*SSLA*, **31**, 93–123. Printed in the United States of America.
doi: 10.1017/S0272263109090044

# NATIVE SPEAKERS' VERSUS L2 LEARNERS' SENSITIVITY TO PARALLELISM IN VP-ELLIPSIS

Nigel G. Duffield and Ayumi Matsuo
*University of Sheffield*

This article examines sensitivity to structural parallelism in verb phrase ellipsis constructions in English native speakers as well as in three groups of advanced second language (L2) learners. The results of a set of experiments, based on those of Tanenhaus and Carlson (1990), reveal subtle but reliable differences among the various learner groups. These differences are interpreted as showing that some L2 learners can acquire sensitivity to parallelism in the absence of surface transfer. Furthermore, the results cast doubt on two conventional theoretical claims: that the parallelism effect has a syntactic basis and that it is uniquely linked to instances of surface anaphora (Hankamer & Sag, 1976).

The research reported here investigates native speakers' (NSs) and second language (L2) learners' knowledge of the acceptable use of verb phrase ellipsis (VPE) constructions in English. Specifically, we investigate the parallelism effect in VPE, which, according to the standard account (Hankamer &

Previous versions of parts of this article were presented at various locations, including the universities of Cambridge (RCEAL), Durham, Essex, Groningen, Utrecht, at Max Planck Institute for Psycholinguistics, Nijmegen, and at various conferences, including NELS 98, Second Language Research Forum, and the Boston University Conference on Language Development 1998, 2000, 2001. We are grateful to these audiences, especially to Melissa Bowerman, Harald Clahsen, Bonnie D. Schwartz, Wolfgang Klein, and John Williams. We would also like to thank Ingrid Leung, Christopher Miller, and Femke Uitdewilligen for research assistance at various stages of the preparation of this article. Finally, we are extremely grateful to Leah Roberts for her help in checking the statistical analyses and to the anonymous *SSLA* reviewers who suggested important revisions and improvements. As ever, we remain solely responsible for all errors or inaccuracies.

Address correspondence to: Nigel Duffield, School of English Literature, Language & Linguistics and Centre for Comparative Linguistic Research, University of Sheffield, Sheffield S10 2TN, England, UK; e-mail: n.g.duffield@sheffield.ac.uk.

© 2009 Cambridge University Press 0272-2631/09 $15.00 

https://doi.org/10.1017/S0272263109090044 Published online by Cambridge University Press

Sag, 1976), is attributable to a syntactic constraint on the form of the antecedent clause. This proposed structural constraint has the effect of degrading VPE sentences with syntactically nonparallel antecedents such as those in (1b) and (1d) in which the relevant antecedent for the elided verb phrase (VP) is a passive VP or derived nominal, respectively.

(1) a. *Someone had to put out the garbage.—But I didn't want to.*
b. *The garbage had to be put out.—*But I didn't want to.*
c. *It annoyed Sally if anyone mentioned her sister's name.—Tom did, out of spite.*
d. *The mention of her sister's name annoyed Sally.—*Tom did, out of spite.*

Whereas there has been a considerable amount of research in first language (L1) acquisition and SLA investigating other aspects of VPE constructions—especially the interpretation of elided pronouns (see Thornton & Wexler, 1999, and references therein)—the parallelism effect has received little attention. To our knowledge, the only works in L1 acquisition to investigate parallelism in ellipsis constructions are those of Matsuo and Duffield (2001) and, more recently, Santos (2006), neither of which addressed the specific issues tackled here.

Matsuo and Duffield's (2001) experiment looked at children's judgments of parallelism violations and concluded that young children are sensitive to the effects of parallelism in ellipsis constructions. However, the stimuli employed in their experiment as well as the age of the children and task demands meant that their results did not arbitrate between theoretical alternatives. Santos's (2006) dissertation is based mostly on an analysis of spontaneous and elicited production data: She did not examine judgment data.

It should be understood from the outset that our main purpose is not to present data that arbitrate among competing theories of SLA (e.g., contrasting theories of full access with those of partial access). Instead, our aim is to use L2 learners' data to help arbitrate among competing explanations of the parallelism effect in ellipsis and anaphora contexts—including syntactic, semantic, and processing explanations—for any group of learners, including NSs. For this reason, we have little to say about particular models of SLA: Nevertheless, we believe that this article is relevant to SLA theory to the extent that it sheds light on the issues of learnability and ultimate attainment.

Ellipsis clauses are standardly contrasted with cases of VP anaphora (VPA), exemplified in (2). Most of the theoretical literature claims that these VPA constructions do not exhibit parallelism effects. Although this is not quite true, there is an undeniable contrast between the two constructions with respect to the perceived strength of these effects.

(2) a. *Someone had to put out the garbage.—But I didn't want to do it.*
b. *The garbage had to be put out.—But I didn't want to do it.*
c. *It annoyed Sally if anyone mentioned her sister's name.—Tom did it, out of spite.*
d. *The mention of her sister's name annoyed Sally.—Tom did it, out of spite.*

https://doi.org/10.1017/S0272263109090044 Published online by Cambridge University Press

The purpose of this article is to examine whether the available experimental evidence, including the results reported here, best supports the traditional structural account of the parallelism effect or whether it is more consistent with an alternative explanation as well as to discover whether the answer to this question varies according to learner group (in this case, NSs vs. L2 learners).

Sentences that involve ellipsis constructions are of interest for several reasons. First, ellipsis clauses are true interface phenomena, whose analysis and interpretation depend on the interaction of conceptual and formal grammatical factors. Unlike regular sentences, the acceptability of an elided sentence is not a sentence-internal property: It can be determined only after the elided material has been reconstructed, and this reconstruction is dependent on the semantic and syntactic properties of the preceding antecedent clause. Were this not the case, all of the ellipsis clauses provided in (1) would be equally acceptable.

A second reason for investigating ellipsis constructions, which Sag (1976) brought to general attention, is that these constructions have been claimed to instantiate purely formal constraints; for example, there is no obvious functional reason why parallelism should not apply equally to VPE expressions and to their VPA counterparts, yet it seems not to do so. That ellipsis constructions are differentially constrained in this way has been taken as prima facie evidence of the existence of autonomous grammatical principles.

A further reason for examining utterances involving parallelism is that they provide a clear illustration of gradient, as opposed to categorical, judgments. As our experiment confirms, violations of structural parallelism cause speakers to disprefer ellipsis clauses to a greater or lesser extent but not to reject them outright. In other words, parallelism is a relatively soft constraint: Depending on how they are interpreted, gradient effects such as these potentially challenge many current grammatical models, which assume, tacitly or explicitly, that grammars consist of categorical rules and principles (Duffield, 2003; see also Bybee & Hopper, 2001). Here, we argue that the gradience observed does not necessarily threaten conventional assumptions about the form of the grammar: It implies, however, a nonsyntactic basis for the parallelism effect itself.

Gradient effects also raise questions of learnability. If NSs accept parallelism violations a significant portion of the time—and if parallelism effects are due to a syntactic constraint—the question arises as to how learners come to acquire this negative constraint in the first place. The standard generativist assumption is that learners do not rely on statistical properties of the input to establish assumptions about grammatical properties (Chomsky, 1957; see also Reali & Christiansen, 2005); instead, negative constraints a priori exclude the production of ungrammatical sentences. In cases in which a constraint is parameterized crosslinguistically, learners have taken to rely on positive evidence to admit a particular structure to their set of

https://doi.org/10.1017/S0272263109090044 Published online by Cambridge University Press

grammatically licensed structures. The gradient effects associated with nonparallel VPEs pose a problem for this parameter-setting logic in two respects. On the one hand, if English speakers produce such sentences with reasonable frequency—and the corpus data suggest that they do—then a child or adult learner is presented with positive evidence for the acceptability of these sentences. Consequently, such sentences should count as grammatical and should be judged fully acceptable, contrary to fact. On the other hand, if a nonparallel VPE is excluded by some innate structural principle—if it is ungrammatical sensu stricto—then the problem is to explain why sentences that violate this principle are nevertheless judged considerably more acceptable than violations of other formal constraints. Additionally, the reason why the constraint fails to apply in a uniform fashion crosslinguistically needs to be explained. The main point to observe, however, is that a learnability problem only arises if the relative acceptability of nonparallel ellipsis is due to a structural constraint; if the effects are explained in another way—in terms of processing, for example—the challenge is to some extent removed.

## A NOTE ON GRAMMATICALITY AND ACCEPTABILITY

Before proceeding further, it is worthwhile to clarify our use of the terms *grammaticality*, *acceptability*, and the related compound term *grammatical acceptability*, so as to preclude misunderstanding. We use *grammatical* in the strict sense of licensed by the grammar, where grammar refers simultaneously to a formal linguistic description as well as to the putatively autonomous internal representation of these rules and constraints (I-language). We use the term *acceptable* to designate utterances that language users would classify—either intuitively or explicitly—as natural well-formed utterances of their language. Hence, this article adopts and endorses the competence-performance distinction (see Newmeyer, 1983, 2003).

Given these definitions—and standard generativist assumptions about grammatical rules—sentences are either grammatical or ungrammatical.[1] However, spoken or written utterances are, of course, more or less acceptable for a wide variety of reasons, including lexical, semantic, contextual, pragmatic, and idiosyncratic factors, in addition to syntactic ones. Our interest here is exclusively with acceptability contrasts among utterances in which all other factors are controlled for as far as possible and where (what is believed to be) grammaticality alone is manipulated. In other words, we are interested in what are commonly but misleadingly termed *grammaticality judgments*. We operationalize *grammatically acceptable* as sentences judged acceptable in virtue of being grammatical and *grammatically unacceptable* as sentences judged unacceptable on account of their ungrammaticality. The specific issue investigated here is the discrepancy between acceptability judgments for allegedly ungrammatical

https://doi.org/10.1017/S0272263109090044 Published online by Cambridge University Press

sentences involving ellipsis or anaphora compared with other types of ungrammatical sentences.

