

# Second Language Research

<http://slr.sagepub.com>

---

## **Factoring out the parallelism effect in VP-ellipsis: English vs. Dutch contrasts**

Nigel Duffield, Ayumi Matsuo and Leah Roberts  
*Second Language Research* 2009; 25; 427  
DOI: 10.1177/0267658309349425

The online version of this article can be found at:  
<http://slr.sagepub.com/cgi/content/abstract/25/4/427>

---

Published by:



<http://www.sagepublications.com>

**Additional services and information for *Second Language Research* can be found at:**

**Email Alerts:** <http://slr.sagepub.com/cgi/alerts>

**Subscriptions:** <http://slr.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.co.uk/journalsPermissions.nav>

**Citations** <http://slr.sagepub.com/cgi/content/refs/25/4/427>

# Factoring out the parallelism effect in VP-ellipsis: English vs. Dutch contrasts

**Nigel Duffield, Ayumi Matsuo** *University of Sheffield* and  
**Leah Roberts** *Max Planck Institute for Psycholinguistics, Nijmegen*

Received August 2007; revised May 2008; accepted May 2008

Previous studies, including Duffield and Matsuo (2001; 2002; 2009), have demonstrated second language learners' overall sensitivity to a parallelism constraint governing English VP-ellipsis constructions: like native speakers (NS), advanced Dutch, Spanish and Japanese learners of English reliably prefer ellipsis clauses with structurally parallel antecedents over those with non-parallel antecedents. However, these studies also suggest that, in contrast to English native speakers, L2 learners' sensitivity to parallelism is strongly influenced by other non-syntactic formal factors, such that the constraint applies in a comparatively restricted range of construction-specific contexts. This article reports a set of follow-up experiments – from both computer-based as well as more traditional acceptability judgement tasks – that systematically manipulates these other factors. Convergent results from these tasks confirm a qualitative difference in the judgement patterns of the two groups, as well as important differences between theoreticians' judgements and those of typical native speakers. We consider the implications of these findings for theories of ultimate attainment in second language acquisition (SLA), as well as for current theoretical accounts of ellipsis.

**Keywords:** processing VP-ellipsis, surface and deep anaphora, syntactic parallelism, finiteness and grammaticality, Fundamental Difference Hypothesis, construction-based learning in SLA

Address for correspondence: Nigel Duffield, Department of English Language and Linguistics, University of Sheffield, Sheffield, S10 2TN, UK; email: n.g.duffield@sheffield.ac.uk

© The Author(s), 2009. Reprints and permissions:

<http://www.sagepub.co.uk/journalspermissions.nav>

10.1177/0267658309349425

## I Theoretical and experimental background

### 1 Theoretical issues

The phenomenon under investigation in this article is the parallelism effect in English VP-ellipsis (VPE), whereby an elided verb-phrase is judged unacceptable by many native speakers (NS) whenever it follows a non-parallel antecedent. The effect is exemplified by the difference in acceptability between (1a) vs. (1b), and (2a) vs. (2b) below (where  $\Delta$  indicates the elided material):

- 1) a. Someone had to clear out the cupboard, so John did  $\Delta$ .
- b. The cupboard had to be cleared out, so John did  $\Delta$ .
- 2) a. Amy wants someone to help her, but Kerry won't  $\Delta$ .
- b. Amy wants some help, but Kerry won't  $\Delta$ .

While the contrasts observed in (1) and (2) are undeniable, what is more open to debate and further empirical investigation is the source of these effects. The standard approach, which derives from the work of Hankamer and Sag (Hankamer and Sag, 1976; Sag and Hankamer, 1984), explains the parallelism effect in purely structural terms: on this approach, the problem with (1b) and (2b) is the syntactic form of the antecedents (*be cleared out/some help*, respectively), which renders them ineligible to serve as proper antecedents for LF-reconstruction of the elided material (or, at least, for a reconstruction with the intended interpretation).<sup>1</sup> This is schematized in (3) below.

- 3) a. \* The cupboard had to [be cleared out ~~the cupboard~~], so John did [be cleared out ~~the cupboard~~].
- b. \* Amy wants some help, but Kerry won't [~~some help~~].

Hankamer and Sag (1976) use this syntactic analysis of parallelism in VP-ellipsis to argue in favour of a broader categorical distinction between 'surface anaphora', where the syntactic form of the antecedent is crucial for well-formedness, and 'deep anaphora', where it is claimed to be irrelevant. The deep anaphora counterpart of VPE is VP-anaphora (VPA), illustrated in (4): here, according to Hankamer and Sag, (4b) and (4d) are perfect.

<sup>1</sup> Alternatively, for PF deletion of this material, depending on the particular analysis adopted. For present purposes it does not matter greatly which of the two syntactic processes – PF deletion or LF reconstruction – is involved in deriving VPE constructions: all that matters is the claim that the constraint is defined syntactically.

- 4) a. Someone had to clear out the cupboard, so John did it.
- b. The cupboard had to be cleared out, so John did it.
- c. Amy wants someone to help her but Kerry won't do it.
- d. Amy wants some help, but Kerry won't do it.

Once again, there is an undisputed difference in the relative acceptability of (1b)/(2b) (non-parallel ellipsis cases) vs. (4b)/(4d) (non-parallel anaphora cases): even though, as we shall show, parallelism plays a role in VPA as well, the observable effects are significantly weaker for VPA than for VPE (at least for English native speakers).

Notice that on this syntactic interpretation of the parallelism effect, (1b) and (2b) are (categorically) ungrammatical, since the underlying (LF or pre-deletion) representation violates a structural constraint. In terms of pure theoretical analysis, treating (1b) as ungrammatical rather than merely dispreferred is not especially problematic, since the proposed analysis in (3) is clearly unacceptable. However, with respect to speakers' intuitions about VPE utterances and – as importantly – for acquisition theory, this is a problem. There are several reasons for this. First, even for comparatively clear violations of structural parallelism, the judgements have long been contested within the theoretical literature. Chao (1987), for example, disputes the unacceptability of Hankamer and Sag's canonical example, given in (5), claiming that many native speakers judge it to be perfectly acceptable (see also Dalrymple *et al.*, 1991):

- 5) The oats had to be taken down to the bin, so Bill did.

Following on from this, Hardt (1993) documents a significant number of recorded examples of blatant violations of parallelism in spontaneous production data, of which the following are representative, the examples in (6) and (7) illustrating ellipsis with passive and nominal antecedents, respectively:

- 6) a. This information could have been released by Gorbachev, but he chose not to. (Daniel Schorr, National Public Radio broadcast, 17 Oct 1992) [Hardt (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 *et al.*, 1991) [Hardt (134)].
- 7) 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, 6 Feb 1992) [Hardt (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 email message) [Hardt (116)].

The point of these examples is that they are uncorrected and, indeed, that they appear at all: standard instances of ungrammaticality such as the EPP (Extended Projection Principle) and ECP (Empty Category Principle) violations in (8) are either virtually non-occurring in spontaneous production, or – where they are found – tend to be repaired by resumptive pronouns, as in (9), or by some grammatical paraphrase:

- 8) a. She believes \*(there) to be only one kind of chocolate worth eating.  
 b. \* She's the sort of writer who everyone wonders when \_\_\_\_ will win a Pulitzer.
- 9) ? She's the sort of writer who everyone wonders when she will win a Pulitzer.

The contrast between the examples in (6) and (7) and those in (8) demonstrates at the very least that violations of parallelism in ellipsis are perceived as more acceptable than violations of other grammatical constraints. The problem is that the standard generative account has no means of capturing this type of gradience in terms of grammaticality: at least since the mid-1970s, grammaticality has been held to be a discrete notion: sentences are either well formed or not; for discussion, see Duffield (2003).<sup>2</sup> Hence, the clear ungrammaticality of the representations in (3) leads to the prediction that there should be no difference in the acceptability judgements for violations of parallelism compared with other grammatical violations, contrary to fact. Of course, many grammatical theories are not concerned with modelling this type of variance in acceptability judgements, and there may well be cases where it is appropriate to gloss over contrasts between more and less acceptable utterances; for discussion, see Newmeyer, 1983; Schütze, 1996. In the present case, however, this is more problematic since the judgements on what are assumed to be the underlying representations for VPE sentences, namely, those as in (3) are clear – these are agreed to be strongly unacceptable – whereas the judgements for

<sup>2</sup> A reviewer takes issue with the claim that grammaticality is assumed to be a discrete notion in generative grammar. This reviewer points out, entirely correctly, that generative grammar has often invoked the notion of 'more severe' vs. 'less severe' violations of grammatical principles: strong vs. weak islands are an example of this. In other words, there are degrees of ungrammaticality, just as there are degrees of unacceptability. However, this does not mean that grammaticality is any less discrete in the sense intended here, since even weak violations are ungrammatical within generative theory. A line is crossed: a sentence either violates one (or more) constraints, or it does not. (Un)grammaticality, in short, is like pregnancy: a sentence can no more be semi-grammatical than a woman may be half-pregnant; this does not preclude the fact that in another sense, some women are more (heavily) pregnant, some sentences are more ungrammatical than others. Acceptability is completely different in this regard.

what is actually heard are much more equivocal. On the assumption that judgements are based on – and inform – grammatical analyses, the mismatch between the relative acceptability of non-parallel VPE and the unacceptability of the explicit analysis of this construction is at least a cause for concern.

The observation that parallelism violations occur relatively frequently in spontaneous production data also raises an important learnability issue. Like almost all syntactic constraints, the Parallelism constraint is a negative one: it excludes ungrammatical forms or interpretations rather than describing occurring forms. Now, at least within the generative literature, there is a consensus that negative constraints can be overridden – or reformulated, in the case of parameterized constraints such as Binding Principles – on the basis of positive evidence; see, for example, Crain (1993). In the case at hand, one might argue that there is at least as much positive evidence for the violability of the Parallelism constraint, as, for example, for the availability of inverse scope readings in English (or other non-isomorphic languages), or for long-distance construal of reflexives in Japanese, or for any other parameterized property that uses positive evidence for determining language-particular settings. In spite of this, violations of parallelism retain their status in end-state competence as marginally acceptable: not perfect, but not so terrible either.

A third factor suggesting that the parallelism effect may not be purely structural in origin is the fact that morpho-syntactic parallelism – and perhaps even stem parallelism, see directly below – is not a necessary condition for total acceptability. In English VPE, the inflectional features of the reconstructed verb-form may be non-identical to those of the antecedent: although agreement in voice is required, tense and agreement features may vary, such that (10b) and (10c) are just as acceptable as (10a):

- 10) 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].

Furthermore, in analyses of languages such as Hebrew and European Portuguese with so-called 'V-stranding VPE', in which the finite verb overtly raises out of the ellipsis phrase, it has been proposed that even stem identity is not required for well-formedness (for Hebrew, see Goldberg, 2005; for Portuguese, Santos, 2006; Cyrino and Matos, 2005). In (11) for example, from Santos (2006), the verbs in the antecedent clause

and ellipsis clause (indicated in bold) are lexically distinct, though they share a common argument-structure and closely related semantics:<sup>3</sup>

- 11) Eu **gostava** de convidar algumas pessoas; mas quem<sub>i</sub> é que tu  
 I like PREP invite some people; but who is it that you  
 want that I invite  
**queres**<sub>j</sub> [ t<sub>j</sub> [ que eu convide t<sub>i</sub> ] ] [Santos, Chapter 2 (120)]

'I would like to invite some people, but who do you want (me to invite)?'

A final argument against a purely structural interpretation of parallelism would be the finding that Deep Anaphora constructions, including VP-anaphora, also display parallelism effects. As noted above, for Hankamer and Sag (1976), and Sag and Hankamer (1984), sensitivity to parallelism is one of the criterial properties that distinguish the two types of anaphora: it is claimed that VPA does not exhibit parallelism effects precisely because the resolution of this type of anaphora does not make reference to syntactic form. Yet, as our previous studies indicate – and the present study confirms – VPA does show parallelism effects, albeit attenuated ones: there are differences between VPE and VPA, but not with respect to the presence or absence of parallelism. From this it follows that if the distinction between Surface and Deep Anaphora is to be maintained, then the locus of the parallelism effect must be in some domain common to both types of constructions: that is to say, it should not be in the syntax.

