Factoring out the Parallelism Effect in VP-Ellipsis

Nigel Duffield and Ayumi Matsuo
McGill University and Max Planck Institute for
Psycholinguistics, Nijmegen

Introduction

Since Sag (1976), also Hankamer and Sag (1976), Sag and Hankamer (1984), there have been three standard assumptions about the Parallelism Effect observed in VP-Ellipsis constructions, such as in the (b) examples in (1) and (2) below. The first assumption is that the effect is *categorical* (that is to say, that ellipsis sentences following non-parallel antecedents are strictly ungrammatical, rather than simply degraded). Second, it is assumed that the effect applies *asymmetrically*, affecting VPE constructions while sparing closely related VP-Anaphora (VPA) constructions, such as those in (3) and (4). The third assumption, related to the first, is that the effect is *uniquely due to syntactic factors*; that is, structural parallelism alone explains the Parallelism Effect.

(1) a. Someone had to put out the garbage, since I hadn't.
     b. ??The garbage had to be put out, since I hadn't.
(2) a. It annoyed Mary when people mentioned her sister's name. Bill did, quite often.
     b. The mention of her sister's name always annoyed Sally. ??Tom did, out of spite.
(3) a. Someone had to put out the garbage, since I hadn't done it.
     b. The garbage had to be put out, since I hadn't done it.
(4) a. It annoyed Mary when people mentioned her sister's name. Bill did it, quite often.
     b. The mention of her sister's name always annoyed Sally. ??Tom did it, out of spite.

The research reported here[1] challenges each of these assumptions, especially the third; see also Hardt (1993). In a set of

---

[1] This work forms part of a larger set of experiments investigating first and

second language learners' sensitivity to the Parallelism Effect. For full details

psycholinguistic experiments, based on the sentence-completion paradigm of Tanenhaus & Carlson (1990), we examined a number of formal factors—logically independent of syntactic parallelism *per se*—that were hypothesized to play a role in determining the Parallelism Effect. Our results, from a combination of online and offline tasks, tend to support an interactional, non-categorical view of the Parallelism Effect.

Formal Factors

In our experiments, we systematically manipulated various formal, non-structural properties of both the antecedent clause and of the ellipsis clause of ellipsis contexts to determine their effects on parallelism, separately and in combination.

The first factor investigated was that of CONSTRUCTION TYPE. Following earlier work by Tanenhaus and Carlson (1990), we contrasted two types of non-parallel antecedent: passive, as in (1b) above, *vs*. nominal, as in (2b). Our prediction here, based on naïve intuitions, was that the Parallelism Effect would be significantly weaker for passive than for nominal antecedents. Notice that the standard structural account predicts no significant difference in acceptability as a function of construction-type, since by hypothesis, construction-specific rules have no theoretical status.

The second antecedent property examined was what we termed RECOVERABILITY. Here, we were in fact concerned with two separate properties, one for each construction type. For passive antecedents, we examined whether the addition of a <u>by</u>-phrase to a passive antecedent made the following ellipsis more acceptable, thus

weakening by the Parallelism Effect. The logic here was that by providing an accessible (though demoted) subject argument in the antecedent clause, the ellipsis clause should be easier to reconstruct. Again, intuition suggests that (5a), with a <u>by</u>-phrase, is more acceptable than (5b), without one. We termed this property CONCEPTUAL RECOVERABILITY.

(5)   a. Mary was busy, so the package was sent by Tom.—?He
        had promised that he would.
        b. When we got back, our driveway had been cleared of
        snow.—??A neighbor told us that Tom had.

Within the set of nominal antecedents we manipulated a different type   of   recoverability,   namely,   MORPHO-SYNTACTIC RECOVERABILITY. Following a suggestion made by Tom Roeper, we examined whether nominal antecedents containing zero-derived nouns, such as (6a), might be more reconstructable than those headed by affixed nominals, as in (6b). (Unlike the other factors, this contrast was motivated by theoretical considerations, rather than by raw intuitions.)

