# Discourse conditions on Verb Phrase Ellipsis and the question of syntactic identity: new evidence from corpus data and acceptability experiments

**Philip Miller***, **Geoffrey K. Pullum**** and **Barbara Hemforth***
*U. Paris Diderot, **U of Edinburgh
philip.miller@univ-paris-diderot.fr

## Abstract

Previous discussions of VP Ellipsis (VPE) with antecedents that fail to match in category have not taken into account the discourse conditions that apply to the construction. Unless rather special conditions are met, examples involving VPE with a category-mismatched antecedent are very likely to violate the discourse conditions. Differences in acceptability may therefore be mistaken for a consequence of the category mismatch, when in fact they are due to a failure to respect the discourse conditions. We report both corpus evidence and experimental results showing that such misdiagnosis has almost certainly been occurring in the literature on VPE.

## 1 Introduction

Verb Phrase Ellipsis (VPE) is standardly taken to be a **surface** anaphor, requiring a syntactically identical antecedent at some level (Hankamer and Sag, 1976; Merchant, 2013). The contrast between (1) and (2) is thus expected:

(1) Jo programmed her computer to control her central heating remotely. Pat did too.

(2) Jo is an excellent computer programmer. #She does professionally.

VPE is fine in (1) because a syntactically identical antecedent is available; (2) is strikingly bad because only an NP antecedent for the VPE is available (category mismatch). In such cases, Verb Phrase Anaphora (VPA, e.g., *do it*) are much better (ok: 'She does it professionally'), because they are **deep anaphors** and simply require an accessible antecedent in the discourse context.

But other studies (e.g. Hardt (1993)) point out acceptable cases of mismatch, e.g.:

(3) I need to know how to force their cooperation or hammer them hard if they won't. (Avatar)

They argue that VPE is simply a proform needing an appropriate antecedent recoverable from the discourse context. This, however, offers no explanation for the unacceptability of (2). Various authors have proposed that the difference between (2) and (3) can be explained in terms of general discourse conditions (e.g., Kehler (2002) and Kertz (2013)).

What has never been noticed in the literature (to our knowledge) is that there are cases with identical judgments to (2), but where there **is** a plausible syntactically matched VP antecedent in the immediate discourse context. The examples in (4) and (5) are based on naturally occurring data.

(4) They said good-bye to Dare, whose boyfriend wasn't feeling well, rubbing a bit of egg from the side of her mouth as they talked briefly. #He did with his napkin in a fatherly way. (Ok: He did it with . . . )

(5) A: How did he get that ball into the hole?
  a. B: He didn't.
  b. B: #He did.
  c. B: He got it into the hole.

Despite the syntactic match, VPE in (4) and (5-b) exhibits an unexpected strong decrease in acceptability, intuitively similar to that of (2). Furthermore, such examples do not violate the discourse conditions of Kehler (2002) and Kertz (2013). These hitherto unnoted cases are crucially important, as they show that acceptability of VPE cannot be fully explained in terms of pure syntax or general discourse conditions. VPE in (4) and (5) are surely not ungrammatical: there are perfectly good antecedents in the immediate discourse context. It is thus clear that an occurrence of VPE can

be grammatical yet quite unacceptable.

This means that the unacceptability of examples like (2) is hard to interpret. We cannot simply assume ungrammaticality, calling into question the idea (Arregui et al., 2006; Frazier, 2013) that mismatch is systematically ungrammatical, but can be more or less acceptable thanks to repair.

## 2 A construction-specific discourse condition on VPE

We propose the following solution to this puzzle: VPE is an anaphor which requires an accessible antecedent in the discourse context without any requirement in terms of syntactic identity. But it is governed by the following **construction-specific** discourse constraint:[1]

(6) If the QUD addressed by the anaphoric clause is entirely inferrable from the antecedent clause VPE is preferred to VPA. If not, i.e., if the anaphoric clause addresses a non-entirely inferrable QUD then VPA is preferred to VPE .

This constraint applies as follows to (7): (7-a) and (7-b) introduce the QUD 'Did Joan clean her room?'. (7-c) and (7-d) address this QUD and VPE (7-c) is clearly preferred to VPA (7-d). On the other hand, (7-e) and (7-f) address the non-entirely inferrable QUD 'How did Joan clean her room?'. In this case, VPA (7-f) is clearly preferred to VPE (7-e).

(7)
a. A: Joan <u>cleaned her room</u>.
b. A: Did Joan <u>clean her room</u>?
c. B: She did. / B: She didn't.
d. ??B: She did it. / ??B: She didn't do it.
e. ??B: She did very thoroughly.
f. B: She did it very thoroughly.

