Bilingualism: Language and Cognition 20 (3), 2017, 494–511 © Cambridge University Press 2016 doi:10.1017/S1366728915000942

Sensitivity to parasitic gaps inside subject islands in native and non-native sentence processing*

OLIVER BOXELL
University of Potsdam
CLAUDIA FELSER
University of Potsdam

(Received: November 25, 2015; final revision received: December 18, 2015; accepted: December 28, 2015; first published online 11 February 2016)

We report the results from an eye-movement monitoring study that investigated late German–English bilinguals' sensitivity to parasitic gaps inside subject islands. The online reading experiment was complemented by an offline scalar judgement task. The results from the offline task confirmed that for both native and non-native speakers, subject island environments must normally be non-finite in order to host a parasitic gap. The analysis of the reading-time data showed that, while native speakers posited parasitic gaps in non-finite environments only, the non-native group initially overgenerated parasitic gaps, showing delayed sensitivity to island-inducing cues during online processing. Taken together, our findings show that non-native comprehenders are sensitive to exceptions to island constraints that are not attested in their native language and also rare in the L2 input. They need more time than native comprehenders to compute the linguistic representations over which the relevant restrictions are defined, however.

Keywords: Parasitic gaps, island constraints, bilingual sentence processing, eye-movement monitoring, English

Introduction

A growing body of second language (L2) processing research has examined non-native speakers' ability to process sentences with non-canonical word orders, or filler-gap dependencies (henceforth, FGDs) (see Dallas & Kaan, 2008, for review). Processing FGDs such as (1) below requires a fronted element to be temporarily stored in working memory until it can be linked to its lexical licenser. Mentally reconstructing *which waiter* at its canonical position following the subcategorizing verb *fire* (indicated by a gap "__"), for example, would establish it as the grammatical object and semantic theme of this verb.

(1) **Which waiter** did the hotel guests think [$_{CP}$ that the receptionist said [$_{CP}$ that the manager might fire __]]?

Whilst several studies have found native (L1) and L2 processing of FGDs to be similar (e.g., Cunnings, Batterham, Felser & Clahsen, 2010; Williams, Möbius & Kim, 2001; Omaki & Schulz, 2011), L2 comprehenders have also sometimes been reported to lack sensitivity to structural detail when processing FGDs (e.g., Marinis, Roberts, Felser & Clahsen, 2005; Felser & Roberts, 2007). The present study uses a time-course sensitive experimental technique – eye-movement monitoring during reading – to investigate L1 and L2 speakers' sensitivity to some fairly subtle restrictions on FGD formation in English.

Although in principle unbounded, FGDs are subject to a number of constraints, including so-called island constraints (Chomsky, 1962; Ross, 1967). 'Islands' are sentence regions within which a filler cannot legitimately search for a gap. Examples (2)–(4) illustrate violations of some types of island, with the filler in bold.

(2) COMPLEX DP ISLAND

*__What__ did the student make [$_{DP}$ the claim that he studied __]?

(3) WH-ISLAND

*__What__ did Jennifer wonder [$_{CP}$ why the girl loved __]?

(4) SUBJECT ISLAND

*__Wha__t was [$_{CP}$ that the teacher confiscated __] considered to be a fact?

From a processing perspective, an interesting question is how island constraints interact with a well-attested economy principle that guides the processing of unbounded dependencies, the so-called Active Filler

* This research was supported by an Alexander-von-Humboldt professorship awarded to Harald Clahsen. We thank Robert Kluender, Matt Wagers and the BLC action editor and reviewers for helpful comments and discussion.

Address for correspondence: Dr. Claudia Felser, University of Potsdam, Potsdam Research Institute for Multilingualism, Karl-Liebknecht-Str. 24–25, 14476 Potsdam, Germany
felser@uni-potsdam.de
Supplementary material can be found online at http://dx.doi.org/10.1017/S1366728915000942

Strategy (AFS) (Clifton & Frazier, 1989). According to the AFS, the processor will seek to minimise the length of FGDs (and thus, memory storage cost) by linking a filler to the earliest potential gap that is comes across. Results from monolingual processing studies indicate that island domains tend to be respected during real-time comprehension (e.g., Traxler & Pickering, 1996). This suggests that encountering an island domain can lower the expectation for a gap in such environments and thus prevent the parser from linking a filler to a potential licenser within an island region.

Island constraints are not necessarily inviolable, though, and there are well-known exceptions to them.[1] In the present paper we investigate the processing of so-called parasitic gaps (PGs) that occur inside sentence regions which are normally considered islands for extraction (Engdahl, 1983). PGs may be deemed acceptable when they co-occur with a licit gap elsewhere in the sentence (on which they are thought to be 'parasitic').[2] The phenomenon is illustrated in (5)–(7) below, where the gap inside the bracketed constituent is a PG.

(5) Which lady did [$_{\text{SUBJECT}}$ the attempt to help __] actually upset __ ?

(6) Which class did [$_{\text{SUBJECT}}$ the teacher's talking to __] calm down __ ?

(7) Which cake did Gemma eat __ [$_{\text{ADJUNCT}}$ before throwing away __]?

Most previous research on the L2 processing of filler-gap dependencies indicates that non-native speakers are sensitive to syntactic islands whose presence is signaled by a highly salient islandhood cue: an overt *wh*-element intervening between the filler and the licit gap (e.g., Felser, Cunnings, Batterham & Clahsen, 2012; Omaki & Schulz, 2011). Little is known however about L2 speakers' online sensitivity to other types of islandhood cue and, to our knowledge, their sensitivity to *exceptions* to island constraints during processing has never been investigated. The current study focuses on PGs inside complex subject islands, which are not attested in our non-native participants' native language and are also underdetermined by the L2 input. This allows us to explore whether language-specific restrictions on filler-gap formation can be acquired and applied during real-time processing despite being extremely rare in the input. Finding that this is the case would be problematic for exposure-based models of L2 acquisition and processing (e.g., Ellis, 1998), and would instead support claims to the effect that restrictions on filler-gap dependency formation at least partly reflect universal (grammatical or processing) constraints that need not be derived from the input.

Building on existing monolingual processing research, the present study examines whether L2 comprehenders are sensitive to potential PG environments during processing, and whether the factors constraining these already guide the initial gap search. Using a highly time-sensitive experimental technique moreover allows us to capture potentially subtle differences between the time-course of L1 and L2 processing.

