

Discourse Structure and Syntactic Parallelism in VP Ellipsis

Christina Kim & Jeffrey Runner

University of Rochester

1. Introduction

This paper presents psycholinguistic data from three magnitude estimation experiments that suggest a syntactic parallelism condition on VP ellipsis (VPE) that can be modulated by discourse coherence, and applies to VPE both within and across sentences. Section 1 will outline some of the arguments for and against syntactic and semantic accounts of VPE, and lay out the predictions of existing proposals with respect to syntactic mismatch in VPE. Section 2 reviews previous psycholinguistic studies on VPE. The three magnitude estimation experiments are presented in sections 3-5. Experiment 1 looks at cases of voice and category mismatch in VPE with sentential coordination, replicating some basic contrasts found by Arregui, Clifton, Frazier & Moulton (2006). Experiment 2 compares voice-mismatched VPE under different discourse coherence relations (Kehler 2000), and Experiment 3 compares voice mismatch in VPE in coordinate structures and across sentence boundaries.

Together, these findings support the existence of a syntactic parallelism condition on VPE which applies at the level of discourse structure. The observation that syntactic mismatch effects are observed to the same extent across discourse and within individual syntactic structures suggests that at least certain syntactic alternations, like passive, are encoded at the level of discourse representations. Empirical consequences of assuming different sources for mismatch-related unacceptability are discussed in section 6.

1.1 Licensing VP ellipsis: The identity condition

How are sentences with ellipsis interpreted? In sentences containing various types of ellipsis—like VPE (1a), sluicing (1b), pseudogapping (1c), comparative deletion (1d)—it's clear that the understood meaning of the elided piece is similar to that of an overt phrase elsewhere in the sentence. We'll refer to this overt phrase as the antecedent.

- (1) a. Jill made fun of Abby, and Matt did ~~make fun of Abby~~, too.

- b. The department really wanted to hire a psycholinguist straight out of grad school, but I don't remember who ~~the department really wanted to hire straight out of grad school~~.
- c. Sameer picked out all the red jelly beans, and Justin did ~~pick out~~ the purple ones.
- d. Jane smokes more cigarettes per day than Nate does ~~smoke x cigarettes per day~~.

In particular, the elided material can't be construed as something else; for example, (1d) cannot mean that Jane smokes more cigarettes per day than Nate drinks cups of coffee. The exact nature of the similarity that must obtain between antecedent and ellipsis site has been under debate.

1.2 Syntactic identity

Before asking what kind of identity must hold between antecedent and elided VPs, we might ask what kinds of information are in the ellipsis site. It's in principle possible that the identity requirement holds across just one of the types of information encoded in the reconstructed VP representation—for instance, semantic or property-level identity could hold of representations that also encode rich syntactic information. However the reverse is not true: syntactic identity could not hold across representations that encode only semantic information.

In fact, there is reason to think the ellipsis site contains unpronounced syntactic structure in sentences like those in (1). When an elided constituent would contain an island violation if its syntactic structure were reconstructed into the ellipsis site, a sentence becomes unacceptable (Haik 1987, Johnson 2001). This is illustrated for an adjunct island in (2b) (from Johnson 2001) and a *wh*-island in (3b) (from Kennedy 2003). If VPE involves deletion or non-pronunciation of a piece of syntactic structure, the unacceptability can be attributed to the island violations in (2c) and (3c).

- (2)
 - a. Sterling criticized every decision that Lou did Δ .
 - b. Sterling criticized every decision that Doug was upset because Lou did Δ .
 - c. Sterling criticized every decision *wh_i* that Doug was upset because Lou did ~~criticize *wh_i*~~.
- (3)
 - a. I know which book Mag read, and which book you hadn't Δ .
 - b. I know which book Mag read, and which book Mr. Yunoshi asked why you hadn't Δ .
 - c. I know which book Mag read, and *which book_i* Mr. Yunoshi asked why you hadn't read ~~*which book_i*~~.

Is it these syntactic representations that must meet the identity condition? Some of the earliest accounts propose that syntactic identity must hold either at surface structure or logical form for VPE to be well-formed (Sag 1976, Williams 1977, Sag & Hankamer 1984, also Hankamer 1979, Tancredi 1992, Wilder 1995). More recent work (e.g. Merchant 2008, Arregui et al. 2006, Kobele 2006, Kennedy & Merchant 2000, Kennedy 2003) has taken up versions of syntactic identity as well. A strict syntactic identity condition accounts for

the unacceptability of (4b), where the antecedent and ellipsis site differ in voice, relative to its matched counterpart (4a).

- (4) a. Jill betrayed Abby, and Matt did ~~betray Abby~~, too.
 b. Jill betrayed Abby, and Matt was ~~betrayed by Jill~~, too.

But, as noted by Kennedy & Merchant (2000), Arregui et al. (2006), Tanenhaus & Carlson (1990), and others, purely syntactic accounts appear to undergenerate, to the extent that certain sentences with syntactic mismatch are acceptable (5).

- (5) a. This information could have been released by Gorbachov, but he chose not to. (*Daniel Shorr, NPR, 10/17/92, from Hardt 1993, cited in Arregui et al. 2006*)
 b. In March, four fireworks manufacturers asked that the decision be reversed, and on Monday the ICC did. (*from Rosenthal 1988; cited in Dalrymple et al. 1991, Kehler 2002*)
 c. This problem was to have been looked into, but nobody did ~~look into it~~. (*from Kehler 2002, cited in Arregui et al. 2006*)

Thus, while sensitivity to syntactic phenomena like binding and passivization suggest that there is syntactic structure in the ellipsis site and that there should be a condition on VPE that is stated in terms of syntactic structure, a strictly syntactic identity condition leaves some apparently acceptable cases of syntactic mismatch unexplained.

1.3 Semantic identity

An alternative approach to licensing VPE allows for such sentences by requiring semantic identity between the ellipsis site and its antecedent (Dalrymple, Shieber & Pereira 1991, Hardt 1993)—on this view the elided VP is syntactically like a proform, in that it has no internal syntactic structure, and picks up its meaning from the context of utterance. For Dalrymple et al. (1991), ellipsis is resolved through higher-order unification, where identity is at the level of property-denoting expressions. Thus (6a-6b) is understood by finding an antecedent property (6c) and substituting it into the second conjunct (6d-6e).

- (6) a. Someone released this information, but Gorbachov didn't Δ .
 b. *release(someone, this information) but not P(Gorbachov)*.
 c. $P = \lambda x. \text{release}(x, \text{this information})$.
 d. *release(someone, this information)*
 but not $\lambda x. \text{release}(x, \text{this information})(\text{Gorbachov})$.
 e. *release(someone, this information)*
 but not $\text{release}(\text{Gorbachov}, \text{this information})$.

The problem is that such accounts will overgenerate—(7) should be resolved in the same way as (6), but here the voice mismatch seems to degrade acceptability.

- (7) This information was released, but Gorbachov didn't Δ .

1.4 Hybrid approaches: Discourse-level alignment

How should the need for strict syntactic identity in some cases and the need for syntactic flexibility in others be resolved? One approach to reconciling this apparent conflict is to take the above observations to indicate that both semantic and syntactic parallelism are available as ways of recovering the elided VP. The problem then becomes figuring out what properties define the cases where the syntactic and semantic recovery mechanisms are used. This is the approach taken by Kehler (1995, 2000, 2002), who proposes that whether parallelism is syntactic or semantic depends on how the meanings of the clauses containing the two VPs are related.

