

A BRIEF REPORT ON VOICE MISMATCH EFFECTS IN VERB PHRASE ELLIPSIS AND SLUICING*

JEFFREY T. RUNNER
TIMOTHY A. DOZAT
University of Rochester

1 Introduction

The research reported here is part of a larger project investigating similarities and differences among varieties of ellipsis constructions, including VP ellipsis, Sluicing, Pseudogapping, Gapping and NP ellipsis. Merchant (2008), and more recently Tanaka (2011), have claimed that while VP ellipsis (VPE) can tolerate a mismatch in voice between antecedent and ellipsis site, Pseudogapping and Sluicing cannot. Merchant derives this from the claim that the [VOICE] feature is within the Pseudogapping and Sluicing ellipsis sites but is above the VPE ellipsis site, so the value of [VOICE] is relevant to the former but not the latter. We present the results of a Magnitude Estimation experiment (Bard et al., 1996) examining this prediction in VP Ellipsis (1) and Sluicing (2).

- (1) Cindi read *War & Peace*, and Sam did, too.
- (2) Someone read *War & Peace*, but I don't know who.

Our results suggest that participants actually judge Sluicing mismatch examples better than VP Ellipsis mismatch examples. However, a new, interesting finding is that voice mismatch has an overall effect on Sluicing-type constructions, with and without ellipsis; in VP ellipsis-type constructions, voice mismatch degrades only actual ellipsis sentences, not no-ellipsis controls, suggesting that voice parallelism is relevant to different constructions at different levels.

2 Background

Based on the observation that voice mismatch seems to be tolerated under certain circumstances in VP Ellipsis (as in the often cited example, *This problem was to have been looked into, but obviously nobody did*, from Kehler, 2000), while in other ellipsis constructions it seems not to, such as Pseudogapping and Sluicing, Merchant (2008) proposes that this difference can be derived

*Thanks go to the summer research assistants in the Runner lab: Quinlan Mitchell, Seth Rosenblatt and Christian Soto; and particular thanks to Christina Kim for her help with the materials, PsyScope and R.

from the “height” of the ellipsis site. If an elided constituent contains the [VOICE] feature, then voice mismatch effects will appear, on the assumption that the material in the elided phrase must be structurally parallel to an antecedent phrase. If it is indeed the case that voice mismatch effects differ based on whether [VOICE] is included in the elided constituent, this would be another important step towards understanding properties of the range of ellipsis constructions. We focus on comparing VP Ellipsis to Sluicing in this report (Kim and Runner, 2011 and Runner and Kim, 2011 report studies examining this prediction in pseudogapping).

Merchant (2008)’s analysis of the VP Ellipsis example in (3a) is illustrated in (3b-3c). In the antecedent clause (3b), the antecedent VP_A [read *War & Peace*] is below the vP, which contains the [VOICE] feature with the value [PASS]. The clause in (3c) contains the [VOICE] feature with the value [ACT], but the elided VP_E, indicated by angle brackets, is below the phrase containing the [VOICE] feature. On the assumption that the trace of the passive subject does not interfere, the two VPs match in the relevant features and the voice mismatch example can be derived. This is in contrast to the derivation of the Sluicing example in (4). Sluicing elides an entire TP. The TP_A in (4b) contains the [VOICE] feature with the value [ACT]. This entire TP is the antecedent of the elided TP_E in (4c). The problem arises because the elided TP contains a [VOICE] feature with the value [PASS]. Thus, there is no possibility of satisfying the structural parallelism requirement on ellipsis in the Sluicing mismatch case.

- (3) a. Cindi read *War & Peace*, and *Othello* was, too.
 b. Cindi [_{VP}v[VOI:ACT] [_{VP_A} read *War & Peace*]], and
 c. *Othello* was [_{VP}v[VOI:PASS] < [_{VP_E} read *t_{NP}*] >], too
- (4) a. Someone read *War & Peace*, but I don’t know by whom.
 b. [_{TP_A} Someone [_{VP}v[VOI:ACT] [_{VP}read *War & Peace*]]], but
 c. I don’t know by whom < [_{TP_E} *War & Peace* was [_{VP}v[VOI:PASS] [_{VP_E} read *t_{NP} t_{PP}*]] >

We tested this basic prediction, that voice mismatch would be better tolerated by VP Ellipsis than by Sluicing.