Island constraints and parasitic gaps in native sentence processing

As noted above, PGs provide well-known exceptions to island constraints, with their licensing conditions showing considerable cross-linguistic variation (Engdahl, 1983). Kurtzman and Crawford (1991) report the results from a series of speeded acceptability judgement experiments which show that English speakers tended to accept sentences such as (8a) below, which contained a PG inside an infinitival subject island. In contrast, sentences such as (8b), in which a subject island contained a PG inside a finite complement clause, were considered less acceptable.

(8) a. Who did [$_{\text{SUBJECT}}$ your attempt to instruct __] confuse __ ?
b. Who did [$_{\text{SUBJECT}}$ your statement that you instructed __] confuse __ ?

Note, however, that end-of sentence judgements of the kind gathered by Kurtzman and Crawford do not tell us anything about when during processing potential PGs are first postulated.

The real-time processing of parasitic gaps inside subject islands was examined by Phillips (2006). Phillips first reported judgement ratings from young adult native speakers of English which showed that sentences containing both a PG and a licit *wh*-gap, such as (9a,b) below, were rated as significantly more acceptable when the verb within the subject island domain was non-finite

[1] The acceptability of island violations is subject to considerable cross-linguistic and within-language variation (e.g., Szabolcsi, 2006). This variability casts doubt on the long-held view that island effects reflect universal grammatical constraints (e.g., Chomsky, 1973), and also raises non-trivial questions about how they might be represented in individual language grammars (see e.g., Haegeman, Jimenez-Fernandez & Radford, 2014). Alternatively, island effects have been argued to reflect (potentially grammaticalized) processing economy constraints (e.g. Deane, 1991; Hofmeister & Sag, 2010; Kluender, 2004). We will remain essentially agnostic with respect to this issue here, however (see e.g., Hofmeister & Sag, 2010, and Phillips, 2013, for extensive discussion).

[2] The availability of a licit gap is not by itself a sufficient licensing condition for PGs. Culicover (2001) lists a number of other conditions thought to constrain the acceptability of PGs. As these were not examined in the current study, we will not discuss them any further.

https://doi.org/10.1017/S1366728915000942 Published online by Cambridge University Press

(9a) compared to when it was finite (9b) (examples adapted from Phillips, 2006: 805).

(9) a. The outspoken environmentalist worked to investigate **what** [SUBJECT the local campaign to preserve __] had harmed __.
b. The outspoken environmentalist worked to investigate **what** [SUBJECT the local campaign that preserved __] had harmed __.

A small effect of finiteness was also observed for sentences containing a PG unaccompanied by a licit gap, which were however rated as considerably less acceptable overall. The finiteness of the embedded verb made no difference to acceptability, on the other hand, for sentences that contained an ordinary *wh*-gap but no PG. This was taken to show that a PG can legitimately occur inside a subject island domain only when the subject island is infinitival (as in 9a) and if a corresponding licit gap is present elsewhere in the sentence.

Having established a clear trend for there to be a finiteness restriction on parasitic gaps inside subjects, Phillips also administered an on-line self-paced reading task, where participants' word-by-word reading times were recorded. He presented sentences like those in (10) below, with two critical manipulations. Firstly, the filler (e.g., *which schools* or *which high school students*) was manipulated so as to be either a plausible or an implausible object of the verb inside the subject island (e.g., *expand*). Secondly, the clause within the island domain was either non-finite (e.g., *to expand*, as in 10a), or finite (e.g., *that expanded,* as in 10b) (examples adapted from Phillips, 2006: 808).

(10) The school superintendent learned . . .

a. {**which schools** / **which high school students**} [SUBJECT the proposal to expand (__) drastically and innovatively upon the current curriculum] would {overburden / motivate} __ during the following semester.

b. {**which schools** / **which high school students**} [SUBJECT the proposal that expanded (__) drastically and innovatively upon the current curriculum] would {overburden / motivate} __ during the following semester.[3]

Note that the sentences in (10) do not in fact include a PG at all, as *expand* is used intransitively here, followed by the prepositional adjunct or argument *upon the current curriculum*. The intention here was simply to see whether the parser would initially posit a temporary PG upon encountering the verb inside the subject island.[4] As implausible fillers typically elicit elevated reading times (in comparison to plausible ones) at or following potential direct object gaps, the plausibility manipulation served as a diagnostic for whether or not a gap was postulated within the island region.

The results showed that reading times for the verb inside the subject island were significantly faster for the 'plausible' condition compared to the 'implausible' one, but only when the verb was infinitival, yielding a statistical interaction between plausibility and finiteness. It was interpreted as demonstrating that the parser only posited a gap inside non-finite subject islands, which according to Phillips' offline results are indeed able to host PGs.[5]

A potential confound in Phillips' (2006) materials concerns the finiteness manipulation, however. First, there is evidence that non-finite clauses are easier to process than finite ones (e.g., Kluender, 1992). Secondly, there is a difference in length between *to inspire* and *that inspired*, as well as the fact that the past tense form is morphologically more complex. Thirdly, infinitival and finite clauses may differ in terms of the functional architecture associated with them (e.g., Bošković, 1996). Fourthly, the relative clauses (RCs) used in the finite conditions are adjuncts while the infinitival clauses are complements. Finally, the RCs used in the finite conditions may contain a covert *wh*-operator (e.g., Chomsky, 1981) whereas the infinitival clauses do not. In short, there are a variety of potential differences between the infinitival and finite conditions which might all contribute to making finite subject islands more difficult to process and to extract from than non-finite ones.

The results from another self-paced reading experiment reported by Wagers and Phillips (2009) suggest that PGs are not necessarily postulated in all environments that support them. They tested sentences such as (11) below which contained adjunct clauses that allowed for a (temporary) PG to be posited after the embedded verb (e.g., *sipping*).

(11) The wines which the gourmets were energetically discussing __ [ADJUNCT before slowly sipping (__) the samples during the banquet] were rare imports . . .

[3] *Overburden* is used with *schools*, and *motivate* with *high school students*.

[4] In example (10) and elsewhere, temporary gaps which are subsequently disproved are put in parentheses.