For Kehler, whether syntactic effects appear depends on the discourse relation between the clauses containing the antecedent and the elided phrase. In particular, an elided expression in a sentence that is part of a Cause-Effect relation only needs to match its antecedent in propositional content, and shouldn't require syntactic information. On the other hand, the elided material in a Resemblance relation relies on aligning its syntactic arguments with those of its antecedent, and therefore should show degradation when there is syntactic mismatch.

Discourse coherence-based accounts predict interactions between Mismatch and Discourse relation type. (4b) and (9) should be worse than their Matched counterparts (4a) and (8), since they are instances of Parallel conjuncts.

- (8) The report criticized Roy, but Kate didn't ~~criticize~~ Roy
- (9) a. The criticism of Roy was harsh, but Kate didn't ~~criticize~~ Roy
b. The report was critical of Roy, but Kate didn't ~~criticize~~ Roy

But (11) and (13) should not be worse than their Matched counterparts (10) and (12) because they are instances of the Cause-Effect. No asymmetry is predicted between Active-Passive and Passive-Active sequences.

- (10) a. Jill invited Abby, so Matt did ~~invite~~ Abby, too.
b. Abby was invited by Jill, so Matt was ~~invited by Jill~~, too.
- (11) a. Jill invited Abby, so Matt was ~~invited by Jill~~, too.
b. Abby was invited by Jill, so Matt did ~~invite~~ Abby, too.
- (12) The report criticized Roy, so Kate didn't ~~criticize~~ Roy
- (13) a. The criticism of Roy was harsh, so Kate didn't ~~criticize~~ Roy
b. The report was critical of Roy, so Kate didn't ~~criticize~~ Roy

Mixed approaches to ellipsis resolution such as Kehler's have the advantage that they allow for both syntactic and semantic recovery mechanisms, and specify the discourse conditions under which one mechanism is used other another. Such a theory links observed unacceptability due to violations of syntactic or semantic identity to the use of a particular (syntactic or semantic) VP recovery mechanism, which is determined categorically by the type of discourse context the ellipsis is embedded in. Thus the strongest form of Kehler's

proposal seems to predict sensitivity to violations of syntactic identity under the Resemblance relation, and complete insensitivity to syntactic structure under Cause-Effect.

2. Experimental evidence

Tanenhaus & Carlson (1990) provide some initial psycholinguistic evidence in favor of a syntactic condition on VPE. In a makes-sense judgment task, they compared how comprehenders' judgments were affected by syntactic mismatch in sentences containing VPE (*surface anaphora*; Hankamer & Sag 1976, Sag & Hankamer 1984) or 'do it' (*deep anaphora*). They discovered that in VPE, pairs of sentences where the antecedent and elided material did not match in voice (14b-14c) or category (15b-14c) were judged to make sense a smaller proportion of times than when there was syntactic match (14a-14c), (15b-14c). On the other hand, the proportion of sentence pairs judged to make sense did not differ between deep anaphora sentences (14d, 15d) where the two sentences either matched or mismatched in voice (14a-14b) or category (15a-15b).

- (14) a. Someone had to take out the trash. ...
b. The trash had to be taken out. ...
c. ... But Bill refused to Δ .
d. ... But Bill refused to do it.
- (15) a. It always annoys Sally when anyone mentions her sister's name. ...
b. The mention of her sister's name always annoys Sally. ...
c. ... However, Tom did Δ anyway out of spite.
d. ... However, Tom did it anyway out of spite.

It appears that in the case of 'do it' anaphora, semantic match between the antecedent and anaphoric VPs is sufficient to license the anaphoric dependency—the 'do it' VP can pick up its meaning from the antecedent whether or not it matches syntactically. In contrast, VPE seems to be sensitive to the syntactic match between the elided material and the antecedent VP.

While these results establish that VPE can show sensitivity to syntactic mismatch, they don't say anything about conditions under which sequences like (14b)-(14c) are acceptable. In fact, Kehler (2002), Hardt (1993), Dalrymple et al. (1991), and others have pointed out that there are naturally-occurring examples of syntactic mismatch in VPE (5). The antecedent VPs in (5) all mismatch their respective elided VPs in voice.

Perhaps more importantly, Tanenhaus & Carlson (1990) provide no account of the gradedness of acceptability, though it seems that acceptability degrades as the extent of the structural mismatch increases. Arregui et al. (2006)'s proposal tries to capture the relationship between acceptability and degree of syntactic mismatch. They asked participants to rate the acceptability of sentences like (16), which were assumed to vary by how syntactically similar the antecedent was to the elided material.

- (16) a. None of the astronomers saw the comet, but John did Δ .

- b. Seeing the comet was nearly impossible, but John did Δ .
- c. The comet was nearly impossible to see, but John did Δ .
- d. The comet was nearly unseeable, but John did Δ .

Because they assume a strict syntactic identity requirement on VPE, all the sentences in (16) are ungrammatical for Arregui et al. (2006) except (16a). To comprehend such sentences, they propose that the comprehender applies special repair operations to the mismatching antecedent, which eventually yield an antecedent that meets the syntactic identity requirement. This addresses the graded acceptability problem by allowing certain ungrammatical sentences to be parsable with the help of repair operations; the greater the number of repair steps required to restore syntactic well-formedness, the less acceptable a sentence sounds.

While both Tanenhaus & Carlson (1990) and Arregui et al. (2006) provide experimental evidence that VPE is sensitive to syntactic non-identity in the form of voice and category mismatches, neither compares the critical ellipsis sentences to their no ellipsis counterparts. As such, the possibility that the observed contrasts are due to a general dispreference for non-parallel conjuncts (instead of an ellipsis-specific parallelism condition) cannot be excluded.

3. Experiment 1: Voice and Category mismatch in VPE

The first experiment looked at cases of VPE where the antecedent was not a perfect structural match for the elided VP. In critical conditions, the antecedent differed in either voice (active vs. passive) or category (nominal or adjectival vs. verbal) from the elided VP in the second conjunct. In addition, each Ellipsis condition had a corresponding No Ellipsis condition, to determine whether any syntactic mismatch effects were general non-parallelism effects, or specific to VPE licensing. The objective of Experiment 1 was to establish an empirical base that any model of VPE would minimally have to account for.

Based on previous experimental results (e.g. Tanenhaus & Carlson 1990, Arregui et al. 2006) we expect structural mismatch to result in degraded acceptability. However there are certain syntactic accounts of ellipsis, like Merchant (2007), that predict voice mismatch will not be subject to syntactic identity in VPE due to the low ellipsis height (compared to TP-level ellipsis). Under such an account, any decrease in acceptability in syntactic mismatch conditions will be unexplained.