This article will present the classic structural account of the parallelism effect (Hankamer & Sag, 1976; Sag & Hankamer, 1984), together with a brief summary of alternative analyses (Frazier, Taft, Roeper, Clifton, & Ehrlich, 1984; Hardt, 1993; Merchant, 2001, 2004). Crosslinguistic variation in the domain of application of parallelism as well as implications of this variability for SLA will also be discussed, followed by a brief presentation of prior experimental work in this area (Tanenhaus & Carlson, 1990) and the details of the current experiment. The concluding section offers an interpretation of these results and suggests future lines of inquiry.

## FORMAL EXPLANATION OF THE PARALLELISM EFFECT

The classic statement of the parallelism effect in ellipsis is the one presented in Hankamer and Sag (1976), an account that saw some revision in Sag and Hankamer (1984) (see also Fiengo & May, 1994; Johnson, 1997, 2001; Wasow, 1972; Williams, 1977). Hankamer and Sag drew a formal distinction between two types of anaphoric expression, termed *surface* and *deep anaphora*. *Surface anaphora* refers to constructions, including VPE, whose grammaticality is claimed to be dependent on the syntactic form of the antecedent clause. Hankamer and Sag assumed that the relevant level of syntactic representation is surface structure, hence the term *surface anaphor*(*a*). For a number of reasons, the authors later modified this claim: Rather than surface structure, ellipsis constructions are claimed to refer to "propositional representation [s] of the kind generally called *logical form*" (Sag & Hankamer, p. 328). The essential claim that remains, however, is that VPE constructions are directly sensitive to the syntactic properties of the antecedent clause. The parallelism constraint is one result of this sensitivity: Indeed, for Hankamer and Sag, it is one of the defining characteristics of surface anaphora. The parallelism effect is thus viewed by Hankamer and Sag as the reflex of a grammatical constraint.

Hankamer and Sag (1976) contrasted instances of surface anaphora with those of deep anaphora, such as VPA, for which it is proposed that the grammaticality and the intended interpretation of anaphoric elements are derived directly from the discourse model. An immediate consequence of this proposed dichotomy is that the syntactic form of the antecedent clause should be irrelevant to the acceptability of deep anaphora constructions.[2] Many of the examples provided by these authors appear to confirm the operation of a parallelism constraint in VPE as well as its absence in VPA contexts. However, our work will show that parallelism effects are also found in VPA contexts, as illustrated in (2b) and (2d), although the observed effects are significantly weaker than in the case of VPE.

https://doi.org/10.1017/S0272263109090044 Published online by Cambridge University Press

## CHALLENGES TO THE DICHOTOMY

This is not the first work to question the standard account. Grinder and Postal (1971), Schachter (1977), and Williams (1977) presented several challenges to Hankamer and Sag (1976), whereas Haïk (1987) further refined Hankamer and Sag's account of surface anaphora (see also Pullum, n.d.). A number of different kinds of evidence speak against the syntactic account (Johnson, 2001). Perhaps the most obvious place to start is with production data. Hardt (1993) documented the availability of nonparallel ellipsis in contrast to some other kinds of ungrammatical sentences. Examples (3) and (4) attest to the productivity of such sentences: Those in (3) illustrate nonparallel ellipsis with passive antecedents, and those in (4) illustrate cases in which the antecedent is a derived (or related) nominal expression.

(3) a. *This information could have been released by Gorbachov, but he chose not to* (Daniel Schorr, National Public Radio broadcast, 10/17/92). (Hardt, p. 131)
b. *A lot of this material can be presented in a fairly informal and accessible fashion, and often I do* (Chomsky, 1982; cited in Dalrymple, Shieber, & Pereira, 1991). (Hardt, p. 134)

(4) a. [*Many Chicago-area cab drivers*]*…sense a drop in visitors to the city. Those who do, they say, are not taking cabs* (Chicago Tribune, 2/6/92). (Hardt, p. 118)
b. *We should suggest to her that she officially appoint us as a committee and invite faculty participation. They won't, of course…* (University of Pennsylvania e-mail message). (Hardt, p. 116)

These examples suggest that ellipsis clauses that follow nonparallel antecedents have a grammaticality status distinct from sentences that NSs reject as categorically unacceptable; otherwise they should not occur in written registers in any significant numbers.[3] These examples also illustrate a different sort of gradience—namely, variability in the strength of the parallelism effect across construction type. Intuitively, among nonparallel antecedents, nominal antecedents are significantly less acceptable than passive antecedents, something that Sag and Hankamer's (1984) approach fails to predict or explain. Until now, this intuition has not been investigated experimentally.

A second factor suggesting that the parallelism effect is not purely structural in origin is the fact that morphosyntactic parallelism—and even stem parallelism—is not necessary for full acceptability. In English VPE, the inflectional features of the reconstructed verb form may be nonidentical to those of the antecedent: Although agreement in voice is usually required, tense and agreement features may vary. For example, (5b) and (5c) are just as acceptable as (5a).

https://doi.org/10.1017/S0272263109090044 Published online by Cambridge University Press

(5) a. *Ben will come to the party, but Bill won't* [*come to the party*].
b. *Ben* ***came*** *to the party, but Bill didn't* [*come to the party*].
c. *Ben* ***comes*** *to every party, but Bill and Mary don't* [*come to every party*].

Moreover, analysis of languages with so-called verb stranding VPE, in which the finite verb overtly raises out of the ellipsis phrase, indicate that even stem identity is not required for well-formedness (see Goldberg, 2005, for Hebrew; Cyrino & Matos, 2007, and Santos, 2006, for Portuguese). In (6), for example, taken from Santos, the verbs in the antecedent clause and ellipsis clause (indicated in bold) are lexically distinct, although they share a common argument-structure and closely related semantics.

(6) *Eu* ***gostava*** *de convidar algumas pessoas; mas quem*$_i$ *é que tu* ***queres***$_j$
I like PREP invite some people but who is-it that you want
*[t*$_j$ *~~[que eu convide t*$_i$~~*]]*
that I invite
"I would like to invite some people, but who do you want (me to invite)?"
(Santos, p. 120)

Such facts tend toward the conclusion that an abstract syntactic constraint is not the ideal explanation of the parallelism effect in VPE.

## Semantic Accounts

A plausible grammatical alternative to the syntactic view has recently been advanced by Merchant (2001, 2004). He proposed that the parallelism effect has a semantic basis (see also Cyrino & Matos, 2007; Santos, 2006). Merchant argued that the effect is due to a semantic identification condition on ellipsis, which, simply stated, requires mutual entailment between the antecedent clause and the ellipsis clause. The stringency of this semantic condition is such that, in practice, only lexically and structurally parallel antecedents are generally able to satisfy the identification condition. However, syntactic parallelism per se is no longer a necessary condition for grammaticality in Merchant's account but rather a concomitant effect of semantic well-formedness.

This semantic account, then, has the advantage of correctly including VPE expressions that diverge from their antecedents with respect to inflectional properties, as in (5)—even with respect to the verb stem itself, as in (6)—while correctly excluding most of the other cases prohibited by the standard syntactic approach. One crucial empirical difference between the two accounts, however, is that the semantic account does not, in principle, exclude passive antecedents to active ellipses, provided that the semantic identification condition is met. Here, the main point is that if parallelism has a semantic basis, there is no reason for the associated acceptability judgments to have

https://doi.org/10.1017/S0272263109090044 Published online by Cambridge University Press

the same character or strength as those for syntactically ungrammatical sentences.

Merchant's (2001, 2004) approach and the traditional analysis, however, both predict a contrast between surface and deep anaphora (VPE vs. VPA). This is because Merchant's account operates on a theoretical distinction—adopted from Rizzi's (1986) analysis of *pro*-drop—between identification conditions and syntactic licensing conditions. Identification conditions only apply to licensed structures: VPA is not syntactically licensed; hence, the identification requirement should not apply. Crosslinguistically, if a given language fails to license VPE in the first place, then this semantic identification requirement will have no structure to apply to, and, again, no parallelism effects should be observable.

An important point to observe with respect to acquisition is that the semantic identification requirement is taken to be innately specified: What must be learned from positive evidence is whether a given language contains licensing structures (e.g., VPE constructions) that contain functional heads with the right type of abstract formal features. If such features are found, the identification requirement should automatically be triggered.

## Parallelism as a Parsing Effect

There is no compelling need, however, for a grammatically based explanation. A plausible alternative is that the parallelism effect is due to parsing preferences rather than to any grammatical constraint, syntactic or semantic. Previous psycholinguistic studies of reading (e.g., Frazier et al., 1984; Frazier, Munn, & Clifton, 2000) have demonstrated clear preferences for parallel over nonparallel conjuncts in overt coordinate structures: (7a), for example, is processed more readily than (7b) although both are equally grammatical.

(7) a. *Jane found a silver coin and a bronze cup.*
b. *Jane found a silver coin and a cup.*

It is reasonable to suppose that reconstruction of covert material in ellipsis contexts is subject to similar processing preferences and that the parallelism effect is a reflex of these preferences rather than of a grammatical constraint. On this interpretation, nonparallel ellipsis contexts are akin to traditional garden-path sentences: The parser initially prefers a parallel over a nonparallel reconstruction of the elided material. When the preferred parse fails to yield a well-formed semantic representation at the discourse level, the initially dispreferred (nonparallel) interpretation may be selected instead. Under this analysis, sentences can be perfectly grammatical (e.g., [1b] and [1d]) but less acceptable in these contexts for processing reasons. Data from L2

https://doi.org/10.1017/S0272263109090044 Published online by Cambridge University Press

learners is extremely informative in helping to tease apart these various analytic options.

## CROSSLINGUISTIC VARIATION AND LEARNABILITY

In the SLA context, the learnability problems raised by gradience are compounded by crosslinguistic variation. Languages vary widely as to whether they permit VPEs. There are also wide discrepancies concerning the construction types subject to parallelism, even in those languages that accept VPEs. As a consequence, the theoretical literature thus far has produced no effective proposal for what might count as a learnable trigger, either for the possibility of VPE in a given language variety or for the application of the parallelism constraint to particular constructions.

