University of Reading

Antecedent contained deletions in native and non-native sentence processing

Article

Accepted Version

It is advisable to refer to the publisher's version if you intend to cite from the work.

To link to this article DOI: http://dx.doi.org/10.1075/lab.15006.box

Publisher: John Benjamins

www.reading.ac.uk/centaur

CentAUR

Antecedent contained deletions in native and non-native sentence processing

Oliver Boxell

Claudia Felser

University of Potsdam, Germany

&

Ian Cunnings

University of Reading, UK

Abstract

We report the results from an eye-movement monitoring study investigating native (L1) and non-native (L2) speakers' real-time processing of antecedent-contained deletion (ACD), a type of verb phrase ellipsis in which the ellipsis gap forms part of its own antecedent. The resulting interpretation problem is traditionally thought to be solved by quantifier raising, a covert scope-shifting operation that serves to remove the gap from within its antecedent. Our L2 group comprised advanced, native German-speaking L2 learners of English. The analysis of the eye-movement data showed that both L1 and L2 English speakers tried to recover the missing verb phrase after encountering the gap. Only the native speakers showed evidence of ellipsis resolution being affected by quantification, however. No effects of quantification following gap detection were found in the L2 group, by contrast, indicating that recovery of the elided material was accomplished independently from the object's quantificational status in this group.

Keywords: L2 processing, eye-movement monitoring, antecedent contained deletion

Non-native (L2) comprehenders have frequently been found to process discontinuous intra-sentential dependencies in a non-nativelike way (see Clahsen & Felser, 2006; Dallas & Kaan, 2008; Felser, 2015, for reviews and discussion). As current models of sentence processing were originally developed to account for adult monolingual (L1) processing data, data from bilinguals or other "non-standard" populations allow us to test the general validity of such models and may ultimately help improve them. Our understanding of L2 dependency processing is still rather patchy, however. Besides studies on morphosyntactic agreement phenomena, most previous research in this domain has focused either on movement dependencies (i.e. non-canonical word orders) or on the resolution of pronominal anaphors. Research on the former has typically investigated whether L2 comprehenders can identify movement-related gaps in the input stream and recover their content, and research on anaphor resolution has examined their ability to identify antecedents for pronominals.

Whilst the identification of gaps during L2 comprehension seems to be problematic only under certain circumstances (Marinis, Roberts, Felser, & Clahsen, 2005; Felser & Roberts, 2007), even highly proficient L2 comprehenders have shown difficulty accessing or navigating previously built syntactic representations during real-time anaphor resolution (e.g. Felser & Cunnings, 2012). L1/L2 differences in the processing of syntactically mediated dependencies might, for example, indicate differences between native and non-native comprehension in the relative weighting of syntactic, semantic or other kinds of interpretation cue, or differences in the relative timing of the mental mechanisms involved in dependency formation. L2 comprehenders' apparent difficulty integrating syntactic and non-syntactic interpretation cues during the resolution of "backwards-looking" dependencies

(Felser, 2015) needs to be explored further, and across different types of anaphoric dependency. The current study investigates a special type of anaphoric dependency whose resolution has been claimed to require the (potentially concurrent) application of syntactic movement: elliptical gaps that are contained in their own antecedent.

Elliptical gaps present a challenge for the human sentence comprehension system because the missing constituent's meaning has to be recovered using information provided by the sentential or discourse context. The current study investigates L1 and L2 English speakers' processing of antecedent contained deletion (ACD), a type of verb phrase ellipsis (VPE) where this task is further complicated by the fact that the gap contains its own antecedent, as illustrated in (1).

(1) John will [$_{VPi}$ read every book that Mary will __$_i$].

This should theoretically give rise to an "infinite regress" problem (i.e. a recursive loop) when a representation of the elided VP is recovered at the gap site. For instance, in (1) the gap refers to the VP *read every book that Mary will*, which itself also contains a gap whose antecedent is the VP *read every book that Mary will*, and so on. In actual fact, ACDs like (1) are perfectly acceptable and readily resolved, but exactly how this is accomplished has been a matter of some debate. Formal solutions to the infinite regress problem include the proposal that covert movement operations such as quantifier raising (QR) (May, 1985; Sag, 1976) or argument movement (Hornstein, 1994) help remove the gap from within its own antecedent, as well as alternative semantic or pragmatic approaches (e.g. Dalrymple, Shieber, & Pereira, 1991; Jacobson, 1998).

Note that from a processing perspective, ellipses as in (1) are covert anaphors, and encountering them will trigger a backwards-looking search for the missing information. L1/L2 differences have previously been observed both in the real-time resolution of overt pronominal anaphors (e.g. Felser & Cunnings, 2012; Patterson, Trompelt, & Felser, 2014; Roberts, Gullberg, & Indefrey, 2008) and for sentences containing covert syntactic constituents (e.g. Dekydtspotter & Miller, 2013; Marinis, et al., 2005; Felser & Roberts, 2007). For VPE, there is evidence that L2 speakers are able to recover the missing information but are less sensitive than L1 speakers to the syntactic parallelism constraint on ellipsis resolution (Duffield & Matsuo, 2009; Duffield, Matsuo, & Roberts, 2009). Little is known about how L2 speakers process elliptical gaps whose resolution is less straightforward than in standard cases of VPE, however.

The present study is the first to examine the real-time processing of ACD structures in non-native speakers. We report eye-movement data for both long and short ACDs and for noun phrases whose quantificational status is systematically altered. These manipulations allow us to investigate two key claims about ACD resolution. First, participants' reading times at and following the ellipsis site should be affected by antecedent size only if the elided material is indeed recovered. Second, our manipulation of the NP's quantificational status allows us to test claims that quantification might trigger mechanisms (such as QR) that facilitate ACD resolution.

Background

Formal and Processing Approaches to ACD

As noted above, ACD represents a special case of VPE. Theoretical approaches to ellipsis phenomena differ with regard to the question of whether the elided material is

reconstructed as a syntactic (e.g. Ross, 1969) or semantic object (e.g. Hardt, 1993).[1] Experimental evidence suggests that native comprehenders generally prefer antecedents that are structurally – rather than just semantically - parallel to the VPE site (e.g. Tanenhaus & Carlson, 1990), which lends support to syntactic reconstruction accounts. There is also a fair amount of evidence suggesting that all or part of the elided constituent's syntactic structure is recovered at ellipsis sites (Arregui, Clifton, Frazier, & Moulton, 2006; Frazier & Clifton, 2001, 2005; Shapiro & Hestvik, 1995; Shapiro, Hestvik, Lesan, & Garcia, 2003; Yoshida, Walsh Dickey, & Sturt, 2013). This might be accomplished either by a copying mechanism or by some form of structure-sharing. In contrast, "pointing" approaches to ellipsis phenomena (e.g. Culicover & Jackendoff, 2005; Ginzburg & Sag, 2000; Martin & McElree, 2008) assume that the missing constituent's meaning is recovered directly from the discourse context, with no recovery of linguistic structure at the gap site. A number of studies have manipulated the size and/or complexity of antecedents to see if this affects ellipsis resolution. Martin and McElree (2008) argue that antecedent size should affect processing difficulty only if a full copy of the antecedent is recovered at the gap site. The results from these studies have been mixed, with some showing antecedent complexity effects (e.g. Murphy, 1985) whilst others do not (e.g. Frazier & Clifton, 2000; Martin & McElree, 2008).

For ACD, fully copying the elided verb phrase into the gap would result in the infinite regress problem pointed out above. Thus some additional mechanism appears to be required to avoid or circumvent this problem. Successful ACD resolution has often been related to the quantificational status of the NP containing the elliptical gap.

---

[1] See Phillips and Parker (2014) for a more detailed review and discussion.

In particular, a distinction is made between "strong" quantifiers that render ACDs

acceptable (2) and "weak" quantifiers that do not (3):

(2) John will read [every/each/most/any] book(s) Mary will __.

(3) *John will read [a/some/few/one, two, three…] book(s) Mary will __.

The difference in acceptability between (2) and (3) can be accounted for by assuming

that strongly quantified expressions as in (2), but not weak quantifiers as in (3),

obligatorily undergo QR (e.g. Diesing, 1992). This is the idea that quantifiers will

raise covertly and adjoin to a functional projection higher up in the syntactic structure,

such as CP in (4), at the level of Logical Form (LF) (May, 1985; Sag, 1976; see Heim

& Kratzer, 1998, for further details and discussion). After QR has taken place, the

ellipsis site (e_j) now refers back to an antecedent VP (VP_j) that no longer includes the

object phrase (DP_i in 4) hosting the ellipsis site.