If the mismatch cases are degraded relative to their syntactic match counterparts, there is a question of whether the syntactically more complex form will be easier to have as an antecedent to the syntactically simpler form, or vice versa. Chung (2006) proposes a “morphological subset” principle that states that each morpheme included in the ellipsis site must have a corresponding morpheme in the antecedent (but not vice versa). According to this principle, Active-Passive would be predicted *worse* than Passive-Active sequences, assuming that the elided constituent in a Passive VP contains more morphological features than the corresponding Active VP. Arregui et al. (2006) in fact find that Active-Passive is rated worse than Passive-Active; they suggest a memory-based explanation involving mis-remembering a mismatching antecedent as a matching one. This mis-remembering is

supposed to be easier when a Passive is being mis-remembered as an Active, than the other way around, presumably because the Active form is more basic.

Both Chung (2006) and Arregui et al. (2006) suggest a connection between the morphosyntactic complexity of the passive form and its relative unacceptability in the ellipsis site, though the explanations provided for this asymmetry are different. We know of no proposal that predicts an asymmetry in the reverse direction (Active-Passive predicted better than Passive-Active order).

3.1 Design

There were 8 cells in the experiment (Ellipsis (Ellipsis, No Ellipsis) x Mismatch (Match, Mismatch) x Mismatch Type (Voice, Category)). Half of the Voice mismatch cases had Active (antecedent)-Passive (ellipsis) order, and half had Passive-Active order. Half of the Category mismatches had a Nominal antecedent-VPE sequence, and half had an Adjectival antecedent-VPE sequence. Example sentences for the experimental conditions are given in Table 1.

		Match	Mismatch
No ellipsis	Voice	Abby invited Bill, and Jill invited him, too	Abby invited Bill, and Jill was invited by her, too
	Category	Abby criticized Bill, and Jill criticized him, too	The criticism of Bill was harsh, but Jill didn't criticize him
Ellipsis	Voice	Abby invited Bill, and Jill did, too	Abby invited Bill, and Bill was, too
	Category	Abby criticized Bill, and Jill did, too	The criticism of Bill was harsh, but Jill didn't

Table 1: Experiment 1 design and example stimuli.

3.2 Method and Procedure

The experimental paradigm used was magnitude estimation, adapted from e.g. Bard, Robertson & Sorace (1996). In this paradigm, participants give numerical ratings to stimuli relative to the rating they gave to some standard, or modulus, at the beginning of the experiment. For language stimuli, the ratings are participants' estimates of the acceptability of the sentence in the current trial compared to the acceptability of the modulus.

The current experiment was run on a Macintosh computer running PsyScope software. Participants first practiced giving estimates of line lengths (cf. Bard et al. 1996), then practiced with sentences. Then they assigned a value to the modulus sentence: *'The children were amused by the cartoon, but their parents weren't'*. On each trial, the modulus appeared on the screen together with the sentence to be rated on that trial. Participants typed their estimates into a text box, then pressed the spacebar to proceed to the next trial.

There was one break halfway through the trials, and the whole experiment took participants approximately 10-15 minutes. 20 native English speakers from the University of Rochester community participated.

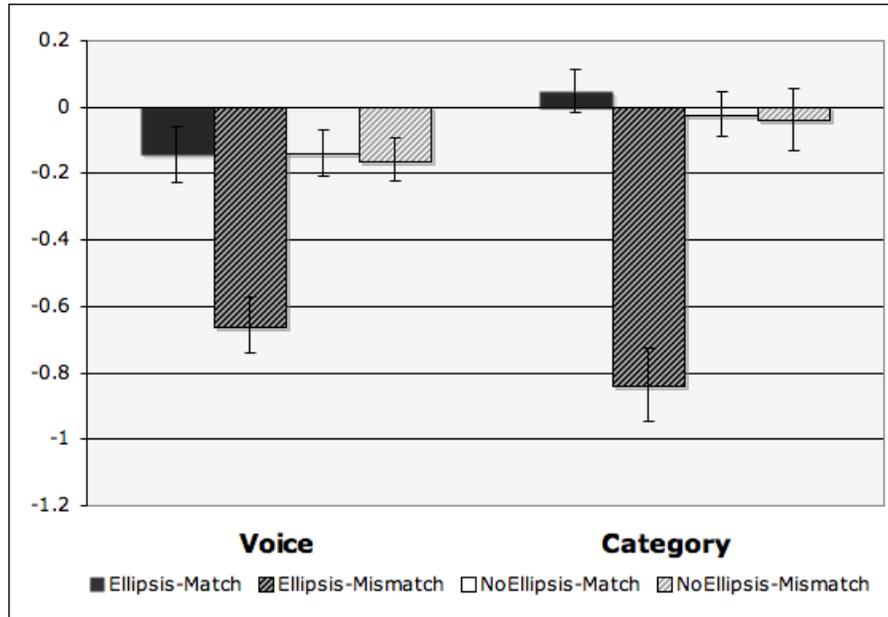


Figure 1: Experiment 1 results (error bars are Standard Error).

3.3 Results

Data analysis

The data were first normalized by dividing each participant's estimates by their modulus value. All analyses were performed on log-transformed values of the normalized data.

The log scores were fitted to a linear mixed-effects model, with three fixed effects (Ellipsis, Mismatch, Mismatch Type), and one random effect (Subject). This model was selected from among other linear models containing subsets of the experimental factors as fixed effects, by comparing the models' log likelihoods. The results of Experiment 1 are plotted in Figure 1.

Main effects and interactions

There were main effects of Ellipsis ($F(1, 19) = 51.0, p < .0001$), Mismatch ($F(1, 19) = 99.2, p < .0001$), and Mismatch Type ($F(1, 19) = 4.5, p < .05$). Ellipsis was worse than NoEllipsis ($t = 10.1, p < .0001$), and Mismatch was worse than Match ($t = 12.0, p < .0001$). In addition, Voice mismatches were worse than Category mismatches ($t = 1.9, p = .05$).

In addition, there was an Ellipsis-Mismatch interaction ($F(1, 19) = 61.7, p < .0001$), such that Mismatch was worse than Match only when there was Ellipsis (for Ellipsis conditions, Mismatch is worse than Match ($t = 16.4, p < .0001$); for NoEllipsis, Mismatch \approx Match ($t = .6, p = 1.0$)). Mean estimates for each factor are given in Table 2, and condition means in Table 3.

	Ellipsis No Ellipsis		Match Mismatch		Voice Category	
<i>mean</i>	-.4	-.09	-.42	-.06	-.27	-.21
<i>st.dev.</i>	.58	.33	.57	.32	.43	.56

Table 2: Experiment 1 main effects.

Voice	No Ellipsis/ Match	No Ellipsis/ Mismatch	Ellipsis/ Match	Ellipsis/ Mismatch
<i>mean</i>	-.14	-.16	-.14	-.66
<i>st.dev.</i>	.30	.29	.37	.47

Category	No Ellipsis/ Match	No Ellipsis/ Mismatch	Ellipsis/ Match	Ellipsis/ Mismatch
<i>mean</i>	-.02	-.04	-.05	-.84
<i>st.dev.</i>	.29	.40	.28	.63

Table 3: Experiment 1, mean log estimates by condition.

Comparisons by Mismatch Type

The Ellipsis trials for Voice and Category conditions were also analyzed separately in post hoc comparisons.