The crosslinguistic variation observed in the native languages of the L2 learners in our experiment (i.e., Dutch, Japanese, and Spanish) are detailed here. The general consensus across the theoretical literature is that none of these languages permits VPE (e.g., Lobeck, 1995; Lopez, 1994).[4] If this is the case and if parallelism has a grammatical basis, then one should be able to treat the L2 learners as a homogeneous group for the purposes of our experiment (given uniform levels of proficiency), although several factors complicate this assumption. First, there is controversy as to whether Japanese permits VPE. Second, even if the standard assumption is true, all of these languages show parallelism effects in other related constructions, and, in each language, these effects apply differently. For these reasons, the languages are best treated separately.

### Dutch

Dutch does not permit VPE, as evidenced by the unacceptability of the sentences provided in (8) (see note 4). Example (8a) illustrates an instance of nonfinite ellipsis, and (8b) shows a finite example. Dutch does, however, have a VPA construction, employing a cognate of English *do* "doen"; examples are provided in (9). Reportedly, the VPA construction displays a mild parallelism effect. (This appears to be also true of English VPA, theoretical claims to the contrary notwithstanding).

(8) a. **Iemand moest het vuilnis buiten zetten, maar Marie wilde niet.*
someone had-to the garbage out set but Marie wanted not
"Someone had to take out the garbage, but Marie didn't want to."

b. **Marie zei dat ze het vuilnis buiten zou zetten maar tot nu toe*
Marie said that she the garbage out would set but till now-till
*heeft ze niet.*
has she not
"Mary said that she would put out the garbage, but so far she hasn't."

https://doi.org/10.1017/S0272263109090044 Published online by Cambridge University Press

(9) a. *Iemand moest het vuilnis buiten zetten, maar Marie wilde het niet doen.*
someone had-to the garbage out set but Marie wanted it not do
"Someone had to take out the garbage, but Marie didn't want to."
b. ?*Het vuilnis moest buiten gezet worden maar Marie wilde het niet doen.*
the garbage had-to out set be but Marie wanted it not do
"The garbage had to be taken out, but Marie didn't want to do it."

Given these differences, if advanced Dutch learners of English can acquire the parallelism effect in English VPE and can reliably distinguish between VPE and VPA with respect to the strength of this effect, it is reasonable to conclude that such sensitivity is acquired or learned rather than transferred.[5]

## Japanese

It is controversial whether Japanese allows VPE—that is, whether a particular Japanese construction is best analyzed as ellipsis or as something else. On the one hand, Otani and Whitman (1991) claimed that Japanese does have VPE; in contrast, Hoji (1998) argued that the constructions in question are better analyzed as null object constructions (NOCs). The analytic problem stems from the fact that it is difficult to distinguish a phonetically empty VP from an elided one. Japanese requires overt verb movement to inflection (I), but it also permits phonetically null objects. As a result, many strings are amenable either to a VPE analysis, as shown in (10a), or to a NOC analysis, as illustrated in (10b) (for a discussion see Goldberg, 2002; Matsuo, 2007).

(10) a. VPE Analysis
[IP ~~[VP NP *t* ]~~ $\text{verb}_i$ ]
b. Null Object Analysis
[IP [VP *pro* $t_i$ ] $\text{verb}_i$ ]

Under either analysis, the Japanese construction that corresponds to English VPE differs from its English counterpart in two respects. First, Japanese VPE involves a repeated lexical verb rather than a dummy auxiliary (such as *do* or *have*). More importantly, in contrast with what is claimed for English, Japanese VPE is compatible with nonparallel passive antecedents; that is, the parallelism effect does not apply across the board. The relevant facts are illustrated in (11).

(11) a. *Gomi-wa dasare-nakereba-naranakatta, demo watasi-wa dasi-taku-nakatta.*
garbage-TOP take-out-PASS-have-to-PST but I-TOP take-out-want-NEG-PST
"??The garbage had to be taken out, but I didn't want to (take out)."
b. **John-wa kisu-ga hosikatta ga, Mary-wa si-taku-nakatta.*
John-TOP kiss-NOM want-PST but Mary-TOP do-want-NEG-PST
"*John wanted a kiss but Mary didn't want to."

https://doi.org/10.1017/S0272263109090044 Published online by Cambridge University Press

Japanese also differs from English with respect to VPA. As (12a) and (12b) illustrate, Japanese VPA is dispreferred with nonparallel (passive or nominal) antecedents.

(12) a. ?*Gomi-wa dasare-nakereba-naranakatta, demo watasi-wa soo*
garbage-TOP take-out-PASS-have-to-PST but I-TOP so
*si-taku-nakatta*.
take-out-want-NEG-PST
"The garbage had to be taken out, but I didn't want to do so."
b. ?*John-wa kisu-ga hosikatta ga, Mary-wa soo si-taku-nakatta*.
John-TOP kiss-NOM want-PST but Mary-TOP so do-want-NEG-PST
"John wanted a kiss but Mary didn't want to do it."

In the construction that corresponds most closely to English VPE, parallelism only applies in one of the two nonparallel constructions under consideration; in VPA contexts, parallelism applies more strongly than it is claimed to do in English. Thus, Japanese presents almost the reverse of the English situation.

## Spanish

It is usually assumed that Spanish lacks VPE. Nevertheless, the closest corresponding construction found in Spanish—null complement anaphora (NCA)—is subject to similar parallelism constraints as English VPE. Only a restricted set of modal predicates license NCA, and the resulting ellipsis is always nonfinite. Examples of Spanish NCA are provided in (13).

(13) a. ??*La basura tiene que sacarse, pero yo no quiero*.
The garbage has COMP take-out but I NEG want
"*The garbage has to be taken out, but I don't want to."
b. **Juan quería un beso, pero Maria no quería*.
John wanted a kiss but Mary NEG wanted
"*John wanted a kiss, but Mary didn't want (to)."

Spanish also has a construction that corresponds to English VPA, but this construction is excluded with nominal antecedents, as shown in (14b). Unlike Japanese or Dutch however, Spanish VPA is compatible with (medio-) passive antecedents, as illustrated in (14a).

(14) a. *La basura tiene que sacarse, pero yo no quiero hacerlo*.
the garbage has COMP take-out-REFL but I NEG want do-it
"The garbage has to be taken out, but I don't want to do it."
b. **Juan quería un beso, pero Maria no quería hacerlo*.
John wanted a kiss but Mary NEG wanted do-it
"John wanted a kiss, but Mary didn't want to do it."

https://doi.org/10.1017/S0272263109090044 Published online by Cambridge University Press

**Table 1.** Summary of crosslinguistic differences in VPE and VPA

| | Antecedent type | | | |
|---|---|---|---|---|
| Language and anaphor type | Active | Passive | Verbal | Nominal |
| English | | | | |
| VPE | √ | * | √ | * |
| VPA | √ | √ | √ | √ |
| Dutch | | | | |
| VPE | NA | NA | NA | NA |
| VPA | √ | ?? | √ | * |
| Japanese | | | | |
| VPE (NOC) | (√) | (√) | (√) | (*) |
| VPA | √ | ? | √ | ? |
| Spanish | | | | |
| VPE (NCA) | (√) | (*) | (√) | (*) |
| VPA | √ | √ | √ | * |

*Note:* NA corresponds to not available.

Table 1 summarizes the relevant crosslinguistic differences highlighted thus far.[6]

The relationship between parallelism and the VPE-VPA contrast is complicated. At the very least, the data show that parallelism effects are not restricted to surface anaphora constructions crosslinguistically because VPA constructions in Dutch, Japanese, and Spanish reportedly exhibit the same effect to varying degrees. This forces the conclusion that the influence of parallelism on English ellipsis and anaphora contexts is a subtle, language-particular, and often construction-specific property, making its acquisition by L2 learners especially challenging. Even if the principle is innate, as the semantic account claims, it has to be associated with different constructions in different languages.

## Summary

Our experiment examines learners' responses to violations of parallelism in English VPE and VPA contrasts. Note that although we were primarily concerned with the comparison between NSs and different groups of L2 learners, the NS controls were no less an experimental group. We did not necessarily expect any group of learners to show the types of categorical responses suggested by the traditional syntactic analysis. Instead, we expected that violations of structural parallelism might yield gradient judgments and that there could be an interaction between parallelism and the construction type of the antecedent clause, which would be consistent with

https://doi.org/10.1017/S0272263109090044 Published online by Cambridge University Press

either of the alternative accounts. With respect to the L1-L2 comparisons, our interest was in whether L2 learners' responses would differ from those of NSs as a function of surface constructional properties of their respective L1s—that is, the availability of VPE or VPA in the L1 and the presence or absence of parallelism effects associated with the two constructions (see note 4).

## EXPERIMENT

### Previous Psycholinguistic Work

The present experiment was modeled on previous psycholinguistic work conducted by Tanenhaus and Carlson (1990), who used the sentence completion task to assess NSs' sensitivity to parallelism violations. In this task, participants were presented with short stories that consisted of two sentences: a context sentence and a subsequent completion sentence. Participants read the context sentence and pressed a button once they understood it. The first sentence was then removed from the screen and replaced by the target completion sentence. At the end, participants were asked to decide as quickly and accurately as possible whether the second sentence formed a sensible completion of the story. To accept a given sentence-pair, participants pressed a button marked *yes* on a button box attached to the computer; to reject it, they pressed *no*. If readers' interpretations of ellipsis sentences were strongly constrained by the syntactic form of the antecedent, VPE completions will not make sense in a nonparallel context—for example, *The garbage had to be taken out / but I didn't want to* [*be taken out*]—and should be rejected.