(4)
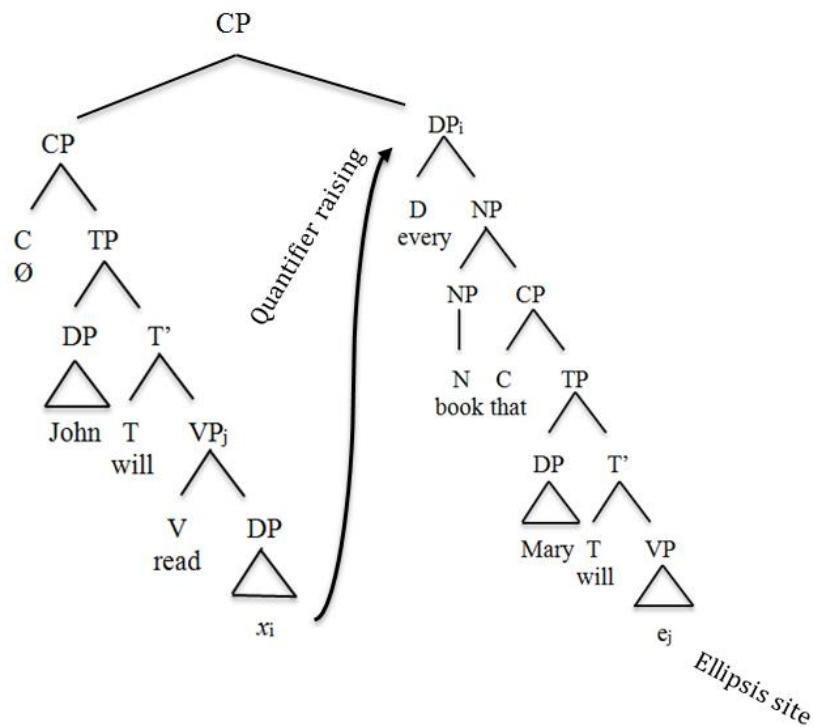
A potential problem with the QR account for ACDs is that it under-generates possible ACDs. That is, an object NP does not necessarily need to be quantificational for ACDs to be deemed acceptable. Take the examples in (5), where the object is (a) a bare plural (Diesing, 1992) or (b) a definite NP (Harley, 2002).

(5) a. John will read Ø books that Mary will __.

  b. John will read the book that Mary will __.

The acceptability of (5a,b) can however be accounted for by assuming that non-quantificational NPs can also potentially be assigned a "strong" reading, in which case they will also be able to undergo QR (e.g. Diesing, 1992).

A more recent alternative to the traditional QR approach to ACD that also accounts for the acceptability of the examples in (5) is the proposal that object

movement serves to remove the VP gap from within its own antecedent (e.g.

Hornstein, 1994). This operation is assumed to be motivated by case checking

requirements and thus applies to both quantified and unquantified NPs alike. As with

QR, the regress problem is solved here by raising the NP containing the problematic

gap to a functional projection above VP such as an agreement-object phrase, as

illustrated in (6).

(6)   John will read [AgrOP [DP every book that Mary will __] [VP ~~read~~

       [DP *every book that Mary will __*]]].

Alternative solutions to ACD involve semantic type-shifting rather than

syntactic movement. Put simply, Cormack (1984) and Jacobson (1998) suggest that

an element in the sentence that is not the object can take on the semantic function of

the object NP. This gives the matrix clause VP the semantics of a full object

complement without a gap inside it. Hence, the self-reference is no longer found

within the semantics of the VP antecedent for the ellipsis site. Such theories operate

on an abstract level of semantic computation that purports to solve ACDs

independently of any antecedent reconstruction (and related quantification effects)

that may or may not occur at the gap site. Unlike the traditional QR approach, neither

Hornstein's (1994) object-movement account nor semantic type-shifting accounts

such as Jacobson's (1998) specifically predict that the quantificational status of the

object NP should facilitate the resolution of ACD gaps.

Hackl et al. (2012) were the first, to our knowledge, to carry out an online

investigation into ACD resolution. They presented mature L1 English speakers with

sentences like (7), manipulating both the quantificational status of the object NP and

the size of the antecedent. Short antecedent conditions were created by using the

dummy auxiliary *did*, whereas corresponding long antecedent conditions involved the

use of *was*, signalling that the elided material included the matrix predicate (e.g.

reluctant to treat…) as well. Finally, no-ellipsis control conditions were created by

replacing the auxiliary with a main verb (e.g. *admitted*).

(7)   The doctor was reluctant to treat…

    a.   the patient that the recently hired nurse {did__/was__/admitted__}

    b.   every patient that the recently hired nurse {did__/was__/admitted__}

    …after looking over the test results.

On the assumption that QR is obligatory for *every* phrases but only applies to definite

NPs as a last resort, the QR account predicts that (i) *every NPs* should take longer to

process that *the NPs* when these are first encountered, but that (ii) quantified objects

should facilitate processing once readers have come across the VP gap.

    Participants' word-by-word reading times were recorded using a self-paced

reading paradigm. Two words after the verb/auxiliary, the authors found an effect of

antecedent size, with more complex antecedents eliciting longer reading times than

less complex ones, and an interaction between quantification and antecedent size. The

interaction was driven by the fact that *every* facilitated reading times relative to *the* in

the short antecedent conditions, whilst the reverse was true for the long antecedent

and the no-ellipsis control conditions. The authors suggest that QR can explain this

interaction on the assumption that QR targets its nearest potential landing site (e.g. the

verb phrase projected by *treat*; e.g. Merchant, 2001), in which case selective

facilitation for the quantified condition might be expected in the short antecedent

condition. Local QR of this sort would undo the antecedent containment in the short antecedent condition but not in the long one, which includes *reluctant* as well.[2]

Alternative interpretations of Hackl et al.'s findings are possible which do not require the assumption of any covert movement, however. Szabolcsi (2014), for example, offers an alternative explanation of Hackl et al.'s results using semantic type shifting. Gibson, Jacobson, Graff, Mahowald, Fedorenko, and Piantadosi (2014) suggest a different alternative interpretation of Hackl et al.'s findings in terms of pragmatic felicitousness. When the event being described by two clauses is the same, there is a known pragmatic-discourse preference for using NPs that reflect *sameness* (e.g. Kaplan, 1984). Specifically, unless a connection between the two clauses is established contextually, there is a preference for DPs to include the modifiers *also* or *same* (e.g. *the same NP)*. The same pressure does not apply to *every NP* forms because the competing *also*/*same* forms are less acceptable than the bare *every NP* for independent reasons (e.g. **every same NP*). Gibson et al. report a series of acceptability rating studies which reveal that (i) *every* improves judgments relative to *the* even when there is no ACD to resolve; (ii) the improvement effect for *every* relative to *the* disappears when either *same* or *also* is added to the definite NP condition; (iii) the improvement effect for *every* disappears when the sentences are presented in a discourse context that already supports the sameness between the two clauses. According to Gibson et al., these results suggest that any acceptability or processing advantage of *every NP* over *the NP* objects in ACD structures is likely to reflect differences in pragmatic felicitousness rather than selective application of QR.

---

[2] Others, however, have found evidence for a preference for long over short antecedents in the processing of antecedent-contained ellipsis in comparative structures (Breakstone, Cremers, Fox, & Hackl, 2013).

In summary, syntactic "copying" and discourse-based "pointing" approaches to ellipsis phenomena primarily differ with regard to the question of whether or not the missing constituent is reconstructed at the gap site. Regarding ACD, examples such as (5a,b) above suggest that explicit quantification is not a necessary condition for this type of ellipsis to be resolvable. Nevertheless, both the QR approach and some alternative semantic or pragmatic approaches predict that the resolution of ACD gaps should be facilitated by quantified object NPs, albeit for different reasons.

VPE and ACD in Language Acquisition

ACDs are likely to be comparatively rare in language learners' input and have not featured very prominently in either L1 or L2 acquisition research. A small number of L1 acquisition studies have examined children's ability to interpret ACDs. Syrett and Lidz (2009, 2011) and Kiguchi and Thornton (2004), for instance, compared children and adults' interpretations of ACDs with regular VPE using truth-value judgment tasks. Their results show that children as young as four are able to interpret ACDs correctly, suggesting that a mechanism for solving the infinite regress problem (such as QR) is part of the children's grammar. Comparing children's ability to interpret ACDs with short vs. long antecedents, Sugawara, Kotek, Hackl, and Wexler (2013) found that children older than 5;5 (but not younger ones) showed antecedent size effects, with ACDs with long antecedents preferred over those with short antecedents.

Previous L2 acquisition research has thus far focused on standard cases of VPE. Ying (2005), for example, presented L1 Chinese-speaking learners of English and English native speakers with sentences containing VPE such as (8) and asked them to choose between two possible interpretations of the gap.

(8)   John defended himself and Bill did __ too.

The elliptical gap here may contain a copy of *himself* that would need to follow

binding condition A in co-referring with the nearest antecedent, namely *Bill*. This is

the "sloppy" identity reading of the reflexive. Alternatively, *himself* may be replaced

by *him*, thus permitting coreference with *John*, which corresponds to the "strict"

reading. Both intermediate and advanced learners patterned with the native controls in

preferring the sloppy reading, a preference that was particularly strong in the lower-

proficiency learners. The author argues that the "sloppy" reading was preferred

because it was easier to process (but cf. Rosales Sequeiros, 2004, for a critique of this

interpretation).