Within the Voice mismatch trials, the antecedent was either an active or a passive VP, yielding four possible active-passive orderings: Active-Active, Passive-Passive, Active-Passive, and Passive-Active. Active-Active sentences were rated better than Passive-Passive (corrected $p < .0001$), which were in turn better than both Active-Passive and Passive-Active (corrected $p < .0001$). The two mismatch conditions did not differ from each other (corrected $p = .29$). The results for Voice conditions only are in Figure 2.

For Category trials, the antecedent was an active VP in Match trials, and either a nominal or an adjectival phrase in Mismatch trials, yielding the following three orderings: VP-VP, NP-VP, and AP-VP. VP-VP sentences were better than NP-VP (corrected $p < .0001$), which were better than AP-VP (corrected $p < .05$). The results for Category conditions only are in Figure 3.

3.4 Discussion

Experiment 1 replicates the basic finding of Tanenhaus & Carlson (1990) and Arregui et al. (2006) that sentences with syntactic mismatch between the first and second conjuncts (or sentences) are judged less acceptable than sentences where conjuncts match syntactically. Importantly, this mismatch effect is only observed when there is ellipsis in the second conjunct; this suggests the sensitivity to syntactic parallelism reflects a constraint on VPE, not a general prohibition on syntactic mismatch in coordinate structures.

To revisit what is predicted by some existing syntactic accounts of ellipsis in light of

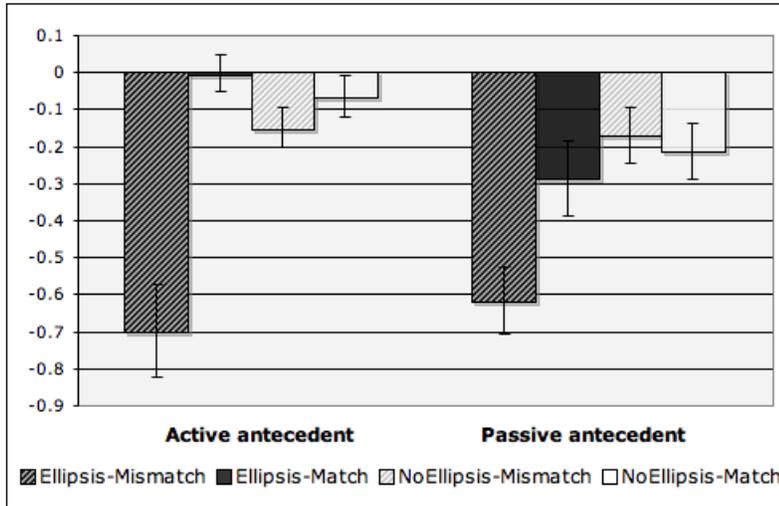


Figure 2: Experiment 1, Voice conditions. Mismatch and Match bars correspond to Active-Passive and Active-Active (left-hand side of graph), and Passive-Active and Passive-Passive (right-hand side).

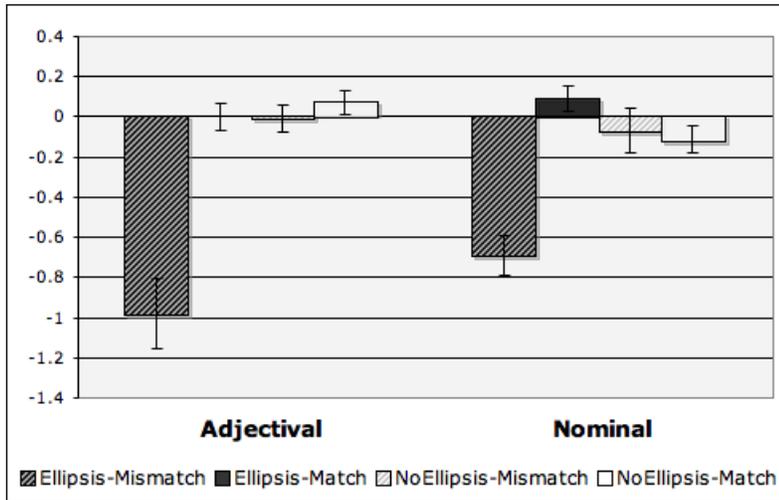


Figure 3: Experiment 1, Category conditions. Mismatch and Match bars correspond to AP-VP and VP-VP (left), and NP-VP and VP-VP (right).

the results from Experiment 1, for Merchant (2007, 2008), asymmetries between VP-level ellipsis and TP-level ellipsis is captured by ellipsis height. In the case of voice mismatch, whether a particular kind of ellipsis will permit it depends on whether the ellipsis applies to a constituent in the syntactic tree that includes the Voice head. If the ellipsis is high, as in sluicing or gapping, and Voice is part of the elided material, it is subject to the syntactic identity condition, and voice mismatch will not be possible. If the ellipsis is low enough that Voice is external to the ellipsis site, as in VPE, the value of the passive feature on the Voice head escapes the syntactic identity condition and can differ between the antecedent and ellipsis conjuncts. While the results of Experiment 1 clearly show that syntactic mismatch results in decreased acceptability, it would be interesting to know whether the strength of the syntactic mismatch effect is in general related to the size of the ellipsis site.

According to both Chung (2006)'s morphological subset principle and Arregui et al. (2006), Active-Passive order should have been rated less acceptable than Passive-Active order in sentences with ellipsis; more generally, the more morphologically or derivationally complex form was predicted to be worse in the ellipsis conjunct. While there was a numerical difference that appeared to be in this direction for voice mismatches (Active-Passive order less acceptable than Passive-Active), this difference was not significant. However for the category mismatches, the Adjective-VP sentences were rated worse than the Nominal-VP sentences. We return to this issue in the general discussion.

A possibility not excluded by Experiment 1 is that syntactic mismatch effects associated with VPE are limited to cases where the conjuncts are in specific discourse relations (Kehler 2000, 2002). In particular, the sensitivity to syntactic parallelism may be restricted to instances of the Resemblance relation. If this is the case, changing the coherence relation to one that doesn't depend on syntactic parallelism, like Cause-Effect, should eliminate any effects of mismatch. Experiment 2 tests this prediction by manipulating discourse coherence (Resemblance vs. Cause-Effect) in addition to syntactic mismatch.

4. Experiment 2: Syntactic effects are modulated by Discourse Coherence

Experiment 2 compares VPE sentences with voice mismatch, when the discourse relation between the antecedent and the ellipsis conjuncts was either Resemblance or Cause-Effect. According to Kehler (2000, 2002), the Resemblance relation relies on the alignment of arguments from one sentence (or conjunct) to the next, and as such should be sensitive to changes in syntactic parallelism. On the other hand, the Cause-Effect relation relates sentences at the propositional meaning level, and should therefore be insensitive to syntactic manipulations that leave the meaning contribution of each conjunct intact.

Frazier & Clifton (2006) have a similar set of experiments where they compare sentences with ellipsis where the relation between the antecedent and ellipsis conjuncts is either Resemblance or Cause-Effect. In both percent 'got it' data and acceptability ratings on a 1 to 5 scale, they failed to find the asymmetry predicted by Kehler: the Resemblance cases were no worse than the Cause-Effect cases. While they did show for the acceptability ratings that corresponding sentences with syntactically matching conjuncts did not differ—that is, the lack of difference in the mismatching sentences was not due to a differ-

ence in the acceptability of the matching forms—they did not show this for the no ellipsis counterparts. In addition, while they themselves (based on Arregui et al. (2006)) would expect mismatches with Active-Passive order to be more unacceptable than Passive-Active order, all the stimuli in Frazier & Clifton (2006) (Experiments 1 and 2) involved passive antecedents. The current experiment, in addition to testing Kehler’s predictions about syntactic parallelism, includes both syntactic match and no ellipsis conditions, as well as both active and passive antecedents.

