Halle & Marantz 1993), for instance, the operation of Vocabulary Insertion is blocked, so that the syntactic terminals in an ellipsis site include only abstract syntactic features and therefore have no pronunciation (e.g., Arregi & Pietraszko 2021). In another version, only a null “elsewhere” morpheme is inserted into the syntactic terminals (Murphy

2016). This type of analysis would have syntactic structure, but only an abstract voice head and, presumably, only an unidentified root for the main verb. This version of the structure generation approach would also not expect higher activation of the actual form *taught* when the antecedent had *teach*, since it does not include that form in the unpronounced structure. However, in this approach, if the root TEACH is actually present in the elided structure, in a local configuration with passive voice, then we might expect that local configuration to activate the actual form *taught*. Hence, we believe that our findings rule out the strongest form of this approach, where the syntactic terminals in an elided constituent contain nothing but abstract syntactic features, but they might be compatible with a version where at least the verb root is present and in a local configuration with voice. Our findings are also most obviously compatible with a structure generation approach that includes actual morphophonological items in the syntactic terminals of the elided phrase, but they are just not pronounced. In this connection, we note that Colley & Bassey (2022) argue that elided material must include prosody; we take their argument, together with our findings, to indicate that ellipsis sites include full morphophonological specification that is just not pronounced.

Turning to the reactivation approach, that approach is not straightforwardly compatible with our results. If no structure is built at the ellipsis site, and all that happens is that the antecedent is reactivated, then our findings are not expected. Moreover, if VP ellipsis is thought to be processed identically to a VP pro-form, then our findings in Experiment 2 are also unexpected. Experiment 2 found VP ellipsis and VP pro-forms behaving differently, in exactly the way that would be expected if VP pro-forms do not include silent structure but VP ellipsis does.

This does not mean that reactivation approaches are absolutely ruled out, however. As discussed above, the reactivation approach has to be augmented with something besides antecedent reactivation. Minimally, this has to be a process of integration with the clause that contains the elided phrase. For instance, if the antecedent has active *teach*, but the elided clause has the passive voice, then *teach* has to be converted into *taught*. For this process of integration to be different from structure generation, a proponent of the reactivation approach will have to explain exactly what it is and how it is different from construction of silent structure. Ideally, this proposal will make different predictions from the structure generation approach, which can then be tested in further research. We hope that our experiments will stimulate exactly this sort of further research.

Data accessibility statement

The data, analysis code and experimental items (critical, fillers, and practice) are available at https://osf.io/6jbyu/?view_only=95ff292635fa4a61ac02fd7ecbd19c92.

Ethics and consent

The experiment was approved by the Institutional Review Board at the University of REDACTED.

Acknowledgements

To be added.

Competing interests

The authors have no competing interests to declare.

Authors' contributions

To be added.

ORCiD IDs

To be added.

References

Arregi, K. & Pietraszko, A. (2021). The ups and downs of head displacement. *Linguistic inquiry,* 52, 241-289.

Arregui, A., Clifton Jr, C., Frazier, L., & Moulton, K. (2006). Processing elided verb phrases with flawed antecedents: The recycling hypothesis. *Journal of memory and language*, *55*(2), 232-246.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, *59*(4), 390-412.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of statistical software*, *67*, 1-48.

Bélanger, S. (2004). Processing of deep and surface anaphors: A replication and extension of Tanenhaus, Carlson & Seidenberg (1985). *Toronto working papers in linguistics,* 22, 1-40.

Bruening, B. (2019). Passive *do so*. *Natural Language and Linguistic Theory*, 37, 1-49.

Chung, S., Ladusaw, W. A., & McCloskey, J. (1995). Sluicing and logical form. *Natural language semantics*, *3*, 239-282.

Clifton, C., Xiang, M., & Frazier, L. (2019). A note on the voice mismatch asymmetry in ellipsis. *Journal of Psycholinguistic Research*, *48*, 877-887.

Colley, J. & Bassi, I. (2022). A prosodic theory of possible ellipsis remnants. *Glossa: A journal of general linguistics*, 7(1), 1-41.

Culicover, P. W., & Jackendoff, R. (2005). *Simpler syntax*. OUP Oxford.

Culicover, P. W., & Jackendoff, R. (2019). Ellipsis in simpler syntax. In van Craenenbroek, J. & Temmerman, T. (eds.), *The Oxford Handbook of Ellipsis*, 162-187. Oxford University Press.

Davies, M. (2008). The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International journal of corpus linguistics, 14(2),* 159-190.

Duffield, N. & Matsuo, A. (2009). Native-speakers' vs. L2 learners' sensitivity to parallelism in VP-Ellipsis. *Studies in second language acquisition*, 31, 1-31.