All of these arguments tend towards the conclusion that the parallelism effect is not best characterized as a syntactic constraint. A plausible alternative, more recently advanced, is that the effect has a semantic, and/or discourse related explanation (Merchant, 2001; Cyrino and Matos, 2005; Santos, 2006). Santos (2006), following Merchant (2001; 2004), argues that parallelism should be understood in terms of a semantic identification requirement on ellipsis. This requirement refers to a particular type of givenness that Merchant terms e-GIVENness (developing earlier proposals of Schwarzschild, 1999):

- 12) Focus condition on VP-ellipsis

A verb-phrase  $\alpha$  can be deleted only if  $\alpha$  is e-GIVEN [Merchant, 2001: 27]

<sup>3</sup> It might be objected that (11) is not really a case of VPE at all, but instead represents some type of (deep) null complement anaphora, in which case the absence of parallelism effects would be irrelevant to the present argument. Santos devotes an extensive section of her dissertation to arguing that elided sentences of this type – those involving *wh*-extraction – are necessarily VPE clauses, at least in European Portuguese. For present purposes, we will assume the validity of her claim.

E-GIVENness is defined as in (13), which involves a double entailment between the antecedent and the elided expression:

13) e-GIVENness

An expression E counts as e-GIVEN iff E has a salient antecedent A, and modulo  $\exists$ -type shifting,

- (i) A entails F-clo(E), and
- (ii) E entails F-clo(A)

On this semantic account, parallelism emerges as the direct consequence of the mutual entailment condition in E-GIVENness: in contrast to a one-way entailment, a condition of mutual entailment sets very narrow limits on the ‘possible semantic distance’ between the antecedent and the ellipsis clauses; typically, they will be truth-conditionally identical. To see this, consider the cases in (14):

14) Alice likes to eat ice-cream, and strawberries, and Harry does  $\Delta$  too.

✓  $\Delta$  (E) = [like to eat ice-cream, and strawberries]

\*  $\Delta$  (E) = [like to eat ice-cream]

If the E-GIVENness condition in (13) only required entailment (i), then (14b) would count as a deletable verb-phrase expression, since e likes to eat ice-cream, and strawberries entails e likes to eat ice-cream. However, by requiring mutual entailment, (14b) is excluded as a candidate for ellipsis in this context, since the converse does not hold (that is, e likes to eat ice-cream does not entail e likes to eat ice-cream, and strawberries).

Merchant’s semantic account, then, has the advantage of correctly including as potential VPE expressions verb-phrases that diverge from their antecedents with respect to inflectional properties, as in (10) – and even with respect to the verb stem itself in ‘V-stranding VPE’, as in (11) – whilst at the same time excluding most of the other cases prohibited by the standard syntactic approach. Merchant’s approach also respects the VPE/VPA distinction, and predicts differences in judgements between the two, since the E-GIVENness constraint only applies to ‘licensed’ VPE structures: deep anaphora constructions are not subject to this condition. That is to say, for Merchant and Santos, the semantic conditions on ellipsis that give rise to the parallelism effect are identification conditions, which only apply to structures licensed by independent syntactic licensing conditions; this is a distinction adopted from Rizzi (1986). If a given language fails to license VP-ellipsis in the first place – as is the case in French, for example – then this semantic



identification requirement will have no structure to apply to, hence no parallelism effects will be observable.

An important point to observe here with respect to language acquisition is that the semantic identification requirement is taken to be innately specified: what must be learned (from positive evidence) is whether or not a given language licenses VPE by having a functional head with the right type of abstract features, after which the identification requirement should automatically be triggered.

Before turning to the experimental issues, it is worth drawing attention to one other significant empirical difference between the two theoretical accounts with respect to active–passive mismatches, as in (1) above: whereas Hankamer and Sag predict clear parallelism effects here, Merchant’s account does not – or at least, not necessarily – since in most (non-scope-related) cases, active and passive paraphrases should be mutually entailing.

In summary, the theoretical literature contains two main proposals for deriving the parallelism effect observed in VPE languages: the original syntactic approach developed by Hankamer and Sag, which explains the effect in terms of identical phrase structure; and a more recent, finessed, semantic approach, where parallelism emerges as a necessary consequence of a mutual entailment condition on the interpretation of the antecedent clause and the elided verb phrase. Both accounts derive canonical parallelism effects equally satisfactorily, and both maintain a clear distinction between VPE and VPA, asserting that parallelism effects are not found with deep anaphors; where they differ is in more marginal cases, where the semantic approach – correctly, it seems – ignores mismatches in morpho-syntactic features and other non-semantic structural properties.

With this theoretical background in place, we turn to consider some previous experimental work examining speakers’ sensitivity to parallelism from a psycholinguistic perspective.

## 2 *Experimental studies*

The first psycholinguistic research investigating the Parallelism Constraint in VPE was conducted by Tanenhaus and Carlson (1990) in an article that provides the model for follow-up studies, including Mauener *et al.* (1995), Duffield and Matsuo (2001), Duffield and Matsuo (2009), as well as the current study. These authors used the Sentence

Completion Judgement (SCJ) task to assess native speakers' sensitivity to the parallelism effect. Since all of the follow-up studies use a modified form of this task, it is worth discussing its methodology in some detail.

In the SCJ task, participants are visually presented on a computer screen with short stories consisting of two sentences: a context sentence, and a following completion sentence. The participants read the first sentence, and press a button when they have understood it; the context sentence is removed from the screen, and replaced by the target sentence. Participants are then asked to decide 'as quickly and accurately as possible' whether the second sentence forms a 'sensible and accurate completion' of the story. The stories manipulate (Structural) Parallelism and Anaphor-Type: the antecedent VP is either structurally parallel to that of the second sentence, or not; the completion involves either VPE or VPA. An example stimulus set is given in Table 1.

To accept the second sentence as a good completion, participants press a button marked 'Yes' on a button box attached to the computer; to reject it, they press a 'No' button. The task yields two dependent measures: an Acceptability Judgement and a Response Latency (reaction time).<sup>4</sup> Participants' sensitivity to parallelism is thus reflected in two ways. On the acceptability measure, participants are expected to accept a significantly higher proportion of target sentences with parallel antecedents than with non-parallel antecedents. With respect to response latencies, if parallelism influences acceptability, participants

**Table 1** Manipulating parallelism and anaphor type in the SCJ task

Condition (parallelism/ anaphor type)	First sentence	Second sentence
Active/VPE	Someone had to put out the garbage.	But John didn't want to.
Passive/VPE	The garbage had to be put out.	But John didn't want to.
Active/VPA	Someone had to put out the garbage.	But John didn't want to do it.
Passive/VPA	The garbage had to be put out.	But John didn't want to do it.

<sup>4</sup> In all cases, the response latency is measured from the onset of the second sentence: if the properties of the antecedent sentence has no influence on the acceptability of the completion, then there should be no significant difference in response times across conditions (except for the difference between response times to accepted vs. rejected items).

are expected to accept ellipses with parallel antecedents significantly more quickly than they accept those with non-parallel antecedents. (Only acceptances are counted in the latency analysis.)

It should be clear that this task is not an explicit grammaticality judgement task: participants are not being asked to make meta-linguistic judgements of grammatical acceptability; rather, the task provides an implicit measure of grammaticality. The quadruplet stimulus sets are so constructed that only formal grammatical properties distinguish the control (parallel) pairs from the test pairs (non-parallel); hence, to the extent that significant differences are found in the responses to the two conditions, the task can be claimed to provide a direct measure of implicit grammaticality.

*a Tanenhaus and Carlson, 1990:* In their 1990 article, Tanenhaus and Carlson report the results of three separate SCJ experiments. Their first two experiments contrast VPE and VPA in active vs. passive and verbal vs. nominal antecedents, respectively.<sup>5</sup> Table 2 presents the main results of those experiments. As the table shows, analysis of the

**Table 2** Summary of Tanenhaus and Carlson (1990): Experiments 1 and 2

Dependent measure	Anaphor type	Antecedent type		Reliable parallelism effect?
		Parallel	Non-parallel	
<i>Experiment 1: Active vs. passive antecedents:</i>				
Accepted completions (%)	VPE	89	70	Yes
	VPA	94	91	No
Latency (msecs)	VPE	2161	2776	Yes <sup>a</sup>
	VPA	2073	2273	
<i>Experiment 2: Verbal vs. nominal antecedents:</i>				
Accepted completions (%)	VPE	89	71	Yes
	VPA	86	86	No
Latency (msecs)	VPE	2556	2923	Yes <sup>b</sup>
	VPA	2686	2954	

*Notes:* <sup>a</sup>For the latency data, there were main effects of Anaphor-type and also of Parallelism, but no reliable interaction between the two variables. <sup>b</sup>In Tanenhaus and Carlson's second experiment, the only significant effect in the latency data was a (small) main effect of parallelism.

<sup>5</sup> The labels 'active' and 'verbal' refer to exactly the same kind of parallel control clause: (active) (verbal) verb-phrases that should function as perfect antecedents for VPE.

acceptability judgements in both experiments revealed a significant effect of structural parallelism in the VPE condition, as well as an interaction between Parallelism and Anaphor Type, with VPE completions eliciting a larger parallelism effect than VPA. Analysis of the response latency results was consistent with the acceptability judgements, though the latter did not generally reach statistical significance. In other words, consistent with standard theoretical claims, both measures confirmed a reliable difference between VPE and VPA, with only VPE completions showing reliable sensitivity to parallelism.

*b Duffield and Matsuo, 2009:* The main aim of the study reported in Duffield and Matsuo (2009) was to determine whether or not advanced second language learners would display comparable effects to those of native speakers; that is, whether L2 speakers of English fully controlled the distinction between surface and deep anaphors with respect to parallelism. In a modified version of Tanenhaus and Carlson's original experiments, Duffield and Matsuo compared the performance of two groups of L2 speakers – Dutch and Japanese learners of English – with that of native speakers.<sup>6</sup> The principal difference between Tanenhaus and Carlson's original experiments and Duffield and Matsuo's replication relates to the experimental design: rather than running two separate experiments, we treated Tanenhaus and Carlson's Experiments 1 and 2 as two levels of one between-items factor, namely, Construction Type.<sup>7</sup>

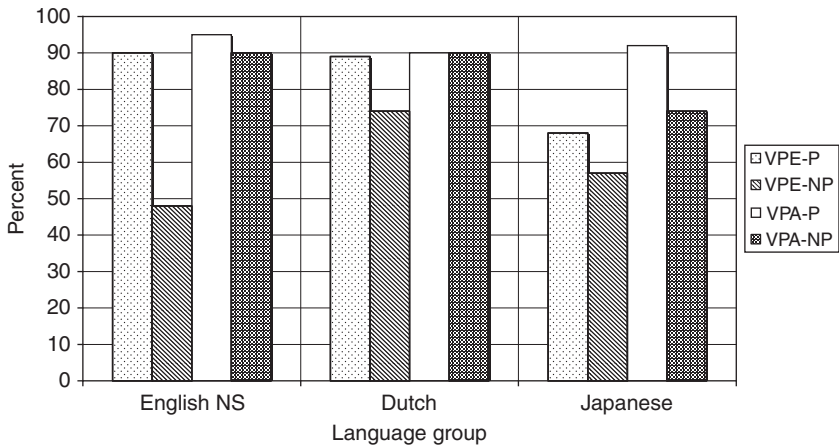
The selection of the two L2 populations was motivated by the properties of their respective first languages. We were primarily interested in examining possible effects of L1 transfer: we wanted to compare Dutch, which is generally agreed not to license VPE, with Japanese, where the theoretical literature is more divided. Otani and Whitman (1991) assert that Japanese has (V-stranding) VPE, whereas Hoji (1998) has

<sup>6</sup> In fact, our previous study also included a group of Spanish learners of English. This group is not discussed further, since, for unclear reasons, Spanish learners generally rejected VPE completions across the board, irrespective of parallelism; hence, their results reveal nothing significant about the interaction of syntactic parallelism with deep and surface anaphors.