4.1 Design

There were 8 cells in the experiment (Ellipsis (Ellipsis, No Ellipsis) x Mismatch (Match, Mismatch) x Discourse Relation (Resemblance, Cause-Effect)). Half of the Match trials had active VPs, and half had passive VPs. Similarly, half of the Mismatch trials have Active-Passive order, and the other half had Passive-Active order. As illustrated in the example stimuli in Table 4, the coherence relation was indicated by connectives like *and* or *but* for Resemblance, and *so* or *because* for Cause-Effect.

		Match	Mismatch
No ellipsis	Resemblance	Abby invited Bill, and Jill invited him, too	Abby invited Bill, and Jill was invited by her, too
	Cause-Effect	Abby invited Bill, so Jill didn’t invite him	Abby invited Bill, so Jill wasn’t invited by her
Ellipsis	Resemblance	Abby invited Bill, and Jill did, too	Abby invited Bill, and Jill was, too
	Cause-Effect	Abby invited Bill, so Jill did, too	Abby invited Bill, so Jill wasn’t

Table 4: Experiment 2 design and example stimuli.

4.2 Method and Procedure

The experimental paradigm and procedure were identical to Experiment 1, except that experimental trials were interspersed with filler sentences, which were either monoclausal or contained a discourse relation that did not appear in the test items.¹ 24 native English speakers from the Rochester community participated in the experiment. Each experimental session took approximately 15 minutes.

4.3 Results

Data analysis

As in Experiment 1, estimates were normalized and log-transformed. The log scores were fitted to a linear mixed-effects model with three fixed effects (Ellipsis, Mismatch, Discourse

¹Experiment 1, which included different types of category mismatches in addition to voice mismatches, had sufficient variation in the test items that we decided against including additional fillers.

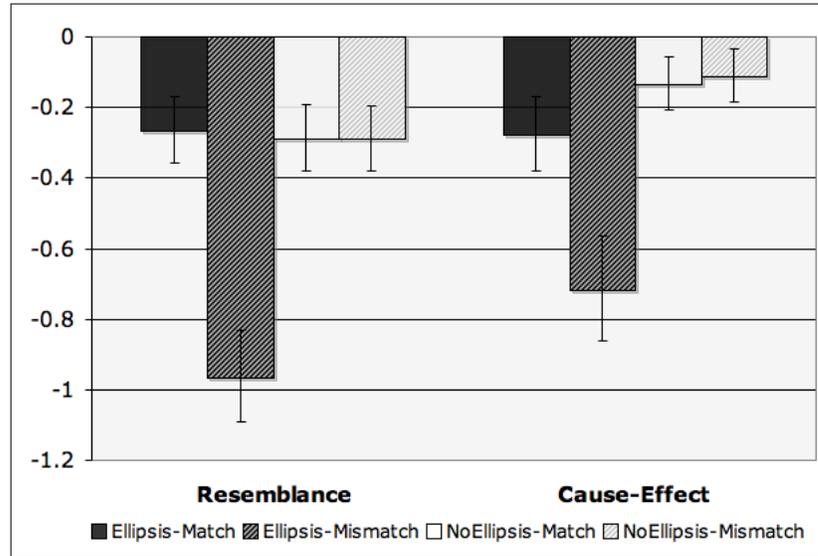


Figure 4: Experiment 2 results (error bars are Standard Error).

Relation), and one random effect (Subject). The results of Experiment 2 are plotted in Figure 4.

Main effects and interactions

There were main effects of Ellipsis ($F(1, 23) = 119.92, p < .0001$), Mismatch ($F(1, 23) = 73.06, p < .0001$), and Discourse Relation ($F(1, 23) = 22.69, p < .0001$). As in Experiment 1, Ellipsis was worse than No Ellipsis ($t = 7.38, p < .0001$), and Mismatch was worse than Match ($t = 5.61, p < .0001$). In addition, Resemblance was worse than Cause-Effect ($t = 1.95, p = .05$). Mean log estimates are given by factor in Table 5.

	Ellipsis	No Ellipsis	Match	Mismatch	Resemblance	Cause-Effect
<i>mean</i>	-.53	-.21	-.25	-.50	-.41	-.31
<i>st.dev.</i>	.33	.53	.35	.54	.46	.47

Table 5: Experiment 2 main effects.

Again as in Experiment 1, there was an Ellipsis-Mismatch interaction ($F(1, 23) = 102.91, p < .0001$), such that Mismatch was worse than Match for Ellipsis trials ($t = 8.58, p < .0001$), but not for No Ellipsis ($t = -.70, p = .49$). In addition, Mismatch interacted with Discourse relation ($F(1, 23) = 4.92, p < .03$)—the Mismatch effect was greater for Resemblance ($t = -5.4, p < .0001$) than for Cause-Effect ($t = -3.28, p = .001$). There was also a three-way Ellipsis-Mismatch-Discourse Relation interaction ($F(1, 23) = 3.95, p < 0.05$)—the Ellipsis-Mismatch interaction was stronger for Resemblance than for Cause-Effect. Condition means are given in Table 6.

Resemblance	No Ellipsis/ Match	No Ellipsis/ Mismatch	Ellipsis/ Match	Ellipsis/ Mismatch
<i>mean</i>	-.29	-.29	-.26	-.96
<i>st.dev.</i>	.36	.35	.49	.28
Cause-Effect	No Ellipsis/ Match	No Ellipsis/ Mismatch	Ellipsis/ Match	Ellipsis/ Mismatch
<i>mean</i>	-.13	-.11	-.28	-.71
<i>st.dev.</i>	.28	.28	.40	.55

Table 6: Experiment 2, mean log estimates by condition.

Active-Passive order

One question brought up by these results is why Frazier & Clifton (2006) failed to find any coherence-related effects on acceptability, in a similar study manipulating voice mismatch in VPE. Recall that all of the syntactic mismatch stimuli in Experiments 1 and 2 of Frazier & Clifton (2006) have Passive-Active order. In fact, when we analyzed only the subset of Passive-Active sentences in our data, there was a numerical but not significant advantage of Resemblance-Ellipsis over Cause-Effect-Ellipsis ($t = .92, p = .36$). This numerical difference still contributes to the main effect of Discourse Relation Type, but it's evident that this effect is being carried by the Active-Passive subset ($t = 2.32, p = .031$). Thus while we are still left with the interesting question of why reconstructing a passive from an active antecedent is more sensitive to the discourse context in which violations of syntactic identity occur, we may be able to reconcile the difference between the current findings and those reported in Frazier & Clifton (2006).