3 Experiment

3.1 Design and Materials

We conducted a magnitude estimation (Bard et al., 1996) experiment (n=20) to test the voice mismatch effects in VP Ellipsis and Sluicing. The experiment had two factors each with two levels: Construction type (VP Ellipsis, Sluicing) and (voice) Match (Match, Mismatch), for 4 conditions, plus their No-Ellipsis counterparts. Figure 1 and the corresponding examples provide sample materials. No-Ellipsis conditions include the material in strike out; the Ellipsis conditions do not.

We created twenty versions of sentences like those in (5)-(8), which were divided among 8 lists, each of which was in a different pseudorandomized order. Fillers consisted of sentences of similar length to the experimental items, and ranged between being (what we expected would be) somewhat better than the experimental items to somewhat worse. This was done so that the experimental materials did not stand out as being particularly bad. To guard against floor effects, our reference sentence (modulus) was a degraded sentence itself (*the woman accused of witchcraft*

Figure 1: Experimental design

| Ellipsis | Construction | Match | Mismatch |
|-------------|--------------|-------|----------|
| No-Ellipsis | VPE | 5a | 6a |
| | | 5b | 6b |
| | Sluicing | 7a | 8a |
| | | 7b | 8b |
| Ellipsis | VPE | 5a | 6a |
| | | 5b | 6b |
| | Sluicing | 7a | 8a |
| | | 7b | 8b |

- (5) a. Cindi read *The Once and Future King*, and Elizabeth did ~~read it~~, too.
 b. *The Once and Future King* was read by Cindi, and *The Lord of the Rings* was ~~read by her~~, too.
- (6) a. Cindi read *The Once and Future King*, and *The Lord of the Rings* was ~~read by her~~, too.
 b. *The Once and Future King* was read by Cindi, and Elizabeth did ~~read it~~, too.
- (7) a. Someone read *The Once and Future King*, but I don't know who ~~read it~~.
 b. *The Once and Future King* was read by someone, but I don't know by whom ~~it was read~~.
- (8) a. Someone read *The Once and Future King*, but I don't know by whom ~~it was read~~.
 b. *The Once and Future King* was read by someone, but I don't know who ~~read it~~.

sank, but I don't know what she sank), and we assigned the reference sentence a fixed score of 100. Previous research has suggested that both having a modulus that is of much higher acceptability dilutes contrasts among less acceptable experimental items, and participants who choose a modulus score that is too low can have trouble assigning contrasts to less acceptable experimental items.

3.2 Results

We modeled participants' (logged normalized) acceptability scores using mixed-model logistic regression, with Construction (VPE, Sluicing), Match (Match, Mismatch) and Ellipsis (Ellipsis, No Ellipsis) as fixed effects, and subject, item and list as random effects. There was one main effect, the effect of Construction ($p < 0.01$): VP Ellipsis-type examples were worse than Sluicing-type ones. This main effect was modulated by a series of interactions. There was a significant ($p < 0.0001$) interaction between Ellipsis and Match, showing that mismatch effects were worse in Ellipsis clauses than in No-Ellipsis controls; this replicates Kim and Runner (2009) and Kim and Runner (2009).

There was also a significant ($p < 0.01$) interaction between Ellipsis and Construction. As Figure 2 illustrates, VP Ellipsis and Sluicing match examples were judged similarly good, while mismatch had a larger effect on VP Ellipsis than on Sluicing. This interaction is unexpected, since the claim of Merchant (2008) is that mismatch effects in VP Ellipsis should be less severe than those in Sluicing; we will come back to this point below. In addition, there was also a significant ($p < 0.03$)