(6)   a. The children had always longed for a visit to the zoo.
        —??But once they had, they were disappointed.
        b. Bill would be helped by a discussion of his poor study
        habits. —??But up to now, nobody has.

In the ellipsis clause, we also manipulated two properties. The first of these, once again replicating Tanenhaus & Carlson's previous experiments, was ANAPHOR TYPE. Keeping the antecedent clause constant, we systematically contrasted pairs of sentences with VPE completions, as in (1) and (2), with those with VPA completions, as in (3) and (4). Here, previous work led us to expect that the parallelism effect does not spare VPA entirely: contrary to standard theoretical assumptions, there is reasonably strong intuitional and experimental evidence that non-parallel antecedents degrade VPA completions (albeit less strongly than they do VPE completions).

At this point, it is worth briefly pointing out why from a theoretical perspective, parallelism should not affect VPA. The standard

account of the VPE *vs.* VPA dichotomy is due to Hankamer & Sag (1976), modified slightly in Sag & Hankamer (1984). Hankamer & Sag propose that a formal distinction be drawn between two types of anaphoric expression, termed SURFACE and DEEP ANAPHORA, respectively. Surface anaphora refers to anaphoric constructions, including VP-ellipsis, whose grammaticality is claimed to be partially determined by the syntactic form of the antecedent clause. In Hankamer & Sag, it is assumed that the relevant syntactic representation is SURFACE STRUCTURE (hence the proposed label).[2] Hankamer & Sag contrast instances of surface anaphora with those of deep anaphora, such as VP-ANAPHORA (VPA). In the latter constructions, it is claimed that the grammaticality and intended interpretation of the anaphoric elements is derived directly from the discourse model, and not *via* any syntactic representation of the antecedent. One consequence of this claim is that the syntactic form of the antecedent clause should be irrelevant to the acceptability of deep anaphora constructions.[3]

In addition to anaphor type, the other ellipsis clause property that we investigated was FINITENESS: in our experiments, we systematically contrasted *finite* ellipsis clauses, headed by a form of the present perfect auxiliary have, with *non-finite* ellipsis clauses, headed by to. This contrast is illustrated by the examples below: see also (1a,b), *vs.* (1e,f) above.

---

[2] For a number of reasons, the authors later modify this claim: rather than surface structure, ellipsis constructions are claimed to refer to 'propositional representations of the antecedent clause'; such representations are approximately equivalent to LF representations in other frameworks. The essential claim, which remains, is that ellipsis constructions are obligatorily sensitive to the *syntactic* properties of the antecedent clause; the structural parallelism constraint is one effect of this.

[3] In their 1984 paper, Sag & Hankamer also revise their characterization of deep anaphora constructions, recasting deep anaphora as 'model-interpretive anaphor[a]'.However, the essential distinction remains, as do the consequent claims for the (non-)availability of parallelism effects.

(7)   a. When we got back, our driveway had been cleared of
        snow—??A neighbor told us that Tom had.
        b. Our driveway needed to be cleared of snow—?But no-
        one wanted to.

In this case, our intuitions—as well as the results of previous work
(Duffield & Matsuo 2001)—indicated that non-finite violations of
syntactic parallelism should be significantly more acceptable than
finite violations (though, again, the standard theory neither predicts
nor accommodates this).

The various factors we manipulated—crossed with structural
parallelism—are summarized in Table 1 below. The number in the
right hand column represents the number of test pairs per cell.

Table 1. Summary of Formal Factors Investigated.