Others have examined L2 speakers' sensitivity to structural parallelism

between VP ellipses and their antecedents. Duffield and Matsuo (2009) used a

sentence completion judgment task (modelled after Tanenhaus & Carlson, 1990) to

investigate advanced Dutch, Spanish and Japanese-speaking learners' sensitivity to the

parallelism constraint in L2 English. Their experimental materials included brief

stories as shown in (9) below, with the second sentence containing either an ellipsis or

an anaphoric pronoun. The degree of parallelism between the antecedent and the

ellipsis was manipulated by changing the active structure in the "parallel" conditions

(9a) into a passive one in the "non-parallel" ones (9b).

(9)   a.   PARALLEL CONDITIONS

                Someone took the wood out of the shed last night. Tom told us that

                Sally did (VPE) / did it (VPA).

        b.   NON-PARALLEL CONDITIONS

The wood was taken out to the shed last night. Tom told us that

Sally did (VPE) / did it (VPA).

Participants had to indicate whether or not the second sentence formed a sensible continuation from the first. Whilst VPE completions should be preferred in parallel (active) contexts, VP anaphora (VPA) should make better completions non-parallel (passive) contexts.

Overall, the results showed that both native English speakers and L2 learners were aware of the parallelism constraint on ellipsis, even if this phenomenon does not exist in the learners' L1 (e.g. Dutch). However, the authors also observed some L1/L2 differences. For example, whilst Spanish and Japanese learners both judged VPE sentences as only marginally acceptable, they still preferred them with parallel antecedents. The authors note that, if the parallelism constraint is taken into account even by learners who have not yet acquired the licensing conditions for VPE, the two may be unrelated in the first place. The Dutch learners' sensitivity to the parallelism constraint seemed to be modulated by finiteness more than the native speakers'. This was subsequently confirmed by a further study reported in Duffield, Matsuo and Roberts (2009), who conclude that the parallelism effect "emerges as essentially a sporadic, construction-specific effect for L2 learners, rather than a generalized structural constraint" (p. 457). Duffield and Matsuo's findings thus indicate that VPE resolution is not necessarily accomplished in the same way in L1 and L2 processing.

As the above studies have used either offline comprehension (Ying, 2005) or acceptability judgement tasks (Duffield & Matsuo, 2009; Duffield et al., 2009), they cannot tell us much about how L2 comprehenders resolve VP gaps in real-time. The present study seeks to address the following two empirical questions: (i) Do both L1

and L2 comprehenders attempt to recover the missing VP after coming across a VP

gap, (ii) does quantification facilitate ACD resolution in both L1 and L2

comprehension?

The Current Study

To obtain a fine-grained picture of the processing time course of ACD resolution, we

carried out an eye-movement-monitoring-during-reading experiment with L1

German-speaking learners of English and native English-speaking controls. We

specifically sought to determine whether both groups were sensitive to the elided VPs

size, and whether ACD resolution was modulated by the quantificational status of the

NP containing the problematic gap during both L1 and L2 processing. Note that

German lacks genuine VP ellipsis, including ACD-type ellipsis (Lobeck, 1995;

Winkler, 1995). Simple transfer of surface structure properties from German to

English should thus be precluded.

Participants

Thirty-two L1 German speakers of L2 English (11 males, mean age: 24 years, range:

19-34) were recruited from the University of Potsdam, Germany. The L2 participants

completed the grammar section of the Oxford Placement Test (Allan, 2004) with an

average score of 81.84% (SD: 7.02, range: 66%-100%), which corresponds to the C1

or "advanced" level of the Common European Framework of Reference for

Languages. All of them started learning English at a mean age of 9.4 years (SD: 1.96)

in a formal educational setting. Thirty-two native speakers of English (8 males, mean

age: 19 years, range: 17-23) recruited from the University of Edinburgh, UK, served

as a control group. All participants had normal or corrected-to-normal vision. They

were offered a small financial incentive or course credits for their participation.

Materials

The current experiment manipulated the size of the antecedent VP and the

quantificational status of the object NP in a 2x2 design, as illustrated in (10).

(10)   PREAMBLE: Australian scouts master many skills.

a.   LONG ANTECEDENT, QUANTIFIED

The scout was taught to catch every snake that the explorer

was __ in the outback in the strong heat.

b.   LONG ANTECEDENT, UNQUANTIFIED

The scout was taught to catch the snake that the explorer

was __ in the outback in the strong heat.

c.   SHORT ANTECEDENT, QUANTIFIED

The scout was taught to catch every snake that the explorer

did __ in the outback in the strong heat.

d.   SHORT ANTECEDENT, UNQUANTIFIED

The scout was taught to catch the snake that the explorer

did __ in the outback in the strong heat.

The design of the stimuli was inspired by, but different from, Hackl et al.'s (2012). In

the long antecedent conditions, the elided material corresponds to the matrix VP (e.g.

taught to catch… in 10a,b), whereas in the short antecedent conditions it corresponds

to the local VP (e.g. the one headed by *catch* in 10c,d). The antecedent VPs all

contained an object NP introduced either by the quantifier *every* or by the definite article.

Twenty-four sentence quadruplets as in (10) (see Appendix) were created and distributed across four presentation lists using a Latin Square design. To ensure that items from the same condition did not appear adjacently, the experimental items were pseudo-randomized. In addition, 60 filler items were added, including 48 items containing pronominal dependencies and a range of other quantifiers, and 12 containing different ellipsis types (including four ACDs with quantifiers other than *every*). Yes/no comprehension questions followed half of the critical items and half of the fillers.

As a reviewer notes, the fact that our antecedent size manipulation involved using VPs that differed not only in length but also semantically might give rise to differences in plausibility between the long and short antecedent conditions. As any inherent plausibility differences between our experimental conditions could potentially affect participants' reading times, we carried out an offline plausibility control study with 18 native English speakers who did not take part in the main experiment. Using an electronic questionnaire, participants were asked to rate the plausibility of VP ellipsis-free versions of our experimental sentences (e.g. *The scout was taught to catch {every/the} snake that the explorer {was taught to catch / caught} in the outback in the strong heat*) on a seven-point scale. The 24 experimental quadruplets were distributed across four lists in a Latin Square design and mixed with 36 plausible and implausible filler sentences of different types that varied in their structural complexity.

All four conditions received mean ratings of 5.39 or above, confirming that they were all deemed plausible. The two non-quantificational conditions were rated

slightly higher than the QP conditions, irrespective of VP size. This is likely due to the fact that quantificational expressions are more complex semantically and thus more difficult to interpret than DPs, other things being equal. Statistical analyses using mixed models showed a significant main effect of quantification (QP vs. DP; estimate = 0.18, SE = 0.07, t = 2.70, p = .007) but not of VP type (long vs. short; estimate = 0.03, SE = 0.07, t = 0.43, p = .666), and no interaction (estimate = 0.05, SE = 0.12, t = 0.43, p = .670). These results confirm that our long and short antecedent conditions did not measurably differ in plausibility.

Predictions

Our main predictions concern the sentence region at which the presence of a VP gap is confirmed (e.g. *in the outback*). If antecedent VPs are mentally reconstructed at this point (e.g. Ross, 1969), we predict an effect of antecedent size, with the contents of long antecedents taking longer than short ones to recover. No such effects are predicted by accounts of ellipsis that do not assume syntactic reconstruction of the antecedent (e.g. Martin & McElree, 2008).

Recovering a full copy of the antecedent VP would normally include reconstructing the ellipsis site itself in the case of ACDs, resulting in a recursive loop. However, if QR is applied to *every NP* (but not to *the NP*) objects when these are first encountered, possibly leading or contributing to longer first-pass reading times for quantified vs. definite NPs, this should ultimately facilitate ACD resolution in the quantified conditions, with shorter reading times for the quantified conditions following readers' initial processing of the VP gap. Gibson et al.'s (2014) pragmatic account also predicts effects of quantification, reflecting the additional inferential processing effort required for sentences containing pragmatically less felicitous (i.e.

definite) objects. No effects of quantification during or following the processing of ACD gaps would be expected from the perspective of approaches according to which the semantic status of the object NP plays no role for ACD resolution (e.g. Hornstein, 1994; Jacobson, 1998 – but cf. Szabolcsi, 2014).

Procedures

Participants were told to read the stimulus sentences silently and answer the end-of-trial questions using "yes" or "no" buttons on a game pad. They were first presented with eight practice items to ensure they had understood the instructions. The text was presented in size 18 black Courier New font on a white background on a computer monitor. The preamble and the target sentence appeared on the same screen but on different lines. We used a desk-mounted EyeLink 1000 system (SR Research, Mississauga, Ontario, Canada) with a sampling rate of 1000 Hz to record participants eye movements as they read through the sentences at their own pace. Participants read binocularly, but we tracked their right eye only. Standard calibration and validation protocols were used to calibrate the EyeLink to participants' eyes.

For the analysis of the reading-time data, the experimental sentences were divided into five regions of interest as is illustrated in (11):

(11)	1.	*Sentence initial region*	The scout was taught to catch
	2	*Object region*	every/the snake that
	3.	*Auxiliary region*	the explorer was/did
	4	*Post-gap region*	in the outback
	5.	*Final region*	in the strong heat.