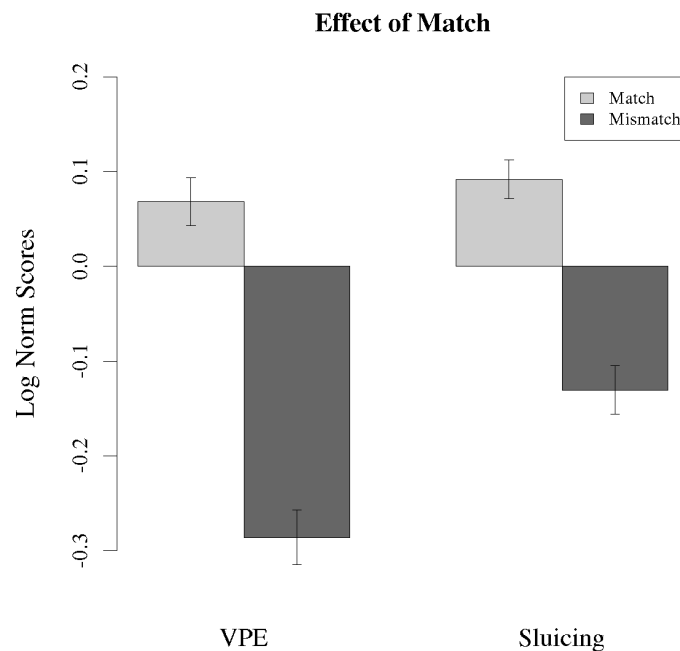


Figure 2: Logged normalized acceptability scores on voice match and mismatch trials in VP Ellipsis and Sluicing ellipsis examples (error bars represent Standard Error).

interaction between Construction and Match, meaning that the mismatch effect was more severe in VP Ellipsis-type sentences than in Sluicing-type structures.

However, before drawing conclusions from those interactions, we report that these effects are modulated by a significant ($p < 0.0001$) three-way interaction among all three variables: Ellipsis, Construction, and Match, illustrated in Figure 3. This interaction reveals that the mismatch penalty for VP Ellipsis is restricted to ellipsis clauses, while the mismatch penalty for Sluicing is the same for both ellipsis and no-ellipsis controls. In the figure that means that the two darkest bars on the left (VP Ellipsis) do not differ, while the two darkest bars on the right (Sluicing) do: this is the effect of mismatch on sentences not containing elided material. What this means is that there is a significant effect of mismatch on Sluicing-types of examples lacking ellipsis. The lighter bars in the graph are simply the data in Figure 3: the effect of mismatch on VP Ellipsis (left) and Sluicing ellipsis (right).

As found by Kim et al. (2011) and Kim and Runner (2009), the effect of mismatch is localized to the ellipsis clauses in VP Ellipsis. What is new is that the effect of mismatch on Sluicing is the same for both No-Ellipsis and Ellipsis clauses (i.e. the difference between the two darker bars on the right is the same as the difference between the two lighter bars on the right). That is, voice mismatch degrades only ellipsis clauses in VP Ellipsis, while voice mismatch degrades both Sluicing and Sluicing no-ellipsis controls equally.

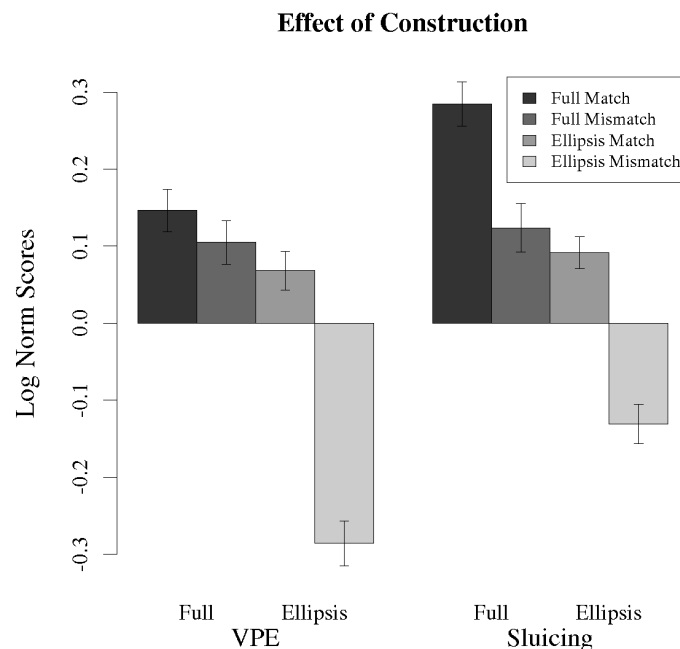


Figure 3: Logged normalized acceptability scores on voice match and mismatch trials in VP and sluicing examples comparing ellipsis and no-ellipsis (error bars represent Standard Error).