Antecedent Properties		Ellipsis Properties	
Syntactic Form	Recoverability (more/less)	Finiteness	Anaphor Type
Parallel: Active/Verbal		Finite	VPE 16
			VPA 16
		Non-finite	VPE 12
			VPA 12
Non-Parallel: Passive	With by-phrase	Finite	VPE 8
			VPA 8
		Non-finite	VPE 6
			VPA 6
	No by-phrase	Finite	VPE 8
			VPA 8
		Non-finite	VPE 6
			VPA 6
Non-Parallel: Nominal	Zero-derived	Finite	VPE 8
			VPA 8
		Non-finite	VPE 6
			VPA 6
	Affixed	Finite	VPE 8
			VPA 8
		Non-finite	VPE 6
			VPA 6

Experiments
Tasks

Our experiments were a modification of the online Sentence Completion Judgment Paradigm presented in Tanenhaus and Carlson (1990); see also Mauner, Tanenhaus and Carlson (1995). In this task, subjects were asked to judge, for pairs of sentences such as those in (1-7) above, whether the second (anaphoric) sentence formed a 'sensible completion' to the first. In our experiment, there were in fact two tasks: an *online* timed anomaly task—replicating Tanenhaus and Carlson—and an *offline* Grammaticality Judgment task using the same materials. The online task yielded two dependent measures: the proportion of sentences accepted in each condition; the time taken to accept or reject the sentences of each condition. The offline task yielded a graded acceptability judgment (on a scale from 0-5, 'totally unacceptable' $\leftrightarrow$ 'perfectly acceptable').[4]

Subjects

We tested 31 native-speakers of North American English, at McGill University and at the Max Planck Institute for Psycholinguistics, Nijmegen.[5] Subjects were divided into four groups: for each sentence-pair, each group received one of the four possible alternants: parallel-VPE, parallel-VPA, non-parallel-VPE, non-parallel VPE. There were a total of 128 test items per subject, plus 56 distracter/control items. For each run of the online experiments, all items were randomized within each block; for the offline

---

[4] Both tasks included a set of control items, of which half were perfectly acceptable—being consecutive sentences from passages of contemporary fiction—and half were not.

[5] We also tested 20 Dutch-speaking (advanced L2 English) learners on the same materials. As noted in footnote 1 above, the native-speaker results reported here form part of a larger comparative study : for details, see especially Duffield & Matsuo (2003).

experiment, the items were randomized separately for each of the four versions of the experiment.

Predictions

Based on the results of previous experiments, the general prediction was that items with non-parallel antecedents should be accepted significantly less frequently than those with parallel antecedents, and that those non-parallel items that *were* accepted should take significantly longer to accept than the corresponding parallel items. In other words, we predicted a main effect of syntactic parallelism for all three dependent measures. Previous work also led us to expect a reliable interaction between syntactic parallelism and anaphor type for all dependent measures: non-parallel antecedents with VPA completions should be significantly more acceptable than those with VPE completions.

The particular design of the present experiments yielded a number of other predictions. Specifically, supposing that the formal factors discussed above contribute to the Parallelism Effect, we predicted significant interactions between syntactic parallelism on the one hand, and finiteness, construction type and recoverability, on the other. In addition, we hypothesized a significant, albeit reduced, parallelism effect for VPA. Finally, we predicted that all effects of parallelism should be gradient, rather than categorical: that is to say, violations of parallelism should yield *less* acceptable, rather than *un*acceptable sentences.

Results and Discussion

(Here, we briefly summarize the principal findings: please see Duffield & Matsuo 2003 for a detailed statistical report and analysis.)

Online Experiment

Analyses of variance of the *judgment* data for the English native-speakers revealed reliable main effects of syntactic parallelism, anaphor type and finiteness, with finiteness showing the largest main effect (all $p < .0001$). (The nested factor, recoverability, was analyzed separately: see below.) There was no main effect of construction type, nor any interaction between construction type and