Similarly, (1) introduces the QUD 'Who programmed her computer ...remotely'.[2] The anaphoric clause addresses this QUD and is felicitous but replacing *did* with *did it* decreases acceptability. In (2) and (4), on the other hand, the idea that 'Joan programs computers' and that 'Dare rubbed a bit of egg from the side of her mouth' are not QUDs. They are backgrounded (by the agent nominalization and participial adjunct status respectively). This explains that the VPA *do it* is preferred to VPE.

In (5), the *wh*-interrogative backgrounds the proposition 'he got that ball into the hole' so that it is not QUD. (5-a) is acceptable because it is coherent for speaker B to force accommodation of that proposition as the QUD in order to contradict it. But — unexpectedly from a syntactic perspective — (5-b) is highly unacceptable, because it makes no sense to accommodate a backgrounded proposition as QUD in order to confirm it. Note crucially that it is not that the content conveyed by (5-b) cannot make sense in the context, as the non elliptical counterpart in (5-c) is acceptable (conveying, e.g., it doesn't matter how).

Condition (6) also explains why VPE is unexpectedly acceptable in (3), despite the nominal antecedent: 'I need to know how to force their cooperation' makes the QUD 'will they or won't they cooperate' inferrable.

## 3 Evidence for the discourse condition from acceptability experiments

### Experiment 1

A first experiment aimed to corroborate the idea that adding an adjunct, and thus going beyond the inferrable QUD, reduces the acceptability of VPE and increases that of VPA. Items were based on 2 binary factors: VPE vs. VPA and the presence or absence of a non contrastive adjunct (Adj+/–), i.e., four conditions as in:

| | |
|---|---|
| VPE/Adj– | Sue didn't write a song. Sam did. |
| VPA/Adj– | Sue didn't write a song. #Sam did it. |
| VPE/Adj+ | Sue didn't write a song. #Sam did for her. |
| VPA/Adj+ | Sue didn't write a song. Sam did it for her. |

20 items of this type were created and distributed across 4 lists in a Latin Square design interspersed with 16 items from a separate experiment as distractors. 48 subjects, recruited on Amazon Mechanical Turk, judged the acceptability of the second clause in the context of the first on a scale of 1 to 7. The experiment was run on Ibex Farm (Drummond, 2014).

The expectation was that that there would be an interaction between the two factors, so that VPE/Adj– and VPA/Adj+ would be judged better than VPE/Adj+ and VPA/Adj–. The results given in Figure 1 perfectly corroborate this: the inter-

[1] This only applies to non comparative VPE.

[2] Note that in general, in the absence of marked intonation, more than one QUD can be inferred from the antecedent clause; e.g., the antecedent clause in (1) could also be interpreted as introducing the QUD 'Did Joan program her computer ...remotely?', in which case the VPE clause *She didn't.* would also be a felicitous continuation.

action between Adj+/– and VPE/VPA was highly significant ($p < 0.0027$).

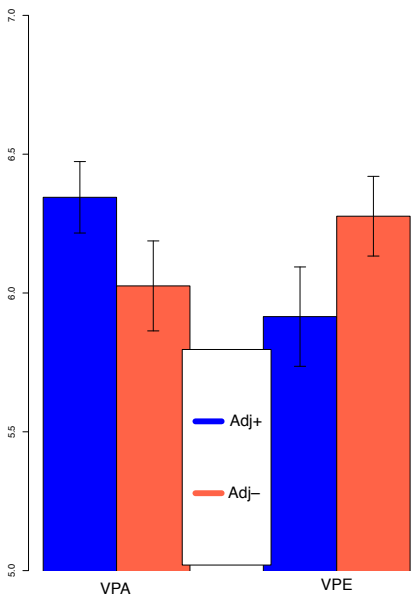


Figure 1: Experiment 1: acceptability judgments

**Experiment 2**

We will report on the results of a second experiment (the materials are ready, but it was impossible to run it before the abstract deadline) aiming to further corroborate the discourse condition. Items will again be based on two binary factors: (i) either the antecedent is asserted (ASS), making it QUD+, or backgrounded (BG), making it QUD–; (ii) the anaphoric clause involves either VPE or VPA. Here is a typical item in its 4 conditions:

| | |
|---|---|
| ASS/VPE | When they were in the study they played bridge. They did? |
| ASS/VPA | When they were in the study they played bridge. They did it in the study? |
| BG/VPE | They were in the study when they played bridge. They did? |
| BG/VPA | They were in the study when they played bridge. They did it in the study? |

Because the first clause involves copular *be*, only the second clause can be the antecedent of the subsequent anaphoric sentence. In the ASS conditions the second clause is the main clause and is thus asserted and QUD+, whereas in the BG conditions, the second clause is a temporal subordinate and thus backgrounded and QUD–. Our expectation is that the anaphoric clause will be judged significantly less acceptable in the BG/VPE condition than in the other three.