3.3 Subanalyses

We built into the design of the study several other comparisons of interest. We will discuss those now.

3.3.1 Chung's generalization

Among the types of VPs included in the VP Ellipsis and Sluicing-type materials, we included some containing PPs, thus allowing us to test Chung (2006)'s generalization about a restriction on Sprouting in Sluicing. Chung (2006) observes that Sprouting of a PP is acceptable as long as the PP is pied-piped with the *wh*-phrase preceding the sluice, as illustrated in (9a) and (9b).

- (9) a. Cindy talked with Billy but I don't know about what ~~Cindy talked *t* with Billy~~.
 b. *Cindy talked with Billy but I don't know what ~~Cindy talked about *t* with Billy~~.

Chung's proposal was the following constraint on Sluicing. Since in (9b) the 'about' in the sluice does not also appear in the antecedent CP, the sentence is predicted to be unacceptable.

- (10) Every lexical item in the numeration of the sluice that ends up (only) in the elided IP must be identical to an item in the numeration of the antecedent CP.

To test the robustness of this condition, we included pairs of sentences similar to (9a) and (9b). As illustrated in Figure 4, voice matched examples similar to (9a) are significantly better ($p < 0.04$)

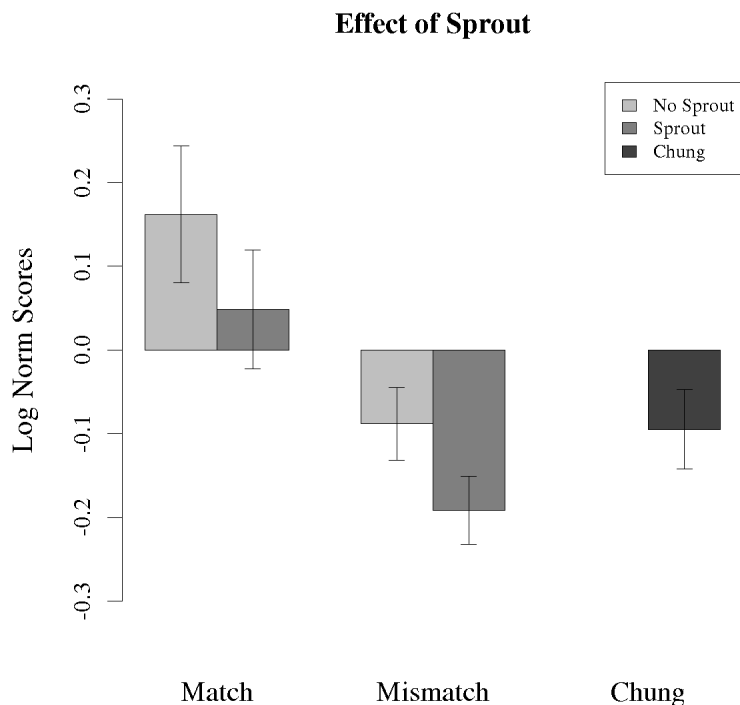


Figure 4: Logged normalized acceptability scores of voice matched and mismatched sluicing ellipsis sentences with and without violations of Chung’s condition (10) (error bars represent Standard Error).

than those that violate (10) (Match/Sprout bar vs. Chung bar in Figure). As a comparison, Figure 4 also shows the basic voice mismatch judgments, which do not differ ($p > 0.9$) from those that violate (10) (NoSprout/Mismatch vs. Chung bar in Figure). In addition, we included examples which contained both mismatch and a violation of Chung’s (10), which are significantly worse ($p < 0.01$), suggesting the violations are additive (Mismatch/Sprout bar in Figure).

The role of (10) in the grammar is not entirely clear. Though Chung seems to envision it as a constraint on the syntactic derivation it also seems possible to think of it as following from the notion of recoverability of deletion, since if the preposition itself is not pied-piped and included in the antecedent CP, its identity cannot be recovered. Processing may also play a role in the degradation, since upon encountering the NP *wh*-phrase, the parser may begin to construct a sluiced TP containing a NP *wh*-gap, which will not be possible given the choice of verb. This kind of predictive structure building in Sluicing has been observed for binding by Yoshida et al. (2008).