4.4 Discussion

Experiment 2 appears to be incompatible with a literal interpretation of Kehler (2000). But we interpret the interaction between discourse relation type and the size of the mismatch effect in the spirit of Kehler's proposal, by noting that the syntactic licensing condition on VPE seems to be more or less strictly enforced depending on the discourse context.²

The results of Experiments 1 and 2 together show that VPE is sensitive to syntactic parallelism, but that the extent of this sensitivity is modulated by the type of discourse coherence relation the antecedent and ellipsis conjuncts are part of. We note, however,

²While we characterize the results of Experiment 2 in terms of Kehler's theory of discourse coherence (i.e. what we intended to manipulate), the current data do not allow us to distinguish this from Frazier & Clifton (2006)'s alternative hypothesis, which invokes a notion of parallelism distinct from the parallelism introduced by a Resemblance coherence relation. Specifically, they suggest that the presupposition introduced by the sentence-final 'too' in many Resemblance sentences is a source of parallelism effects in VPE. Since information carried by the connective and 'too' become available at different locations in the sentence, future experiments using online measures may help separate out the timecourses of these potentially different sources of parallelism.

that Kehler's theory of coherence types was intended to characterize relations between sentences, not two conjuncts in a coordinate structure. While it seems an intuitive extension of Kehler's original proposal to consider coherence relations between clauses within a sentence,³ there may be reasons not to do so.

A typical assumption in theoretical syntax is that core syntactic operations and principles are confined to the sentence domain—for instance, when considering possibilities for linguistic coreference, candidate referents outside the current sentence are not thought of as being in a c-command relationship with anaphors inside the sentence. Instead, discourse structural considerations are often assumed to play a more important role in relating sentences to each other in terms of their informational organization. Under this view, while the internal syntactic structure of one sentence may not affect the interpretation of subsequent sentences, it may contribute to the discourse structure by establishing what is given, focused, the topic or question under discussion. This in turn influences subsequent interpretation.

This view is taken up by Frazier & Clifton (2005), who claim on the basis of acceptability experiments that syntactic effects are limited in their domain to the sentence; once outside the sentence domain, a different set of interpretive constraints (based on extracting the main assertion from a sentence) applies at the level of discourse structure. They compared sentences like (17a), where an elided VP and its antecedent are in a single sentence with coordination, with pairs of sentences like (17b), where the antecedent VP is in one sentence and an elided VP is in another.

- (17) a. John said that Fred went to Europe and Mary did, too.
b. John said that Fred went to Europe. Mary did, too.

Frazier & Clifton (2005) find that in the coordination condition, people are more likely to construe the 'go to Europe' VP as the antecedent of the elided VP, while in the two-sentence condition, they are less likely to consider it the antecedent.

If Frazier and Clifton are right that syntactic effects are confined to the sentence domain because comprehenders do not retain detailed syntactic representations across sentences, the syntactic mismatch effects observed in Experiments 1 and 2 are predicted to go away if the dependency between an elided VP and its antecedent crosses a sentence boundary. But following the basic intuition in Kehler's work—that the relations between meanings in a discourse context influence whether structural or semantic information is important for interpreting ellipsis given an antecedent—we might expect such discourse-modulation to extend straightforwardly from clause sequences to sentence sequences. Experiment 3 tests this prediction.

5. Experiment 3: Syntactic effects above the 'syntactic' domain

Experiment 3 compares cases of VPE in a coordinate structure (Coordination) with VPE across a sentence boundary (Cross-sentential). Antecedent and ellipsis clauses were related

³See e.g. Rohde, Levy & Kehler (2008), who show that discourse coherence between clauses of a sentence can influence syntactic attachment preferences for a relative clause in the same sentence.

by Resemblance or Cause-Effect, as in Experiment 2. The Coordination conditions are identical in structure to Experiment 1 (Voice conditions) and Experiment 2; as such we expect to see the same Ellipsis-Mismatch interaction pattern in the current experiment. The critical question is whether there is an effect of syntactic mismatch when the antecedent of VPE is in the sentence preceding the sentence with ellipsis.

5.1 Design

There were 8 cells in the experiment (Ellipsis Type (Coordination, Cross-sentential) x Discourse Relation (Resemblance, Cause-Effect) x Mismatch (Match, Mismatch)). All of the Mismatch trials had voice mismatch; half of these had Active-Passive order, and half had Passive-Active order. All sentences contained ellipsis. The design of Experiment 3 and example stimuli are given in Table 7.

		Match	Mismatch
Coordination	Resemblance	Abby invited Bill, but Jill didn't.	Abby invited Bill, but Jill wasn't.
	Cause-Effect	Abby invited Bill, so Jill did, too.	Abby invited Bill, so Jill was, too.
Cross-sentential	Resemblance	Abby invited Bill. Jill did, too.	Abby invited Bill. Jill was, too.
	Cause-Effect	Abby invited Bill. So Jill did, too.	Abby invited Bill. So Jill was, too.

Table 7: Experiment 3 design and example stimuli.

5.2 Method and Procedure

The procedure and method were identical to Experiments 1 and 2. Approximately half of the practice and filler items contained two sentences. 14 native English speakers from the Rochester community participated. Each experimental session took approximately 15 minutes.

5.3 Results

Data analysis

As before, data were normalized and log-transformed. Log scores were fitted to a linear mixed-effects model with three fixed effects (Ellipsis Type, Discourse Relation, Mismatch) and one random effect (Subject).

Main effects and interactions

As in the previous two experiments, there was a main effect of Mismatch ($F(1, 13) = 187.69, p < .0001$): Mismatch was worse than Match ($t = 11.87, p < .0001$). There was

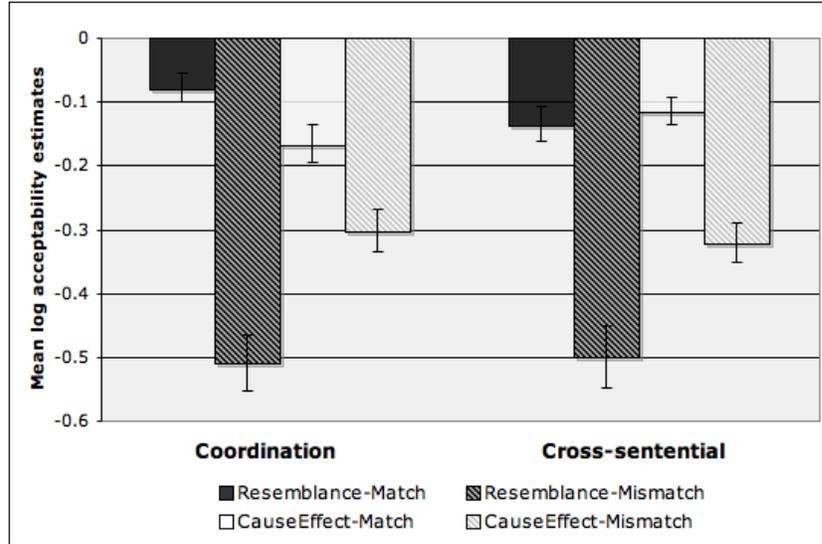


Figure 5: Experiment 3 results (error bars are Standard Error).

also a main effect of Discourse Relation ($F(1, 13) = 14.56, p = .0001$), as in Experiment 2: Resemblance was worse than Cause-Effect ($t = 3.08, p < .005$). There was no main effect of Ellipsis Type ($F(1, 13) = .034, p = .85$): VPE in Coordination did not differ from Cross-sentential VPE.