[5] This is not to suggest that PGs are never permitted inside finite subject islands. As an anonymous reviewer points out, examples such as the following may actually be deemed acceptable:

(i) Here is the boy who everyone who has met __ thinks __ is clever. (Engdahl, 1983, who attributes this example to Janet Fodor)

(ii) She is the kind of person that everyone who meets __ ends up falling in love with __. (Kayne, 1983)

https://doi.org/10.1017/S1366728915000942 Published online by Cambridge University Press

Unlike for similar sentences involving coordination, which actually required a gap in the second conjunct, there was no evidence in the reading-time data that a PG was posited inside adjunct islands in sentences like (11). Wagers and Phillips conclude that once the grammatical requirement that the filler be thematically interpreted has been met, further optional gaps may not always be computed. This finding raises the question of whether, or how, the possibility of interpreting the filler thematically at an earlier PG, as in the stimulus sentences used in Phillips' (2006) study, might affect filler integration at the licit gap downstream (which was located in a sentence region that Phillips did not analyze).

Even though the self-paced reading paradigm has proven very useful for identifying sentence regions that give rise to processing difficulty, this technique also has some limitations. For one thing, presenting the experimental materials in chunks (usually, word by word) does not allow for particularly natural reading. Moreover, with only a single reading time measure obtained per word or other region of interest, it is not possible to distinguish between effects of initial (or 'first-pass') parsing decisions and later, 'second-pass' or reprocessing effects. The eye-movement monitoring technique used in the present study allows for a more fine-grained record of the processing time-course to be obtained than is possible with self-paced reading. Furthermore, our materials were designed so as to allow for possible effects of PGs on the processing of the terminal gap to be examined as well.

Sensitivity to island constraints in L2 processing

Both native and non-native comprehenders appear to be guided by an active filler strategy during the processing of FGDs. That is, like native speakers, L2 speakers attempt to link a filler to the first potential subcategoriser they come across (e.g., Williams et al., 2001).

Most relevant to the current study are findings showing that L2 speakers also respect extraction islands in processing tasks (Aldwayan, Fiorentino & Gabriele, 2010; Cunnings et al., 2010; Felser et al., 2012; Juffs & Harrington, 1995; Omaki & Schulz, 2011). Juffs and Harrington (1995), for example, report the results from a reading-time-plus-grammaticality-judgement study with Chinese-speaking learners of English and native English-speaking controls. Ungrammatical stimuli included several types of island violation, including subject island violations as in *Who did a story by please the children?* Both the L2 group and the native controls showed sensitivity to island violations, with subject island violations being rejected even more frequently by the Chinese participants (95.5%) than by the native controls (83.3%) in experiment 1. However, as pointed out above, end-of-sentence judgements are not informative about the application of island constraints during real-time.

Also using a self-paced reading task, Aldwayan et al. (2010) found no evidence of L2 speakers postulating gaps inside subject noun phrases in sentences such as *My sister wondered who [the boring comments about John's used car] were intended to entertain*, although their experimental design did not include any closely matched non-island control conditions.

The precise timing of island effects was examined by Felser et al. (2012) in an eye-movement monitoring study. They carried out two online reading experiments to investigate whether, and when during processing, proficient German-speaking learners of English would show sensitivity to relative clause islands as in (12a) below. These were contrasted with sentences like (12b) that contained a licit potential gap after the embedded verb.

(12) a. Everyone liked **the magazine / the shampoo** that the hairdresser [$_{RC}$ who read (__) extensively and with such enormous enthusiasm] bought __ before going to the salon.
b. Everyone liked **the magazine / the shampoo** that the hairdresser [$_{VP}$ read (__) extensively and with such enormous enthusiasm about __] before going to the salon.

Their first experiment manipulated the filler's plausibility as a direct object of the embedded verb (e.g., *read*) as a diagnostic for dependency formation (following Traxler & Pickering, 1996). The L2 group showed evidence for island sensitivity even earlier during processing than did the native controls here, with plausibility effects at the embedded verb region restricted to non-island sentences such as (12b). Similar results have been reported by Omaki and Schulz (2011) for native Spanish-speaking learners of English using self-paced reading, and by Cunnings et al. (2010), who showed that even learners from a *wh*-in-situ background (Chinese) were sensitive to English RC islands during processing. In their second experiment, Felser et al. (2012) used filled gaps as a diagnostic for dependency formation instead, with otherwise similar materials.[6] Here effects of island sensitivity were found to be comparatively delayed in the L2 group.

Together, the above findings show that L2 speakers are sensitive to at least some types of island during processing. Interpreting these findings is not entirely straightforward, however. They would follow naturally from the assumption that island effects reflect general

[6] "Filled gap" effects may occur during sentence processing when a syntactic position that could potentially host a gap is discovered to be already occupied, giving rise to elevated processing times at or following this point (Stowe, 1986).

https://doi.org/10.1017/S1366728915000942 Published online by Cambridge University Press

cognitive or processing constraints (e.g., Hofmeister & Sag, 2010) but are also compatible with the assumption that they reflect universal grammatical constraints (e.g., Chomsky, 1973), or possibly, language-specific restrictions that can be learned. Entering into the debate of whether island effects ultimately reflect grammatical or processing constraints (see e.g., Hofmeister & Sag, 2010; Phillips, 2013) is beyond the aims and scope of the current study. We merely note that, whatever the ultimate origin of island effects may be, for an island region to become recognisable as such, comprehenders must be able to process the relevant island-inducing cues (whether these are syntactic, semantic or pragmatic in nature). Such cues may be easy to notice if they take the shape of another *wh*-filler intervening between the original *wh*-filler and its associated gap (e.g., Felser et al., 2012; Omaki & Schulz, 2011), but less obvious in the case of complex subject islands, for example. Given earlier claims to the effect that L2 comprehenders sometimes have difficulty processing morphosyntactic detail or abstract structural information in real-time (cf. Clahsen & Felser's, 2006, Shallow Structure Hypothesis), it is thus conceivable that they might violate certain island constraints during processing even if these constraints have a universal basis, and despite being aware of their existence.

Showing that non-native speakers are sensitive to constraints on filler-gap dependency formation that are not attested in their L1 and also underdetermined by the L2 input would provide evidence that some kind of universal (grammatical or processing) constraints are involved (compare e.g., Schwarz & Sprouse, 2000). This is the case, for example, for parasitic gaps of the kind under investigation here.