3.3.2 Short vs. Long Passive

Tanenhaus and Carlson (1990)’s original experiments showed that voice- (and category-) mismatched VP ellipsis sentences were judged less acceptable than their voice- (and category-) matched counterparts, while judgments on mismatched and matched VP anaphora (‘do it’) did

not differ based on voice- and (category-) match. However, the decision latency did show an effect of voice/category mismatch on both the VP ellipsis and VP anaphora examples. This led Tanenhaus and Carlson (1990) to conclude that there was at least a small effect of parallelism even for VP anaphora. Mauener, Tanenhaus and Carlson (1995) pointed out that the original Tanenhaus and Carlson (1990) materials included a mix of long (containing an overt 'by'-phrase) and short (lacking a 'by'-phrase) passive antecedents. Indeed, when the data were reanalyzed taking this into account the effect of parallelism disappeared for the short passive mismatch VP anaphora sentences. This raises the question of passive-length might in even VP ellipsis.

To examine this we included passive antecedents in our VP Ellipsis materials that either contained a 'by'-phrase (long), as in (11) or did not (short), as in (12).¹ Before turning to the results, let's consider the predictions. In the Mauener et al. (1995) study the voice mismatch effects were reduced on VP anaphora when the antecedent was a short passive. Thus, one possibility would be that in our VP Ellipsis sentences, short passive antecedents would reduce the mismatch effect. However, another possibility seems reasonable. Kertz (2008) has shown that contrast structure is important in understanding ellipsis structures, and indeed she points out that at least part of what goes wrong in voice mismatch VP Ellipsis sentences is that the relevant contrast created in the antecedent VP cannot easily line up with the subsequent elided active VP. To the degree that overt material can be part of a contrast, the long passives actually provide more material that can create a contrast with the subsequent VP Ellipsis clause.

Consider the long passive example in (11). In the first clause both arguments of the laughing event are mentioned, and then the second clause contrasts a new NP with one of the previously introduced ones. In (11a) this is easily achieved because the subject of the antecedent clause and the subject of the ellipsis clause are being contrasted; in (11b), however, it's the NP in the 'by'-phrase that is contrasting with the subject of the ellipsis clause, which is nonparallel. However, things are worse in (12), where only one NP is mentioned in the first clause. This is not problematic in the matching (12a), but it could lead to further difficulty in the mismatching (12b). The subject of the ellipsis clause wants to be contrastive with the unexpressed 'by'-phrase from the antecedent, something that is not possible. From this point of view of contrast structure long passive may actually be easier to interpret (while still being degraded because of the voice mismatch, which has the effect of forcing the contrasting arguments not to align to the same grammatical function).²

- (11) a. Fred was laughed at by Billy and Sandy was, too.
b. Fred was laughed at by Billy and Sandy did, too
- (12) a. Fred was laughed at and Sandy was, too.
b. Fred was laughed at and Sandy did, too.

The means for this subanalysis appear in Figure 5. While there was not a significant (or even marginal) main effect of Match ($p > .12$), there was a marginal main effect of Length ($p < .09$). However, there was a significant ($p < .02$) interaction between the two factors, carried mainly by the fact that short passive mismatch was judged less acceptable than long passive mismatch. This

¹Our design also included short and long sluicing examples, but that was the way the sprouting/Chung manipulation was included; thus, we do not have exactly comparable long and short passive antecedent examples in order to compare these effects in Sluicing to the effects in VP Ellipsis

²We thank Christina Kim for discussion here; this line of thinking is what she is developing in her on-going work on ellipsis constructions.