## 4 Teasing apart the effects of mismatch and discourse conditions

As discussed in the introduction, it has traditionally been assumed by most linguists working on VPE that the usually clear decrease in acceptability with mismatched antecedents is caused by the resulting violation of the syntactic identity constraint. However, we have seen intuitively, and corroborated through experiments, that VPE with matched antecedents can be significantly degraded in acceptability if the discourse condition (6) is not satisfied. Because category mismatch typically leads to backgrounding the antecedent (as in (2)), the question arises as to whether the decrease in acceptability is due to the mismatch as such, or whether it is a consequence of the correlated violation of the discourse condition (6).

In order to tease apart the respective effects of mismatch and discourse conditions, it is necessary to investigate those cases of mismatch where, in spite of the mismatch, a QUD can be inferred from the antecedent, as in (3) above. To do this, we used 'polar nouns' as antecedents. These nouns, which we have discussed in previous work, are exceptional in that they can be the head of an NP which can be interpreted in a way very similar to an indirect question, when placed in certain contexts taking indirect interrogatives. This is illustrated in the following examples:

(8) Mubarak's <u>survival</u> is impossible to predict and, even if he <u>does</u>, his plan to make his son his heir apparent is now in serious jeopardy. (COCA)
(≈ Whether or not Mubarak will survive is impossible to predict . . . )

(9) Your <u>success</u> depends on your <u>presence</u> in class and your active <u>participation</u> in class discussions and activities. (iWeb)
(≈Whether or not you succeed depends on whether or not you are present in class and whether or not you participate actively . . . )

Thus, in appropriate contexts, NPs headed by these nouns allow the inference of a QUD ('will Mubarak survive?' in (8)), which in turn makes them very good antecedents for VPE, as is illustrated in the attested example (8). We exploited this possibility in the construction of the materials for Experiment 3.

## Experiment 3

Items were based on two binary factors, category match vs mismatch (CM vs CMM) and whether the discourse condition (6) is satisfied, i.e., whether the elliptical clause addresses a QUD that is entirely inferrable from the antecedent (QUD+ vs QUD–). Here is a typical item in its 4 conditions:

| | |
|---|---|
| CM/QUD+ | We are uncertain whether he will participate in the study. It will improve the results if he does. |
| CMM/QUD+ | We are uncertain of his participation in the study. It will improve the results if he does. |
| CM/QUD– | We are uncertain whether he will participate in the study. It will improve the results if he does actively. |
| CMM/QUD– | We are uncertain of his participation in the study. It will improve the results if he does actively. |
| CONTROL | We are uncertain whether he will participate in the study. It will improve the results if he participates actively. |

The CMM conditions involve a polar noun in an indirect interrogative environment ('uncertain of his participation' in this item). The corresponding CM conditions involve the obvious paraphrase with an actual indirect interrogative clause. As in Experiment 1, we create the QUD– environments by adding a non contrastive adjunct ('actively' in this item). Crucially, this design allows us to compare the respective effect on acceptability of category mismatch and discourse conditions with minimal changes within each item.

Clearly, we expect the CM/QUD+ condition to be the most acceptable and the CMM/QUD– condition to be the least acceptable. The question is whether there is a difference in acceptability between CMM/QUD+ and CM/QUD–, i.e., does a violation of the syntactic identity constraint affect acceptability more or less than a violation of the discourse constraint?

In order to be sure that a decrease in acceptability in the CM/QUD– is not due to a bad choice of adjunct, we also included a control condition, in which the verb is simply repeated before the adjunct.

**Results** The overall results are shown in Figure 2.[3] There was no significant difference between the CONTROL and the CM/QUD+ conditions, confirming that any decrease in acceptability in the QUD– conditions is not due to the choice of the adjunct. Individual violation of the match condition (CMM/QUD+) and of the discourse condition (CM/QUD–) both led to a significant decrease in acceptability compared to the control condition ($p<0.001$), with no significant difference between them. In other words, **violation of the discourse condition can have as much effect on acceptability as category mismatch**. When both the match and discourse condition are violated, acceptability is significantly lower than when a single condition is violated, but interestingly we find an underadditive interaction: the decrease in acceptability with respect to CM/QUD+ is less than the sum of the individual decreases (in other words, the difference between the 4th condition and the 2nd and 3rd (0.94 and 1.02 respectively) is less than half of that between the latter and the 1st condition (0.44 and 0.37 respectively).