To our knowledge, L2 speakers' real-time sensitivity to PGs inside island domains has never been investigated. PGs have also received relatively little attention in traditional L2 acquisition research. One exception is a grammaticality judgement study reported by Felix (1988), which examined (*inter alia*) the acceptability of PGs inside finite adjunct clauses as in *a person they spoke to because they admired*. L1 German-speaking learners of English judged these as grammatical around 30% of the time, whereas the native controls uniformly rejected them. This finding could be taken to suggest that L2 speakers sometimes posit PGs even in contexts in which they are not normally permitted.

Felix's findings are interesting because the occurrence of PGs is much more restricted in German than in English. According to Parker (1999), they can normally occur only in tenseless adjunct clauses, whereas Kathol (2001) has argued that standard German lacks true PGs altogether. Even though Southern dialectal varieties appear to be more liberal (Felix, 1985), PGs are normally excluded from infinitival subject islands of the kind that license PGs in English. Thus we may assume that native German-speaking L2 learners of English have had very little prior exposure to PGs inside non-finite subject islands, and that any direct L1/L2 mapping of potential PG environments is precluded.[7] From the point of view of language acquisition, their language-specific licensing conditions and comparative rarity render PGs a classical 'poverty-of-stimulus' phenomenon (Felix, 1988; Parker, 1999).

In what follows, we report the results from both an untimed acceptability judgement task and an online reading task which examine and compare L1 and L2 comprehenders' sensitivity to PGs inside English subject islands.

Experiment 1: Offline judgements

A scalar acceptability rating task was administered in order to examine whether our non-native participants were aware of the licensing criteria for PGs inside subject islands that have been reported for native English speakers elsewhere (e.g., Phillips, 2006).

Participants

Twenty-four native speakers of English (12 females, mean age: 23.9 years; SD: 5.71) were recruited from the Universities of Essex (UK) and Potsdam (Germany) and their surrounding communities. Twenty-four native German-speaking L2 learners of English (18 females, mean age: 25.6 years; SD: 5.4) were recruited from the University of Potsdam. The L2 speakers were given the most advanced grammar section of the Oxford Placement Test (Allan, 2004). They achieved a mean score of 77% (SD: 5.01), indicating that they were advanced learners of English. Finally, the participants reported having initially learnt English as part of their formal education in Germany (mean starting age: 10 years, SD: 1.68), with 19 of the 24 participants reporting that they had subsequently spent some time (at least six months) living in an English-speaking country. All participants had normal or corrected-to-normal vision, and no linguistic or general cognitive disorders. They were offered either eight Euros or course credits for their participation.

Materials

The questionnaire experiment had a 3x2 design, manipulating the factors Gap Type and Finiteness. Gap Type refers to whether the PG co-occurred with a licit

[7] Note that German does permit ordinary (i.e. non-parasitic) *wh*-gaps inside nominal subjects under certain conditions (e.g. Haider, 1993). Extraction from non-finite sentential subjects is not normally considered acceptable, however (Jurka, Nakao, & Omaki, 2011).

gap (= 'both gaps' conditions), alone (= 'bad gap' conditions), or whether the licit gap occurred by itself (= 'good gap' conditions). Each of these three conditions came in two versions, one containing a non-finite clause within the subject island and the other a finite one, yielding the six experimental conditions shown in (13).

(13) a. BAD GAP, INFINITIVAL
It was not clear which animals [the plan to look after __] would protect the forest.
b. BAD GAP, FINITE
It was not clear which animals [the plan that looked after __] would protect the forest.
c. GOOD GAP, INFINITIVAL
It was not clear which animals [the plan to look after the forest] would protect __.
d. GOOD GAP, FINITE
It was not clear which animals [the plan that looked after the forest] would protect __.
e. BOTH GAPS, INFINITIVAL
It was not clear which animals [the plan to look after __] would protect __.
f. BOTH GAPS, FINITE
It was not clear which animals [the plan that looked after __] would protect __.

Twenty-four item sets as in (13a–f) were created and distributed across six presentation lists using a Latin Square design and were then pseudo-randomised.

Thus, each participant would see only one version of each sentence but an equal number of sentences per condition, and no item of the same condition occurred directly after another. Forty fillers (half of which were acceptable, half unacceptable) were added to the experimental items, including legitimate and illegitimate PG structures of different types, various types of island violation, and heavy or light NP shift and non-shift in different syntactic contexts.

Procedure

Participants were given a printed booklet each containing a total of 64 sentences, and were asked to rate them on a 10-point scale of acceptability, where ten was the most structurally well-formed and semantically meaningful, and one was the least. No time limit was given for the task, but participants were asked to give their ratings as quickly as possible, and were told not to return and make any changes once a rating had been given. The paper-and-pencil task took around 15–20 minutes to complete.

Predictions

We predicted that participants' ratings should be higher when a PG occurred inside an infinitival subject island compared to when it occurred in a finite one. But finiteness should only affect participants' ratings when a PG is present, and specifically in the critical 'both gaps' conditions. In the 'good gap' conditions, finiteness should not have an effect since no PG is present.

Table 1. *Average acceptability ratings by group (SDs in parentheses)*

Condition	Native speakers	L2 speakers
Bad gap, infinitival	4.6 (1.78)	4.83 (1.96)
Bad gap, finite	3.61 (1.69)	5.16 (2.29)
Good gap, infinitival	5.84 (2.27)	6.39 (2.71)
Good gap, finite	5.33 (2.1)	6.85 (2.36)
Both gaps, infinitival	6.96 (1.59)	7.66 (1.93)
Both gaps, finite	4.53 (1.83)	4.97 (2.41)

Results

Table 1 provides an overview of the L1 and L2 groups' raw average ratings across the six conditions for all participants.