other factors. As predicted, a significant two-way interaction was observed between syntactic parallelism and anaphor type (*F*1, *F*2 p < .0001). This is illustrated in Figure 1. (Planned comparisons based on the subject ANOVA showed a significant effect of parallelism in the VPE condition, but no parallelism effect in the VPA condition overall).
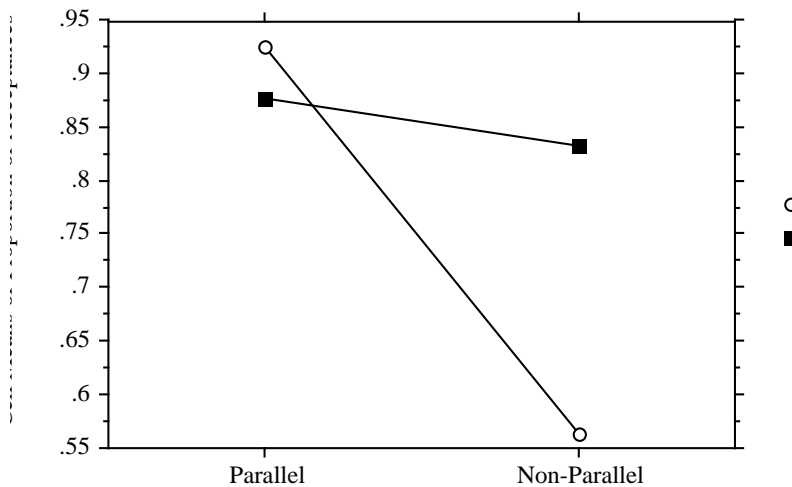

Figure 1. Interaction of Parallelism and Anaphor Type

A statistically reliable interaction was also observed between syntactic parallelism and finiteness (*F*1 p < .0002; *F*2 p < .05): as Figure 2 shows, the parallelism effect was significantly weaker for non-finite than for finite ellipsis. However, planned comparisons (based on the by-subject ANOVA) revealed a significant effect of parallelism even in the non-finite condition.
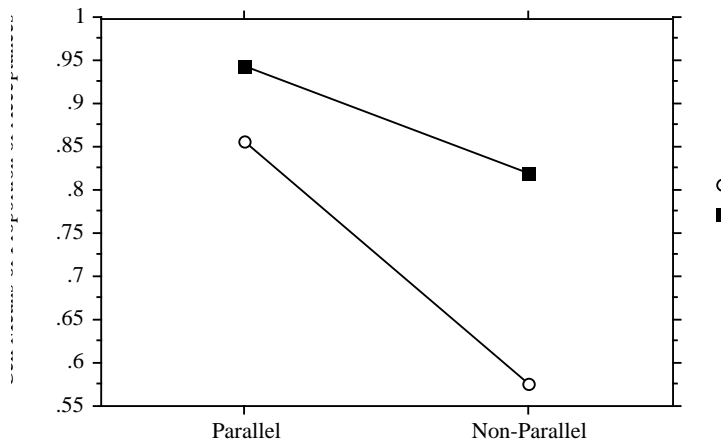
Figure 2. Interaction of Syntactic Parallelism and Finiteness

As for the *latency* data, separate analyses of variance revealed a significant main effect of finiteness (*F*1 p < .0001, *F*2 p < .0005) as well as a marginal effect of syntactic parallelism. No other main effects or interactions were significant in the item analysis (in contrast to the subject analysis).

As for recoverability, the set of analyses of variance *within* the non-parallel conditions failed to show any reliable main effects, either in the judgment data, or in the latency data. That is to say, sentence-pairs where the non-parallel antecedent contained a by-phrase were accepted overall no more frequently—nor more quickly—than passive antecedents without a by-phrase; zero-derived nominals were no better antecedents than affixed nominal antecedents.