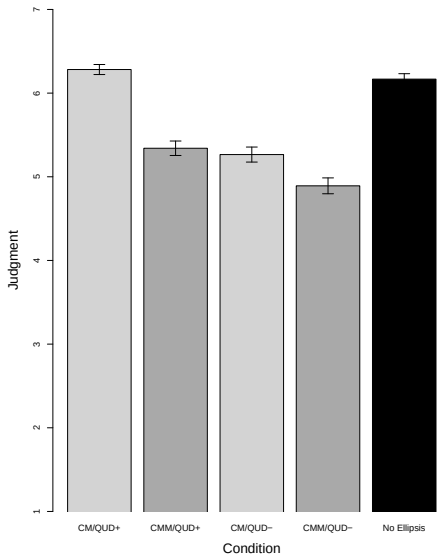


Figure 2: Experiment 3: acceptability judgments

[3]The data were analyzed with R version 3.5.0. Ratings were submitted to a linear mixed effects regression (Baayen, 2008) using the lmer-function from the R-package lme4, version 1.1-17 (Bates et al., 2015). We first ran a model to compare all ellipsis conditions to the 'no ellipsis' condition. We used the factor CONDITION as a within participant and within items predictor. In addition to random intercepts of items and participants, we included random slopes for CONDITION for items and for participants. P-values were computed with the lmerTest-package (Kuznetsova et al., 2017), using Satterthwaite approximation of degrees of freedom. The no ellipsis condition was defined as the reference condition for this model using treatment coding. In a second model, we used the factors MATCH and the factor QUD as predictors. Predictors were centered for this model. In addition to random intercepts of items and participants, we included random slopes for both factors for items and for participants. For all analyses, we excluded correlations between random intercepts and slopes to avoid overprediction. We used the optimx package to optimise model fit (Nash, 2014).

**Discussion** As previously mentioned, the central result is that not respecting the discourse condition with a matched antecedent can decrease acceptability as much as having a mismatched antecedent with the discourse condition satisfied. However, one does not simply find that mismatch has no effect at all on acceptability: even when the discourse condition is satisfied, a nominal antecedent leads to a decrease in acceptability. We argue that this is not due to a syntactic identity requirement but to processing constraints.

Specifically, it is simply not the case that classical deep anaphors exhibit no acceptability effects linked to syntactic mismatches with their antecedents. This has been known in the literature for a very long time (McKoon et al., 1993), as evidenced by examples of like the following:

(10) Kim is from France, but he's never lived there as an adult.

(11) Kim is a Frenchman, but he's never lived there as an adult.

Although *there*, as a deep anaphor, is assumed to simply require that an antecedent be accessible in the discourse context, clearly a case like (10), where the necessary antecedent 'France' is directly mentioned, is more acceptable than (11), where it must be inferred from from 'Frenchman'.

Similarly, corroborating various studies (Bélanger, 2004; Woodbury, 2011) we have shown in previous experimental results that the VPA *do it* is slightly (but significantly) less acceptable with a mismatched antecedent (cf. also the classical results of Tanenhaus and Carlson (1990), who find no difference in acceptability but *do* find increased processing difficulty in cases of *do it* with mismatch).

Thus following the argumentation developed by various others (Kim et al., 2011; Kim and Runner, 2018; Kertz, 2013) we suggest that mismatch has no effect on grammaticality, but affects acceptability because processing is easier with matched antecedents (the processor makes use of all available information including what is present in short term syntactic memory).

## 5 Conclusion

The position we defend is that VPE is an anaphoric construction requiring an accessible antecedent in the discourse context, without any requirement of syntactic identity. This allows us to explain numerous phenomena such as exophoric uses of VPE, which (contra the classic claims of Hankamer and Sag) is copiously attested (Miller and Pullum, 2014) and VPE with split antecedents (Frazier and Duff, 2019).

We suggest that one of the central reasons this idea has not been embraced by many specialists on ellipsis is linked to the fact that examples like (4), (5-b), (7-e), which are degraded in acceptability despite the presence of a matched antecedent, have not been noted in the literature. Thus the importance of discourse constraints on acceptability has been underestimated, and likewise the necessity of disentangling their effects from those of mismatch *per se*. More specifically, though general (non-construction-specific) discourse constraints have been proposed on VPE (Rooth, 1992; Kehler, 2002; Kertz, 2013), these predict that a VPA like *do it* should be subject to the same constraints. Rooth (1992) applies his theory of focus and alternatives to VPE, using the idea of semantic redundancy. This leads one to expect that *do it* should function in precisely the same way, because it exhibits the same kind of redundancy. Similarly, Kertz (2013) argues that a large part of the decrease in acceptability due to argument structure mismatch is also present in the corresponding non-elliptical sentences, and one would expect the same to be true of the corresponding *do it* sentences. Our data raise problems for such accounts, because, generally, when the discourse condition (6) is not satisfied, replacing VPE by VPA strongly increases acceptability.