The raw data were analyzed using linear mixed effects (LME) modeling in the IBM SPSS 21 MIXED procedure (e.g., Norusis, 2011). Participant and item random slopes were fitted for each fixed effect and interaction of fixed effects and random intercepts for participants and items were used since this maximal random effects structure yielded a convergent model which also fits with the recommendations of Barr, Levy, Scheepers and Tily (2013). The fixed effect factors entered into the model were Gap Type, with three levels (BAD GAP, GOOD GAP and BOTH GAPS), and Finiteness, with two levels (INFINITIVAL, FINITE). This revealed a marginal main effect of Gap Type (Estimate: 0.91 (SE: 0.48), $t = 1.89$, $p = .06$), and a significant interaction between Gap Type and Finiteness (Estimate: 1.32 (SE: 0.61), $t = 2.16$, $p < .05$). The factor Group (L1 vs. L2) was added but failed to yield either a main effect or any interactions, and there was also no main effect of Finiteness (ps > .05). To explore the interaction between Gap Type and Finiteness further, planned comparisons using t-tests were run comparing each of the pairs for each gap type. The difference between the infinitival condition and the finite one was significant for the 'both gaps' pair (t_1 (47) = 3.43, $p < .01$, t_2 (23) = 2.94, $p < .01$), but was not significant for either the 'good gap' or 'bad gap' pairs (ps > .05).

Discussion

Finiteness affected the 'both gaps' condition only here, with the infinitival version being rated much higher

https://doi.org/10.1017/S1366728915000942 Published online by Cambridge University Press

than the finite one. Our results confirm that the criteria for a PG being hosted inside a subject island include that (i) the subject clause is infinitival, and (ii) there is an accompanying licit gap downstream. Furthermore, it would appear that both our L1 and L2 speakers are aware of these criteria.

In Standard German, the translation equivalents of our 'bad gap' conditions involving extraction from subject islands (13a,b) (=14a,b) are also unacceptable, whereas equivalents of the 'good gap' conditions (14c,d,) are acceptable. The two languages differ with regard to the availability of PGs within subject islands, however. Both the equivalent of the licit PG structure in (13e) (=14e) and its finite counterpart in (14f) are unacceptable in German:

(14) a. **Es war nicht klar, welche Tiere [der Plan __ zu behüten] den Wald schützen würde.*
it was not clear which animals the plan to look.after the forest protect would (cf. 13a)

b. **Es war nicht klar, welche Tiere [der Plan, der __ behütete] den Wald schützen würde.*
it was not clear which animals the plan that looked.after the forest protect would (cf. 13b)

c. *Es war nicht klar, welche Tiere [der Plan den Wald zu behüten] __ schützen würde.*
it was not clear which animals the plan the forest to look.after protect would (cf. 13c)

d. *Es war nicht klar, welche Tiere [der Plan, der den Wald behütete] __ schützen würde.*
it was not clear which animals the plan that the forest looked.after protect would (cf. 13d)

e. **Es war nicht klar, welche Tiere [der Plan __ zu behüten] __ schützen würde.*
it was not clear which animals the plan to look.after protect would (cf. 13e)

f. **Es war nicht klar, welche Tiere [der Plan, der __ behütete,] __ schützen würde.*
it was not clear which animals the plan that looked.after protect would (cf. 13f)

Our results thus suggest that non-native PG licensing conditions are acquirable, and overall show a similar pattern to that reported by Phillips (2006). However, one difference between our and Phillips' results is that our 'both gaps, infinitival' condition was rated much better than our 'good gap' conditions, whereas in Phillips' study the ratings were about the same. We might speculate that the presence of a legitimate PG in 'both gaps, infinitival' can boost the activation of the filler, thus facilitating its integration at the tail of the dependency, a possibility we will return to in our general discussion.

Experiment 2: Eye-movement monitoring

With participants' offline sensitivity to possible and impossible PG environments established, we now turn to the real-time processing of sentences containing subject islands. Recording participants' eye movements during reading provides detailed information about moment-by-moment processing while allowing for fairly natural reading (e.g., Staub & Rayner, 2007). Experiment 2 sought to investigate the following questions:

- Do the subject island constraint and, in particular, the finiteness restriction on PGs restrict the initial gap search in both L1 and L2 comprehension?
- How does postulating a PG affect filler integration at the licit gap site further downstream?

The same participants that completed Experiment 1 also took part in Experiment 2.[8]

Materials

The experiment had a 2x2 design, manipulating the plausibility of the relationship between the fronted *wh*-phrase and the embedded verb on the one hand, and the finiteness of the clause within a subject island on the other, as illustrated in (15) below. All critical sentences were preceded by a lead-in sentence that served to situate them within an appropriate discourse context.

(15) Every effort is being made to get people who have committed crimes back to a decent living.

a. PLAUSIBLE, INFINITIVAL
The policeman knew **which prisoners** [the activities **to inspire** (__) massively and cleverly the sensible criminals who want a life inside of the law] would help __ because the training programmes available included a gardening club.

b. PLAUSIBLE, FINITE
The policeman knew **which prisoners** [the activities **that inspired** (__) massively and cleverly the sensible criminals who want a

[8] Experiment 2 was in fact administered before Experiment 1, to prevent participants from developing expectations about the experimental stimuli used in the online reading task.

https://doi.org/10.1017/S1366728915000942 Published online by Cambridge University Press

life inside of the law] would help __ because the training programmes available included a gardening club.

c. IMPLAUSIBLE, INFINITIVAL
The policeman knew **which houseplants** [the activities **to inspire** (__) massively and cleverly the sensible criminals who want a life inside of the law] would help __ because the training programmes available included a gardening club.

d. IMPLAUSIBLE, FINITE
The policeman knew **which houseplants** [the activities **that inspired** (__) massively and cleverly the sensible criminals who want a life inside of the law] would help __ because the training programmes available included a gardening club.

All experimental sentences contained a *wh*-complement clause (with the fronted *wh*-phrase in (15) indicated in bold) whose subject was complex in that it contained an embedded clause of its own. This clause was either infinitival (e.g., *to inspire*) or finite (e.g., *that inspired*), and the *wh*-phrase was either a pragmatically plausible or implausible direct object of the verb in the subject clause (e.g., *to inspire prisoners* vs. *to inspire houseplants*).

Note that all potential object gaps inside the subject clause subsequently proved to be filled by an overt direct object (e.g., *the sensible criminals who want a life inside of the law*), which means that a PG could only be sustained temporarily here. The direct object was deliberately made ‘heavy’ so as to be able to undergo Heavy NP Shift, which then allowed us to insert additional padding material between the verb and its object. A three-word adjunct phrase (e.g., *massively and cleverly*) was inserted to help ensure that the parser would not be able to disconfirm a PG immediately after processing the verb. Since all experimental sentences were actually ‘good gap only’ sentences, they were all globally plausible. For the sentence quadruplet in (15), for instance, both prisoners and houseplants are plausible objects of the verb *help*, which marks the location of the licit *wh*-gap.