The task yielded two dependent measures: an acceptability judgment and a reaction time. In all cases, the response time was measured from the onset of the second sentence: If the properties of the antecedent clause had no influence on the acceptability of the completion, then there should be no reliable difference in acceptability ratings or in latencies to accepted items across conditions. With respect to the acceptability judgments, participants were expected to accept a significantly higher proportion of target sentences with parallel antecedents than with nonparallel antecedents. As for response latencies, if parallelism influenced acceptability, then participants should accept ellipses with parallel antecedents significantly faster than those with nonparallel antecedents. (Only acceptances were counted in the analysis of latency data.)

Tanenhaus and Carlson (1990) reported the results of three separate experiments. The first two experiments contrasted VPE and VPA completions following active versus passive or verbal versus nominal antecedents, respectively. Table 2 presents the main results of these experiments.

As shown in Table 2, the analysis of the acceptability judgments in both experiments revealed a reliable effect of structural parallelism in the VPE condition as well as an interaction between parallelism and anaphor type,

https://doi.org/10.1017/S0272263109090044 Published online by Cambridge University Press

**Table 2.** Summary of Tanenhaus and Carlson's (1990) experiments 1 and 2

| | Acceptability judgment (%) | | Latency (ms) | |
|---|---|---|---|---|
| Experiment | Parallel | Nonparallel | Parallel | Nonparallel |
| Active versus passive[a] | | | | |
| VPE | 89 | 70 | 2,161 | 2,776 |
| VPA | 94 | 91 | 2,073 | 2,273 |
| Verbal versus nominal[b] | | | | |
| VPE | 89 | 71 | 2,556 | 2,923 |
| VPA | 86 | 86 | 2,686 | 2,954 |

*Note:* Parallelism effects were observed for the VPE in both acceptability judgments and latency data. However, no parallelism effects were observed for the VPA acceptability judgments in either experiment.
[a] For the latency data, there was an effect of type of anaphor as well as of parallelism; however, the interaction between anaphor type and parallelism was not significant.
[b] In Tanenhaus and Carlson's experiment 2, the only significant effect in the latency data was a (small) main effect of parallelism.

with VPE completions eliciting a larger effect than VPA. Analysis of the latency data was consistent with that of the acceptability judgments although the latter results did not often reach statistical significance.

One aim of the present experiment was to determine whether advanced L2 learners are able to converge on the same types of subtle judgments—that is, whether they come to control the distinction between surface and deep anaphora in English. (Our initial assumption about the theoretical distinction between surface and deep anaphors is called into question by the present results and suggests directions for future research [see Duffield, Matsuo, & Roberts, in press]).

## Method

The experiment reported here is modeled on Tanenhaus and Carlson's (1990) first two experiments.The main difference between this experiment and those of Tanenhaus and Carlson is that we treated the previous authors' two experiments as two levels of one variable—namely, construction type.[7]

## Participants

Our experiment compared the behavior of English NSs ($n$=22) with three groups of L2 learners: Dutch ($n$=20), Japanese (n=19), and Spanish ($n$=20). The experiment was run in three universities in the Netherlands, the United States, and Canada. All of the Japanese and Spanish participants were nonnative graduate students in English-medium universities in North America, in

https://doi.org/10.1017/S0272263109090044 Published online by Cambridge University Press

their second or third year of doctoral study, who had acquired English as a foreign language in their home countries prior to commencing graduate study. Although their English was not independently assessed, the entry requirements for graduate study in the North American institutions included a minimum TOEFL score of 550 (on the paper version). The Dutch L2 participants were undergraduate and graduate students at a Dutch university who were members of a L2 subject pool of the nearby research institute. Participants were paid a small fee for their participation.[8]

## Measures and Procedure

The sentence completion task used in the experiment was designed using PsyScope 1.2.5 (Cohen, MacWhinney, Flatt, & Provost, 1993) and run on Macintosh computers, with a serial button box attached. Aside from this technical (hardware and software) difference, we followed the procedure used by Tanenhaus and Carlson (1990).

The test materials were presented as two separate blocks: The first block included 20 quadruplet pairs of active versus passive VPE versus VPA sentences; the second block included 20 quadruplet pairs of verbal versus nominal VPE versus VPA sentences. An example quadruplet set from the first block is provided in (15) (see the Appendix for a full list of sentences). There were also 40 filler items. Four versions of the experiment were prepared, using a Latin square design, such that each participant received only one of the four pairs from each quadruplet set. The trial order of sets within each block was randomized for each participant.

(15) a. Active / VPE / (VPA).
*Someone took the wood out to the shed last night / Tom told us that Sally did / (it).*
b. Passive / VPE / (VPA).
*The wood was taken out to the shed last night / Tom told us that Sally did / (it).*

Thus, there were three between-items factors, each with two levels: structural parallelism (parallel vs. nonparallel), anaphor type (VPE vs. VPA), and construction type (active-passive vs. verbal-nominal). The between-subjects factor was native language: This factor had four levels in the main analysis (English, Japanese, Spanish, and Dutch).

## Predictions

In light of the results in Tanenhaus and Carlson (1990), we predicted a reliable main effect of syntactic parallelism overall as well as an interaction

https://doi.org/10.1017/S0272263109090044 Published online by Cambridge University Press

between parallelism and anaphor type (VPE vs. VPA), such that a significantly weaker effect was expected for VPA than for VPE.

The alternative theoretical accounts of parallelism yielded distinct predictions with respect to the discreteness of the acceptability judgments as well as the possible effects of construction type. Whereas the traditional syntactic account predicted clear rejection of ellipsis following nonparallel antecedents—with no significant contrast between passive and nominal antecedents—the semantic account was consistent with a much weaker rejection of nonparallel ellipsis overall as well as a preference for passive antecedents over nominal ones (passive antecedents being more likely to satisfy the mutual entailment condition). Although both grammatical accounts predicted full acceptance of nonparallel VPA, the parsing account is also consistent with parallelism effects in VPA (albeit weaker ones) because deep anaphora allows for—but does not require—a syntactic analysis of the anaphorically interpreted material.

With respect to possible L1-L2 contrasts, once more, the different accounts make various predictions. On the syntactic account, the parallelism effect is intimately tied to the structural properties of the VPE construction itself: Sensitivity to parallelism presupposes knowledge of the construction. Logically, it should be impossible to reject—or not to recognize—the English VPE construction but still show sensitivity to parallelism. On the semantic and processing accounts, however, sensitivity to parallelism and syntactic knowledge of particular constructions are dissociable: L2 learners might show a preference for parallel over nonparallel structures independently of their knowledge of particular syntactic constructions.

Under either theoretical account, we can investigate whether L2 learners' judgment patterns converge on those of NSs and to what extent any divergences can be attributed to the constructions available in the L1s of those learners. It should be clear that the semantic account predicts fewer language-specific effects because it rests on a putatively innate (universal) semantic principle. Nevertheless, the linking problem remains: The problem of correctly associating parallelism effects with particular constructions might be influenced by L1 properties.

## RESULTS AND DISCUSSION

The participants' responses yielded two measures: an acceptability judgment and a response time. Separate analyses were conducted on the proportion of sentence-pairs judged to make sense (the judgment data) and on the latencies for those sentences (the latency data). Note that both datasets come from the same set of responses. All latencies more than two standard deviations above a participant's mean were replaced with the appropriate two-standard-deviation cutoff, calculated separately for each participant.

https://doi.org/10.1017/S0272263109090044 Published online by Cambridge University Press

Participants reported no difficulty with the task: Nevertheless, one of the Japanese participants with very long response latencies was excluded from both analyses.

## Overall Results

The overall results, collapsing across the different learner groups, are presented in Table 3.[9]

***Judgment Data.*** An analysis of variance on the proportion of sentence-pairs judged to make sense revealed a main effect of parallelism, $F_1(1, 82) = 73.11, p < .0001$; $F_2(1, 152) = 41.11, p < .0001$. There were also reliable main effects of the three other factors: language, $F_1(1, 82) = 3.57$, $p < .05$; $F_2(3, 152) = 13.12$, $p < .0001$; construction type, $F_1(1, 82) = 9.53$, $p < .005$; $F_2(1, 152) = 4.53$, $p < .05$; and anaphor type, $F_1(1, 82) = 112.89$, $p < .0001$; $F_2(1, 152) = 107.82$, $p < .0001$. As Tanenhaus and Carlson (1990) also found, a reliable interaction was observed between syntactic parallelism and anaphor type, $F_1(1, 82) = 13.32$, $p < .0005$; $F_2(3, 152) = 8.56$, $p < .0001$. Nonparallel antecedents exerted a significantly stronger effect on the acceptability of VPE completions than on VPA completions, as illustrated in Figure 1a.

Crucially, reliable two-way interactions were observed between language and parallelism, $F_1(3, 82) = 6.49$, $p < .0005$; $F_2(3, 152) = 8.56$, $p < .0001$, and between language and anaphor type, $F_1(3, 82) = 3.24$, $p < .05$; $F_2(3, 152) = 5.6$, $p < .005$. There was also a three-way interaction among language, anaphor type, and parallelism, $F_1(3, 82) = 7.97$, $p < .0001$; $F_2(3, 152) = 7.77$, $p < .0001$. These

**Table 3.** Experiment overall means by factor

| Factor | Judgment (%) | Latency (ms) |
|---|---|---|
| Parallelism | | |
| Parallel | 84 | 3,180 |
| Nonparallel | 71 | 3,388 |
| Anaphor type | | |
| VPE | 67 | 3,382 |
| VPA | 88 | 3,185 |
| Construction type | | |
| Active-passive | 79 | 3,112 |
| Verbal-nominal | 74 | 3,457 |
| Language | | |
| English | 78 | 2,509 |
| Dutch | 84 | 2,188 |
| Japanese | 72 | 4,506 |
| Spanish | 75 | 3,911 |

*Note:* The reported means are based on a subject ANOVA.

https://doi.org/10.1017/S0272263109090044 Published online by Cambridge University Press