For each sentence region, we report five eye-movement measures. *First-pass* reading time is the sum of all initial fixations in a region; *regression path* time is the time it takes to read and re-read text within a region before exiting it to the right for the first time, potentially also including rereading of previous regions. *Rereading times* includes all fixations in a region after it was first exited and is thus thought to reflect later (or second-pass) processing stages (Rayner, 1998). We also report the number of *regressions into* each region, which are looks originating from later sentence regions, and *regressions out* of each region, which are looks towards earlier sentence regions.

Prior to the statistical analysis of the reading-time data, fixations of 80ms or lower and within one degree of another fixation were automatically merged with neighboring fixations. Fixations of 80ms or lower which could not be merged were removed. Data points in excess of 1.5 SDs from a participant's mean and/or the item's mean for each measure for each region were also excluded. The total amount of data removed from the raw data set was 4.33% for the L1 group and 5.98% for the L2 speakers. The remaining data were analyzed using repeated-measures ANOVAs, with Cohen's *d* effect sizes reported for each statistic.

Results

The native English group answered 88.34% of the end-of-trial comprehension questions correctly overall, with an accuracy of 89.29% on the experimental items. The L2 group scored 85.14% correct overall, and 87.71% on the experimental items. This indicates that both groups attended to the reading comprehension task. Skipping rates across all sentence regions were below 3.6% for the L1 and below 2.3% for the L2 speakers.

//TABLE 1 ABOUT HERE//

Table 1 reports the two participant groups' mean reading times per condition for each region of analysis.[3] To establish whether the two groups' reading-time patterns differed statistically, we carried out a series of preliminary mixed ANOVAs with the within-subjects factors Quantification (*quantified, unquantified*) and Antecedent Size (*long, short*) and the between-subjects factor Group (*L1, L2*) for each eye-movement measure and for all sentence regions. At each of the five regions this yielded main effects of Group for all reading time measures reported (ps < .001). Furthermore, we found a number of significant or marginally significant interactions between Group and Quantification across the data set, including: regressions into the sentence-initial region in the by-participants analysis (F_1 (1,62) = 2.97, p = .09, d = .31; F_2(1,22) = 2.61, p > .05, d = .48), first-pass reading times in the object region (F_1(1,62) = 8.84, p < .01, d = .53; F_2(1,22) = 8.84, p < .01, d = .88), rereading times for the auxiliary region (F_1 (1,62) =5.49, p < .05, d = .42; F_2(1,22) = 5.45, p < .05, d = .69), the post-gap region's regression path (F_1(1,62) = 4.73, p < .05, d = .39; F_2(1,22) = 4.71, p < .05, d = .64) and rereading times (F_1 (1,62) = 3.46, p = .06, d = .33; F_2(1,22) = 3.39, p = .08, d = .54), as well as regressions out of the final region (F_1 (1,62) = 3.07, p = .08, d = .31; F_2(1,22) = 3.00, p = .09, d = .51). We also found a significant interaction between the factors Group and Antecedent Size for rereading

---

[3] There are no regressions into the final region since there were no subsequent regions of text from which such regressions might originate. Regressions out of the sentence-initial region are directed towards the preamble.

times in the auxiliary region ($F_1(1,62) = 26.98$, p < .001, $d = .93$; $F_2(1,22) = 21.72$, p < .001, $d = 1.37$).

The observed interactions, in the presence of main effects of Group, are indicative of different reading-time patterns among the L1 and L2 groups, particularly with regard to the way in which these were influenced by quantification. We therefore went on to analyse each group's data separately. In the following, only significant or marginally significant effects or interactions will be reported.

Reading times at and following the gap. *Auxiliary region.* At the region containing the manipulated auxiliary (e.g. *the explorer was/did*), a main effect of Quantification was found for the L1 group's rereading times ($F_1(1,31) = 8.96$, p < .01, $d = .76$; $F_2(1,23) = 8.20$, p < .01, $d = .84$), reflecting the fact that the quantified conditions were read more slowly than the unquantified ones. Both groups showed significant main effects of Antecedent Size (L1: $F_1(1,31) = 9.99$, p < .01, $d = .80$; $F_2(1,23) = 9.88$, p < .01, $d = .93$; L2: $F_1(1,31) = 17.57$, p < .001, $d = 1.06$; $F_2(1,23) = 17.10$, p < .001, $d = 1.22$). Note, however, that the directionality of this effect differed for the two groups. Whilst the L1 speakers read the long antecedent conditions faster than the short ones, the reverse pattern was seen in the L2 group.

Regressions into the auxiliary region yielded a main effect of Quantification for the L1 group only ($F_1(1,31) = 6.39$, p < .05, $d = .64$; $F_2(1,23) = 5.48$, p < .05, $d = .69$) reflecting significantly fewer regressions for the quantified compared to the unquantified conditions. Both groups again showed a main effect of Antecedent Size, with the long antecedent conditions receiving more regressions in than their short counterparts (L1: $F_1(1,31) = 33.20$, p < .001, $d = 1.46$; $F_2(1,23) = 29.89$, p < .001, $d = 1.61$; L2: $F_1(1,31) = 20.69$, p < .001, $d = 1.16$, $F_2(1,23) = 18.78$, p < .001, $d = 1.28$).

For regressions out we found a marginal main effect of Antecedent Size for the L1 speakers ($F_1(1,31) = 3.42$, p = .07, $d = .47$; $F_2(1,23)= 3.78$, p = .06, $d = .57$), due to the slightly higher number of outbound regressions for the long antecedent conditions.

//TABLE 2 ABOUT HERE//

Post-gap region. The statistical results for the region immediately following the auxiliary (e.g. *in the outback*), which was our primary region of interest, are summarized in Table 2. The native speakers' first-pass reading times showed a marginal main effect of Antecedent Size, due to reading times for the long antecedent conditions being slightly higher than for the short ones.

Regression path times showed only the L1 speakers reading the quantified conditions significantly more slowly than the unquantified ones. Both groups showed main effects of Antecedent Size, however, with the long antecedent conditions being read more slowly than the short ones in the L1 group, and the reverse pattern seen in the L2 group.

The analysis of participants' rereading times again revealed a main effect of Quantification for the L1 group only, with the quantified conditions now being read more quickly than the unquantified ones. A main effect of Antecedent Size was found for both groups, reflecting longer rereading times for the long antecedent conditions (Figure 1). A significant Quantification by Antecedent Size interaction was found only for the L1 group. Planned comparisons showed that in the long antecedent pair, the unquantified condition took the L1 speakers significantly longer to read than the quantified one ($t_1(31) = -2.85$, p = .005, $d = .72$; $t_2(23) = -2.64$, p < .05, $d = .78$). For

the short antecedent pair, this difference was not statistically significant ($t_1(31) = .20$,

$p > .05$, $d = .05$; $t_2(23) = .18$, $p > .05$, $d = .05$). The source of the interaction seen in

the L1 group, then, was the larger difference found between the long compared to the

short antecedent conditions.

//FIGURE 1 ABOUT HERE//

Analysis of the regressions into this region again showed a main effect of

Quantification in the L1 group, reflecting relatively fewer regressions for the

quantified conditions. The L2 group showed a main effect of Antecedent Size instead,

with significantly more regressions into the post-gap region for the long compared to

the short antecedent conditions.

For regressions out of this region, both groups showed main effects of

Antecedent Size, with the long antecedent condition triggering significantly more

regressions than the short antecedent conditions.

Final region. Since participants' reading times at the final sentence region

(e.g. *in the strong heat*) may partly reflect end-of-sentence wrap-up processes (Just &

Carpenter, 1980), statistical results at this region should be interpreted with caution, if

at all. They are reported here for completeness. The L1 group showed a marginal

main effect of Antecedent Size in their regression path times by participants ($F_1(1,31)$

$= 2.88$, $p = .09$, $d = .43$; $F_2(1,23) = 1.94$, $p > .05$, $d = .41$), whilst this effect was

significant in the L2 group ($F_1(1,31) = 7.57$, $p < .01$, $d = .70$; $F_2(1,23) = 6.74$, $p < .05$,

$d = .77$). For both groups, the long antecedent conditions elicited longer regression

path times than their short antecedent counterparts.

Similarly, both groups show a main effect of Antecedent Size (marginal in the L1 speakers) in their rereading times, with the long antecedent conditions taking longer to read than the short ones (L1: $F_1(1,31) = 3.51$, p $= .06$, $d = .48$; $F_2(1,23) = 3.34$, p $= .08$, $d = .54$; L2: $F_1(1,31) = 10.13$, p $< .01$, $d = .81$; $F_2(1,23) = 9.86$, p $< .01$, $d = .93$). Regressions out of the final region showed a main effect of Quantification for the native speakers ($F_1(1,31) = 4.51$, p $= .03$, $d = .54$; $F_2(1,23) = 4.21$, p $= .05$, $d = .61$), with fewer regressions out for the quantified than the unquantified conditions.