Recall that, as in Phillips' (2006) original study, our finiteness manipulation might result in sentence pairs that differ in their structural and/or processing complexity beyond a mere [±tense] difference. In order to control for this potential confound, we added two control conditions to our stimulus materials as in (16a,b) below, which were gapless variants of the experimental sentences and contained either an infinitival or a finite subject clause. This should allow us to measure any possible differences in processing difficulty between our critical infinitival (15a,c) and finite (15b,d) conditions that are unrelated to the availability of PGs.

(16) a. The policeman knew that [the activities **to inspire** massively and cleverly the sensible criminals who want a life inside of the law] would help particular prisoners because the training programmes available included a gardening club.

b. The policeman knew that [the activities **that inspired** massively and cleverly the sensible criminals who want a life inside of the law] would help particular prisoners because the training programmes available included a gardening club.

Twenty-four sentence sets were created (each comprising the four experimental and two control conditions) and distributed across six presentation lists using a Latin Square design.[9] A full list of experimental items is available as Supplementary Materials. The stimulus items were pseudo-randomised so that no two items from the same condition occurred directly next to each other. These were then mixed with 50 filler items, which included 15 sentences with Heavy NP Shift, 15 containing PGs of different types, plus 20 further items which varied in their syntactic structure, length and complexity. The fillers were designed so as to prevent participants from developing any strategic reading patters when encountering our experimental stimuli, and the majority of filler items were of similar structural complexity (and also often stylistically marked) as the critical stimuli. All critical items and half of the fillers were followed by *yes/no* comprehension questions targeting different parts of the preceding sentence, to monitor whether participants paid attention to the task and to provide motivation for reading the stimuli carefully for meaning.

Predictions

Given the results from Experiment 1, which suggested that PGs are possible inside infinitival but not inside finite subject islands, we expected effects of our plausibility manipulation to be restricted to the infinitival conditions, yielding an interaction between the factors Finiteness and Plausibility. That is, reading times at or immediately

[9] Adding two control conditions required us to spread the experimental items across six rather than four presentation lists. As a result each participant was exposed to four items per condition only, which is a lower number of items than normal in studies of this kind. We did not want to include more than 24 experimental items in each list as the structure of our stimulus sentences was highly complex and marked, so that we would have needed more fillers of similar complexity, and ultimately the experiment would have simply been too long and exhausting. We felt that the benefits of controlling for some confounding factors by adding the control conditions outweighed the potential costs of reduced by-item power.

https://doi.org/10.1017/S1366728915000942 Published online by Cambridge University Press

following the embedded verb (e.g., *inspire*) should be longer if the filler is implausible as a direct object of this verb compared to when it is plausible for infinitival subject islands (15a,c) only.

If both the subject island constraint and the finiteness restriction on PGs constrain the initial gap search, then this interaction should already be visible during participants' first-pass reading times. If, on the other hand, the island constraint only serves to filter out illicit gaps at a later processing stage, as has been suggested e.g., by Clifton and Frazier (1989), then PGs might initially be postulated in both infinitival and finite environments, giving rise to an early main effect of Plausibility not modulated by Finiteness.

A further empirical question that our study seeks to address is how the availability of a PG might affect participants' ability to integrate the filler at the ultimate, licit gap (i.e., at the verb *help* in (15)). One possibility is that positing an intermediate gap, even if only temporarily, facilitates filler integration later on (e.g., Gibson & Warren, 2004; Marinis et al., 2005). This should be reflected in faster reading times for the infinitival compared to the finite conditions at the licit gap region. Another possibility is that the attempt to interpret the filler thematically within the subject phrase interferes with the processing of the licit gap, which itself is also in a thematic position. This should give rise to a reading-time disadvantage for the infinitival conditions instead, possibly modulated by the filler's semantic fit at the earlier PG.

Procedures

Participants were tested in a controlled laboratory setting. The experiment began with three practice items to make certain participants had understood the task. All text was presented on a computer screen in black Courier New font on a white background. Participants' eye movements were recorded using a desk-mounted EyeLink 1000 setup, with a sampling rate of 1000 Hz, while participants' heads were kept in a fixed position using a chin rest. Although participants read binocularly, only their dominant eye was tracked. The EyeLink was calibrated to each participant's eye using a standardised nine-point tracking test, whereby participants had to fixate dots around the screen in random succession. The calibration was automatically re-checked before each trial. Participants were asked to read each item silently and at their regular reading speed before pressing a button to indicate they were done. When comprehension questions were asked, participants had to respond either 'yes' or 'no' by pressing the appropriate buttons on a marked game pad.

The native English speakers completed the task in around 35–40 minutes, together with Experiment 1. The L2 participants additionally completed Part 2 of the grammar section of the Oxford Placement Test as well as a printed vocabulary checklist which included the critical nouns and verbs used in the experimental sentences. Overall, an experimental session lasted approximately one hour for native speakers, and one hour and 15 minutes for the non-native speakers.

Data analysis

We selected five regions from the critical sentences for analysis: (i) the pre-critical region (the noun immediately preceding the critical region, e.g., *activities*), (ii) the subject-internal verb (e.g., *inspire*/*inspired*), (iii) the post-verbal adjunct immediately following it (e.g., *massively and cleverly*), (iv) the post-critical region (the head noun of the direct object, e.g., *criminals*), and (v) the verb at which the FGD is terminated (e.g., *help*). Four eye-gaze measures will be reported. Firstly, we will report first-pass reading time, which is the sum of the initial fixations within a region once it has been entered for the first time until it is exited to the left or right. The second eye-movement measure we are reporting here is regression path reading time. This is the sum of all fixations (first-pass and rereading) until the region is initially exited to the right. The third measure is rereading time. This is the sum of all fixations within a region after it was initially exited to the left or right following the first-pass. Since rereading occurs after participants' initial reading of a given sentence region, this measure is commonly thought to reflect a later (or 'second-pass') processing stage, while first-pass times reflect earlier (or 'first-pass') processing. Finally, we also report the total overall reading times for each region of interest.