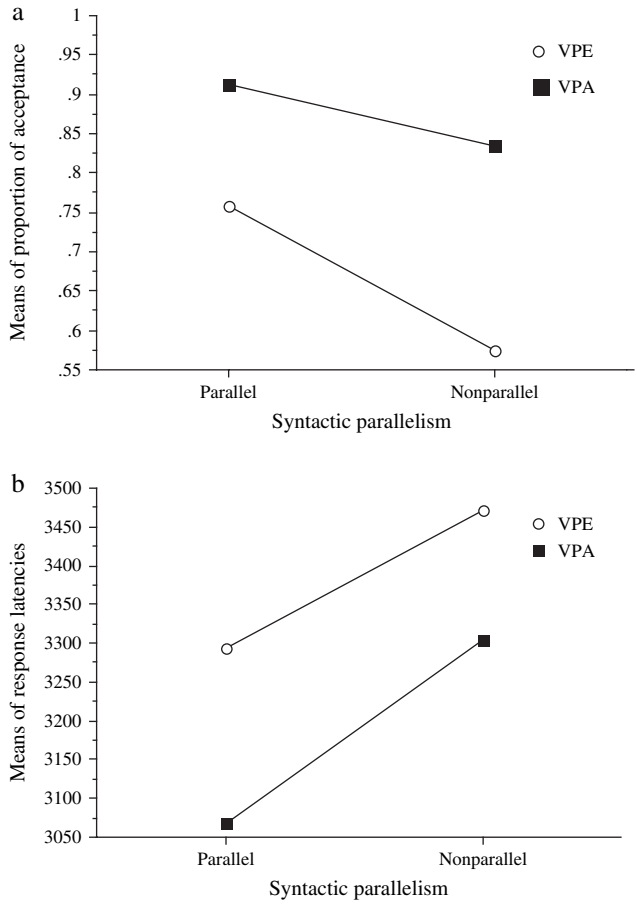



**Figure 1.** Interaction of syntactic parallelism and anaphor type: (a) judgment data and (b) latency data.

interactions clearly demonstrate that the various learner groups behaved differently from one another in this experiment.

***Latency Data.*** The latency data were largely consistent with the judgment results. Analysis of variance on the response latencies for sentence-pairs judged to make sense revealed significant main effects of construction type, $F_1(1, 81) = 23.94$, $p < .0001$; $F_2(1, 149) = 4.53$, $p < .05$, and of language, $F_1(3, 81) = 20.28$, $p < .0001$; $F_2(3, 149) = 126.8$, $p < .001$; that is to say, overall, active-passive sentences were accepted more quickly than verbal-nominal ones. Moreover, as predicted, some groups of learners responded more quickly than others (i.e., Dutch faster than English faster than Spanish faster than Japanese). Notice that the Dutch group responded on average faster than the NS control group.

https://doi.org/10.1017/S0272263109090044 Published online by Cambridge University Press

The main effects of parallelism and anaphor type were reliable by subject but not by item. In contrast to the judgment data analyses, no interaction was observed between language and any other factor (ignoring a marginal interaction of language and anaphor type, $p < .1$). Similar to Tanenhaus and Carlson (1990), we found no interaction between anaphor type and parallelism in the latency data: Increased latencies were observed across the board for pairs with nonparallel antecedents for both anaphor types (VPE and VPA), as is illustrated in Figure 1b.

In the absence of any reliable effects of parallelism in the latency data, our analysis focused solely on the judgment data.[10] Separate analyses of variance were performed on the judgment data in each language group.

## Native Speakers' Results: Judgment Data

Analysis of variance of the NSs' judgment data revealed significant main effects of parallelism, $F_1(1,21) = 72.16, p < .0001$; $F_2(1,152) = 69.47, p < .0001$, and anaphor type, $F_1(1,21) = 75.36, p < .0001$; $F_2(1,152) = 85.31, p < .0001$. NSs accepted pairs with parallel antecedents more often than those with nonparallel antecedents, and VPE completions more often than VPA completions. The effect of construction type was significant by subject although not by item, $F_1(1,21) = 5.18, p < .05$; $F_2(1,152) = 3.45, p = .065$. NSs tended to accept active-passive pairs more frequently than verbal-nominal pairs (see Table 4a).

As predicted by the syntactic account, there was also a significant interaction in the judgment data between parallelism and anaphor type, $F_1(1,21) = 38.36, p < .0001$; $F_2(1,152) = 31.0, p < .0001$; that is to say, the parallelism effect was significantly stronger for VPE than for VPA completions. This finding is supported by the results for the individual conditions. English NSs reliably differentiated between parallel and nonparallel antecedents to VPE completions in both constructions: active versus passive antecedents, $F_1(1,20) = 48.10, p < .0001$; $F_2(1,152) = 46.98, p < .0001$; verbal versus nominal antecedents, $F_1(1,20) = 43.82, p < .0001$, $F_2(1,152) = 42.52, p < .0001$. This shows that structural parallelism strongly influences the acceptability of VPE completions, as claimed by the standard theory. However, contrary to theoretical expectations, nonparallel VPA sentences also showed effects of parallelism, even though these effects were not statistically reliable in this experiment.

## L2 Learners' Results: Judgment Data

Analyses revealed divergences between the Dutch subjects on one hand—who behaved very similarly to the NS group—and the Spanish and Japanese groups on the other hand. Tables 4b–4d present the results for the Dutch, Japanese, and Spanish groups, respectively.

https://doi.org/10.1017/S0272263109090044 Published online by Cambridge University Press

**Table 4.** Acceptability judgments by group

| | Anaphor type | |
|---|---|---|
| Antecedent type | VPE | VPA |
| a. English NSs | | |
| First condition | | |
| Active | 90 | 95 |
| Passive | 48 | 90 |
| Statistical difference | $p < .0001$ | *ns* |
| Second condition | | |
| Verbal | 80 | 97 |
| Nominal | 40 | 87 |
| Statistical difference | $p < .0001$ | *ns* |
| b. Dutch NSs | | |
| First condition | | |
| Active | 89 | 90 |
| Passive | 74 | 90 |
| Statistical difference | $p < .05$ | *ns* |
| Second condition | | |
| Verbal | 82 | 95 |
| Nominal | 68 | 83 |
| Statistical difference | $p < .05$ | $p = .06$ |
| c. Japanese NSs | | |
| First condition | | |
| Active | 68 | 92 |
| Passive | 57 | 74 |
| Statistical difference | $p < .05$ | $p < .001$ |
| Second condition | | |
| Verbal | 62 | 87 |
| Nominal | 58 | 80 |
| Statistical difference | *ns* | *ns* |
| d. Spanish NSs | | |
| First condition | | |
| Active | 68 | 90 |
| Passive | 62 | 88 |
| Statistical difference | *ns* | *ns* |
| Second condition | | |
| Verbal | 69 | 85 |
| Nominal | 56 | 79 |
| Statistical difference | $p < .05$ | *ns* |

***Dutch.*** Separate analysis of the Dutch participants' judgment data reveals a main effect of parallelism, $F_1(1, 20) = 15.73, p < .001$; $F_2(1, 152) = 9.6, p < .005$, and anaphor type, $F_1(1, 20) = 11.66, p < .005$; $F_2(1, 152) = 13.18, p < .0005$. In contrast to the English NS results, no interaction was observed between parallelism and anaphor type (Table 4b). This is due to the fact that Dutch participants' data show roughly the same size of parallelism effect for either VPE

https://doi.org/10.1017/S0272263109090044 Published online by Cambridge University Press

or VPA completions overall. With respect to the effect of parallelism on the individual conditions, this was reliable for VPE in both constructions but not for VPA (although it approached significance in the verbal-nominal condition).

***Japanese.*** Analysis of the Japanese judgment data also reveals a significant main effect of parallelism, $F_1(1,23)=8.43, p<.01$; $F_2(1,152)=6.82, p<.01$, as well as a sharp effect of anaphor type, $F_1(1, 23)=33.03$, $p < .0001$; $F_2(1,152)=34.73, p<.0001$. No interaction was observed between parallelism and anaphor type. In contrast to the Dutch and English results—but similar to the Spanish results—the main effect of anaphor type here is due to a clear dispreference for VPE constructions generally; that is, it is independent of parallelism.

The Japanese results also showed in the subject analysis a reliable interaction between parallelism and construction type, $F_1(1, 23)=4.74$, $p < .05$; $F_2(1,152)=2.32, p=.1299$. Collapsing across anaphor types, the parallelism effect was observed for the active-passive contrast, although not for the verbal-nominal pairs. This is reflected in the analysis of the individual conditions shown in Table 4c. A clear parallelism effect was observed for active versus passive VPA pairs, with a milder effect for active versus passive VPE pairs; in contrast, no effect of parallelism was observed for either anaphor type in the verbal versus nominal condition although the trend was in the expected direction in both cases.[11]

***Spanish.*** Analysis of the Spanish participants' data also showed reliable main effects of structural parallelism, $F_1(1,18)=6.21, p<.05$; $F_2(1,152)=3.98$, $p<.05$, and anaphor type, $F_1(1,18)=9.31, p<.005$; $F_2(1,152)=46.72, p<.001$. As was the case for the Japanese group, the sharp effect of anaphor type was independent of parallelism, so that no interaction was obtained between these variables: Like the Japanese group, Spanish participants simply dispreferred VPE constructions in general.

The Spanish participants diverged from the Japanese in one respect: Whereas the Japanese group showed reliable effects of parallelism for both anaphor types in the active-passive condition, the only significant individual effect in the Spanish data was for verbal versus nominal VPE.

## Comparison with Previous Results

The extent to which these results match those of Tanenhaus and Carlson (1990) depends on the level of analysis. When one considers the overall results, there is substantial agreement between Tanenhaus and Carlson's results and our own. As in Tanenhaus and Carlson's experiments, significant main effects of structural parallelism were observed as well as a reliable interaction between parallelism and anaphor type. For both NSs and L2 learners overall, parallelism appears to exert a significantly stronger effect on VPE

https://doi.org/10.1017/S0272263109090044 Published online by Cambridge University Press

completions than on VPA. Furthermore, a significant main effect of parallelism was observed for each language group considered separately. This is clearly consistent with Tanenhaus and Carlson's results. Equally important, for NSs, it is consistent with the proposed theoretical contrast between deep and surface anaphora—at least when considered in purely statistical terms—inasmuch as the NSs display a reliable parallelism effect for VPE, but none for VPA. (There are distinct trends, but these are not statistically reliable in this experiment on the by-item analysis.) One might interpret these results, then, as providing validation of the theoretical distinction and as offering suggestive evidence that advanced L2 learners are able to acquire knowledge of the distinction between VPE and VPA.