	Match	Mismatch	Resemblance	Cause-Effect	Coord.	Cross-sentence
<i>mean</i>	-.28	-.94	-.70	-.52	-.61	-.62
<i>st.dev.</i>	.53	.86	.90	.64	.79	.79

Table 8: Experiment 3 main effects.

There was a Discourse Relation-Mismatch interaction ($F(1, 13) = 30.25, p < .0001$), as in the previous experiment. VPE was more degraded under Resemblance than Cause-Effect, when there was Mismatch between the two clauses ($t = 4.89, p < .0001$), but not when there was Match ($t = 1.37, p = .17$). No two-way or three-way interactions involving Ellipsis Type were significant. In other words, whether VPE occurred within a single sentence or across two sentences had no effect on acceptability. The data are plotted in Figure 5. Mean scores are given by factor in Table 8; condition means are in Table 9.

Active-Passive order

We again separately analyzed the subset of Mismatch trials by Active-Passive order; the Mismatch data were fit to a linear mixed-effects model with Active-Passive order and Discourse Relation as fixed effects, and Subject as a random effect. As in Experiment 2, there was an effect of order: Active-Passive order was worse than Passive-Active order

Coordination	Resemblance/ Match	Resemblance/ Mismatch	Cause-Effect/ Match	Cause-Effect/ Mismatch
<i>mean</i>	-.18	-1.17	-.38	-.69
<i>st.dev.</i>	.46	.92	.62	.68
Cross-sentential	Resemblance/ Match	Resemblance/ Mismatch	Cause-Effect/ Match	Cause-Effect/ Mismatch
<i>mean</i>	-.31	-1.15	-.26	-.74
<i>st.dev.</i>	.55	1.03	.45	.65

Table 9: Experiment 3, mean log estimates by condition.

when there was Mismatch ($F(1, 13) = 3.87, p < .01$). Interestingly, this order effect interacted with Discourse Relation ($F(1, 13) = 13.04, p < .0001$): Active-Passive order showed greater degradation under Resemblance relative to Cause-Effect (mean difference = .53), compared to Passive-Active order (mean difference = .36).

5.4 Discussion

Experiment 3 finds that the modulation of structural mismatch by discourse coherence when interpreting VPE (Experiment 2) extends to cases where the antecedent-ellipsis dependency is across sentences. This result argues against a syntax/discourse divide as proposed by Frazier & Clifton (2005): the Ellipsis-Mismatch interaction was unaffected by whether antecedent and ellipsis site were in the same sentence, or separated by a sentence boundary.

The fact that we find syntactic mismatch degrades VPE even across sentences seems to indicate that we have access to syntactic information at the level of discourse: as a sentence containing VPE is interpreted, the antecedent-ellipsis relation is constrained by syntactic parallelism irrespective of whether the antecedent is part of the same syntactic structure or a different structure. If sentence interpretation were limited in such a way that only one unit of syntactic structure could be attended to at a time, there would be no way to compare the structure of an antecedent in a previous sentence to structure in the current sentence. In fact, it appears that any view of the relationship between syntactic and discourse structures where discourse representations contain no or very impoverished syntactic information will be unable to explain how syntactic identity can be enforced across discourse.

6. General discussion

6.1 Reconstructing elided VPs under syntactic identity

The results from Experiments 1-3 (and previous work by Tanenhaus & Carlson (1990) and Arregui et al. (2006), and work in the syntax literature (e.g. Kennedy 2003; Johnson

2001; Merchant 2007, 2008) show that there has to be a syntactic licensing condition on VPE. But even if we accept that there is a syntactic identity condition on VPE, there is the further question of what we assume happens when there fails to be perfect syntactic match between the antecedent and elided VPs. Since we observe that even in the face of syntactic mismatch, comprehenders are able to understand what a sentence means, there must be a way to parse and interpret such a sentence that involves a mechanism for restoring or preserving syntactic identity.

Arregui et al. (2006)'s proposal tries to capture the relationship between acceptability and degree of syntactic mismatch. Because they assume a strict syntactic identity requirement on VPE, all the sentences in (16) are ungrammatical for Arregui et al. (2006) except (16a). To comprehend such sentences, they propose that the comprehender applies special repair operations to the mismatching antecedent, which eventually yield an antecedent that meets the syntactic identity requirement. This addresses the graded acceptability problem by allowing certain ungrammatical sentences to be parsable with the help of repair operations; the greater the number of repair steps required to restore syntactic well-formedness, the less acceptable a sentence sounds.

In contrast, Kobele, Kim, Hale & Runner (2008) assume that cases of syntactic mismatch are grammatical in that they are generated by the grammar. Instead, they link degradations in acceptability to the size of the elided constituent in the derivation tree. Syntactic mismatch will generally have the consequence that a smaller subpart of the antecedent and elided VP structures will be identical; this, combined with a parsing preference in the spirit of MaxElide (cf. Merchant 2008, Takahashi & Fox 2005) that prefers larger constituents to be elided, predicts the graded pattern of acceptability observed, with greater degrees of mismatch yielding greater degradation. A key difference from the approach in Arregui et al. (2006) is that syntactically non-parallel structures can be generated in a normal way, and as such do not have to invoke anything outside of the usual mechanisms for computing meanings from syntactic structures.

The empirical predictions of both repair and the ellipsis size accounts will ultimately depend on the grammar assumed to generate the structures in question. (Note that Arregui et al. (2006) claim all repair operations must correspond to existing syntactic operations specified by the grammar.) But in light of the findings to date, we seem to have at least one argument in favor of a repair-based account, and one in favor of a non-repair account.

One reason to prefer a repair account of VPE is that it provides a way to make sense of the asymmetry between Active-Passive and Passive-Active mismatches. Arregui et al. (2006), Kobele et al. (2008), and the results we report here all show that sentences with Active-Passive order are more degraded due to syntactic mismatch than their Passive-Active counterparts. Indeed, Chung (2006) and the Recycling Hypothesis proposed in Arregui et al. (2006) predict this asymmetry, since Active-Passive sentences require reconstructing a matching Passive structure from an Active antecedent. This draws on the observation that active syntax is more *basic* than passive syntax; under some analyses, a passive structure contains its corresponding active structure. While this is also true of the grammar assumed in Kobele et al. (2008), the absence of a repair process leaves no source for the observed mismatch asymmetry.

On the other hand, there appear to be no easy adjustments to repair accounts that would enable them to account for the discourse coherence effects we report here and in Kobele et al. (2008). From the point of view of Kobele et al. (2008), the discourse context that an instance of VPE appears in influences the strength of the ellipsis size constraint; since sentences containing syntactic mismatch in VPE are generated by the grammar, there is no relationship between the extent to which this constraint is satisfied and the interpretability of the sentence. It is less obvious why changing the discourse context would affect the number of repair steps needed to restore syntactic identity (Arregui et al. 2006), and since syntactic identity is assumed to be strict, presumably VPE can only be interpreted after the full sequence of repair operations has been completed.