Two native English speakers scored below 60% on the comprehension questions and were consequently excluded from further analysis. All remaining participants scored at least 65% (L2 speakers: 67%), which indicated that they paid attention to the task and actively tried to read the stimulus items for meaning. The data from one native and one non-native speaker were removed because their average overall reading time at the critical verb was below 100ms. 2.73% of the total data set was also lost because regions of interest were skipped altogether. Fixations of 80ms or shorter and within one degree of another fixation were automatically merged with neighbouring fixations. Fixations of 80ms or shorter which could not be merged were removed, as were any fixations above 800 ms.

Eighteen individual trials were removed as a result of eye-tracking drift, and a further 11 trials were removed because participants had marked some critical vocabulary items contained within them as unfamiliar on our vocabulary checklist. Outlier data points in excess of 2.5 SDs from a participant's mean and/or the item's mean for each measure for each region were also excluded. This led to a total data removal of 5.87% of the overall raw data set.

The remaining reading-time data were again analyzed using linear mixed effects (LME) modeling with participant and item random slopes and intercepts. Separate analyses for each language group, containing the two main effects and their interaction, were carried out only when the preliminary omnibus analysis yielded a reliable three-way interaction.

Results

Table 2 shows the average first-pass and rereading times for each condition at each region of analysis by group. Below we report the statistical results for each region of interest in turn. A full summary of the model outputs is provided in Table 3.

Pre-critical region (e.g., activities)

No statistical effects or interactions amongst the experimental conditions were found for first pass reading times. The L2 group generally showed longer regression path, rereading, and total reading times than the native group, which was reflected in significant main effects of Group.

Subject-internal verb (e.g., inspire/inspired)

Even though reading times tended to be longer for finite than for non-finite forms, with the shortest reading times seen in condition (15a) (= plausible, infinitival), there were no statistically significant main effects or interactions (all ps $> .05$) in first pass reading times. There were only main effects of Group for regression path, rereading and total reading times.

Post-verbal adjunct (e.g., massively and cleverly)

Similar numerical patterns were seen at the post-verbal adjunct (e.g., *massively and cleverly*) as for the verb preceding it. In first-pass reading times the native speakers read the plausible condition faster than the implausible one for the infinitival pair, but not for the finite one. The L2 group, meanwhile, read the plausible condition faster than the implausible one for both the infinitival and finite pairs. A preliminary omnibus analysis revealed main effects of Finiteness, Plausibility and Group, as well as an interaction between Plausibility and Finiteness. Most interestingly, there was also a significant three-way interaction between Finiteness, Plausibility and Group, suggesting that the two participant groups showed statistically different reading-time patterns across the four experimental conditions. To explore this interaction further, the two groups' first-pass reading times for this region were analyzed separately (compare Table 4).

For the native speakers, this yielded a main effect of Finiteness and a marginal main effect of Plausibility, as well as an interaction between the two. To explore the source of this interaction, we conducted pairwise planned comparisons using t-tests, which revealed that it was driven by the implausible condition being read slower than its plausible counterpart only for the infinitival pair ($t_1(20) = 2.79$, $p < .05$, $t_2(23) = 2.64$, $p < .05$), whilst there was no significant difference in reading times for the finite pair ($t_1(20) = 1.69$, $p > .05$, $t_2(23) = 1.38$, $p > .05$). The L2 speakers, in contrast, only showed the two main effects but no interaction, reflecting the fact that the plausible conditions were read faster than the implausible ones, and the infinitival conditions faster than the finite ones.

Preliminary analysis of the regression path measurements of this region also showed main effects of Finiteness, Plausibility and Group, an interaction between Finiteness and Plausibility and a three way interaction between Finiteness, Plausibility and Group. Once again we therefore did a group-wise analysis, reported in Table 5.

This time we found interactions between Plausibility and Finiteness in both the L1 and L2 groups. We carried out some planned comparisons to investigate the source of the interaction. Both groups read the plausible condition significantly faster in the infinitival pair (L1: $t_1(20) = 2.47$, $p = .02$; $t_2(23) = 2.22$, $p < .05$, $d = .97$; L2: $t_1(22) = 3.01$, $p < .01$; $t_2(23) = 2.78$, $p = .01$, $d = 1.21$), whilst there were no significant differences in the finite pair (L1: $t_1(20) = 1.54$, $p > .05$; $t_2(23) = 1.28$, $p > .05$, $d = .04$; L2: $t_1(22) = 1.63$, $p > .05$; $t_2(23) = 1.71$, $p > .05$, $d = .14$). What, then, is the source of the interaction? Whilst the numerical (and statistical) trends for the two groups are the same, notice the effects are bigger both numerically and statistically for the L2 group. A measure of effect size, Cohen's d, is given above for each of the planned comparisons we carried out. These confirm that the effect of Plausibility in the infinitival pair is indeed bigger for the non-natives ($d = 1.21$, classified as 'a *very* large effect') than for the natives ($d = .98$, 'a large effect').

An omnibus analysis of the rereading times for this region revealed main effects of Finiteness, Plausibility and Group. In addition, we found a significant two-way interaction between Finiteness and Plausibility that was not modulated by Group. Planned comparisons confirmed that this was driven by the plausible condition being read faster than the implausible one in the non-finite pair ($t_1(43) = 2.78$, $p < .01$, $t_2(23) = 2.84$, $p < .01$) but not their finite counterparts ($t_1(43) = 1.21$, $p > .05$, $t_2(23) = 1.19$, $p > .05$).

Total reading times again showed main effects of Finiteness, Plausibility and Group, and an overall interaction between Finiteness and Plausibility not modulated by Group. Planned comparisons confirmed the plausible condition was read faster than the implausible one in the non-finite pair ($t_1(43) = 6.72$, $p < .0001$, $t_2(23) = 6.31$, $p < .0001$) but not the finite one ($t_1(43) = 1.43$, $p > .05$, $t_2(23) = 1.33$, $p > .05$).