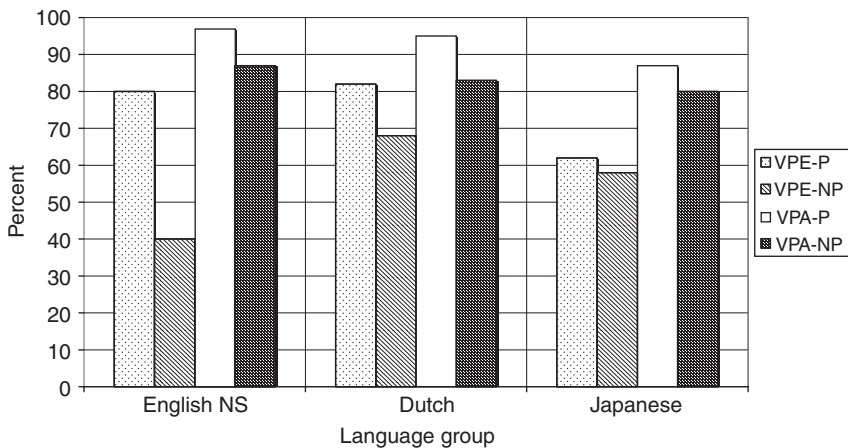
<sup>7</sup> There were two other points of divergence between Duffield and Matsuo (2009) and Tanenhaus and Carlson's Experiments 1 and 2. First of all, we slightly modified some of the test sentences in order to lessen the lexical difficulty for second language learners. Second, we were obliged to create our own filler and distracter items, since Tanenhaus and Carlson contains no information on the original fillers. For these reasons, our experiment was only a partial replication. 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.

claimed that structures corresponding to English VPE should instead be analysed as Null Object Constructions.

As it turned out, the results of the earlier study did not yield any definitive conclusion with respect to transfer effects, since it was the Dutch L2 learners (without VPE in their L1) that performed better in the experiment, reliably distinguishing VPE from VPA, and displaying clear sensitivity to parallelism. As shown in the Figures 1 and 2, Dutch learners performed very similarly to English native speakers overall.



**Figure 1** Interaction of parallelism with anaphor type: active vs. passive antecedent clauses



**Figure 2** Interaction of parallelism with anaphor type: verbal vs. nominal antecedents

Separate analysis of the Dutch participants judgement data revealed a main effect of Parallelism [ $F_1(1,20) = 15.73, p < .001; F_2(1, 152) = 9.6, p < .005$ ] and Anaphor Type [ $F_1(1,20) = 11.66, p < .005; F_2(1, 152) = 13.17, p < .0005$ ]. As for the effects of parallelism in the individual conditions, this was reliable for VPE in both constructions ( $p < .05$ ), but not for VPA, although this approached significance in the Dutch data in the verbal–nominal condition ( $p = .06$ ). The Japanese learners, by contrast, only showed reliable sensitivity to parallelism in the active vs. passive condition.<sup>8</sup>

In summary, our initial study confirmed the hypothesis that advanced L2 learners of English can come to show sensitivity to parallelism in VP-ellipsis constructions, even in cases where the L1 in question does not license VPE. On closer examination, however, these results raised at least as many issues as they resolve, since the L2 learners appeared to diverge in interesting respects from those of native speakers. Not only that: the native speakers' results, though broadly consistent with traditional theoretical analyses, displayed a number of unexpected trends that challenge standard assumptions, especially those concerning the Surface/Deep Anaphora distinction.

### 3 Discussion

The first point of contrast between native speakers and L2 learners is the apparent effect of construction type in the L2 data: whereas the native speakers and Dutch learners show roughly the same size of parallelism effect for either type of non-parallel antecedent (passive or nominal) in the VPE condition, the Japanese learners only showed a reliable parallelism effect ( $p < .05$ ) for passive antecedents. Although other explanations are of course possible – see below – this at least hints at the possibility that some L2 learners refer to construction-specific rules or templates in judging well-formedness.

Related to this is the observation that all groups show some parallelism effects in VPA, as well as VPE, contrary to the predictions of either theoretical analysis. In the case of English native speakers, these effects are so small relative to those in the VPE condition that there is a reliable interaction between Parallelism and Anaphor Type ( $F_1, F_2, p < .001$ ); this replicates the findings in Tanenhaus and

<sup>8</sup> In this study, there were again no reliable effects in the latency data. For further details of the data and analyses, see Duffield and Matsuo (2009).

Carlson's original experiment. For the Dutch learners, however, no similar interaction is observed between Parallelism and Anaphor Type ( $F_1 p = .1428$ ,  $F_2 p = .1528$ ), since the Dutch participants' data show roughly the same size of parallelism effect for either VPE or VPA with nominal antecedents.<sup>9</sup> The same is true for the Japanese learners, who display at least as large parallelism effects for VPA as for VPE (indeed, the effects are larger with nominal antecedents, though neither one is reliable in this condition).

A third finding, which emerged from inspection of the responses to individual stimulus items, was an apparent effect of finiteness on the acceptability of non-parallel antecedents. Standard theoretical accounts do not distinguish between non-finite ellipsis clauses: those involving *to* as in (15a) vs. finite ellipsis involving *do* (or some other auxiliary verb, such as *has*, as in (15b):

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

That is to say, the parallelism effect is generally claimed to affect finite and non-finite ellipsis clauses equally. However, based on the results of the earlier study, as well as our own intuitions, it was felt that the observed parallelism effect was weaker with non-finite ellipsis. Although the results of the first experiments were consistent with this intuition, we were unable to confirm this statistically since the materials in that experiment did not properly control for the finiteness distinction. Therefore, for the current study, we constructed materials with balanced sets of finite and non-finite ellipsis clauses in each condition.

A final point to observe about the first experiment, which was also true of Tanenhaus and Carlson's original experiment, is that all the effects here – even the statistically reliable ones – are gradient: however large the effects of parallelism might be for particular participant groups or individual conditions, the non-parallel versions are never categorically rejected, and remain reliably more acceptable than the unacceptable control items. As discussed earlier, this fact seems to speak against the traditional syntactic analysis, which predicts more categorical rejection of ungrammatical items.

<sup>9</sup> Though not with passive antecedents, where Dutch learners behave like English native speakers; indeed, better than native speakers (see Table 2: Experiment 1). This again points to the involvement of construction-specific templates.

The current study, then, systematically manipulates the non-syntactic formal factors of construction-type and finiteness in an effort to identify more precisely the sources of the parallelism effect, and to determine the nature and extent of L1 vs. L2 contrasts.

In addition to these factors, the present study also explores the role of recoverability in affecting acceptability judgements. As discussed earlier, the semantic analysis of the parallelism effect does not actually require any structural parallelism between the antecedent and the elided material: it is simply the case that structural identity is the optimal means of satisfying the mutual entailment requirement. If this is correct, it is reasonable to suppose that there should be no parallelism effect for syntactically non-parallel antecedent clauses, such as the passive antecedents of an active VPE, just in case the intended VPE interpretation is semantically recoverable from the antecedent clause (and the mutual entailment requirement is satisfied). In the current study, this notion of semantic recoverability was operationalized in terms of *by*-phrase addition, as illustrated by the contrast in (16). We created two sets of items, with and without *by*-phrases, nested within the structurally non-parallel passive condition, to see whether improving the chances of mutual entailment in this way would lead to an increase in acceptability, independently of syntactic parallelism:

- 16) a. Mary was busy, so the package was sent by someone. (with *by*-phrase) – ?  
A neighbour told us that Tom had.
- b. When we got back, our driveway had been cleared of snow. (no *by*-phrase). – ??  
A neighbour told us that Tom had.

Within the nominal condition, we investigated a separate kind of recoverability, namely, morpho-syntactic recoverability. Following a suggestion originally due to Tom Roeper, we examined whether non-parallel antecedents containing zero-derived nominals, such as *mention*, might be more easily reconstructed as stem-identical verbs in the ellipsis clause than affixed nominals, such as *discussion*. If this were the case, we would predict that the former non-parallel antecedents should be more acceptable than the latter (holding syntactic parallelism constant): that is to say, sentence pairs such as (17a) below should be more acceptable than those such as (17b). To test this, half of the non-parallel nominal antecedents contained zero-derived nominals, the other half contained affixed nominals.

- 17) 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.



(Since both types of recoverability were nested within the non-parallel level of the parallelism condition and apply to different construction-types, these are analysed and reported separately from the main analysis; see Appendix 1.)

The final modification in the current study was the replacement of the finite auxiliaries *do/does/did* by *has/have/had*. This was to control for the possibility that the relative success of Dutch L2 learners in the earlier experiment, compared to Japanese learners, was due to their re-analysing English VPE as VPA. When the ellipsis clause involves the preterite *do*, which was the case in the earlier studies, the only difference between VPE and VPA is the presence or absence of the pro-forms (*it* or *so*). It is conceivable that Dutch L2 learners might have analysed this as an instance of VPA, especially since Dutch also has VPA with a cognate form of *do* (*doen*). Such an interpretation is clearly less likely if the ellipsis clause contains a perfect auxiliary. For this reason, we changed the grammatical aspect of both sentences in the test items, replacing preterite forms of *do* with present perfect auxiliaries (*have, has*), as in (17b), both to test this potential confound of the first experiment and, at the same time, to test the generality of the constraint: that is, to determine whether it applies equally to all types of finite ellipsis clause, regardless of tense and aspect.

- 18) a. The garbage had to be taken out. Yesterday, Tom did.  
 b. The garbage had to be taken out. Till now, no one has.

## II Experiment

The current study comprised two separate experiments: the first using the timed methodology employed in the previous study, the second, a traditional pen-and-paper acceptability judgement task (using the same materials). The pen-and-paper task was administered immediately following the SCJ, and involved participants' judging the acceptability of the completion clause on an interval scale from 1–5 (1 being the least acceptable). In what follows, we refer to the two tasks – for convenience, if somewhat misleadingly – as ‘online’ and ‘offline’ tasks, respectively.<sup>10</sup>

<sup>10</sup> We are well aware that the SCJ task is not an online task in the strict sense. The distinguishing characteristic of a true online task is that it is concerned with the time-course of language comprehension: the interest is not only in what types of implicit information determine a given response, but also in when these different types of information make their contribution. Strictly speaking, the experiment reported here is only ‘pseudo-online’, since the (only) probe point is at the end of each stimulus pair.

### 1 Method

In total 256 experimental items were constructed (64 quadruplet sets), half of which involved the Construction Type 1 stimuli – active vs. passive antecedents – and the other half the Construction Type 2 stimuli, i.e. the verbal vs. nominal antecedents. The passive and the nominal antecedent types constituted the structurally non-parallel, and the active and verbal types the structurally parallel conditions (Syntactic Parallelism). For each antecedent type, there were a balanced number of VPE vs. VPA completions (Anaphor Type). We counterbalanced the finiteness of the ellipsis clause within items, leading to half the parallel and half of the non-parallel items with a finite ellipsis clause and the other half with a non-finite clause (Finiteness). Finally, within the non-parallel conditions were the two nested factors of Recoverability: antecedents with and without *by*-phrases in the case of passive antecedents; and zero-derived vs. non-zero-derived nominals in the nominal conditions. The overall experimental design is schematized in Table 3.

The stimuli were organized in a Latin Square design such that each participant received only one version of each quadruplet set (with all versions being presented an equal number of times); that is, each participant was presented with a quarter of the 256 experimental items (= 64). In addition, there were a matching number of distractor/filler items, of which half involved sensible completions to the first sentence, while the other half did not.

Twenty-nine English speakers and 20 Dutch L2 learners took part in the online task; in the offline task, there was one fewer Dutch participant ( $n = 19$ ).

### III Analysis and results

The results for the online and offline tasks were analysed separately; within the online experiment, the two dependent variables – acceptability rate and response latency – were also treated separately. For all of the data, we ran Analyses of Variance with four within-participants factors, each with two levels: Syntactic Parallelism (parallel/non-parallel), Construction Type (active–passive/verbal–nominal), Anaphor Type

---

This contrasts, for example, with tasks such as self-paced reading, or eye-tracking, where measures are taken at different points in stimulus presentation. However, it has become common practice to refer to any computer-based task involving reaction times as online, and we adopt this practice here.