However, more detailed analysis points at a different interpretation—namely, that the parallelism effect is orthogonal to the syntactic distinction between VPE and VPA, at least for L2 learners. This interpretation is supported by two sets of results. First, several groups of learners, including the Dutch and the Japanese—as well as the English NSs on the by-subject analysis—showed reliable parallelism effects in VPA contexts in at least one of the two constructions tested. For the Japanese learners, the parallelism effect for VPA with passive antecedents was actually stronger than the effect for the corresponding VPE condition. This shows that parallelism effects are not exclusively linked to cases of surface anaphora (VPE), as claimed by both grammatical accounts (syntactic and semantic).

The second consideration speaking against the syntactic interpretation of parallelism is the fact that Japanese and Spanish learners showed statistically reliable parallelism effects in certain individual conditions, even though they rejected VPE constructions overall. That is, even when implicitly considering all VPE sentences as only marginally acceptable, Japanese and Spanish learners nonetheless preferred English VPE sentences with parallel antecedents over those with nonparallel antecedents. Intuitively, the high levels of rejection of parallel VPE sentences suggest that these learners have not mastered the construction, yet they still show parallelism effects. The very least that one can conclude from this is that, for L2 learners, sensitivity to parallelism is not necessarily linked to accurate knowledge of the structural properties of the VPE construction. A less conservative conclusion would be that these results support the idea that parallelism itself has a nonsyntactic basis. The results are certainly more consistent with the alternative semantic approach advocated by Merchant (2001) than with the traditional syntactic approach (although the fact that effects are found with VPA speaks against Merchant's account as well).

## Interpreting L1 Versus L2 Differences

Analysis of the individual L2 learner groups also suggests that at least some L2 learners can acquire sensitivity to parallelism effects in English ellipsis

https://doi.org/10.1017/S0272263109090044 Published online by Cambridge University Press

independently of the surface syntactic properties of their L1. The comparison of these results with the crosslinguistic differences summarized in Table 1 shows that certain results cannot be explained by surface transfer. In particular, the Dutch learner group was remarkably successful in converging on the English NSs' performance in this experiment, in spite of the fact that Dutch does not have VPE constructions—and that it displays a moderate parallelism effect for VPA.[12]

Notwithstanding these overall similarities, at a more detailed level of analysis, the parallelism effect may have been influenced by different weightings of factors for English and Dutch groups. Specifically, the finiteness of the ellipsis clause appeared to interact with the construction type of the antecedent, playing a larger role in determining acceptability for the Dutch participants than for the English NSs, for whom syntactic parallelism per se was a sufficient condition for a reliable effect across conditions. In this experiment, however, we were unable to establish this interaction statistically because our materials were not constructed to control for finiteness. These results motivated a set of follow-up investigations, reported elsewhere (Duffield et al., in press).

As for the Japanese and Spanish groups, whereas many individual results are consistent with an account in terms of surface transfer, some are not. In particular, the Japanese subjects' results showed a significant parallelism effect for passive VPE sentences in English, even though no such effect is reported for the corresponding Japanese construction. Moreover, even for those conditions in which the parallelism effect was not statistically reliable, the L2 learners' results trend toward more closely approximating English NSs' judgments than those that would be predicted on the basis of the corresponding L1 properties.

## Effects of Gradience

The results of this experiment demonstrate rather clearly that the parallelism effect is gradient rather than categorical. Contrary to standard theoretical claims, participants did not categorically reject nonparallel antecedents for VPE completions. Instead—at least when the nonparallel antecedent was a passive VP—participants overall accepted such sentences more than half of the time (48% acceptance for NSs and 74% for the advanced Dutch participants). Although this is significantly less often than when the antecedent is parallel—the parallelism effect is statistically reliable in this condition—the acceptance rate for the nonparallel VPE is much higher than for the unacceptable versions of the filler sentences and certainly within the range of acceptability for other fully grammatical structures.

The traditional syntactic explanation for the parallelism effect fails to predict marginal acceptability: Indeed, it trades on the categorical unacceptability

https://doi.org/10.1017/S0272263109090044 Published online by Cambridge University Press

of such structures. In contrast, Merchant's (2001) semantic account makes no strong predictions of categorical rejection; if anything, it predicts that nonparallel structures should be judged more acceptable than they seem to be for NSs (more like the Dutch learners' responses). Thus, gradience speaks against the syntactic view and is more consistent with either a semantic interpretation or one in terms of parsing.

## A Methodological Objection

It might be argued that the gradient effect observed here is due to the task itself. In the sentence completion task, participants did not see the antecedent clause and the ellipsis clause simultaneously: When they judged the acceptability of the ellipsis, the antecedent clause had already disappeared. It is conceivable, therefore, that, some of the time, participants mentally transformed the nonparallel antecedents so that the mental representation of the antecedent clause was in fact parallel to the ellipsis. Obviously, they cannot always do this, otherwise no parallelism effect would have been observed in the judgment data. One way to test this would be to administer a corresponding offline acceptability judgment task in which both sentences are presented simultaneously and for which grammatical acceptability is judged on an interval scale. In follow-up experiments, therefore, participants were presented with both the computer-based task and a pencil-and-paper grammaticality judgment task using the same materials. The results of these follow-up experiments—reported in Duffield et al. (in press)—showed no significant differences between computer-based and fully offline presentations of the test stimuli.[13]

## Summary

In conclusion, the results of this experiment reveal that NSs and L2 learners diverge in a number of theoretically interesting ways in their sensitivity to violations of parallelism in English VPE. Whereas the judgments of NSs are largely consistent with the standard theoretical account of ellipsis constructions—including a fairly rigid distinction between deep and surface anaphora—the judgments of L2 learners suggest that the parallelism effect is independent of ellipsis and may not even be essentially syntactic in nature. Furthermore, the L2 learners' results imply that specific construction types are a significant factor in determining judgments, something that is unexpected on the standard account. These findings suggest future lines of investigation, some of which are taken up in Duffield et al. (in press; see also Duffield, Matsuo, & Roberts, 2007).

https://doi.org/10.1017/S0272263109090044 Published online by Cambridge University Press

## THEORETICAL IMPLICATIONS

The results of the present study have theoretical and methodological implications both for grammatical theory and, to a lesser extent, for theories of SLA. Considering the theoretical perspective first, our results speak against the traditional structural explanation for the parallelism effect proposed by Hankamer and Sag (1976). The gradient response to violations of parallelism shown by all participants as well as the apparent dissociation of parallelism effects from knowledge of the VPE construction observed in some L2 learners' results clearly suggest that the parallelism effect is independent of structural parallelism. In other words, syntax does not determine acceptability in this instance.

It is not clear, though, that these results fully support the alternative grammatical theory either (i.e., the semantic theory). Our results are consistent with Merchant's (2001) analysis, if one restricts attention to NSs' responses to VPE conditions. The semantic account is certainly compatible with gradient effects overall and with comparatively stronger parallelism effects for nominal over passive antecedents. However, the semantic account fails to explain other trends observed in the L2 learners' data, such as effects of finiteness and construction type. More important, however, is the fact that parallelism effects are also observed in VPA contexts for all participants, including NSs; for some L2 learners, these effects are stronger than for VPE, a result that is incompatible with either grammatical account.

In principle, there are two ways to resolve the discrepancy between our results and the predictions of Merchant's (2001) theory. The first is to adopt Merchant's semantic explanation of the parallelism effect but to detach it from claims about structure—that is, to treat it as a purely semantic constraint rather than as an identification condition requiring a particular structure to license it. The adoption of this explanation would then allow the semantic constraint to apply in VPA contexts as well, as suggested by the data. Notice that this move allows for the retention of the grammatical distinction between surface and deep anaphora, albeit at the cost of its primary diagnostic.

The more radical alternative would be to reject the notion that the parallelism effect has any grammatical basis whatsoever and to suggest instead that parallelism is better explained in terms of parsing preferences. A parsing explanation would be consistent with most of the results reported here. First, it would explain the gradient nature of the parallelism effect: If nonparallel ellipsis is not ungrammatical but only dispreferred, there is no reason to predict categorical rejection. Second, it explains the apparent effects of construction type (of the antecedent clause): Nominal antecedents inhibit the correct parse more strongly than passive VPs, which differ only in voice from the VP that is reconstructed in the ellipsis site. It is thus understandable if nominal antecedent VPEs are rejected more often than VPEs following passives. Third, a parsing explanation is consistent with the presence of (weaker) parallelism effects in VPA contexts as well: Just

https://doi.org/10.1017/S0272263109090044 Published online by Cambridge University Press

because deep anaphora does not require syntactic reconstruction does not mean that the parser may not attempt a syntactic parse; if it does, there should be effects of parallelism. Finally, and most significantly, the parsing explanation is consistent with the dissociation between parallelism and knowledge of the VPE construction shown by both Japanese and Spanish groups: A processing account explains why these learners were able to show sensitivity to parallelism while strongly rejecting the carrier construction (i.e., the VPE), irrespective of antecedent type.

Whatever appeal this processing account may have, however, it does not provide a complete explanation. The fact that other languages vary with respect to the strength of the parallelism effect associated with particular construction types cannot be explained by a parsing preference alone. These language-particular, construction-specific associations must somehow be encoded in speakers' linguistic competence. Thus, if further investigation reveals reliable interactions between parallelism and construction type in these other languages—something not tested in this experiment—and reliable effects of construction type in English for English NSs, both of these phenomena, then such results would suggest a more radical reconsideration of the relationship between grammaticality and acceptability as well as of the status of construction-specific rules in grammatical descriptions.[14] The current results, however, do not yet force such a revision, and we leave the matter open at this point (although see Duffield et al., 2007, in press).