Reading times at the precritical regions. *Sentence-initial region.* At the region containing the matrix subject and the beginning of the antecedent predicates (e.g. *The scout was taught to catch*), the analysis of participants' rereading times showed a significant interaction between Quantification and Antecedent Size for the L1 speakers only ($F_1(1,31) = 4.83$, p $= .03$, $d = .56$; $F_2(1,23) = 4.71$, p < 0.05, $d = .64$). Planned comparisons revealed that the source of the interaction was longer rereading times for the unquantified relative to the quantified condition in the long antecedent pair ($t_1(31) = 2.85$, p $< .01$, $d = .72$; $t_2(23) = 2.64$, p $< .05$, $d = .78$), with no significant difference between the short conditions. The L1 group furthermore showed a significant main effect of Quantification ($F_1(1,31) = 7.43$, p $< .01$, $d = .69$, $F_2(1,23) = 7.31$, p $< .05$, $d = .80$) in the number of regressions into this region, with the unquantified conditions attracting significantly more of them than the quantified ones.

Object region. First-pass reading times at the beginning of the object phrase (e.g. *every/the snake that*) showed a main effect of Quantification for both participant groups, with the quantified conditions being read more slowly than the unquantified ones (L1: $F_1(1,31) = 12.41$, p $< .001$, $d = .89$; $F_2(1,23) = 12.01$, p $< .01$, $d = 1.02$; L2: $F_1(1,31) = 28.29$, p $< .001$, $d = 1.35$; $F_2(1,23) = 27.65$, p $< .001$, $d = 1.55$).

The L1 speakers' regression path times are similar to their first-pass times at this region, with a main effect of Quantification ($F_1(1,31) = 4.14$, p < .05, $d = .52$; $F_2(1,23) = 3.99$, p = .06, $d = .59$) reflecting longer reading times for the quantified conditions.

A main effect of Quantification was also found in the L1 group's rereading times ($F_1(1,31) = 4.69$, p < .05, $d = .55$; $F_2(1,23) = 4.57$, p < .05, $d = .63$). However, the direction of this effect was now reversed, with the quantified conditions being read significantly faster than the unquantified ones. Both L1 and L2 speakers also showed a main effect of Antecedent Size at this region (L1: $F_1(1,31) = 7.04$, p < .01, $d = .67$; $F_2(1,23) = 6.99$, p < .05, $d = .78$; L2: $F_1(1,31) = 6.68$, p < .05, $d = .66$; $F_2(1,23) = 6.55$, p < .05, $d = .75$), caused by the long antecedent conditions eliciting longer rereading times than their short antecedent counterparts.

As with the sentence-initial region, regressions into the object region showed a main effect of Quantification for the L1 group only ($F_1(1,31) = 6.79$, p < .05, $d = .66$; $F_2(1,23) = 6.48$, p < .05, $d = .75$), reflecting fewer regressions for the quantified compared to the unquantified conditions. The L1 group showed a marginal main effect of Antecedent Size in the by-participants analysis instead ($F_1(1,31) = 2.87$, p = .09, $d = .43$; $F_2(1,23) = 2.78$, p > .05, $d = .49$), due to the long antecedent conditions receiving slightly more regressions than the short ones.

Discussion

The present study sought to investigate whether the elided material is recovered at ACD gap sites, and whether ACD resolution is affected by the object NP's quantificational status, in both L1 and L2 comprehension. The analysis of the reading-time data revealed both similarities and differences between the L1 and L2 participant

groups. After coming across the VP gap, both groups showed sensitivity to our antecedent size manipulation, replicating an effect also observed by Hackl et al. (2012) for L1 speakers. The main difference between the L1 and L2 participant groups that we observed was that, following their detection of a VP gap, only the native speakers' eye-movement patterns were affected by the object NP's quantificational status. Our findings and their implications will be discussed in more detail below.

Antecedent Recovery Effects

As noted by Martin and McElree (2008), ACD resolution should be affected by the antecedent VP's complexity only if the elided material is actually recovered. Our results provide strong evidence that readers did indeed attempt to recover the missing VP's contents after encountering an elliptical gap. This is shown by longer reading times and higher proportions of regressions (both in and out of an interest region) for syntactically and semantically more complex antecedents across several eye-movement measures and sentence regions.

Significant effects of antecedent size were seen in both participant groups, indicating that both L1 and L2 speakers tried to recover the missing VPs' contents after coming across the gap. Effects of antecedent size were visible slightly earlier in the L1 than in the L2 group's data. In the L1 group, long antecedents triggered more regressive eye movements than short ones out of the auxiliary region (e.g. *the explorer was/did*), where a VP gap might first be postulated, as well as longer first-pass reading times at the region at which the gap was confirmed (e.g. *in the outback*). This indicates that the L1 group initially found long antecedents more difficult to

recover than short ones, which is inconsistent with the prediction made by "pointing" approaches to ellipsis resolution (e.g. Martin & McElree, 2008).

In the L2 group, reliable antecedent size effects were first observed in regression path times at the post-gap region, but with the short antecedent conditions taking longer to read than the long antecedent conditions. We will return to this observation below. Note that the direction of this effect reversed at later processing stages, however, with the L2 speakers showing more regressions out of the post-gap region, and longer rereading times, for the long antecedent conditions.

Effects of Quantification

As noted earlier, theoretical approaches differ with regard to the question of whether or not antecedent recovery should be affected by the object NP's quantificational status. On the assumption that the default interpretation of unquantified definite object NPs is weak, QR is assumed to apply to these only as a last resort, triggered by encountering an otherwise unresolvable VP gap. The processing effort associated with the *ad hoc* raising of an unquantified object NP in order to help resolve ACD-type ellipsis should be reflected in longer reading times for the unquantified compared to the quantified conditions during or following readers' processing of the gap. Gibson et al.'s (2014) pragmatic account also predicts facilitating effects of quantification at later sentence regions and/or processing stages, making the two accounts difficult to distinguish empirically.

The results from our L1 group partly confirm the above prediction. Following the L1 speakers' encountering of the VP gap, the quantified object conditions elicited longer regression path times than the unquantified ones at the critical post-gap region initially, though, and longer rereading times than the unquantified ones at the

preceding auxiliary region. Considering that Hackl et al. (2012) also observed numerically longer reading times for *every NPs* compared to *the NPs* two words after the auxiliary in their long antecedent conditions, this unexpected processing advantage for the unquantified conditions requires an explanation. We might speculate that the additional processing cost associated with quantified compared to unquantified object NPs following native readers' identification of a VP gap might reflect an initial attempt to recover the missing predicate phrase verbatim, with quantified but not unquantified objects then giving rise to a local type-mismatch effect.

Quantification had facilitating effects elsewhere for the L1 group, however. Regions one to four attracted fewer regressive eye movements in the quantified compared to the unquantified conditions, and quantified NPs elicited shorter rereading times than unquantified ones at the object phrase itself. The quantified conditions also triggered relatively fewer regressive eye movements out of the sentence-final region. Interactions between the factors antecedent size and quantification were seen in the L1 group's rereading times at the sentence-initial and post-gap regions, with quantification facilitating processing significantly in the long (but not the short) antecedent conditions. This particular finding differs from Hackl et al.'s, who found facilitating effects of quantification restricted to their short antecedent conditions. From the perspective of a QR approach, our results (contra Hackl et al.'s, 2012) suggest that QR readily takes matrix clause scope (e.g. Kennedy, 1997).[4] Although

---

[4] Note that a preference for maximal scope (i.e. larger antecedent) interpretations for VPE was previously noted by Frazier and Clifton (2001) and Sugawara et al. (2013), but see Gibson et al. (2014) for some arguments as to why the benefits of quantification should in principle affect both long and short antecedent conditions.

the question of whether QR preferentially takes minimal or maximal scope is clearly

of theoretical interest, resolving this issue was not one of the aims of the current

study. There are several possible reasons as to why the interaction observed in ours

and Hack et al.'s studies showed a different direction, and lack of space prevents us

from speculating about these any further here.[5] Irrespective of these issues, the crucial

finding, from the point of view of the current study, in both our and Hackl et al.'s

studies is that L1 speakers' reading times were measurably influenced by the

quantificational status of the object noun phrase.

Unlike the L1 group, our L2 group did not show any effects of quantification

following their encountering of the VP gap. Both groups took longer to read

quantified compared to unquantified NPs when these were first encountered, however.

Although this early effect of quantification is consistent with the QR account and

unexpected from the point of view of Gibson et al.'s (2014) pragmatic account, it

could also be due to quantified expressions having more complex semantic

representations than unquantified ones and thus taking longer to encode (Carminati,

Frazier, & Rayner, 2002); or it could simply be a length effect (*every* vs. *the*).

L1/L2 Differences in ACD Processing

Whilst our non-native participants showed evidence of antecedent recovery in the

shape of antecedent size effects, there was no evidence of quantified object NPs either

slowing down or facilitating ACD resolution in this group.

Given the partial discrepancy between ours and Hackl et al.'s (2012) results, no firm

conclusions seem warranted here as to possible target sites for QR, however.