**Table 3** Experimental design (summary of stimulus items)

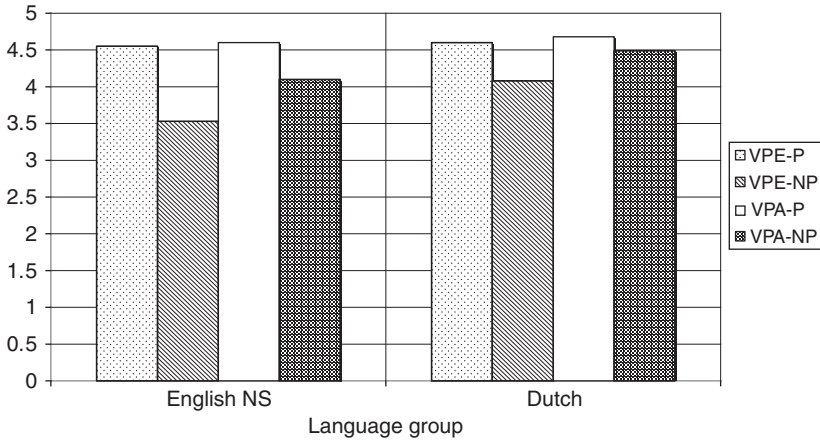
Structural parallelism	Recoverability (nested)	Anaphor type	Finiteness
<i>Construction Type 1:</i>			
Active (Parallel) (64)	–	VPE (32)	Finite (20) Non-finite (12)
		VPA (32)	Finite (20) Non-finite (12)
Passive (Non-parallel) (64)	<i>by</i> -phrase (32)	VPE (16)	Finite (10) Non-finite (6)
		VPA (16)	Finite (10) Non-finite (6)
No <i>by</i> -phrase (32)	No <i>by</i> -phrase (32)	VPE (16)	Finite (10) Non-finite (6)
		VPA (16)	Finite (10) Non-finite (10)
<i>Construction Type 2:</i>			
Verbal (Parallel) (64)	–	VPE (32)	Finite (20) Non-finite (12)
		VPA (32)	Finite (20) Non-finite (12)
Nominal (Non-parallel) (64)	Zero-derived (32)	VPE (16)	Finite (10) Non-finite (6)
		VPA (16)	Finite (10) Non-finite (6)
	Non-zero derived (32)	VPE (16)	Finite (10) Non-finite (6)
		VPA (16)	Finite (10) Non-finite (6)

(VPA/VPE) and Finiteness (+/–), which was treated as a between-items factor in the items analysis. The between-participants factor of Language Group (English natives/Dutch L2 learners) was treated as a within-participants factor in the items analysis. We first consider the results of the offline task, turn attention to the online data.

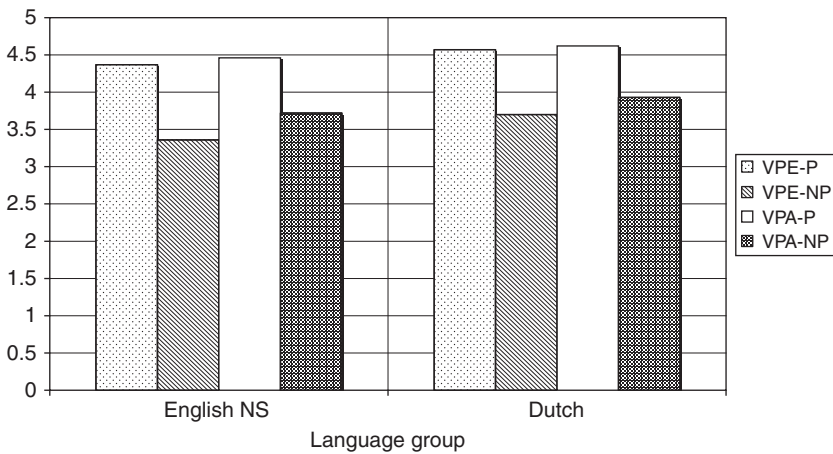
### 1 Offline task: judgement data

Figures 3 and 4 show the mean judgement scores for English native speakers and Dutch learners in the two Construction Types. For the moment – and for comparability with Figures 1 and 2 – we collapse across levels of finiteness, and also ignore the nested recoverability factors.

*a Overall analysis (groups collapsed):* As can be seen from the mean judgement scores, for both groups there is an overall parallelism



**Figure 3** Interaction of parallelism with anaphor type: active vs. passive antecedent clauses



**Figure 4** Interaction of syntactic parallelism with anaphor type: verbal vs. nominal antecedent clauses

effect with sentences following active antecedents considered more acceptable than those following passive ones: similarly, those with verbal antecedents are considered more acceptable than those with nominal antecedents. This difference is reflected in the main effects of Syntactic Parallelism, by participant and by item ( $F_1(1, 46) = 160.33$ ;  $p < 0.001$ ;  $F_2(1, 29) = 89.18$ ;  $p < 0.001$ ). Furthermore, for both groups the sentences with nominal antecedents were judged significantly less

acceptable than those with passive sentences, as indicated by the main effect of Construction Type by participants ( $F_1(1, 46) = 34.40; p < 0.001$ ;  $F_2(1, 29) = 1.55; p > 0.2$ ). There was also a significant main effect of Anaphor Type (VPE/VPA) ( $F_1(1, 46) = 16.14; p < 0.001$ ;  $F_2(1, 29) = 16.36; p < 0.001$ ), given the across-the-board judgement advantage for the VPA vs. VPE. Finiteness also had a similar overall effect for both groups ( $F_1(1, 46) = 121.37; p < 0.001$ ;  $F_2(1, 29) = 12.49; p < 0.005$ ) with non-finite ellipsis completions considered more acceptable than finite.

While the parallel items were considered more acceptable than the non-parallel items in general, the parallelism effect was considerably stronger in the VPE condition (mean difference .82) than with VPA completions (mean difference .05): this is confirmed by the Syntactic Parallelism \* Anaphor Type interaction, which was significant by participants and marginal by items:  $F_1(1, 46) = 9.78; p < 0.01$ ;  $F_2(1, 29) = 3.72; p < 0.07$ ). There was also a significant interaction between Parallelism and Construction Type ( $F_1(1, 46) = 13.43; p > 0.01$ ;  $F_2(1, 29) = 4.54; p < 0.05$ ) reflecting the stronger parallelism effect found in the comparison between the verbal and nominal (mean difference 0.83) vs. the active and passive constructions (mean difference 0.54), as well as a reliable interaction by items among these three factors, Parallelism, Anaphor and Construction Type ( $F_1(1, 46) = 2.85; p > 0.09$ ;  $F_2(1, 29) = 9.17; p < 0.01$ ).

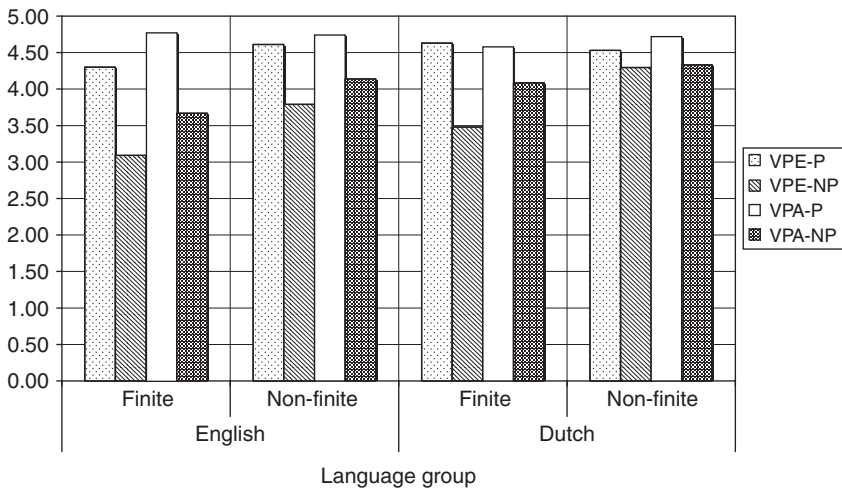
As predicted then, the strength of the parallelism effect varied as a function of both Construction Type (significantly stronger in the verbal–nominal contrast) and Anaphor Type (stronger in the VPE conditions), with the least acceptable non-parallel condition being the nominal/VPE combination.

Furthermore – and contrary to theoretical accounts, which do not distinguish finite from non-finite ellipsis – the finiteness of the ellipsis clause had a reliable effect on the strength of the syntactic parallelism effect, at least by participant (Syntactic Parallelism \* Finiteness:  $F_1(1, 46) = 30.34; p < 0.001$ ;  $F_2(1, 29) = 2.82; p > 0.1$ ): specifically, the parallelism effect was significantly in the finite (mean difference .87) than in the non-finite conditions (mean difference .50).<sup>11</sup> For both groups, the

<sup>11</sup> As well as two-way interactions significant by participants between Finiteness and Anaphor Type by ( $F_1(1, 46) = 4.18; p < 0.05$ ;  $F_2(1, 29) = .67; p > 0.4$ ) and Finiteness and Construction Type ( $F_1(1, 46) = 7.80; p < 0.01$ ;  $F_2(1, 29) = .80; p > 0.4$ ), there were three-way interactions between Parallelism, Anaphor Type and Finiteness ( $F_1(1, 46) = 20.35; p < 0.001$ ;  $F_2(1, 29) = .17; p > 0.6$ ) as well as Construction Type, Anaphor Type and Finiteness, both significant by participants:  $F_1(1, 46) = 6.17; p < 0.05$ ;  $F_2(1, 29) = 1.43; p > 0.2$ ).

presence of non-finite, as opposed to finite, ellipsis significantly raised the acceptability of the completion clause – hence attenuated the parallelism effect – in both the nominal/VPE conditions (finite: mean 3.07; non-finite: 4.02) as well as the nominal/VPA condition (finite: mean 3.50; non-finite: 4.07), in this offline judgement task. These rather striking effects of finiteness are diagrammed in Figure 5 (which collapses across Construction Type): as this figure indicates, finiteness exerts a much larger influence on the L2 data than on native speakers’ judgements, a point addressed directly below.

*b Language group differences:* Despite these broadly similar general trends across the two groups, the L2 learners’ judgement patterns did differ from the native speakers’ in certain important respects. Overall, the L2 learners reported elevated acceptability scores compared to native speakers, as indicated by the significant main effect of Group ( $F_1(1, 46) = 4.76; p < 0.05; F_2(1, 29) = 108.35; p < 0.001$ ). More importantly, there were significant interactions between Group and other independent variables: two-way interactions between Syntactic Parallelism and Group ( $F_1(1, 46) = 4.53; p < 0.05; F_2(1, 29) = 4.51; p < 0.05$ ), Finiteness and Group ( $F_1(1, 46) = 8.92; p < 0.01; F_2(1, 29) = 16.96; p < 0.001$ ) and Anaphor Type and Group by items ( $F_1(1, 46) = .42; p > 0.5; F_2(1, 29) = 15.73; p < 0.001$ ); as well as three-way interactions with Group – significant by participants – between Syntactic Parallelism



**Figure 5** Effects of finiteness of syntactic parallelism

and Construction Type ( $F_1(1, 46) = 10.42$ ;  $p < 0.01$ ;  $F_2(1, 29) = 1.03$ ;  $p > 0.3$ ), Syntactic Parallelism and Finiteness ( $F_1(1, 46) = 4.27$ ;  $p < 0.05$ ;  $F_2(1, 29) = .34$ ;  $p > 0.5$ ), with marginally significant four-way interactions with Syntactic Parallelism, Anaphor Type and Finiteness ( $F_1(1, 46) = 3.99$ ;  $p > 0.06$ ;  $F_2(1, 29) = .085$ ;  $p > 0.7$ ) and Construction, Anaphor Type and Finiteness ( $F_1(1, 46) = 3.91$ ;  $p > 0.06$ ;  $F_2(1, 29) = .04$ ;  $p > 0.9$ ).