Nevertheless, asymmetries in recoverability *were* observed in certain conditions: at least for conceptual recoverability, the trend was consistently in the predicted direction. Specifically, as predicted, the size of the parallelism effect was smaller for more recoverable non-finite VPE conditions (p = .062) than for the less recoverable non-finite VPE condition (p < .005): an example of a relevant pair is

provided in (8) below.   A similar (non-significant) trend was observed in the offline results; see next section.[6]

(8)   a. Mary was busy, so the package was sent by Tom.—?He had promised to.
(p=.06, non-significant parallelism effect for pairs of this kind)
b. When we got back, our driveway had been cleared of snow.—??Tom had promised to.
(p<.005, significant parallelism effect for pairs of this kind)

Offline Experiment

The results from the offline experiment largely confirmed those obtained in the online task. Separate analyses of variance of the native-speakers' judgment data revealed reliable main effects for syntactic parallelism, finiteness, and anaphor type (all $p < .0001$). The main effect of construction type was reliable by subject ($p < .0001$), though not by item ($p > .05$). Although reliable two-way interactions were observed in the by-subject ANOVA—between anaphor type and syntactic parallelism ($p < .005$), construction type and finiteness ($p < .005$), and syntactic parallelism and finiteness ($p < .005$), respectively—none of these interactions were reliable in the by-item ANOVA. However, the interactions between anaphor type and syntactic parallelism, and syntactic parallelism and finiteness approached significance in the latter analysis. There were no effects for recoverability, although once again, conceptual recoverability showed a trend in the predicted direction in certain subconditions.

Figure 3 below shows a breakdown of the parallelism effect by construction type, finiteness and recoverability in the offline task.[7]

---

[6] By contrast, to the extent that any pattern is observed for morpho-syntactic recoverability, the tendency is in the opposite direction: zero-derived nominals tend to be less, not more, acceptable than affixed forms. Again, parallel results are observed in the offline data.

[7] The figure collapses across anaphor type: as might be expected, these effects are more marked for VPE than for VPA.

The top line of the chart shows the acceptance rate for items with structurally parallel antecedents; the bottom two lines show the rates for structurally non-parallel antecedents (less and more recoverable); on the x-axis, construction 1 refers to the active-passive contrast, construction 2 to verbal-nominal pairs. The figure shows the striking influence of finiteness on the strength of the parallelism effect: across both construction types—and across anaphor types (though this is not shown here)—non-finite ellipsis elicits a significantly weaker parallelism effect than does finite ellipsis.
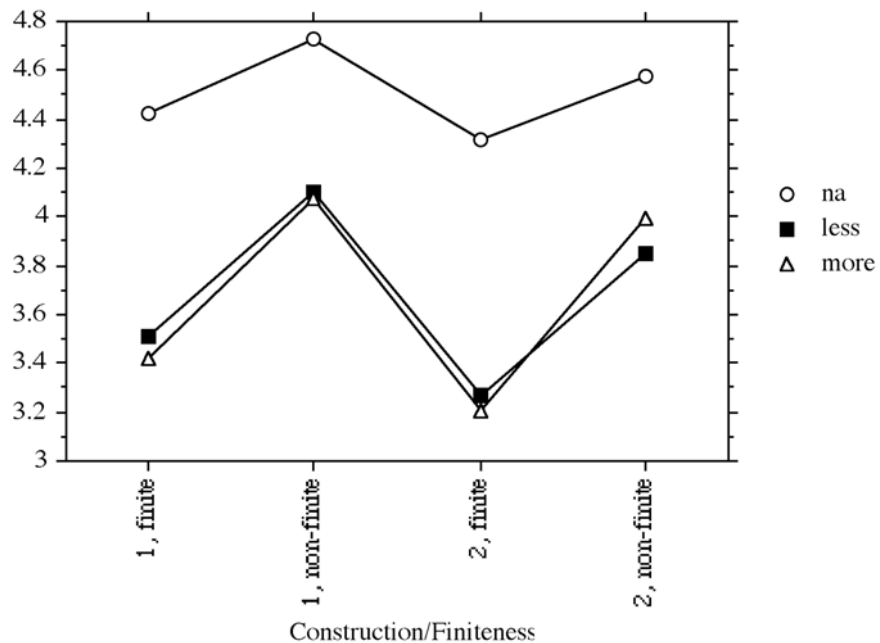

Figure 3. Effects of Finiteness, Recoverability and

Construction Type on Parallelism (offline GJ task)