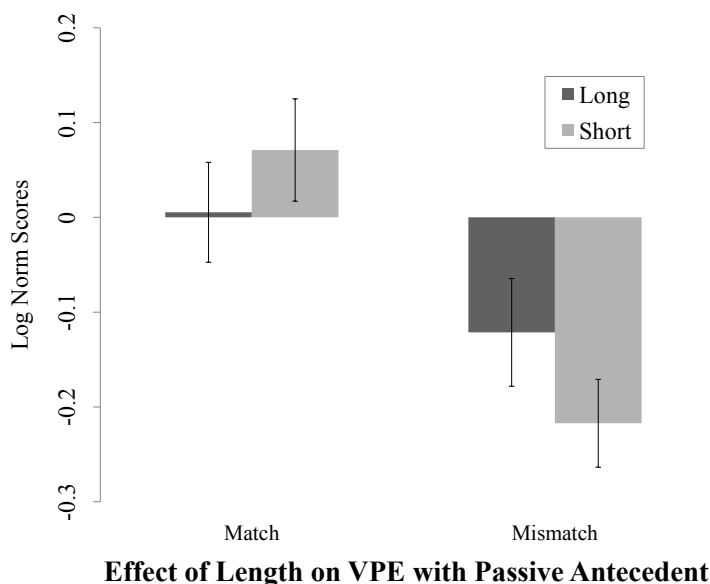


Figure 5: Logged normalized acceptability scores of long and short passive antecedents with matched and mismatched VP Ellipsis (error bars represent Standard Error).

is consistent not with the predictions based on the Mauner et al. (1995) study, but rather with the predictions that long passive provides the relevant arguments for the subsequent contrast created by the VP Ellipsis, while short passive does not.

3.4 Discussion

We found that VP Ellipsis and Sluicing received approximately the same ratings for voice match examples, but that VP Ellipsis was less tolerant than Sluicing to voice mismatch (see Figure 2). This was the basic comparison we designed the experiment to give us. This result is not immediately consistent with Merchant (2008)’s analysis, laid out above, as it predicts that voice mismatch in Sluicing should be ungrammatical, while voice mismatch in VP Ellipsis is grammatical, though perhaps degraded. There are several points about this that should be made. First, Merchant does not claim that mismatch in VP Ellipsis should be acceptable and mismatch in Sluicing should be unacceptable. His analysis is just that the former is grammatical (syntactically derivable) while the latter is not. Thus, we have to ask what accounts for the unacceptability of mismatch in some cases of VP Ellipsis and the improvement in others. One suggestion is that of Kehler (2000), who suggests that the discourse relations the two clauses involved in matters. His account predicts that “resemblance” (or parallel) discourse relations increase the degree to which syntactic parallelism is relevant in VP Ellipsis interpretation, while in “cause-effect” discourse relations syntactic parallelism is decreased. Kim and Runner (2009) examined Kehler’s predictions in a series of magnitude estimation experiments. Their results do suggest that mismatch effects are

stronger in resemblance relations than in cause-effect relations. The VP Ellipsis materials we used in the current experiment are similar to those used in the Kim and Runner (2009) resemblance conditions. Thus, the conditions in our experiment were those that should have led to the most robust voice mismatch effects in VP Ellipsis. It is possible that the increase in sensitivity to voice brought on by the resemblance (or parallel³) relation can account for the larger mismatch effect in the VP Ellipsis compared to Sluicing. This is something we will be exploring in future work.

Secondly, we saw that the interaction between Construction and Match just discussed was actually part of a 3-way interaction among Construction, Match and Ellipsis itself (as illustrated in Figure 3). This interaction revealed an interesting difference between VP Ellipsis and Sluicing. As Kim and Runner (2009) and Kim et al. (2011) found previously, the No Ellipsis controls did not show any sensitivity to the voice manipulation. The mismatch effect was localized entirely to the VP Ellipsis clauses. However, in Sluicing we found that the mismatch effect was the same on the No Ellipsis controls as on the Sluicing Ellipsis sentences themselves. This result suggests that the sensitivity to voice is qualitatively different for the two constructions. Indeed a qualitatively different sensitivity to voice is what Merchant's account predicts, but not in the way we have found. This is because it relies on the [VOICE] feature value being relevant only in Ellipsis (indeed, as our results suggest for VP Ellipsis). For Sluicing voice alignment seems to be important regardless of whether Ellipsis has taken place. Ongoing work will have to continue to examine the overall effects of mismatch on both the ellipsis and unelided controls in ellipsis constructions.

We also found that short passive antecedents triggered a more severe mismatch effects than long passive antecedents. This seems to fit with the view that one part of participants' judgment is made up of the difficulty with which they can assign the proper contrast required by the ellipsis construction. In long passives, all of the relevant arguments are available, though the contrast requires mismatching the grammatical functions of the pair of phrases. Short passive antecedents do not even provide one of the required contrast members, which could contribute to their further degradation. This direction (discussed in Kertz, 2008, and subsequent work) may also lead to a better understanding of the role of mismatch in Sluicing as well, and how the effects seem to be present even in sentences without ellipsis itself.