Table 4 presents the results of *t*-test comparisons for each of the two groups, conducted to probe these differences in the strength of the parallelism effect across the various conditions. What is clear from this closer analysis is a marked contrast in the distribution of reliable parallelism effects for the two groups. Restricting attention to the *t*-tests by participants, it can be seen that the English native speakers' results show reliable effects of syntactic parallelism of roughly the same magnitude across the board: irrespective of Finiteness, Construction Type and, interestingly, of Anaphor Type – a point we shall return to below – syntactically parallel antecedents were reliably preferred over syntactically non-parallel antecedents in the offline task. For the Dutch

**Table 4** Summary of *t*-test comparisons: offline data

Construction type	Finiteness	Participants			Items		
		<i>t</i>	<i>df</i>	Signifi- cance	<i>t</i>	<i>df</i>	Signifi- cance
<i>English: VPE:</i>							
Active vs.	Finite	9.12	28	.001	3.22	17	.010
passive	Non-finite	5.33	28	.001	3.28	12	.007
Verbal vs.	Finite	7.62	28	.001	1.40	17	.100
nominal	Non-finite	5.60	28	.001	1.93	12	.073
<i>English: VPA:</i>							
Active vs.	Finite	4.40	28	.001	1.82	17	.087
passive	Non-finite	3.55	28	.001	1.01	12	.331
Verbal vs.	Finite	3.79	28	.001	2.32	17	.050
nominal	Non-finite	4.81	28	.001	1.97	12	.070
<i>Dutch: VPE:</i>							
Active vs.	Finite	5.91	18	.001	1.59	17	.131
passive	Non-finite	0.60	18	.554	-1.07	12	.306
Verbal vs.	Finite	5.82	18	.001	0.76	17	.456
nominal	Non-finite	3.04	18	.010	0.68	12	.507
<i>Dutch: VPA:</i>							
Active vs.	Finite	2.35	18	.031	1.14	17	.270
passive	Non-finite	1.28	18	.216	-0.52	12	.610
Verbal vs.	Finite	5.20	18	.001	3.82	17	.001
nominal	Non-finite	4.09	18	.010	3.27	12	.007

learners by contrast, Finiteness and Construction Type were crucial factors in determining the acceptability of syntactically non-parallel antecedents: although the Dutch learners generally pattern with the English native speakers in finite conditions, the only reliable parallelism effects observed with non-finite items came in Construction Type 2 (that is to say, in the contrast between verbal and nominal antecedents).

*c Summary of offline results:* In summary, the results of the offline task confirm that advanced Dutch learners can indeed acquire sensitivity to parallelism effects in VPE, as claimed in Duffield and Matsuo (2009): the results also show that Dutch learners can appropriately distinguish – in certain constructions at least – between VPE and VPA with respect to the size of this effect. At the same time, these data suggest an important difference between English native speakers and L2 learners: namely, whereas non-syntactic factors have only a marginal effect on native speakers' judgements – syntax is pre-eminent for native speakers – for L2 learners construction-specific properties are key determinants of acceptability. Indeed, there is no evidence in these data that L2 judgements are influenced by the sort of generalized (trans-derivational) syntactic constraints assumed by mainstream syntactic theory. As we shall see, this conclusion is further reinforced by the results of the online task.

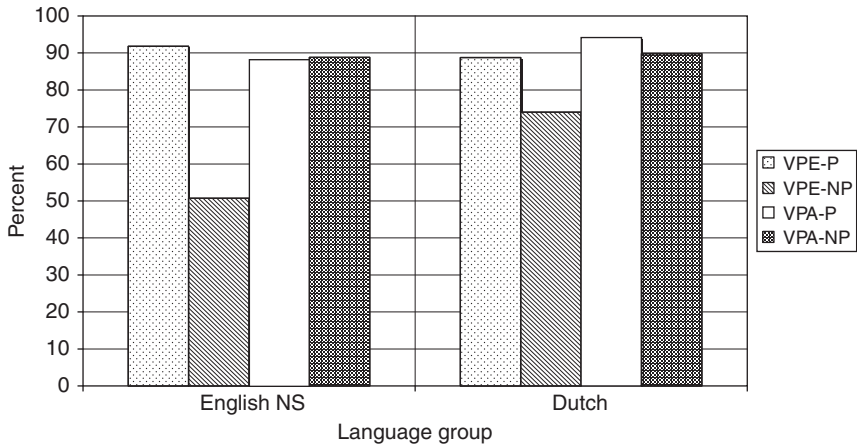
## 2 Online task

*a Data treatment:* In the SCJ task, both accuracy and response latency are taken into account. Both sets of data were cleaned before proceeding with the statistical analyses. For the response latency data, individual responses that were more than two standard deviations away from a participant's mean were replaced with the mean response time for that participant. For items analyses, any missing data points were replaced by the mean per item per condition.

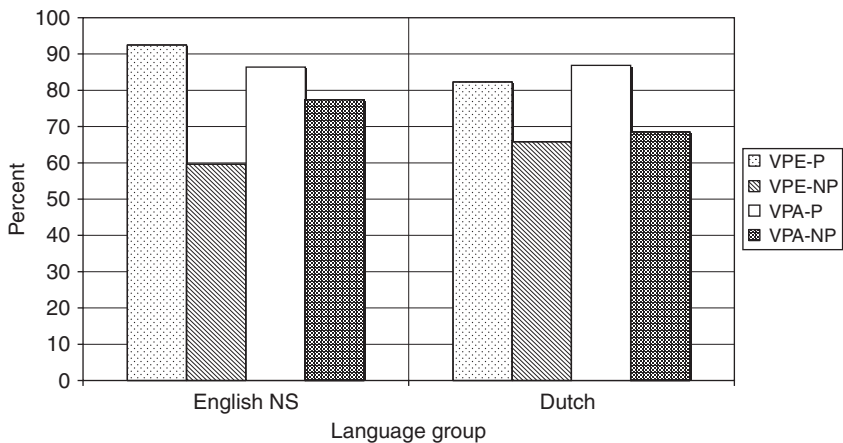
*b Judgement data: overall results:* Figures 6 and 7 show the mean judgement scores for English native speakers and Dutch learners across the two Construction Types. Once again, we collapse across levels of finiteness, and ignore the nested recoverability factors.

As with the judgement scores in the offline task, there were robust effects of syntactic parallelism overall: native speakers and L2 learners were more likely to judge a completion acceptable if it followed a structurally parallel antecedent than a passive or nominal one; here,





**Figure 6** Interaction of parallelism with anaphor type: active vs. passive antecedent clauses



**Figure 7** Interaction of parallelism with anaphor type: verbal vs. nominal antecedent clauses

once again, nominal antecedents were considered less acceptable than passives. This was supported by main effects of Syntactic Parallelism ( $F_1(1, 47) = 123.81; p < 0.001; F_2(1, 29) = 36.77; p < 0.001$ ) and Construction Type – marginal by items – ( $F_1(1, 47) = 22.02; p < 0.001; F_2(1, 29) = 3.37; p < 0.08$ ). In general, acceptance was also higher for VPA vs. VPE items shown by the reliable main effect of Anaphor Type ( $F_1(1, 47) = 28.07; p < 0.001; F_2(1, 29) = 20.18; p < 0.001$ ). Further to this, and again consistent with the offline results,

the strength of the parallelism effect was greater in the verbal–nominal (mean difference 19.1) vs. the active–passive contrast (mean difference 15.0): this was directly reflected in the interaction between Syntactic Parallelism \* Construction Type:  $F_1(1, 47) = 2.28$ ;  $p > 0.1$ ;  $F_2(1, 29) = 8.21$ ;  $p < 0.01$ ). The effect was also modulated by the type of anaphor in the completion clause (Syntactic Parallelism \* Anaphor Type:  $F_1(1, 47) = 26.53$ ;  $p < 0.01$ ;  $F_2(1, 29) = 17.70$ ;  $p < 0.001$ ), with the difference between the parallel and non-parallel VPA items being much smaller (7.9) than the difference between the parallel and non-parallel VPE items (26.3). As in the offline data, there was a significant three-way interaction between Syntactic Parallelism, Construction Type and Anaphor Type ( $F_1(1, 47) = 11.10$ ;  $p < 0.01$ ;  $F_2(1, 29) = 4.38$ ;  $p < 0.05$ ): the least acceptable Construction Type being the non-parallel nominal, with acceptability increasing when this antecedent type was followed by a VPA ellipsis clause (VPA: 73 vs. VPE: 63).

In this task also, the finiteness of the ellipsis clause affected the acceptability of sentence completions for both groups. As well as a main effect of Finiteness, with non-finite items being considered more acceptable than non-finite ( $F_1(1, 47) = 103.62$ ;  $p < 0.01$ ;  $F_2(1, 29) = 8.20$ ;  $p < 0.01$ ), there was a significant interaction between Finiteness and Parallelism ( $F_1(1, 47) = 19.47$ ;  $p < 0.001$ ;  $F_2(1, 29) = 5.42$ ;  $p < 0.01$ ): the parallelism effect was again much stronger with finite VPE or VPA (mean difference 23.7) than for the non-finite items (mean difference 10.4).

*c Judgement data: language group differences:* Overall, the patterns are rather similar to those found in the offline judgement task, with gradient judgements apparent for both groups. And, as with the earlier task, the two language groups performed differently on closer analysis. In addition to significant two-way interactions between Group and Parallelism ( $F_1(1, 47) = 5.47$ ;  $p < 0.05$ ;  $F_2(1, 29) = 11.23$ ;  $p < 0.01$ ), Group and Construction Type ( $F_1(1, 47) = 15.77$ ;  $p < 0.001$ ;  $F_2(1, 29) = 11.61$ ;  $p < 0.01$ ), and Group and Finiteness, significant by participants ( $F_1(1, 47) = 4.95$ ;  $p < 0.05$ ;  $F_2(1, 29) = 2.23$ ;  $p > .1$ ), there was a reliable three-way interaction between Parallelism, Anaphor Type and Group ( $F_1(1, 47) = 15.72$ ;  $p < 0.001$ ;  $F_2(1, 29) = 17.09$ ;  $p < 0.001$ ) and a reliable four-way interaction (by participants) between Group and all the other experimental variables, Parallelism, Construction Type, Anaphor Type and Finiteness ( $F_1(1, 47) = 4.31$ ;  $p < 0.05$ ;  $F_2(1, 29) = 1.07$ ;  $p > 0.3$ ).

Table 5 shows the results of *t*-test comparisons run on the mean acceptability judgement scores per group. The data for English in

**Table 5** Summary of *t*-test comparisons: online judgement data

Construction type	Finiteness	Participants			Items		
		<i>t</i>	<i>df</i>	Signifi- cance	<i>t</i>	<i>df</i>	Signifi- cance
<i>English: VPE:</i>							
Active vs.	Finite	8.633	28	<b>.000</b>	5.252	17	.001
passive	Non-finite	5.585	28	<b>.000</b>	4.462	12	.001
Verbal vs.	Finite	7.810	28	<b>.000</b>	5.447	17	.001
nominal	Non-finite	2.994	28	<b>.006</b>	3.114	12	.009
<i>English: VPA:</i>							
Active vs.	Finite	1.225	28	.231	0.880	17	.391
passive	Non-finite	-2.117	28	.043	-1.251	12	.235
Verbal vs.	Finite	1.777	28	.086	2.037	17	.057
nominal	Non-finite	1.612	28	.118	1.771	12	.102
<i>Dutch: VPE:</i>							
Active vs.	Finite	3.687	19	<b>.002</b>	2.900	17	.010
passive	Non-finite	1.159	19	<b>.261</b>	0.142	12	.889
Verbal vs.	Finite	3.210	19	<b>.005</b>	4.407	17	.001
nominal	Non-finite	1.223	19	<b>.236</b>	0.987	12	.346
<i>Dutch: VPA:</i>							
Active vs.	Finite	0.653	19	.522	0.531	17	.603
passive	Non-finite	1.228	19	.234	0.485	12	.636
Verbal vs.	Finite	3.433	19	.003	2.886	17	.010
nominal	Non-finite	2.274	19	.035	2.169	12	.051