[5] Note that a direct comparison between Hackl et al.'s (2012) and our results is

precluded due to differences in the experimental method, materials and design.

From the perspective of the QR approach to ACD resolution, the absence of any post-gap quantification effects might indicate that the L2 group either (i) QR-ed both quantified and unquantified objects NPs indiscriminately, (ii) did not apply any QR at all in the first place, or (iii) QR-ed quantified objects but then resolved subsequent ACD gaps by some mechanism unaffected by the object's quantificational status. We think that the first possibility is unlikely given that the L2 group did show sensitivity to our quantification manipulation during their first reading of the object region containing the determiner or quantifier, and on the assumption that definite determiner's normally have a non-quantificational interpretation and thus do not undergo QR as readily as quantified NPs (e.g. Heim & Kratzer, 1998; Szabolcsi, 2014). As to the second possibility, there is little reason to suspect that L2 learners have any difficulty with QR as such. Although Ionin, Luchkina and Stoops (2014) observed a non-nativelike general preference for surface scope readings in L1 English/L2 Russian speakers for sentences containing more than one quantifier, other studies have shown that sufficiently advanced L2 learners can compute inverse scope readings and are even able to acquire L2 specific constraints on quantifier scope (Chung, 2013; Marsden, 2004). Note also that even though German, the native language of our L2 participants, lacks ACD-type ellipsis, the existence of quantifier scope ambiguities in German suggests that L1 German speakers do not have any problem applying covert scope-shifting operations.

This leaves us with possibility (iii), the idea that the L2 group recovered the missing content in some way independently of whether or not the object NP was initially QR-ed. After encountering the post-auxiliary gap, our L2 speakers may have recovered a syntactically shallow or purely semantic representation of the elided VP which lacked the problematic recursive gap altogether. In this case, QR would be

unnecessary for interpreting ACD-type ellipsis. A possible formalisation of this idea is Jacobson's (1998) proposal that VPE is actually transitive verb ellipsis, with the auxiliary functioning like a free pronoun in search for an antecedent: the missing verb or verbal complex's meaning, which can be picked up from the discourse.

That is, there may in fact be two alternative mechanisms available for resolving ACDs, with the QR-independent "semantic" route more likely to be applied by L2 than by L1 comprehenders. An interpretation of our L2 data along these lines would be consistent with Clahsen and Felser's (2006) shallow structure hypothesis, according to which L2 comprehenders may have difficulty recovering abstract structural detail during real-time processing. It also fits with Duffield and Matsuo's (2009) and Duffield et al.'s (2009) observation that L2 speakers are more likely to accept syntactically non-parallel antecedents for VPE, suggesting that VPE resolution is subject to a semantic rather than a syntactic identity constraint (Merchant, 2001).

From the perspective of Gibson et al.'s (2014) pragmatic approach, the lack of quantification effects on ACD resolution in the L2 group might suggest the absence (during online comprehension) of the additional inferential processing thought to be triggered by the pragmatically infelicitous unquantified conditions. Whilst the above possibilities are difficult to distinguish empirically, they have in common the idea that L2 speakers have difficulty integrating different types of information during online processing (compare e.g. Hopp, 2009; Sorace, 2011). In the case of ACD resolution, simultaneously recovering the elided material and building or accessing abstract LF representations which include scope-shifted operators may have exceeded the L2 speakers' processing resources. Alternatively, the L2 group's lack of post-gap quantification effects may reflect difficulty taking into account pragmatic information

(i.e. the absence of overt 'sameness' indicators in the unquantified conditions) during antecedent recovery.

Another L1/L2 difference we observed concerns the direction of the antecedent size effect seen during the two groups' initial processing of the post-gap region (e.g. *in the outback*). Whilst the L1 group took less time to process the post-gap region in the short than in the long antecedent conditions, the L2 group showed an initial processing advantage for the long antecedent conditions in their regression path times. This could be taken to indicate that, when trying to recover the missing information, the L2 speakers initially focused on the more salient matrix verb. Alternatively, the reversal of the antecedent size effect could simply be a reflex of the fact that the sequence of words being processed at the time was locally grammatical in the long antecedent conditions (e.g. *the explorer was in the outback*) but not in the short antecedent conditions (e.g. **the explorer did in the outback*). Put differently, the L2 participants may have been more likely to miss the presence of a VP gap during their initial reading of this region in the long compared to the short antecedent conditions. As this reading-time pattern was not seen in any of the other reading-time measures, it should however be interpreted with caution.

Concluding Remarks

Our results showed both similarities and differences between L1 and L2 comprehenders' online resolution of ACD gaps. Both groups showed evidence of trying to recover the missing information after encountering a VP gap, even though German, the native language of our L2 participants, lacks genuine VPE. Only the native speakers' post-gap processing patterns were affected by the object NP's quantificational status, however. Whereas the L1 results are consistent with

theoretical accounts which predict that quantification should facilitate ACD

resolution, our L2 speakers did not measurably benefit from quantification.

      Whilst our study has revealed some interesting L1/L2 differences in real-time

ACD resolution, it also leaves several questions unanswered. One concerns the

possible influence of learners' L1 on their processing of ACD, which can only be

assessed by comparing learners from different L1 backgrounds. Another empirical

question is whether or not L2 speakers' reduced sensitivity to quantification extends to

phenomena other than ACD. A further issue that the current study did not investigate

is the role of proficiency in non-native ACD resolution. Other, more general questions

concern the extent to which L2 learners are capable of carrying out syntactic and non-

syntactic processes concurrently, and the precise nature of the representations

computed during ellipsis resolution. It is conceivable that syntactic and semantic or

pragmatic routes to ellipsis resolution co-exist, with some populations more likely to

employ one or the other. Future research may want to explore these issues further.

References

Allan, D. (2004). *The Oxford Placement Test*. Oxford, UK: Oxford University Press.

Arregui, A., Clifton, C., Frazier, L., & Moulton, K. (2006). Processing elided verb

phrases with flawed antecedents: The recycling hypothesis. *Journal of Memory

and Language, 55,* 232–246.

Breakstone, M.Y., Cremers, A., Fox, D., & Hackl, M. (2011). On the analysis of

scope ambiguities in comparative constructions: Converging evidence from real-

time sentence processing and offline data. In N. Ashton, A. Chereches, & D. Lutz

(Eds.), *Proceedings of SALT* 21 (pp. 712-731). Cornell University Department of

Linguistics.

Carminati, M., Frazier, L., & Rayner, K. (2002). Bound variables and c-command.

Journal of Semantics, 19, 1-34.

Chung, E.S. (2013). Sources of difficulty in L2 scope judgments. *Second Language

Research, 29,* 285–310.

Clahsen, H. & Felser, C. (2006). How native-like is non-native language processing?

Trends in Cognitive Sciences, 10, (12), 564-570.

Cormack, A. (1984). VP anaphora: variables and scope. In F. Landman & F. Veltman

(Eds.), *Varieties of formal semantics* (pp. 81–102). Dordrecht: Foris Publications.

Culicover, P., & Jackendoff, R. (2005). *Simpler syntax*. Oxford: Oxford University

Press.

Dallas, A. & Kaan, E. (2008). Second language processing of filler-gap dependencies

by late learners. *Language and Linguistics Compass, 2.*

Dalrymple, M., Shieber, S., & Pereira, F. (1991). Ellipsis and higher order unification.

Linguistics and Philosophy, 14, 399-452.

Dekydtspotter, L., & Miller, A.K. (2013). Inhibitive and facilitative priming induced by traces in the processing of wh-dependencies in a second language. *Second Language Research, 29,* 345–372.

Diesing, M. (1992). *Indefinites*. Cambridge, MA: MIT Press.

Duffield, N., & Matsuo, A. (2009). Native-speakers' vs. L2 learners sensitivity to parallelism in VP-Ellipsis. *Studies in Second Language Acquisition, 31,* 1-31.

Duffield, N., Matsuo, A., & Roberts, L. (2009). Factoring out the parallelism effect in VP-ellipsis: English vs. Dutch contrasts. *Second Language Research*, *25*, 427-467.

Felser, C. (2015). Native vs. non-native processing of discontinuous dependencies. *Second Language, 14*.

Felser, C., & Cunnings, I. (2012). Processing reflexives in English as a second language: The role of structural and discourse-level constraints. *Applied Psycholinguistics, 33,* 571-603.

Felser, C., & Roberts, L. (2007). Processing wh-dependencies in a second language: A cross-modal priming study. *Second Language Research, 23,* 9-36.

Frazier, L., & Clifton, C. (2000). On bound variable interpretations: The LF-only hypothesis. *Journal of Psycholinguistic Research, 29,* 125–140.

Frazier, L., & Clifton, C. (2001). Parsing coordinates and ellipsis: Copy alpha. *Syntax, 4,* 1-22.

Frazier, L., & Clifton, C. (2005). The syntax-discourse divide: Processing ellipsis. *Syntax, 8,* 121-174.