Summary and Discussion

In summary, the results from both experiments demonstrate rather clearly that structural (syntactic) parallelism is not the sole determinant of the Parallelism Effect: other non-structural formal factors, most notably, finiteness, have a significant influence on the

strength of this effect.[8]   Furthermore, our results suggest that the Parallelism Effect does not spare VPA, as the standard theory predicts: although VPA constructions did not show significant parallelism effects overall, closer examination of particular subconditions—for example, finite VPA following a nominal antecedent—revealed reliable parallelism effects for 'deep anaphora' also. Finally, in all of the experiments, the parallelism effect was shown to be gradient, rather than categorical: even the 'least acceptable' violations of parallelism—i.e., pairs with less recoverable, nominal antecedents and finite VPE completions— proved to be reliably more acceptable than the 'unacceptable' control items.

The picture of the parallelism effect that emerges from these experiments challenges the standardly-held assumptions: assuming the results obtained are valid, then parallelism is gradient, rather than categorical, affects both VPE and VPA, and is significantly influenced by non-structural, as well as purely structural, factors.

The final question—perhaps the most important—is how best to interpret these results.   It is clear that they cannot easily be accommodated under the standard theory, which does not naturally handle gradient, interactional effects of this kind. It is less clear, however, whether the results should be interpreted as supporting an outright rejection of the standard account, or as suggesting a more moderate revision of that account, one that allows 'parallel access' to syntactic and conceptual representations for both types of anaphora. While we tend toward the latter interpretation—see Duffield & Matsuo (2003) for details—further investigations of this type may support the more radical conclusion.

---

[8] The effects of non-structural factors—and the interactions among them—were even more striking for the L2 learner groups discussed in Duffield & Matsuo (2003). For example, for Dutch learners, non-finite ellipsis effectively canceled the parallelism effect with passive antecedents; while finite *VP-anaphora* with nominal antecedents elicited a stronger parallelism effect than non-finite VPE in the same construction.

References

Duffield, Nigel and Matsuo, Ayumi. 2001. A Comparative Study of Ellipsis and Anaphora in L2 Acquisition. *Proceedings of the 25th Boston University Conference on Language Development*, ed. by Anna H.-J. Do, Laura Domínguez and Aimee Johansen: 238-49. Somerville, MA: Cascadilla Press.

—. 2002. Finiteness and Parallelism: assessing the generality of knowledge about English ellipsis in SLA. *Proceedings of the 26th Boston University Conference on Language Development*, ed. by Barbora Skarabela, Sarah Fish and Anna H.-J. Do: 197-207. Somerville, MA: Cascadilla Press.

—. 2003. Acquiring competent gradience: factoring out the Parallelism Effect in VP-ellipsis. *McGill Working Papers in Linguistics* 18. [in submission to *Language*]

Hankamer, Jorge and Sag, Ivan. 1976. Deep and surface anaphora. *Linguistic Inquiry* 7.

Hardt, Daniel. 1993. *Verb phrase ellipsis: form, meaning and processing.* Computer and Information Science, University of Pennsylvania: Ph.D. dissertation.

Matsuo, Ayumi and Duffield, Nigel. 2001. VP-Ellipsis and anaphora in first language acquisition. *Language Acquisition* 9: 301-27.

Mauner, G, Tanenhaus, Michael K and Carlson, Greg N. 1995. A note on parallelism effects on processing verb phrase anaphors. *Language and Cognitive Processes* 10: 1-12.

Sag, Ivan. 1976. *Deletion and logical form*. MIT: Ph.D. dissertation.

Sag, Ivan and Hankamer, Jorge. 1984. Towards a theory of anaphoric processing. *Linguistics and Philosophy* 7: 325-45.

Tanenhaus, Michael and Carlson, Greg N. 1990. Comprehension of deep and surface verbphrase anaphors. *Language and Cognitive Processes* 5: 257-80.