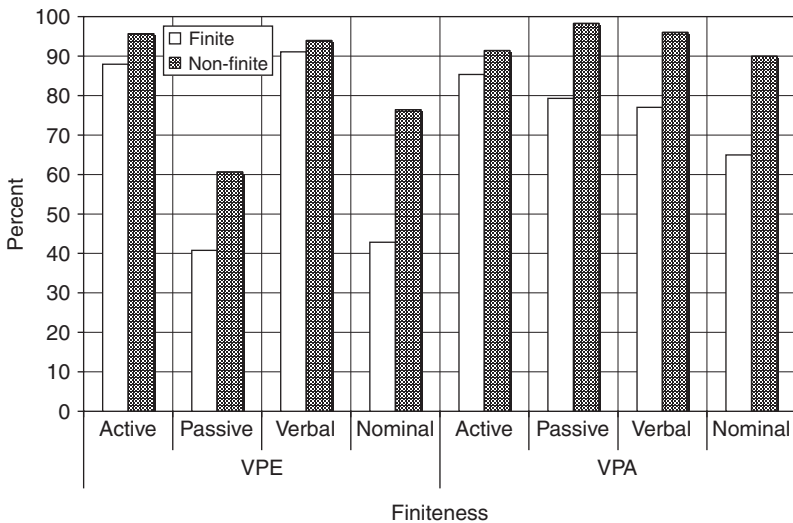
Table 5 reveal a judgement pattern for native speaker controls that is almost entirely consistent with the classical syntactic approach to the parallelism effect: with reliable effects of syntactic parallelism for all non-parallel VPE conditions, largely irrespective of Finiteness or Construction Type, and the absence of any corresponding effects with VPA completions. These data, then, basically support the traditional view that parallelism is structurally, rather semantically determined, and that it applies in an across-the-board fashion to VPE only, in accordance with the Surface Anaphora/Deep Anaphora distinction: statistically at least, there is no parallelism effect for VPA. (Note, however, that there are nevertheless trends towards parallelism effects in the English VPA data, and that we have not yet considered the effects of recoverability: see Appendix 1.)

The Dutch L2 learners, however, present a distinct pattern of judgements, as Table 5 shows. Compare, in particular, the entries in bold in the table. The most striking fact is the overwhelming effect of Finiteness on Syntactic Parallelism: for the Dutch learners, the parallelism effect

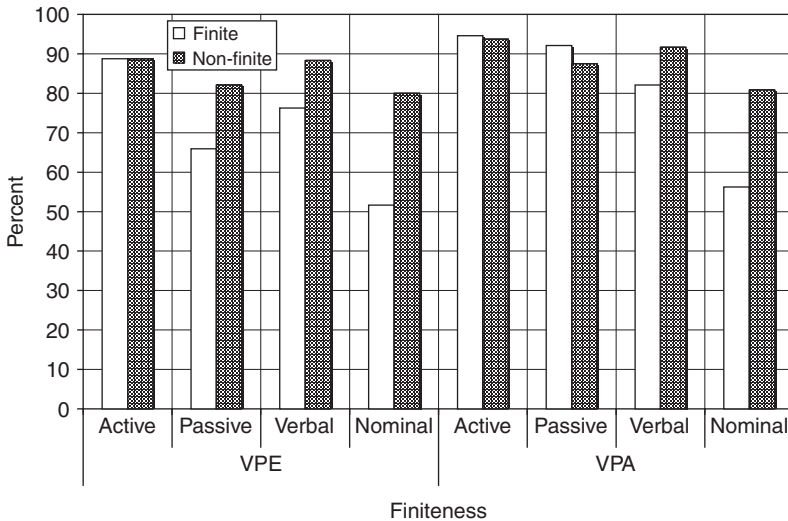
is essentially eliminated with non-finite VPE (and only shows up in the verbal vs. nominal VPA condition). In other words, Dutch learners don't know – at least, don't show evidence of knowing – that syntactic parallelism is a generalized constraint: what this evidence instead suggests is that they know that certain constructions are unacceptable. Notice also that in certain conditions, Dutch learners show stronger parallelism effects for VPA completions than for VPE, contrary to the predictions of the classical theory. Figures 8 and 9 illustrate this L1 vs. L2 contrast with respect to finiteness:

*d Summary of online judgement data:* In sum, the trends in the judgement data in the SCJ task were similar to those found in the participants' offline task, but they were somewhat magnified in the offline task, perhaps because the participants were under less time pressure. What the online results confirm is that the overall picture of L1/L2 convergence is misleading: the data suggest instead that non-structural factors play a qualitatively different role in determining L2 learners' judgements than is the case for native speakers, for whom structural considerations are paramount.

Before discussing the theoretical implications of these results, we need to consider the other dependent measure in the online task, namely, response latencies, as well as the nested factor of recoverability (discussed in the Appendix 1).



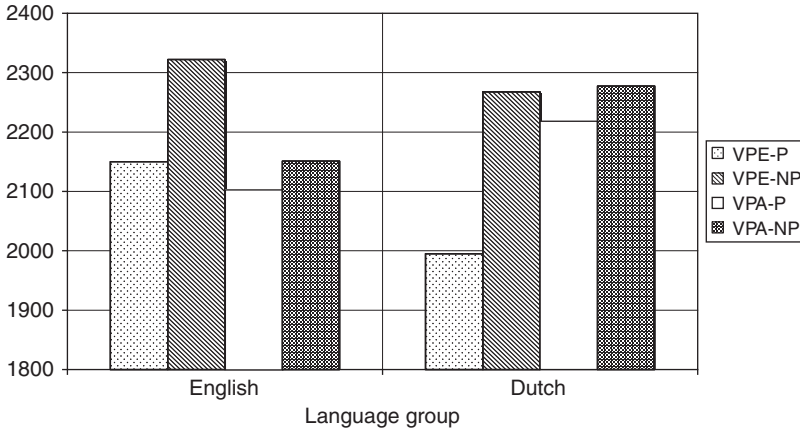
**Figure 8** Effects of finiteness on parallelism: English



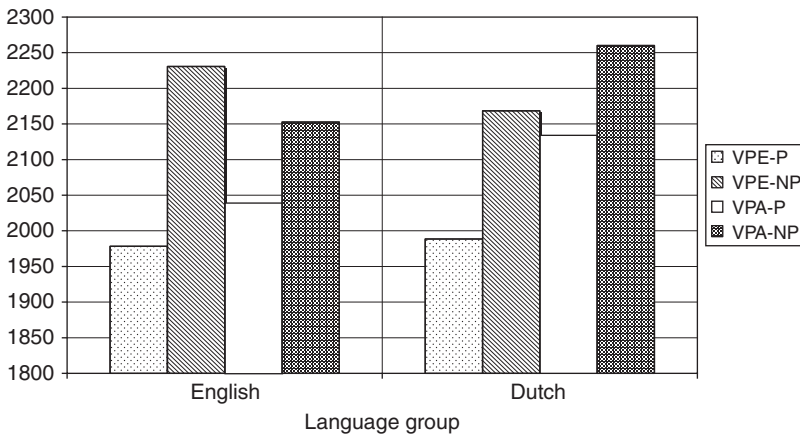
**Figure 9** Effects of finiteness on parallelism: Dutch

*e Online task: latency data:* Figures 10 and 11 show the mean response times for the English native speakers and the Dutch L2 learners, respectively.<sup>12</sup> The overall pattern of results presented here is of particular interest in that it is not merely a replication of earlier findings: whereas in previous studies (Tanenhaus and Carlson, 1990; Duffield and Matsuo, 2009) latency data had shown (non-significant) trends consistent with the judgement data, in the current study the majority of these effects are statistically significant. The latency results show that the overall parallelism effect pattern is directly reflected in the time taken for the participants to make their sentence completion decisions. Specifically, there was a main effect of Syntactic Parallelism ( $F_1(1, 47) = 22.60$ ;  $p < 0.001$ ;  $F_2(1, 29) = 4.67$ ;  $p < 0.05$ ) with structurally non-parallel items eliciting longer reaction times (average 2229 ms) than structurally parallel items (average 2076 ms). There was the predicted interaction between Syntactic Parallelism and Anaphor Type (significant by participants) ( $F_1(1, 47) = 7.54$ ;  $p < 0.01$ ;  $F_2(1, 29) = 2.79$ ;  $p > 0.1$ ), with the overall parallelism effect for VPE completions being reliably larger (average 219 ms) than for VPA completions (average 86 ms). It is

<sup>12</sup> Recall that these means are for the accepted items across experimental conditions, the expectation being that where they are accepted, VPE items with non-parallel antecedents should take longer to accept than those with parallel ones.



**Figure 10** Interaction of parallelism with anaphor type: active vs. passive antecedent clauses



**Figure 11** Interaction of parallelism with anaphor type: verbal vs. nominal antecedent clauses

worth noting, however, that both groups showed parallelism effects for VPA too: the statistically reliable interaction is due to the relative size of this effect, rather than to its presence in VPE vs. absence in VPA. A main effect of Finiteness was also observed with overall latencies shorter for items with non-finite ellipsis clauses (2043) than with finite ellipsis clauses (2262) ( $F_1(1, 47) = 93.56; p < 0.001; F_2(1, 29) = 11.08; p < 0.01$ ). Also observed was a two-way interaction between Finiteness and Construction Type (significant by participants) ( $F_1(1, 47) = 11.21; p < 0.01; F_2(1, 29) = 3.02; p > 0.09$ ) as well as a three-way interaction

between Construction Type, Anaphor Type and Finiteness ( $F_1(1, 47) = 4.29$ ;  $p < 0.05$ ;  $F_2(1, 29) = 2.80$ ;  $p > 0.1$ ). However, there was no interaction between Finiteness and Syntactic Parallelism. Interestingly, for both groups, finiteness failed to have any effect in the active/VPE conditions, where both the finite and the non-finite conditions elicited the same response latencies (English, finite: 2133 ms; non-finite: 2167; Dutch finite: 2031 ms; non-finite: 1959 ms). This directly mirrors the judgement data in the same condition, in which the finite and non-finite active/VPE conditions were judged equally acceptable (floor, rather as opposed ceiling effects).

*f Latency data: language group differences:* As the figures above show, the two groups performed very similarly to each other, even in their overall response times, and this was borne out statistically with the factor Group interacting only with Anaphor Type ( $F_1(1, 47) = 10.40$ ;  $p < 0.01$ ;  $F_2(1, 29) = 10.95$ ;  $p < 0.01$ ) since for the learners, overall, the VPA conditions elicited the longer latency times (2222 ms) than VPE conditions (2105 ms), whereas there was no difference in latencies between the native speakers (2112 vs. 2170 ms), though this effect may have been driven by the very short responses to the Active/VPE condition.

With respect to individual conditions, the latency data corresponded quite closely with the judgement data. As Table 6 shows, planned *t*-test comparisons reveal reliable differences in reaction time following parallel vs. non-parallel antecedents in many – though not all – of the expected conditions:

*g Summary: offline task and sentence completion task combined:* Our results may be summarized as follows:

- The results of both tasks pattern with those of earlier studies. The strength of the parallelism effect was affected by the Construction Type of the non-parallel antecedent – greater for nominal vs. passive antecedents – as well as by the Anaphor Type – greater for VPE than VPA, but still present for VPA, contrary to theoretical models. Therefore, the effects found earlier are shown to be robust, and to generalize across different aspectual auxiliaries in the ellipsis clause (*do* vs. *have*), even for Dutch L2 learners.
- Even though theoretical accounts of the parallelism effect do not distinguish between finite and non-finite ellipsis, both the offline and the

**Table 6** T-tests (latency data)

Construction type	Finiteness	Participants		
		<i>t</i>	<i>df</i>	Significance
<i>English: VPE:</i>				
Active vs. passive	Finite	-2.41	28	<b>.01</b>
	Non-finite	-1.30	28	.101
Verbal vs. nominal	Finite	-2.54	28	<b>.0008</b>
	Non-finite	-3.06	28	<b>.0002</b>
<i>English: VPA:</i>				
Active vs. passive	Finite	0.005	28	.49
	Non-finite	-0.96	28	.17
Verbal vs. nominal	Finite	-1.86	28	<b>.036</b>
	Non-finite	-0.432	28	.33
<i>Dutch: VPE:</i>				
Active vs. passive	Finite	-3.20	19	<b>.0002</b>
	Non-finite	-1.78	19	<b>.045</b>
Verbal vs. nominal	Finite	-1.48	19	.0977
	Non-finite	-1.57	19	.067
<i>Dutch: VPA:</i>				
Active vs. passive	Finite	0.114	19	.455
	Non-finite	-1.34	19	.098
Verbal vs. nominal	Finite	1.27	19	.108
	Non-finite	-1.06	19	.150

online results demonstrate unequivocally that the strength of the effect differs as a function of the finiteness of the ellipsis clause. Overall, non-finite VPE/VPA completions were considered more grammatical and acceptable than finite completions, and they were responded to faster. Interestingly, for both groups, finiteness interacted with structural parallelism, such that the relative acceptability of completions following non-parallel passive antecedents increased significantly when the ellipsis clause was non-finite. Once again, this result fails to square with existing theoretical models.