The present study also has possible implications for models of SLA, even though it has been not the primary purpose of this article to address such issues. What the study shows, at least for this domain of grammatical inquiry, is that it is quite possible for L2 learners to come to behave like NSs in making very subtle acceptability judgments about complex syntactic phenomena without necessarily having any reliable knowledge about the structures themselves. The implications are reasonably clear: Convergence on subtle judgments about an uninstructed phenomenon at the syntax-semantics interface—be it sensitivity to minimal interpretive contrasts (see Dekydtspotter, Sprouse, & Anderson, 1997), binding (see Bruhn de Garavito, 1997), and aspectual differences (see Montrul & Slabakova, 2003), among others—does not in itself license any inference to L2 learners' access to Universal Grammar, unless we can unambiguously determine the underlying basis of those judgments. On this question there is some, albeit weak, consensus (Schütze, 1996; compare Allen & Seidenberg, 1999): To quote former United States Secretary of Defense Donald Rumsfeld (2002), this is a "known unknown."[15]

(*Received 23 January 2008*)

## NOTES

1. Grammaticality was not always regarded as a categorical notion: As Levelt, van Gent, Haans, and Meijers (1977) noted, theoretical researchers in the 1960s and early 1970s

https://doi.org/10.1017/S0272263109090044 Published online by Cambridge University Press

developed theories on degrees of ungrammaticality (Levelt et al. cited Chomsky, 1964, Katz & Postal, 1964, Lakoff, 1971, and Ziff, 1964). Note that the issue of categorical well-formedness judgments for sentences is separate from the question of discreteness of lexical categories (e.g., Aarts, 2004; Ross, 1973a, 1973b).

2. Sag and Hankamer (1984) revised their characterization of deep anaphora constructions, recasting deep anaphora as model-interpretive anaphora. Nevertheless, the essential distinction remains, as do the consequent claims for the (non)availability of parallelism effects in VPA.

3. In response to an anonymous *SSLA* reviewer's comment, this should not be taken to imply that ungrammatical sentences are never found in naturally occurring production data: It is well accepted that performance data is full of ungrammatical imperfections of various kinds. What distinguishes Hardt's (1993) examples from other types of production data is that these are generally taken from formal spoken and written registers where the surrounding context is agreed to be uniformly grammatical and where there is little evidence of processing complexity and no evidence of subsequent correction: This is not the case for other instances of ungrammatical usage.

4. The acceptability judgments of the foreign language data in this section were obtained from theoretical linguists who are NSs of the relevant languages or from the existing theoretical literature (e.g., Carretero, 1999; Otani & Whitman, 1991; Santos, 2006).

5. An anonymous *SSLA* reviewer has objected that this remark implies a naïve view of transfer. It is not our intention to claim that L2 learners cannot, or do not, transfer properties of their L1 grammar that are much more abstract than those associated with particular constructions or phrase-structure configurations. However, the point is that for whatever reason, parallelism effects are associated with different constructions in different languages—at least according to the theoretical literature (e.g., Johnson, 1997, 2001; Lobeck, 1995; Lopez, 1994)—and it is reasonable to suppose that learners may transfer these associations to their L2 grammar, just as they transfer particular lexical-semantic associations—for example, incorrectly assuming that the German word (*jemanden*) *erschießen* means "to shoot someone" rather than "to shoot someone dead" (a long persisting error of one of the authors)—particularly when such distinctions are not taught. Furthermore, on Hankamer and Sag's (1976) syntactic theory, the parallelism effect is the criterial property of the VPE construction, not a side effect of it. Therefore, it seems reasonable to ask whether not having that construction in one's L1 or L2 interlanguage grammar means that one is insensitive to parallelism effects. If the parallelism effect and VPE constructions are dissociable, it speaks against Hankamer and Sag's syntactic position. Thus, it seems justified to us to adopt an apparently superficial view of transfer that considers knowledge of particular surface constructions.

6. We remain agnostic as to whether Japanese has VPE: The tick marks here indicate acceptability of the corresponding Japanese construction in these contexts. Moreover, we temporarily assume the correctness of the theoretical claims that nonparallel VPE is ungrammatical and unacceptable in English, and that VPA is insensitive to parallelism.

7. There were two other points of divergence between our experiment and Tanenhaus and Carlson's (1990) experiments 1 and 2. First, we slightly modified some of the test sentences in order to lessen the lexical difficulty for L2 learners. Second, we were obliged to create our own filler items because Tanenhaus and Carlson contained no information on the original items. However, assuming that Tanenhaus and Carlson's results were intended to generalize beyond their particular stimulus set to other sentences of the same type, the comparison should still be valid.

8. An anonymous *SSLA* reviewer suggested that the absence of an independent proficiency measure for all participants invalidates the study and precludes any conclusions about L1 effects. Although the study may have been improved by the addition of some proficiency pretest—although this is debatable—there are three points to be made. First, it is unclear which standard instrument (vocabulary test, offline grammatical sentence completion test, or reading comprehension) would provide an appropriate measure of this kind of linguistic competence. Second, the L2 group that performed in the most nativelike fashion—the Dutch group—had the least favorable prospects in terms of informal objective proficiency measures: These were predominantly undergraduate students of English in a Dutch-medium university with limited daily use of spoken English. The Japanese and Spanish participants, in contrast, were typically older postgraduate students of linguistics, social, or political science who had spent at least a year immersed in English-medium academic study in an English-speaking environment. It is likely—although admittedly unproven—that the three groups would not have been distinguished by standard proficiency measures. Third, and most importantly, we do not make any specific claims about the relative influence of proficiency versus L1 grammar in explaining the results; we were simply interested in assessing the extent to which L2 learners and NSs converge in their judgments and in trying to understand any

https://doi.org/10.1017/S0272263109090044 Published online by Cambridge University Press

points of divergence. Given these aims, it seems useful to take the L2 speakers' L1 properties into account, particularly where they differ from those of English grammar.

9. The analyses were all carried out as subjects-and-items ANOVAs: $F_1$ refers to the ANOVA by subject; $F_2$ refers to the ANOVA by item. As is standard, an effect was only deemed to be statistically reliable where the corresponding *p*-values for both analyses were below the .05 criterion.

10. It is worth reiterating that the latency data in Tanenhaus and Carlson's (1990) experiment also only showed consistent trends, as opposed to reliable differences. The results of follow-up experiments reported elsewhere reveal reliable latency differences (see Duffield et al., in press).

11. These results may also speak to the theoretical debate on the proper analysis of Japanese VPE constructions. If these constructions were really instances of VPE, as opposed to NOC (or some other construction), it would seem reasonable to expect Japanese L2 learners to accept English VPE constructions more readily than the other two language groups. The fact that Japanese learners reject this construction as much as Spanish learners—which does not have VPE—could be taken to support the idea that Japanese does not have VPE either—that is, that Hoji's (1998) NOC analysis (or some other alternative) is correct.

12. In fact, as shown by our follow-up experiments, English speakers also display reliable parallelism effects for VPA: Thus, the Dutch participants were in fact better representative of the theoretical target judgments than the NSs (see Duffield et al., in press, for details).

13. We also presented a subset of the same stimuli in an experiment using eye-tracking to chart online reactions to ellipsis and anaphora completions. The results of this experiment suggest that readers reliably distinguish between VPE and VPA and that the results obtained here do not arise through transformation or misanalysis of one construction as the other (see Duffield et al., 2007).

14. Space constraints preclude elaboration here. For a specific theoretical proposal that could accommodate these results, see Duffield (2003).

15. "There are known knowns. There are things we know that we know. There are known unknowns. That is to say, there are things that we now know we don't know. But there are also unknown unknowns. There are things we do not know we don't know" (Rumsfeld, 2002).

## REFERENCES

Aarts, B. (2004). Modelling linguistic gradience. *Studies in Language*, *28*, 1–49.

Allen, J., & Seidenberg, M. S. (1999). The emergence of grammaticality in connectionist networks. In B. MacWhinney (Ed.), *The emergence of language* (pp. 115–152). Mahwah, NJ: Erlbaum.

Bruhn de Garavito, J. (1997). Verb complementation, coreference and tense in the acquisition of Spanish as a second language. In A. T. Perez-Lerouz & W. Glass (Eds.), *Contemporary perspectives on the acquisition of Spanish* (Vol. 1, pp. 167–188). Somerville, MA: Cascadilla Press.

Bybee, J., & Hopper, P. J. (Eds.). (2001). *Frequency and the emergence of linguistic structure*. Amsterdam: Benjamins.

Carretero, L. L. (1999). VP-ellipsis in English and Spanish and the features of auxiliaries. *Probus*, 11, 263–298.

Chomsky, N. (1957). *Syntactic Structures*. Berlin: Mouton de Gruyter.

Chomsky, N. (1964). Degrees of grammaticalness. In J. A. Fodor & J. J. Katz (Eds.), *The structure of language: Readings in the philosophy of language* (pp. 384–389). Englewood Cliffs, NJ: Prentice Hall.

Chomsky, N. (1982). *Some concepts and consequences of the theory of government and binding*. Cambridge, MA: MIT Press.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavior Research Methods, Instruments and Computers*, *25*, 257–271.

Cyrino, S., & Matos, G. (2007). Local licenser and recovering in VP ellipsis construction: Variation across languages and language varieties. *Journal of Portuguese Linguistics*, *4*, 79–112.

Dalrymple, M., Shieber, S., & Pereira, F. (1991). Ellipsis and higher-order unification. *Linguistics and Philosophy*, *14*, 399–452.

Dekydtspotter, L., Sprouse, R. A., & Anderson, B. (1997). The interpretive interface in L2 acquisition: The process-result distinction in English-French interlanguage. *Language Acquisition*, *6*, 297–332.

https://doi.org/10.1017/S0272263109090044 Published online by Cambridge University Press

Duffield, N. G. (2003). Measures of competent gradience. In R. van Hout, A. Hulk, F. Kuiken, & R. Towell (Eds.), *The lexicon-syntax interface in second language acquisition* (pp. 97–127). Amsterdam: Benjamins.