Gibson, E., Jacobson, P., Graff, P., Mahowald, K., Fedorenko, E., & Piantadosi, S.T. (2014). A pragmatic account of complexity in definite Antecedent-Contained-

Deletion relative clauses. *Journal of Semantics*. Published online,

doi:10.1093/jos/ffu006

Ginzburg, J., & Sag, I.A. (2000). *Interrogative investigations*. Stanford, CA: CSLI

Publications.

Hackl, M., Koster-Hale, J., & Varvoutis, J. (2012). Quantification and ACD: evidence

from real-time sentence processing. *Journal of Semantics, 29,* 145–206.

Hardt, D. (1993). VP ellipsis and contextual interpretation. Technical Reports (CIS).

Paper 417. http://repository.upenn.edu/cis_reports/417.

Harley, H. (2002). ACD, WCO and QR of DPs. *Linguistic Inquiry, 33,* 659-664.

Heim, I., & Kratzer, A. (1998). *Semantics in Generative Grammar*. Blackwell.

Malden, MA.

Hopp, H. (2009). The syntax-discourse interface in near-native L2 acquisition: Off-

line and on-line performance. *Bilingualism: Language and Cognition, 12,* 463-

483.

Hornstein, N. (1994). An argument for minimalism: The case of antecedent-contained

deletion. *Linguistic Inquiry, 25,* 455–480.

Ionin, T., Luchkina, T., & Stoops, A. (2014). Quantifier scope and scrambling in the

second language acquisition of Russian. In U. Minai et al. (Eds.), *Selected*

proceedings of the 5th Conference on Generative Approaches to Language

Acquisition–North America (pp. 169-180). Somerville, MA: Cascadilla

Proceedings Project.

Jacobson, P. (1998). Direct compositionality and variable-free semantics: The case of

antecedent contained deletion. In K. Johnson (Ed.), *Topics in ellipsis* (pp. 30-68).

Cambridge, UK: Cambridge University Press.

Just, M.A., & Carpenter, P. (1980). A theory of reading: From eye fixations to

    comprehension. *Psychological Review, 87,* 329-354.

Kaplan, J. (1984). Obligatory too in English. *Language, 60,* 510-518.

Kennedy, C. (1997). Antecedent contained deletion and the syntax of quantification.

    Linguistic Inquiry, 28, 662-688.

Kiguchi, H., & Thornton, R. (2004). Binding principles and ACD constructions in

    child grammars. *Syntax, 7,* 234-271.

Lobeck, A. (1995). *Ellipsis: Functional Heads, Licensing and Identification*. New

    York: Oxford University Press.

Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language

    sentence processing. *Studies in Second Language Acquisition, 27,* 53-78.

Marsden, H. (2004). Quantifier scope in non-native Japanese: A comparative

    interlanguage study of Chinese, English, and Korean-speaking learners. PhD

    dissertation, University of Durham.

Martin, A.E., & McElree, B. (2008). A content-addressable pointer mechanism

    underlies comprehension of verb-phrase ellipsis. *Journal of Memory and*

    Language, 58, 879-906.

May, R. (1985). *Logical form: its structure and derivation*. Cambridge, MA: MIT

    Press.

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

    Oxford, UK: Oxford University Press.

Murphy, G.L. (1985). Processes of understanding anaphora. *Journal of Memory and*

    Language, 24, 290–303.

Patterson, C., Trompelt, H., & Felser, C. (2014). The online application of binding condition B in native and non-native pronoun resolution. *Frontiers in Psychology, 5*:147. doi: 10.3389/fpsyg.2014.00147.

Phillips, C., & Parker, D. (2014). The psycholinguistics of ellipsis. *Lingua*, *151,* 78-95.

Rayner, K. (1998). Eye-movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372-422.

Roberts, L., Gullberg, M., & Indefrey, P. (2008). Online pronoun resolution in L2 discourse: L1 influence and general learner effects. *Studies in Second Language Acquisition, 30,* 333-357.

Rosales Sequeiros, X. (2004). Interpretation of reflexive anaphora in L2 VP-ellipsis: Relevance Theory and paradigms of explanation. *Second Language Research 20,* 256-280.

Ross, J.R. (1969). Guess who? In R.I. Binnick, A. Davidson, G.M. Green & J.L. Morgan (Eds.), *Proceedings of the Fifth Annual Meeting of the Chicago Linguistics Society* (pp. 252-286). Chicago, IL: CLS.

Sag, I.A. (1976). *Deletion and Logical Form*. MIT dissertation.

Shapiro, L.P., & Hestvik, A. (1995). On-line comprehension of VP-ellipsis: Syntactic reconstruction and semantic influence. *Journal of Psycholinguistic Research, 24,* 517-532.

Shapiro, L.P., Hestvik, A., Lesan, L., & Garcia, A.R. (2003). Charting the time-course of sentence processing: Reconstructing missing arguments in VP-ellipsis constructions. *Journal of Memory and Language, 49,* 1-19.

Sorace, A. (2011). Pinning down the concept of "interface" in bilingualism. *Linguistic Approaches to Bilingualism, 1,* 1-33.

Sugawara, A., Kotek, H., Hackl, M., & Wexler, K. (2013). Long vs. short QR: Evidence from the acquisition of ACD. *Proceedings of Boston University Conference on Language Development (BUCLD) 37*. Somerville, MA: Cascadilla Press.

Syrett, K., & Lidz, J. (2009). Quantifier Raising in 4-year-olds. *Language Acquisition, 16,* 67-81.

Syrett, K., & Lidz, J. (2011). The locality of QR: Evidence from children's interpretations of antecedent contained deletion. *Linguistic Inquiry, 42,* 305-337.

Szabolcsi, A. (2014). Quantification and ACD: What is the evidence from real-time processing evidence for? A reply to Hackl et al. (2012). *Journal of Semantics, 31*, 135–145.

Tanenhaus, M., & Carlson, G.N. (1990). Comprehension of deep and surface verb phrase anaphors. *Language and Cognitive Processes*, *5*, 257 – 280.

Winkler, S. (2005). *Ellipsis and focus in generative grammar.* Berlin/New York: de Gruyter.

Ying, H.G. (2005). Relevance and second language learners' interpretation of reflexive anaphora in VP-ellipsis. *Language Sciences*, *27,* 551-570.

Yoshida, M., Walsh Dickey, M., & Sturt, P. (2013). Predictive processing of syntactic structure: Sluicing and ellipsis in real-time sentence processing. *Language and Cognitive Processes, 28,* 272-302.

Appendix: List of experimental items

1. Coursework is often done during vacations. Jack was encouraged to read {every/the} book that Mary {did/was} during the study break in the summer.

2. Hospitals worry about people's safety. The nurse was told to check {every/the} patient that the doctor {did/was} throughout the day during a storm.

3. The building needed brightening up. The decorator was asked to paint {every/the} wall that the painter {did/was} before the opening of the new building.

4. Visitors like to feed zoo animals. The tourist was instructed to feed {every/the} penguin that the zookeeper {did/was} during the feeding time on Monday.

5. Campaigners want to stop smoking. The TV show was made to display {every/the} health warning that the studio {did/was} when actors smoke during the day.

6. A local store wanted to sell more candy. The chef was asked to create {every/the} sweet that the shop {did/was} after a survey of people's favourites.

7. Some people like to test their IQ. Sally was required to calculate {every/the} sum that the computer {did/was} during the test to see who was smarter.

8. Many people travel around Europe. Tom was inclined to drink {every/the} beer that Linda {did/was} during the festival in southern Germany.

9. Some teachers don't watch television. The teacher was persuaded to see {every/the} cartoon that the student {did/was} after the visit to the common room.

10. Each new-year people throw bigger parties. Gemma was asked to attend {every/the} party that Mary {did/was} for the new-year celebrations in December.

11. Theatres suffer when actors get sick. The substitute was told to learn {every/the} line that the actor {did/was} just in case someone became very ill.

12. Politicians make hard choices. The new president was required to keep {every/the} promise that the treasurer {did/was} throughout the period of uncertainty.

13. The Government liked skills training. The university was told to teach {every/the} course that Oxford {did/was} following a review of academic standards.

14. Pets can do impressive stunts. The dog was made to perform {every/the} trick that Lassie {did/was} after the success of certain movies.

15. Australian scouts master many skills. The scout was taught to catch {every/the} snake that the explorer {did/was} in the outback in the strong heat.

16. Many people get nervous on planes. The pilot was requested to comfort {every/the} passenger that the attendant {did/was} during a bad and bumpy flight.

17. Some people do things they later regret. Rachel was persuaded to eat {every/the} cake that Simon {did/was} even though the food was rather disgusting.

18. Some bands get carried away. The guitarist was tempted to perform {every/the} song that the singer {did/was} at the lively and exciting concert.

19. Writers are competitive. The best journalist was asked to write {every/the} story that the newspaper {did/was} about the big upcoming sports event.

20. Schools promote their best teachers. The professor was asked to present {every/the} lecture that the department {did/was} at the open day event in August.

21. Plants often need a lot of water. The gardener was told to water {every/the} plant that the florist {did/was} during the warm weather last spring.