- Close analysis of the data reveals important differences between native speakers and second language learners. Whereas for English native speakers syntactic non-parallelism is generally a sufficient condition for eliciting reliable parallelism effects across the board, for L2 learners it is only a necessary one: other non-structural factors, especially finiteness, play a determining role. One consequence of this is the parallelism effect emerges as essentially a sporadic, construction-specific effect for L2 learners, rather than a generalized structural constraint.



#### IV Theoretical implications

Finally, let us consider the implications of these findings for theories of ultimate attainment in SLA, as well as for current theoretical accounts of ellipsis: the results of these experiments seem to us to pose an empirical challenge in both theoretical domains.

With respect to ultimate attainment, the conclusions to be drawn depend entirely on the granularity of the analysis. When viewed at a distance, the achievement of the Dutch L2 learners seems extremely impressive: faced with canonical instances of VPE, Dutch learners' sensitivity to parallelism clearly shows that they have acquired knowledge of a subtle negative constraint applying to a non-native construction, and that – quantitatively at least – they can successfully distinguish the constraints applying to this construction (VPE) from structurally and semantically similar instances of VPA (in the judgement data). Moreover, the results of the latency data show that L2 learners achieve almost exactly the same pattern of reaction times as do native speakers. These overall results, then, are consistent with those who believe that L2 learners can attain levels of implicit competence and online processing comparable to those of native speakers, and that they are not absolutely constrained by the parameter-settings of their L1; see, for example Schwartz and Sprouse (1996); White (2003).<sup>13</sup>

On the other hand, a more detailed examination of the three sets of data supports a different conclusion, namely, that the Dutch learners are 'faking it': the overall appearance of convergence on native speaker judgements is illusory, and belies a very different weighting of structural and non-structural factors. Specifically, as discussed above, purely structural factors play a much less dominant role in predicting parallelism effects for Dutch learners than for native speakers. Conversely, the non-structural factors of finiteness and construction-type – which merely 'add colour' to native speakers' judgements and response times – are key determinants of Dutch learners' response patterns. As a result, Dutch learners' data fails to show any knowledge of a generalized syntactic constraint, and is much more consistent with a grammar consisting of construction-specific rules.<sup>14</sup> At this level of granularity, then, our results broadly support those who believe that L2 learners'

<sup>13</sup> Though, in contrast to the claims of the authors just cited, these results do not necessarily speak for or against access to specifically innate knowledge.

<sup>14</sup> Of course, these results do not preclude a Strong Continuity interpretation, which might then attribute the observed differences to performance rather than competence limitations. However,

internal representations are fundamentally different from those of native speakers; for example, Bley-Vroman *et al.* (1988); Clahsen and Muysken (1989); see also Ullman (2001); Clahsen and Felser (2006).<sup>15</sup>

With regard to purely theoretical issues, our results are similarly confirmatory and challenging at the same time. Restricting attention to the native speaker participants, if one only looks at the various VPE conditions, then the results from both the offline and online versions of the experiment – and from both dependent measures of the online task – are quite compatible with either of the two standard theoretical explanations of parallelism: all relevant measures show reliable interactions between Syntactic Parallelism and Anaphor Type, as predicted by the syntactic account (and consistent with the semantic analysis); and no other major independent variable has a significant influence on these effects. So far, so good.

On the other hand, our results yield two conclusions that run counter to theoretical expectations. The first, that the finiteness of the ellipsis clause determines the strength of the parallelism effect, is more surprising than disturbing (at least in the case of the native speakers' results): although this finding does not have any obvious theoretical explanation, it does not in itself challenge either of the traditional accounts. However, the second conclusion, namely, that parallelism affects VPA as well as VPE, would seem to pose a more serious challenge. Our results, especially those of the offline task, show unequivocally that reliable parallelism effects are found for both anaphor types (albeit the effects for VPA are weaker). These empirical findings directly challenge the theoretical distinction between Surface and Deep Anaphora proposed by Sag and Hankamer – but also subscribed to by Merchant – since they show that the syntactic form of the antecedent clause significantly influences the acceptability of deep anaphors as well as ellipses.

On this point, then, our results appear to arbitrate between the two theoretical accounts, lending greater support to the semantic approach

---

given that we have three sets of convergent data from two different implicit judgement tasks – four measures, three tasks if the earlier experiments are taken into account – we suggest that the burden of proof is on proponents of 'full access' to show that Dutch learners can acquire a fully generalized parallelism constraint.

<sup>15</sup> In contrast to these researchers, however, our findings suggest that L2 learners' grammatical rules are syntactically just as sophisticated as those of native speakers: they are simply less generalized. There is, for example, no evidence of 'shallow syntactic processing' in these L2 data, as proposed by Clahsen and Felser (2006); quite the contrary, otherwise the L2 learners would not display target-like discrimination of VPE and VPA, which are semantically and thematically identical.

of Merchant (2001) over Hankamer and Sag's syntactic account. This is because, for Hankamer and Sag the presence vs. absence of a structural parallelism effect is the key defining property of the Surface/Deep Anaphor distinction. For Merchant, on the other hand, the Surface/Deep Anaphora distinction can still be motivated independently: although he assumes that the identification requirement does not apply to VPA, since the constraint is ultimately a semantic one, it can be relaxed to apply to VPA as well, without necessarily doing injury to the Surface/Deep dichotomy.<sup>16</sup>

As ever, more research is necessary to determine whether the differences observed here between English native speakers and Dutch second-language learners generalize to other grammatical phenomena, and/or other learner groups. However, if this turns out to be the case, it suggests a marked shift in the interpretation of the Fundamental Difference Hypothesis, from the traditional (implicitly) 'deficit-based' approach to something more like a 'complementary model' of L2 competence: second language learners are neither more or less competent than native speakers, they do things differently.<sup>17</sup>

## V References

- Bley-Vroman, R., Felix, S. and Ioup, G.** 1988: The accessibility of universal grammar in adult language learning. *Second Language Research* 4, 1–32.
- Chao, W.** 1987: On ellipsis. Unpublished PhD dissertation, Department of Linguistics, University of Massachusetts, Amherst, MA.
- Chomsky, N.** 1982: *Some concepts and consequences of the theory of Government and Binding*. Linguistic Inquiry Monograph, number 6. Cambridge, MA: MIT Press.
- Clahsen, H. and Felser, C.** 2006: Grammatical processing in language learners. *Applied Psycholinguistics* 27, 3–42.
- Clahsen, H. and Muysken, P.** 1989: The UG paradox in L2 acquisition. *Second Language Research* 5, 1–29.
- Crain, S.** 1993: Language acquisition in the absence of experience. In Bloom, P., editor, *Language acquisition: core readings*. New York: Harvester/Wheatsheaf, 364–409.

<sup>16</sup> The analysis of the 'recoverability' data (see Appendix 1) provides additional support for the semantic view: the fact that the parallelism effect fails to show up in (non-finite) passive VPE contexts where a *by*-phrase enhances recoverability is consistent with the semantic approach, but unexpected on the syntactic view.

<sup>17</sup> For one specific proposal on how to model this difference, see Duffield, 2003; see also Ullman, 2001.

- Cyrino, S. and Matos, G.** 2005: Local licenser and recovering in VP ellipsis construction: variation across languages and language varieties. *Journal of Portuguese Linguistics* 1, 177–95.
- Dalrymple, M. Shieber, S. and Pereira, F.** 1991: Ellipsis and higher-order unification. *Linguistics and Philosophy* 14, 399–452.
- Duffield, N.** 2003: Measures of competent gradience. In Van Hout, A., Hulk, A., Kuiken, F., and Towell, R., editors, *The lexicon–syntax interface in second language acquisition*. Amsterdam and Philadelphia: John Benjamins, 97–127.
- Duffield, N. and Matsuo, A.** 2001: A comparative study of ellipsis and anaphora in L2 acquisition. In Do, A.H.-J., Domínguez, L., and Johansen, A., editors, *Proceedings of the 25th Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, 238–49.
- 2002. Finiteness and parallelism: assessing the generality of knowledge about English ellipsis in SLA. In Skarabela, B., Fish, S., and Do, A.H.-J., editors, *Proceedings of the 26th Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, 197–207.
- 2009: Native-speakers' vs. L2 learners' sensitivity to parallelism in VP-ellipsis. *Studies in Second Language Acquisition* 31, 1–31.
- Goldberg, L.** 2005: Verb-stranding VP-ellipsis. Unpublished PhD dissertation, Department of Linguistics, McGill University.
- Hankamer, J. and Sag, I.** 1976: Deep and surface anaphora. *Linguistic Inquiry* 7, 391–428.
- Hardt, D.** 1993: Verb phrase ellipsis: form, meaning and processing, computer and information science. Unpublished PhD dissertation, University of Pennsylvania, Philadelphia, PA.
- Hoji, H.** 1998: Null object and sloppy identity in Japanese. *Linguistic Inquiry* 29, 127–51.
- Maurer, G., Tanenhaus, M. and Carlson, G.** 1995: A note on parallelism effects on processing verb phrase anaphors. *Language and Cognitive Processes* 10, 1–12.
- Merchant, J.** 2001: *The syntax of silence: sluicing, islands, and the theory of ellipsis*. Oxford: Oxford University Press.
- 2004. Fragments and ellipsis. *Linguistics and Philosophy* 27, 661–738.
- Newmeyer, F.** 1983: *Grammatical theory: its limits and possibilities*. Chicago, IL: University of Chicago Press.
- Otani, K. and Whitman, J.** 1991: V-raising and VP-ellipsis. *Linguistic Inquiry* 22, 345–58.
- Rizzi, L.** 1986: Null Objects in Italian and the theory of pro. *Linguistic Inquiry* 17, 505–57.
- Sag, I. and Hankamer, J.** 1984: Towards a theory of anaphoric processing. *Linguistics and Philosophy* 7, 325–45.

- Santos, A.L.** 2006: Minimal answers: ellipsis, syntax and discourse in the acquisition of European Portuguese. Unpublished PhD dissertation, University of Lisbon.
- Schütze, C.** 1996: *The empirical base of linguistics*. Chicago, IL: University of Chicago Press.
- Schwartz, B. and Sprouse, R.** 1996: L2 cognitive states and the full transfer/full access model. *Second Language Research* 12, 40–72.
- Schwarzchild, R.** 1999: GIVENness, AVOIDF, and other constraints on the placement of accent. *Natural Language Semantics* 7, 141–77.
- Tanenhaus, M. and Carlson, G.** 1990: Comprehension of deep and surface verb-phrase anaphors. *Language and Cognitive Processes* 5, 257–80.
- Ullman, M.** 2001: The neural basis of lexicon and grammar in first and second language: the declarative/procedural model. *Bilingualism: Language and Cognition* 4, 105–22.
- White, L.** 2003: *Second language acquisition and Universal Grammar*. Cambridge: Cambridge University Press.

## Appendix 1 Effects of recoverability

In post-hoc analyses of the data from the two non-parallel construction types, we investigated potential effects of recoverability, distinguishing two kinds: conceptual and morpho-syntactic recoverability.

### 1 Results: recoverability in passive items

To investigate the role of conceptual recoverability, we analysed the passive items separately. Data from these conditions were entered into an ANOVA with the within-participants factors of Anaphor Type (VPE/VPA), Recoverability (+/–) and Finiteness (+/–) and the between-participants factor of Group (English native speakers/Dutch L2 learners), this latter factor being treated as a within-participants factor, and Finiteness and Recoverability as between-participants factors in the items analysis.