Duffield, N. G., Matsuo, A., & Roberts, L. (2007, November). *Knowing what to leave out, online: Evidence from eye-tracking*. Paper presented at the Second Workshop on Language Processing in First and Second Language Learners. Nijmegen, The Netherlands.

Duffield, N. G., Matsuo, A., & Roberts, L. (in press). Factoring out the parallelism effect in VP-ellipsis: English vs. Dutch contrasts. *Second Language Research*.

Fiengo, R., & May, R. (1994). *Indices and identity*. Cambridge, MA: MIT Press.

Frazier, L., Munn, A., & Clifton, C. (2000). Processing coordinate structures. *Journal of Psycholinguistic Research*, *29*, 343–370.

Frazier, L., Taft, L., Roeper, T., Clifton, C., & Ehrlich, K. (1984). Parallel structure: A source of facilitation in sentence comprehension. *Memory and Cognition*, *12*, 421–430.

Goldberg, L. (2002). *On the syntax of null direct objects in Modern Hebrew*. Unpublished manuscript, McGill University, Montreal.

Goldberg, L. (2005). *Verb-stranding VP-ellipsis*. Unpublished doctoral dissertation, McGill University, Montreal.

Grinder, J., & Postal, P. M. (1971). Missing antecedents. *Linguistic Inquiry*, *2*, 269–312.

Haïk, I. (1987). Bound vps that need to be. *Linguistic Inquiry*, *10*, 503–530.

Hankamer, J., & Sag, I. (1976). Deep and surface anaphora. *Linguistic Inquiry, 7*, 391–428.

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

Hoji, H. (1998). Null object and sloppy identity in Japanese. *Linguistic Inquiry*, *29*, 127–151.

Johnson, K. (1997). *When verb phrases go missing*. Unpublished manuscript, University of Massachusetts, Amherst.

Johnson, K. (2001). What VP-ellipsis can do, what it can't, but not why. In M. Baltin & C. Collins (Eds.), *The handbook of contemporary syntactic theory* (pp. 439–479). Oxford: Blackwell.

Katz, J., & Postal, P. (1964). *An integrated theory of linguistic descriptions*. Cambridge, MA: MIT Press.

Lakoff, G. (1971). Presuppositions and wellformedness. In D. D. Steinberg & L. A. Jakobovitz (Eds.), *Semantics: An interdisciplinary reader in philosophy, linguistics and psychology* (pp. 329–340). New York: Cambridge University Press.

Levelt, W. J. M., van Gent, J. A. W. M., Haans, A. F. J., & Meijers, A. J. (1977). Grammaticality, paraphrase, imagery. In S. Greenbaum (Ed.), *Acceptability in language* (pp. 87–101). Berlin: Mouton de Gruyter.

Lobeck, A. (1995). *Ellipsis*. Oxford: Oxford University Press.

Lopez, L. (1994). The syntactic licensing of VP-ellipsis: A comparative study of Spanish and English. In M. L. Mazzola (Ed.), *Issues and theory in Romance linguistics: Selected papers from the lingusitic symposium on Romance languages XXIII* (pp. 333–354). Washington, DC: Georgetown University Press.

Matsuo, A. (2007). Differing interpretations of empty categories in English and Japanese VP-ellipsis contexts. *Language Acquisition*, *14*, 3–29.

Matsuo, A., & Duffield, N. G. (2001). VP-ellipsis and anaphora in child language acquisition. *Language Acquisition*, *9*, 301–327.

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

Merchant, J. (2004). Fragments and ellipsis. *Linguistics and Philosophy*, *27*, 661–738.

Montrul, S., & Slabakova, R. (2003). Competence similarities between native and near-native speakers: An investigation of the preterite / imperfect contrast in Spanish. *Studies in Second Language Acquisition*, *25*, 351–398.

Newmeyer, F. (1983). *Grammatical theory: Its limits and possibilities*. Chicago: University of Chicago Press.

Newmeyer, F. (2003, January). *Grammar is grammar and usage is usage*. Address presented at Linguistic Society of America, Atlanta, Georgia. Retrieved April 15, 2007, from University of Washington Web site: http://faculty.washington.edu/fjn/Newmeyer_LSA_address.pdf.

Otani, K., & Whitman, J. (1991). V-raising and VP-ellipsis. *Linguistic Inquiry*, *22*, 345–358.

Pullum, G. (n.d.). Hankamer was! In S. Chung, J. McCloskey, & N. Sanders (Eds.), *Jorge Hankamer Webfest*. Retrieved April 15, 2007, from the University of California, Santa Cruz Web site: http://ling.ucsc.edu/Jorge/index.html.

Reali, F., & Christiansen, M. H. (2005). Uncovering the richness of the stimulus: Structure dependence and indirect statistical evidence. *Cognitive Science*, *29*, 1007–1028.

https://doi.org/10.1017/S0272263109090044 Published online by Cambridge University Press

Rizzi, L. (1986). Null objects in Italian and the theory of *pro*. *Linguistic Inquiry*, *17*, 505–557.
Ross, J. R. (1973a). A fake NP-squish. In C.-J. N. Bailey & R. Shuy (Eds.), *New ways of analyzing variation in English* (pp. 96–140). Washington, DC: Georgetown University Press.
Ross, J. R. (1973b). Nouniness. In O. Fujimura (Ed.), *Three dimensions of linguistic theory* (pp. 137–258). Tokyo: TEC Company, Ltd.
Rumsfeld, D. (2002, February). *Defense department briefing*. Retrieved June 16, 2008, from http://en.wikipedia.org/wiki/Unknown_unknown.
Sag, I. (1976). *Deletion and logical form*. Unpublished doctoral dissertation, MIT, Cambridge, MA.
Sag, I., & Hankamer, J. (1984). Towards a theory of anaphoric processing. *Linguistics and Philosophy*, *7*, 325–345.
Santos, A. L. (2006). *Minimal answers: Ellipsis, syntax and discourse in the acquisition of European Portuguese*. Unpublished doctoral dissertation, Universidade de Lisboa, Lisbon, Portugal.
Schachter, P. (1977). Does she or doesn't she. *Linguistic Inquiry*, *8*, 763–767.
Schütze, C. T. (1996). *The empirical base of linguistics*. Chicago: University of Chicago Press.
Tanenhaus, M., & Carlson, G. N. (1990). Comprehension of deep and surface verb phrase anaphors. *Language and Cognitive Processes*, *5*, 257–280.
Thornton, R., & Wexler, K. (1999). *Principle B, VP ellipsis, and interpretation in child grammar*. Cambridge, MA: MIT Press.
Wasow, T. (1972). *Anaphoric relations in English*. Unpublished doctoral dissertation, MIT, Cambridge, MA.
Williams, E. (1977). Discourse and logical form. *Linguistic Inquiry*, *8*, 101–139.
Ziff, P. (1964). On understanding utterances. In J. A. Fodor & J. J. Katz (Eds.), *The structure of language: Readings in the philosophy of language* (pp. 390–399). Englewood Cliffs, NJ: Prentice Hall.

# APPENDIX

## TEST STIMULI

1. Someone took the wood out to the shed last night.
   The wood was taken out to the shed last night.
   Tom told us that Sally did (it).
2. Mike should really change the oil in his car.
   The oil should really be changed.
   However he / Mike won't take the time to (do it).
3. John broke the antique vase which belonged to Mrs. Jones.
   The valuable antique vase which belonged to Mrs. Jones was broken by John.
   She was furious that he did (it).
4. Someone broke our basement window last night.
   Our basement window was broken last night.
   Sam thinks he knows who (did it).
5. An architect designed the elaborate conference room.
   The elaborate conference room was designed by an architect.
   He was paid a lot of money to (do it).
6. Jim and his friends attended the picnic.
   The picnic was attended by Jim and his friends.
   They were glad that they did (it).

https://doi.org/10.1017/S0272263109090044 Published online by Cambridge University Press

7. We had to replace our broken garbage disposal.
   Our broken garbage disposal had to be replaced.
   Jim tried to (do it), but he failed.
8. Al asked Sally to the concert.
   Sally was asked to the concert by Al.
   For weeks he had been planning to (do it).
9. When Jane returned from work, she found that someone had stolen her stereo.
   When Jane returned from work, she found her stereo had been stolen.
   She had no idea who (did it).
10. Someone shoveled our driveway last night.
    Our driveway was shovelled by someone last night.
    A neighbor told us that Tom did (it).
11. We asked the plumber if he could fix the sink.
    We asked the plumber if the sink could be fixed.
    He said he would try to (do it).
12. Because Mary was busy Tom sent the package.
    Because Mary was busy the package was sent by Tom.
    He had promised her that he would (do it).
13. Mary washed the clothes every weekend.
    The clothes were washed by Mary every weekend.
    Her older sister Sally refused to (do it).
14. Nobody had ever climbed the mountain before.
    The mountain had never been climbed before.
    Yesterday my friend Sally did (it).
15. Tom hadn't finished the report yet.
    The report hadn't been finished yet.
    Tom said that he hadn't been able to (do it).
16. Bill was supposed to feed the cat last night.
    The cat was supposed to have been fed by Bill last night.
    He fell asleep and forgot to (do it).
17. My brother Sam wrote a best-selling novel.
    A best-selling novel was written by my brother Sam.
    Our family was proud he did (it).
18. Somebody had to paint the garage.
    The garage had to be painted.
    Finally my younger sister Carol agreed to (do it).
19. The police are curious about who robbed the jewelry store.
    The jewelry store robbery is a mystery.
    They still don't know who did (it).
20. At the sports event the opera star sang the national anthem.
    At the sports even the national anthem was sung by an opera star.
    People were surprised he agreed to (do it).

https://doi.org/10.1017/S0272263109090044 Published online by Cambridge University Press