Duffield, N., Matsuo, A., & Roberts, L. (2009). Factoring out the parallelism effect in VP-ellipsis: English vs. Dutch contrasts. *Second language research*, 25, 427-467.

Fiengo, R., & May, R. (1994). *Indices and identity* (No. 24). MIT press.

Flambard, G. (2018). English VP anaphors: *do it, do this, do that*. PhD dissertation, Université Sorbonne-Paris Cité.

Ginzburg, J., & Sag, I. (2000). *Interrogative investigations*. Stanford: CSLI publications.

Grant, M., Clifton Jr, C., & Frazier, L. (2012). The role of non-actuality implicatures in processing elided constituents. *Journal of Memory and Language*, *66*(1), 326-343.

Halle, M. & Marantz, A. (1993). Distributed Morphology and the pieces of inflection. In Hale, K. & Keyser, S.J. (eds.), The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger, 111-176. Cambridge, MA: MIT Press.

Hardt, D. (1993). Verb phrase ellipsis: Form, meaning and processing. Philadelphia, PA: University of Pennsylvania dissertation.

Hardt, D. (1999). Dynamic interpretation of verb phrase ellipsis. *Linguistics and philosophy*, 22, 185-219.

Kim, C., Kobele, G., Runner, J., & Hale, J. (2011). The acceptability cline in VP ellipsis. *Syntax*, *14*, 318-354.

Kim, J.-B., & Runner, J. (2022). Pseudogapping in English: A direct interpretation approach. *The Linguistic Review,* 39, 457-494.

Lasnik, H. (2001). When can you save a structure by destroying it? In *Proceedings of the North East Linguistic Society (NELS) 31,* pp. 301-320). Amherst, MA: GLSA, University of Massachusetts.

Martin, A.E., & McElree, B. (2008). A content-addressable pointer mechanism underlies comprehension of verb-phrase ellipsis. *Journal of memory and language*, 58, 879-906.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of memory and language*, *94*, 305-315.

Mauner, G., Tanenhaus, M.K., & Carlson, G. (1995). A note on parallelism effects in processing deep and surface verb-phrase anaphora. *Language and cognitive processes*, 10:1, 1-12.

Merchant, J. (2001). *The syntax of silence: Sluicing, islands, and the theory of ellipsis*. Oxford University Press.

Merchant, J. (2013). Voice and ellipsis. *Linguistic Inquiry*, *44*(1), 77-108.

Murphy, A. (2016). Subset relations in ellipsis licensing. *Glossa: A journal of general linguistics*, 1(1):44, 1-34.

Murphy, G. (1985). Processes of understanding anaphora. *Journal of memory and language*, 24, 290-303.

Murphy, G. (1990). Interpretation of verb phrase anaphora: Influences of task and syntactic context. *The quarterly journal of experimental psychology*, 42A(4), 675-692.

Nykiel, J., & Kim, J. B. (2022). Fragments and structural identity on a direct interpretation approach. *Journal of linguistics*, *58*(1), 73-109.

Phillips, C., & Parker, D. (2014). The psycholinguistics of ellipsis. *Lingua*, *151*, 78-95.

Roberts, L., Matsuo, A., & Duffield, N. (2013). Processing VP-ellipsis and VP-anaphora with structurally parallel and nonparallel antecedents: An eye-tracking study. *Language and cognitive processes*, 28:1-2, 29-47.

Ross, J. R. (1969). Guess who? *Papers from the Fifth Regional Meeting of the Chicago Linguistic Society*, 252-286. Chicago, IL: University of Chicago.

Sag, I. A. (1976). *Deletion and logical form*. Cambridge, MA: Massachusetts Institute of Technology dissertation.

Shapiro, L. P., Hestvik, A., Lesan, L., & Garcia, A. R. (2003). Charting the time-course of VP-ellipsis sentence comprehension: Evidence for an initial and independent structural analysis. *Journal of Memory and Language*, *49*(1), 1-19.

Stockwell, R. (2024). Ellipsis, contradiction and voice mismatch. *Glossa: a journal of general linguistics*, *8*(1), 1-21.

Tanenhaus, M.K., & Carlson, G.N. (1990). Comprehension of deep and surface verbphrase anaphors. *Language and cognitive processes*, 5:4, 257-280.

Xiang, M., Grove, J., & Merchant, J. (2019). Structural priming in production through 'silence': An investigation of verb phrase ellipsis and null complement anaphora. *Glossa: a journal of general linguistics*, *4*(1), 1-25.

Zehr, J., & Schwarz, F. (2018). PennController for internet based experiments (IBEX).