*a Off-line task:* Figures 12 and 13 show the mean acceptability judgement scores given to the non-parallel passive experimental items, broken down by Recoverability, in the offline task. As these figures show, the two groups' response patterns are highly similar. Although there was no main effect of Recoverability, there was a significant interaction between Recoverability and Anaphor Type in the participants' analysis ( $F_1(1, 46) = 10.77; p < 0.01; F_2(1, 27) = 1.05; p > 0.3$ ): as predicted

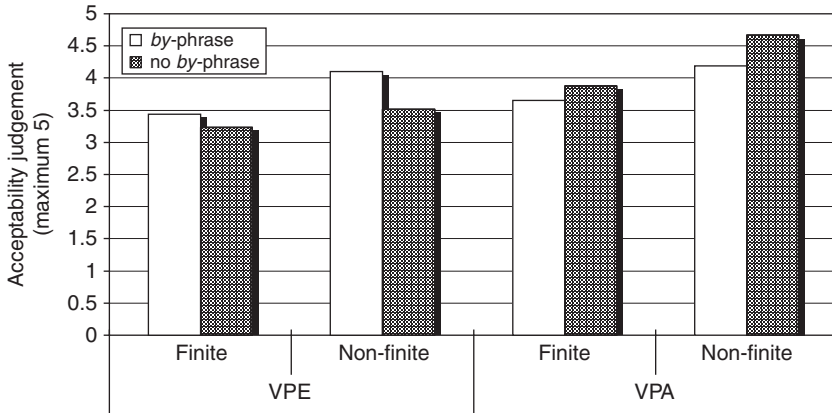


Figure 12 Effects of conceptual recoverability: offline data (English)

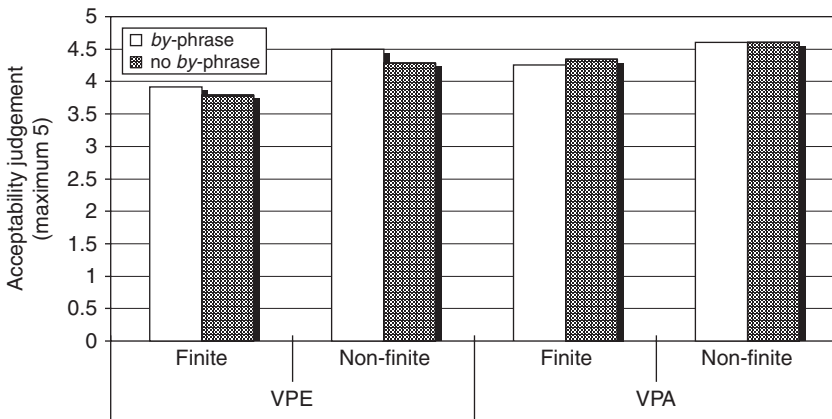


Figure 13 Effects of conceptual recoverability: offline data (Dutch)

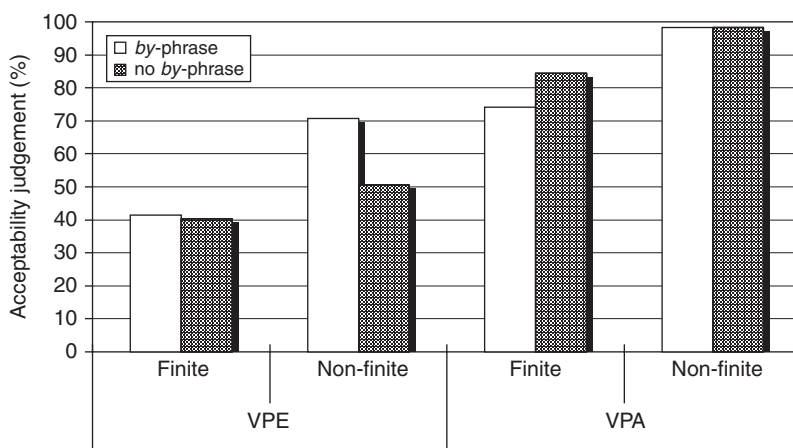
by the semantic analysis, whether or not there was a *by*-phrase in the antecedent clause affected acceptability. Specifically, the presence of the *by*-phrase marginally affected the acceptability score in the VPE conditions. Interestingly, the opposite pattern obtained with VPA.

*b Online task:*<sup>18</sup> Analysis of the judgement data in the online task found a significant three-way interaction between Recoverability, Finiteness

<sup>18</sup> Analysis of the response latency data, however, revealed no main effect of Recoverability, nor any interaction between Recoverability and any other factor.

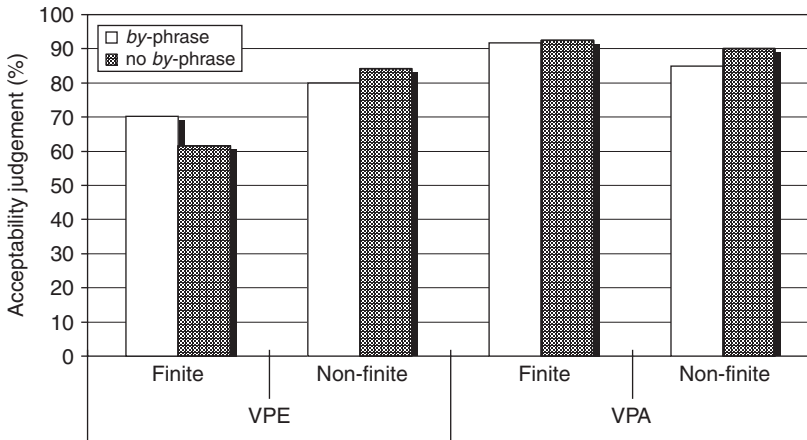
and Group ( $F_1(1, 47) = 4.15$ ;  $p < 0.05$ ;  $F_2(1, 27) = 6.24$ ;  $p < 0.05$ ), reflecting the difference in the two groups' mean acceptability ratings. Specifically, whereas the presence of a *by*-phrase facilitated acceptability of the non-finite items for the native English speakers (recoverable: 84.5% vs. non-recoverable: 74.4%; VPE/VPA collapsed), the opposite was found in their responses to the finite items (recoverable: 57.8% vs. non-recoverable: 62.4%). In both the finite and the non-finite cases, the opposite pattern was found for the Dutch L2 learners, but here the differences overall were very much smaller than found in the natives (finite, recoverable: 80.9%; finite, non-recoverable: 77.1%; non-finite, recoverable: 82.5%, non-finite, non-recoverable: 87.1%).

Figures 14 and 15 show the mean acceptability judgement scores in the online task broken down by Group and Anaphor Type. In terms of theoretical, rather than statistical, significance the most interesting data-points in these figures occur in the English VPE conditions, which clearly show an interaction between the non-structural factors of finiteness and recoverability.<sup>19</sup> Note that although all of the columns in these two figures represent judgements for completions following non-parallel antecedents – where we should observe uniform levels of rejection according to the standard syntactic analysis – one group of



**Figure 14** Effects of conceptual recoverability: online data (English)

<sup>19</sup> Given these high overall means, it seems plausible that the absence of an effect of recoverability in the Dutch data is, in part at least, due to ceiling effects in non-finite completions: for Dutch learners, the effects of finiteness swamp those of recoverability.



**Figure 15** Effects of conceptual recoverability: online data (Dutch)

items is basically accepted as grammatical by native-speakers: these results indicate that non-finite VPE following a passive antecedent with a *by*-phrase is acceptable, whereas the same completion is unacceptable without a *by*-phrase. In other words, the data here supports the judgements assigned to the sentences in (19).

- 19) a. The garbage had to be taken out by someone in the house. But John didn't want to.  
 b. ?? The garbage had to be taken out. But John didn't want to.

This contrast clearly speaks against the traditional idea that the parallelism effect is solely driven by structure; however, it does not provide unequivocal support for the alternative semantic account. See conclusion.

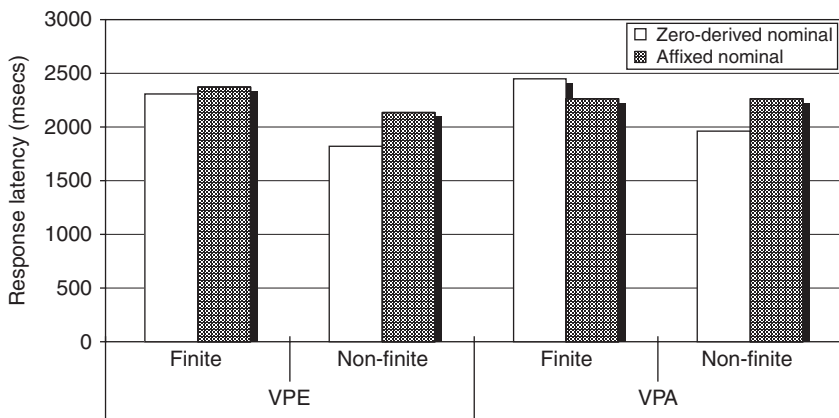
## 2 Recoverability in nominal items: results and discussion

The same analyses were run on the non-parallel nominal items. As discussed above, we examined whether non-parallel antecedents containing zero-derived nominals, such as *mention*, might be more easily reconstructed (as stem-identical verbs in the ellipsis clause) than affixed nominals, such as *discussion*. If any effect were to be found, we predicted that the zero-derived non-parallel antecedents should be more acceptable than affixed nominals.

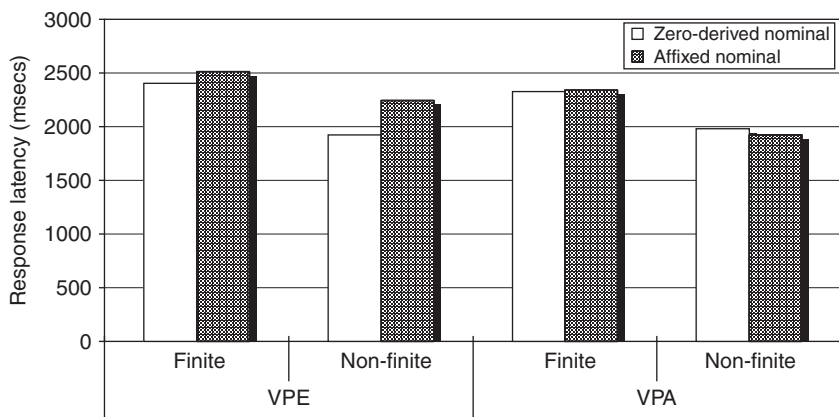
Analyses of the judgement data in the both online and offline tasks revealed no main effects or interactions with the factor Recoverability.



Remarkably, though, there were effects in the latency data. Figures 16 and 17 show the mean response latencies for each of the passive antecedent conditions of the online task, distinguishing Recoverability, Anaphor Type and Finiteness. Analysis of the latency data reveals both a significant main effect by participants of Recoverability ( $F_1(1, 47) = 5.49; p < 0.05; F_2(1, 26) = .11; p > 0.7$ ) and an interaction between Recoverability and Finiteness ( $F_1(1, 47) = 4.88; p < 0.05; F_2(1, 27) = .48; p > 0.4$ ). For both the native speakers and L2 learners, completions following zero-derived nominal antecedent were accepted more quickly (2147 ms) than those following less recoverable (affixed) nominals (2257 ms).



**Figure 16** Effects of morphological recoverability: online data (English)



**Figure 17** Effects of morphological recoverability: online data (English)

However, finiteness influenced the recoverability effect, since there was a difference of 220 ms between the non-finite recoverable (1922 ms) vs. the non-finite less recoverable (2142 ms), whereas there was no difference at all between the more and less recoverable finite conditions, both eliciting equally high response latencies (2373 vs. 2372 ms).

In sum, for the non-parallel nominal items, morpho-syntactic recoverability had no effect on the types of acceptability/grammaticality judgments made by either group. However, it did influence the length of time participants took to make their responses in the SCJT, beneficially affecting only the (comparatively short) non-finite conditions.