6.2 The relationship between syntax and the discourse context

A possible alternative explanation of the mismatch effects reported here that does not involve a syntactic identity condition is that what we are calling syntactic mismatch effects are really discourse structural mismatch. Since changes to syntactic structure (active vs. passive, nominal or adjectival vs. verbal) are likely to have corresponding effects on discourses (by means of changes in information structure, or predication structure), any of the structural mismatch effects reported here can be interpreted as discourse effects which are only indirectly related to syntactic structure. In fact, such an alternative is considered by Tanenhaus & Carlson (1990) from the point of view of comparing deep and surface anaphora.

It is difficult to pull apart effects of syntactic and discourse structure when they co-occur as they do here, but we note that whatever level of structure is responsible for the sensitivity to voice (or category) mismatches in VPE, it must be at least structurally rich enough that actives and passives (or nominalizations, deverbal adjectives, and verb phrases) have distinct representations. While the discourse might encode structure in other ways additional to the kinds of structural representations that distinguish actives from passives, we consider it plausible that the mismatch effects in Experiment 3 are syntactic effects, meaning that the discourse structure must also include syntactic structural representations. What we'll need in future work is a model of the discourse context that allows us to represent coherence relations, focus structure, and enough syntactic structure to capture the kinds of effects reported in this paper.

6.3 Interpreting VPE in discourses

Other issues that are outside the scope of this paper come up particularly when considering ellipsis across sentences. One is that there may be a locality bias for antecedents of ellipsis in situations where there is more than one candidate antecedent. An example of this 'local antecedent' bias might be (18). But note that the preference for the local antecedent ('*apply for a job in Taipei*') seems to be reduced in (19), which differs minimally from (18) in that the matrix verb has a more clearly eventive interpretation.

(18) I considered applying for a job in Taipei after Sameer did \triangle .

(19) I decided to apply for a job in Taipei after Sameer did Δ .

While the factors that influence the choice of antecedent in multiple antecedent situations should be investigated further, it's easy to imagine how such a bias might interact with syntactic mismatch effects. To take an example, consider (20a); to the extent that it's easier to interpret it with the local antecedent, the local bias seems to reverse in (20b), where the local antecedent is a nominalization, but the non-local one is a normal VP.

- (20) a. I thought about complaining to the registrar's office after Sameer did Δ .
b. I thought about a complaint to the registrar's office after Sameer did Δ .

Another discourse-level consideration that affects the strength of structural parallelism constraints is information structure: Kertz (2008) shows that, in addition to discourse coherence relations modulating lower level syntactic effects, the focus or information structure of a sequence of sentences may affect the relative ease of resolving ellipsis. Unlike the case of coherence relations, information structural effects found by Kertz appear to operate independently of ellipsis—in other words, there is a general preference for parallelism of coordinate structures or pairs of sentences with respect to focus structure. We leave to future research the interesting question of what the relationship is between various parallelism effects that result from general discourse pressures as opposed to narrower constraints specific to particular constructions or linguistic contexts.

References

- Arregui, Ana, Charles Clifton, Lyn Frazier & Keir Moulton (2006). Processing elided verb phrases with flawed antecedents: the recycling hypothesis. Journal of Memory and Language 55, pp. 232–246.
- Bard, Ellen Gurman, Dan Robertson & Antonella Sorace (1996). Magnitude estimation of linguistic acceptability. Language 72:1, pp. 32–68.
- Chung, Sandra (2006). Sluicing and the lexicon: The point of no return. Cover, R. T. & Y. Kim (eds.), Proceedings of BLS 31: General Session and Parasession on Prosodic Variation and Change, Berkeley Linguistics Society, Berkeley, CA, pp. 73–91.
- Dalrymple, Mary, Stuart M. Shieber & Fernando C.N. Pereira (1991). Ellipsis and higher-order unification. Linguistics and Philosophy 14, pp. 399–452.
- Frazier, Lyn & Charles Clifton (2005). The syntax-discourse divide: processing ellipsis. Syntax 8:2, pp. 121–174.
- Frazier, Lyn & Charles Clifton (2006). Ellipsis and discourse coherence. Linguistics and Philosophy 29, pp. 319–346.
- Haik, Isabel (1987). Bound VPs that need to be. Linguistics and Philosophy 10, pp. 503–530.
- Hankamer, Jorge (1979). Deletion in Coordinate Structures. Garland, New York.
- Hankamer, Jorge & Ivan Sag (1976). Deep and surface anaphora. Linguistic Inquiry 7, pp. 391–428.

- Hardt, Daniel (1993). Verb phrase ellipsis: Form, meaning, and processing. Ph.D. thesis, University of Pennsylvania.
- Johnson, Kyle (2001). What VP ellipsis can do, and what it can't, but not why. Baltin, M. & C. Collins (eds.), The Handbook of Contemporary Syntactic Theory, Blackwell, pp. 439–479.
- Kehler, Andrew (1995). Interpreting cohesive forms in the context of discourse inference. Ph.D. thesis, Harvard University.
- Kehler, Andrew (2000). Coherence and the resolution of ellipsis. Linguistics and Philosophy 23:6, pp. 533–575.
- Kehler, Andrew (2002). Coherence, reference and the theory of grammar. CSLI Publications.
- Kennedy, Chris (2003). Ellipsis and syntactic representation. The Syntax-Semantics Interface: Interpreting (Omitted) Structure, John Benjamins.
- Kennedy, Chris & Jason Merchant (2000). Attributive comparative deletion. Natural Language and Linguistic Theory 18:1, pp. 89–146.
- Kertz, Laura (2008). Verb phrase ellipsis in context. Annual Meeting of the LSA, Chicago.
- Kobele, Gregory, Christina Kim, John Hale & Jeffrey Runner (2008). A processing model of ungrammatical VP ellipsis. Talk presented at 21st CUNY Conference on Human Sentence Processing, Chapel Hill.
- Kobele, Gregory M. (2006). Generating Copies: An investigation into Structural Identity in Language and Grammar. Ph.D. thesis, UCLA.
- Merchant, Jason (2007). Voice and ellipsis. Ms, University of Chicago.
- Merchant, Jason (2008). An asymmetry in voice mismatches in VP-ellipsis and pseudo-gapping. Linguistic Inquiry 39:1, pp. 169–179.
- Rohde, Hannah, Roger Levy & Andrew Kehler (2008). Implicit causality biases influence relative clause attachment. Poster presented at 21st CUNY Conference on Human Sentence Processing, Chapel Hill.
- Sag, Ivan (1976). Deletion and Logical Form. Ph.D. thesis, MIT.
- Sag, Ivan & Jorge Hankamer (1984). Toward a theory of anaphoric processing. Linguistics and Philosophy 7, pp. 325–345.
- Takahashi, Shoichi & Danny Fox (2005). MaxElide and the Re-binding Problem. Georgala, E. & J. Howell (eds.), Proceedings of SALT 15, CLC Publications, Ithaca, NY, pp. 223–240.
- Tancredi, Christopher (1992). Deletion, deaccenting, and presupposition. Ph.D. thesis, MIT.
- Tanenhaus, Michael & Greg Carlson (1990). Comprehension of deep and surface verb phrase anaphors. Language and Cognitive Processes 5, pp. 257–280.
- Wilder, Chris (1995). Some properties of ellipsis in coordination. Geneva Generative Papers 2, pp. 23–68.
- Williams, Edwin (1977). Discourse and logical form. Linguistic Inquiry 8, pp. 101–139.