Finally, we tested Chung's condition on sluicing (in (10)) and found that it violated the degree of unacceptability was comparable to voice mismatch effects in sluicing. We also noted that sluicing ellipsis sentences that had both voice mismatch and violated (10) were rated significantly lower, suggesting that these effects are additive.

4 Conclusions

This report is intended to contribute to the growing experimental literature on ellipsis constructions (Tanenhaus and Carlson, 1990, Arregui et al., 2006, Frazier and Clifton, 2006, Kertz, 2008, Kim et al., 2011 and many others). Our results suggest that while voice mismatch is an ellipsis-specific feature for VP Ellipsis (replicating Kim et al., 2011 and Kim and Runner, 2009), voice feature values are relevant to Sluicing-type sentences independent of ellipsis itself. This does not follow from the Merchant (2008) account, which predicts voice to matter only in ellipsis clauses. Instead,

³It is still unclear whether the increased sensitivity to structural parallelism is due to the resemblance discourse relation or simply the increased expectation of parallelism triggered by the types of sentences that appear in resemblance relations; see Frazier and Clifton (2006) for discussion.

it raises new and interesting questions about the differing effects of voice parallelism in varieties of constructions with and without ellipsis.

References

- Arregui, A., C. Clifton Jr, L. Frazier, and K. Moulton. 2006. Processing elided verb phrases with flawed antecedents: The recycling hypothesis. *Journal of memory and language* 55:232–246.
- Bard, Ellen Gurman, Dan Robertson, and Antonella Sorace. 1996. Magnitude estimation of linguistic acceptability. *Language* 72:32–68.
- Chung, Sandy. 2006. Sluicing and the lexicon: The point of no return. In *BLS General Session and Parasession on Prosodic Variation and Change*, ed. Rebecca T. Cover and Yuni Kim, Vol. 31, 73–91. Berkeley, CA: Berkeley Linguistics Society.
- Frazier, Lyn, and Charles Clifton. 2006. Ellipsis and discourse coherence. *Linguistics and Philosophy* 29:319–346.
- Kehler, Andrew. 2000. Coherence and the resolution of ellipsis. *Linguistics and Philosophy* 23:533–575.
- Kertz, Laura. 2008. Focus structure and acceptability in verb phrase ellipsis. In *Proceedings of the 27th West Coast Conference on Formal Linguistics*, ed. Natasha Abner and Jason Bishop.
- Kim, Christina, and Jeffrey T. Runner. 2009. Strict identity, coherence, and parallelism in VP ellipsis. In *Proceedings of Semantics and Linguistic Theory (SALT)*, ed. Ed Cormany, Satoshi Ito, and David Lutz, Vol. 19, 275–287. eLanguage.
- Kim, Christina, and Jeffrey T. Runner. 2011. Syntactic identity isn't enforced blindly: VP ellipsis and pseudogapping. In *Linguistic Society of America*.
- Kim, Christina S., Greg M. Kobele, Jeffrey T. Runner, and John T. Hale. 2011. The acceptability cline in VP ellipsis. *Syntax* 14.
- Maurer, Gail, Michael K. Tanenhaus, and Greg N. Carlson. 1995. A note on parallelism effects in processing deep and surface verb-phrase anaphora. *Language and Cognitive Processes* 10:1–12.
- Merchant, Jason. 2008. An asymmetry in voice mismatches in VP-ellipsis and pseudogapping. *Linguistic Inquiry* 39:169–179.
- Runner, Jeffrey T., and Christina Kim. 2011. Two (potential) differences between VP ellipsis and pseudogapping. In *Architectures and Mechanisms of Language Processing*.
- Tanaka, Hidekazu. 2011. Syntactic identity and ellipsis. *Linguistic Inquiry* 42:470–490.
- Tanenhaus, Michael K., and Greg N. Carlson. 1990. Comprehension of deep and surface verb phrase anaphors. *Language and Cognitive Processes* 5:257–280.
- Yoshida, Masaya, Michael Walsh Dickey, and Patrick Sturt. 2008. Sluicing and syntactic prediction. In *Architectures and Mechanisms of Language Processing*.