22. Shops compete with each other. The shop owner was persuaded to reduce {every/the} price that the rival shop {did/was} when sales dropped in the winter.

23. Party games seem better these days. The adult was tempted to play {every/the} game that the kid {did/was} at the birthday party a few weeks ago.

Networking is often required in business. Nick was invited to attend {every/the}

meeting that the boss {did/was} at the important conference in Paris

		L1 speakers					L2 speakers				
		First-pass reading time	*Regression Path*	*Rereading time*	*Regressions in (%)*	*Regressions out (%)*	*First-pass reading time*	*Regression path*	*Rereading time*	*Regressions in (%)*	*Regressions out (%)*
Region 1	L, + Q	920 (322)	1059 (473)	441 (291)	46.4 (18)	12.0 (13)	1367 (572)	1545 (452)	821 (395)	54.7 (38)	10.0 (10)
	L, - Q	879 (253)	1003 (435)	628 (219)	52.6 (26)	11.6 (7)	1374 (472)	1503 (399)	918 (412)	55.3 (44)	12.1 (9)
	S, + Q	849 (394)	1040 (502)	485 (288)	41.7 (18)	11.5 (18)	1316 (333)	1551 (431)	833 (302)	53.1 (28)	11.5 (10)
	S, - Q	853 (424)	1014 (422)	463 (221)	53.1 (29)	10.5 (14)	1328 (437)	1535 (482)	796 (297)	55.0 (31)	10.6 (7)
Region 2	L, + Q	439 (224)	529 (475)	302 (271)	47.4 (14)	6.4 (10)	776 (372)	808 (522)	628 (402)	59.4 (37)	2.1 (7)
	L, - Q	406 (194)	477 (487)	404 (305)	59.9 (25)	4.9 (6)	633 (312)	751 (572)	654 (374)	61.1 (46)	5.9 (3)
	S, + Q	464 (187)	525 (407)	236 (196)	46.9 (31)	5.6 (4)	739 (258)	843 (521)	488 (219)	53.7 (42)	6.3 (6)
	S, - Q	402 (266)	466 (393)	266 (214)	51.6 (22)	4.8 (5)	628 (278)	778 (595)	452 (263)	54.5 (39)	6.5 (5)
Region 3	L, + Q	299 (212)	472 (372)	398 (124)	60.4 (19)	24.6 (17)	488 (315)	709 (660)	759 (342)	69.3 (28)	19.3 (10)
	L, - Q	294 (228)	505 (477)	339 (179)	69.8 (21)	31.6 (25)	503 (341)	682 (549)	858 (441)	66.8 (22)	18.1 (13)
	S, + Q	323 (274)	440 (295)	486 (254)	40.1 (28)	20.7 (18)	532 (296)	683 (514)	484 (362)	47.9 (21)	18.1 (11)
	S, - Q	346 (245)	501 (368)	411 (265)	47.4 (24)	23.2 (15)	534 (281)	694 (503)	468 (325)	56.6 (31)	17.1 (14)
Region 4	L, + Q	512 (371)	963 (531)	497 (288)	46.9 (31)	41.9 (28)	701 (319)	1047 (746)	816 (332)	58.3 (37)	35.9 (21)
	L, - Q	552 (263)	683 (493)	724 (321)	57.3 (34)	45.0 (21)	703 (421)	1020 (721)	881 (276)	61.1 (31)	34.0 (18)
	S, + Q	478 (289)	779 (711)	312 (143)	45.8 (21)	16.8 (19)	674 (357)	1121 (509)	547 (298)	43.2 (29)	23.3 (24)
	S, - Q	503 (255)	686 (741)	376 (189)	47.9 (27)	22.6 (21)	654 (494)	1078 (603)	532 (206)	48.7 (34)	26.1 (22)
Region 5	L, + Q	525 (311)	642 (452)	266 (103)	--	67.5 (37)	739 (393)	921 (427)	463 (421)	--	77.6 (37)
	L, - Q	468 (297)	663 (471)	302 (174)	--	75.5 (41)	744 (276)	917 (452)	566 (383)	--	76.8 (42)
	S, + Q	477 (304)	621 (411)	228 (132)	--	66.1 (28)	742 (205)	877 (466)	333 (372)	--	73.2 (29)
	S, - Q	464 (237)	634 (398)	212 (109)	--	72.0 (35)	712 (266)	884 (439)	338 (322)	--	72.4 (25)

Table 1. Participants' mean reading times (in milliseconds) and proportions of regressions for all sentence regions, SD in parentheses (L = long antecedent, S = short antecedent, +Q = quantified, -Q = unquantified).

		1st Pass Times	Regression Path	Rereading Times	Regressions In	Regressions Out
L1 speakers						
Antecedent Size	F1 (1, 31)	2.95(*) (.44)	6.14* (.63)	28.39** (1.35)	n.s.	33.42** (1.47)
	F2 (1, 23)	3.52(*) (.55)	5.99* (.72)	27.67** (1.55)	n.s.	28.99** (1.59)
Quantification	F1 (1, 31)	n.s.	28.58** (1.36)	16.84** (1.04)	3.69(*) (.49)	n.s.
	F2 (1, 23)	n.s.	27.89** (1.56)	16.22** (1.19)	4.56* (.63)	n.s.
Interaction	F1 (1, 31)	n.s.	n.s.	5.76* (.61)	n.s.	n.s.
	F2 (1, 23)	n.s.	n.s.	5.61* (.70)	n.s.	n.s.
L2 speakers						
Antecedent Size	F1 (1, 31)	n.s.	6.47* (.65)	17.34** (1.06)	11.69** (.87)	8.62** (.75)
	F2 (1, 23)	n.s.	6.18* (.73)	16.78** (1.21)	10.88** (.97)	7.90** (.83)
Quantification	F1 (1, 31)	n.s.	n.s.	n.s.	n.s.	n.s.
	F2 (1, 23)	n.s.	n.s.	n.s.	n.s.	n.s.
Interaction	F1 (1, 31)	n.s.	n.s.	n.s.	n.s.	n.s.
	F2 (1, 23)	n.s.	n.s.	n.s.	n.s.	n.s.

Table 2. Summary of per-group ANOVA results for the post-gap region ($p < .1 =$ (*), $p < .05 = *$, $p < .01 = **$;

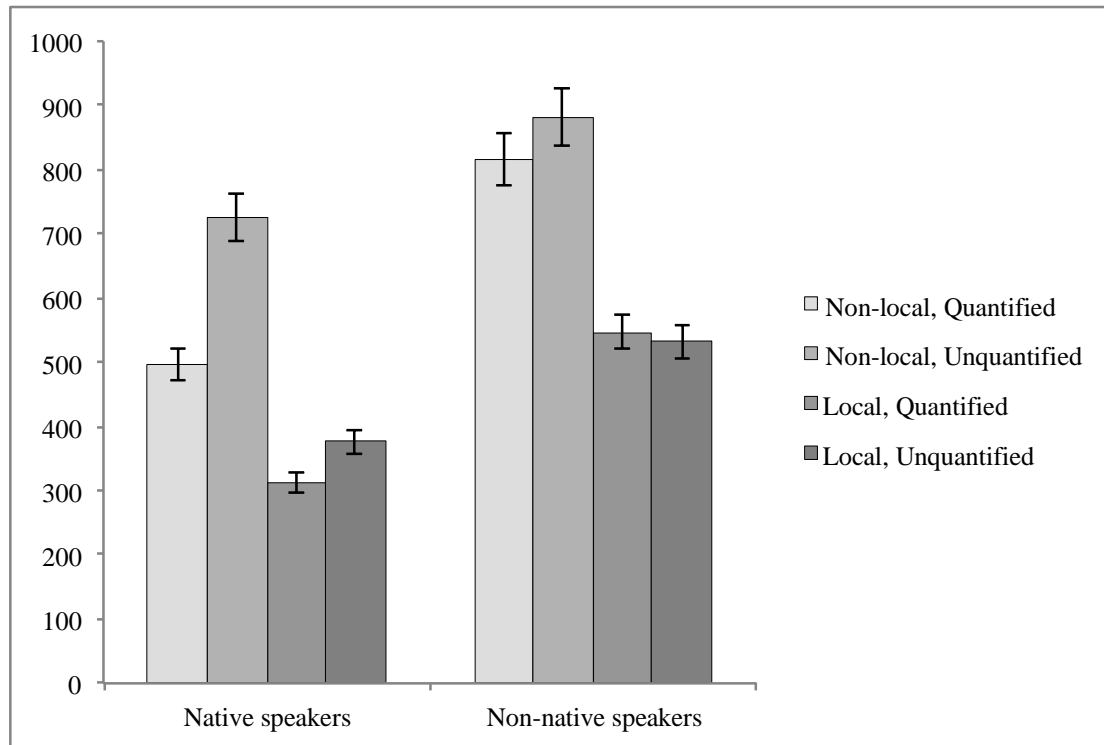Cohen's *d* effect sizes are reported in parentheses for significant effects).

Figure 1. Overview of native and non-native speakers' rereading times (in milliseconds) at the post-gap region across the four experimental conditions.