Notice that this discussion says nothing about the issue of unpronounced syntactic structure in ellipsis sites (as advocated e.g. by Merchant (2013) and denied e.g. by Jacobson (2016)). Though these two hypotheses are often treated conjointly, the fact that a VP antecedent can be contextually inferred does not preclude the idea that it could actually appear in the syntactic structure without being pronounced. The thesis of this paper is neutral with regard to the Merchant–Jacobson theoretical dispute.

VPE demands (i) a pragmatically accessible unit of meaning and (ii) compliance with certain construction-specific discourse constraints. When conditions (i) and (ii) are satisfied, VPE will happily tolerate absence of a syntactic antecedent altogether (exophoric VPE) or an antecedent constituent of mismatched category.

## References


Ana Arregui, Charles Clifton, Jr., Lyn Frazier, and Keir Moulton. 2006. Processing elided verb phrases with flawed antecedents: The recycling hypothesis. *Journal of Memory and Language*, 55:232–246.

R. Harald Baayen. 2008. *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge University Press, Cambridge.

D. Bates, M. Maechler, B. Bolker, and S. Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67:1–48.

Suzanne Michelle Bélanger. 2004. Processing of deep and surface anaphors: A replication and extension of Tanenhaus, Carlson and Seidenberg (1985). In *Toronto Working Papers in Linguistics (Vol 22)*.

Alex Drummond, 2014. *Ibex Farm*. http://spellout.net/latest_ibex_manual.pdf.

Lyn Frazier and John Duff. 2019. Repair or accommodation? split antecedent ellipsis and the limits of repair. *Glossa*, to appear.

Lyn Frazier. 2013. A recycling approach to processing ellipsis. In Lisa Lai-Shen Cheng and Norbert Corver, editors, *Diagnosis in Syntax*, pages 485–501. Oxford University Press, Oxford.

Jorge Hankamer and Ivan A. Sag. 1976. Deep and surface anaphora. *Linguistic Inquiry*, 7:391–426.

Daniel Hardt. 1993. *Verb Phrase Ellipsis: Form, Meaning, and Processing*. Ph.D. thesis, University of Pennsylvania. Distributed as IRCS Report 93-23.

Pauline Jacobson. 2016. The short answer: Implications for direct compositionality (and vice versa). *Language*, 92:331–375.

Andrew Kehler. 2002. *Coherence, Reference, and the Theory of Grammar*. CSLI Publications, Stanford.

Laura Kertz. 2013. Verb phrase ellipsis: The view from information structure. *Language*, 89:390–428.

Christina S. Kim and Jeffrey T. Runner. 2018. The division of labor in explanations of Verb Phrase ellipsis. *Linguistics and Philosophy*, 41:41–85.

Christina S. Kim, Gregory M. Kobele, Jeffrey T. Runner, and John T. Hale. 2011. The acceptability cline in VP ellipsis. *Syntax*, 14:318–354.

Alexandra Kuznetsova, Per B. Brockhoff, and Rune H. B. Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82:1–26.

Gail McKoon, Gregory Ward, and Roger Ratcliff. 1993. Morphosyntactic and pragmatic factors affecting the accessibility of discourse entities. *Journal of Memory and Language*, 32:56–75.

Jason Merchant. 2013. Voice and ellipsis. *Linguistic Inquiry*, 44:77–108.

Philip Miller and Geoffrey K. Pullum. 2014. Exophoric VP ellipsis. In Philip Hofmeister and Elisabeth Norcliffe, editors, *The Core and the Periphery: Data-driven Perspectives on Syntax Inspired by Ivan A. Sag*, pages 5–32. CSLI Publications, Stanford, CA.

John C. Nash. 2014. *Nonlinear Parameter Optimization Using R Tools*. Wiley, Chichester.

Mats Rooth. 1992. Ellipsis redundancy and reduction redundancy. In Steve Berman and Arild Hestvik, editors, *Proceedings of the Stuttgart Ellipsis Workshop*.

Michael K. Tanenhaus and Greg N. Carlson. 1990. Comprehension and deep and surface verbphrase anaphors. *Language and Cognitive Processes*, 5:257–280.

Rebecca R. Woodbury. 2011. *Behavioral and Neural Correlates of "Deep" and "Surface" Anaphora*. Ph.D. thesis, MIT, Cambridge, MA.