Table 2. *Native and non-native speakers' reading times at five regions of analysis in milliseconds (SDs in parentheses)*

		Native speakers				Non-native speakers			
		First-pass reading time	Regression path	Rereading time	Total reading time	First-pass reading time	Regression path	Rereading time	Total reading time
Pre-critical region	Infinitival, plausible	247 (125)	394 (346)	321 (253)	568 (306)	232 (160)	424 (387)	426 (424)	658 (483)
	Infinitival, implaus.	213 (97)	410 (319)	327 (150)	540 (275)	252 (167)	451 (324)	408 (489)	660 (436)
	Finite, plausible	223 (212)	387 (328)	362 (292)	585 (228)	250 (207)	442 (355)	407 (433)	657 (488)
	Finite, implaus.	236 (198)	401 (354)	327 (315)	563 (341)	216 (226)	429 (326)	433 (328)	649 (421)
Subject verb	Infinitival, plausible	197 (104)	246 (152)	122 (110)	319 (356)	207 (122)	291 (143)	194 (101)	401 (275)
	Infinitival, implaus.	212 (143)	253 (118)	132 (126)	344 (285)	221 (102)	308 (207)	210 (153)	431 (383)
	Finite, plausible	219 (113)	260 (144)	137 (94)	356 (275)	235 (120)	326 (174)	215 (138)	450 (406)
	Finite, implausible	222 (143)	269 (135)	134 (85)	356 (291)	245 (91)	319 (188)	205 (126)	450 (395)
Post-verbal adjunct	Infinitival, plausible	111 (77)	241 (98)	357 (163)	468 (322)	265 (131)	424 (181)	317 (126)	582 (376)
	Infinitival, implaus.	214 (102)	358 (144)	455 (236)	669 (371)	340 (144)	657 (213)	615 (156)	955 (505)
	Finite, plausible	263 (124)	371 (169)	578 (335)	841 (311)	396 (98)	683 (219)	683 (210)	1079 (398)
	Finite, implausible	246 (134)	365 (173)	560 (298)	806 (472)	458 (121)	651 (236)	628 (204)	1086 (412)
Post-critical region	Infinitival, plausible	254 (103)	318 (328)	263 (210)	517 (403)	285 (149)	361 (186)	262 (568)	547 (319)
	Infinitival, implaus.	265 (137)	341 (258)	238 (198)	503 (326)	299 (161)	378 (144)	270 (103)	569 (376)
	Finite, plausible	247 (215)	332 (276)	245 (201)	492 (351)	280 (143)	367 (138)	247 (133)	527 (412)
	Finite, implausible	222 (198)	324 (243)	257 (219)	479 (337)	279 (156)	359 (175)	255 (247)	534 (392)
Tail of dependency	Infinitival, plausible	228 (82)	240 (212)	83 (148)	311 (217)	267 (153)	299 (183)	163 (301)	430 (442)
	Infinitival, implaus.	329 (157)	438 (269)	226 (229)	555 (266)	360 (312)	484 (221)	394 (341)	754 (438)
	Finite, plausible	257 (105)	221 (178)	122 (169)	379 (294)	297 (242)	327 (242)	196 (329)	493 (388)
	Finite, implausible	241 (103)	233 (194)	115 (167)	356 (325)	273 (170)	345 (261)	177 (245)	450 (410)

Table 3. *Summary of statistical analyses for four eye-movement measures at five regions of text in Experiment 2, (*) marks p < .06, * marks p < .05, ** marks p < .01, *** marks p < .001.* F = Finiteness, P = Plausibility, G = Group

		First pass reading time			Regression path			Rereading time			Total reading time		
		Est.	*SE*	*t*	*Est.*	*SE*	*t*	*Est.*	*SE*	*t*	*Est.*	*SE*	*t*
Pre-critical region	F	14	26	0.54	20	34	0.58	38	44	0.86	28	42	0.66
	P	28	35	0.80	44	51	0.86	33	48	0.69	56	43	1.30
	G	26	69	0.37	154	75	2.05*	276	109	2.53*	295	125	2.36*
	F*P	1	11	0.09	24	44	0.55	47	66	0.71	45	49	0.91
	F*G	41	53	0.77	58	64	0.91	20	47	0.43	21	55	0.38
	P*G	1	23	0.04	23	57	0.40	25	41	0.61	31	43	0.72
	F*P*G	0	27	0.01	22	69	0.32	33	46	0.72	27	41	0.66
Subject verb	F	84	56	1.51	76	54	1.41	33	54	0.61	117	64	1.82
	P	42	52	0.80	26	42	0.62	14	33	0.42	55	63	0.87
	G	58	47	1.23	216	94	2.29*	299	128	2.34*	357	125	2.86**
	F*P	89	50	1.78	83	58	1.43	86	62	1.39	42	55	0.76
	F*G	34	29	1.17	49	61	0.80	44	51	0.86	32	42	0.76
	P*G	23	34	0.67	19	44	0.43	10	22	0.45	29	36	0.80
	F*P*G	39	29	1.34	77	63	1.22	15	29	0.52	93	53	1.75
Post-verbal adjunct	F	433	184	2.35*	390	146	2.67*	703	222	3.16**	998	204	4.89***
	P	223	100	2.23*	312	122	2.55*	323	138	2.34*	546	142	3.84**
	G	626	215	2.91**	896	344	2.60*	292	126	2.32*	918	255	3.60**
	F*P	136	61	2.23*	164	75	2.18*	199	92	2.16*	518	144	3.59**
	F*G	74	61	1.21	61	84	0.72	57	44	1.30	96	58	1.65
	P*G	55	46	1.20	27	34	0.79	14	26	0.54	64	69	0.92
	F*P*G	131	63	2.09*	124	59	2.10*	19	26	0.73	128	74	1.73
Post-critical region	F	48	44	1.09	16	42	0.38	44	51	0.86	104	55	1.89
	P	20	27	0.74	24	54	0.44	64	43	1.48	12	31	0.38
	G	155	71	2.18*	150	76	1.97(*)	43	52	0.83	186	67	2.77*
	F*P	52	48	1.08	74	64	1.16	79	51	1.55	101	78	1.29
	F*G	0	12	0.01	33	63	0.52	66	48	1.38	71	59	1.20
	P*G	34	51	0.66	38	58	0.65	48	54	0.88	67	84	0.79
	F*P*G	14	19	0.74	28	54	0.52	34	42	0.81	64	51	1.25
Tail of dependency	F	117	56	2.09*	335	147	2.28*	256	108	2.37*	372	121	3.07**
	P	−154	73	−2.11*	−413	152	−2.72*	−349	132	−2.64*	−502	158	−3.17**
	G	141	86	2.10*	323	149	2.17*	384	167	2.30*	526	251	2.09*
	F*P	−255	109	−2.34*	−347	133	−2.61*	−270	111	−2.43*	−549	129	−4.25***
	F*G	0	24	0.00	18	84	0.22	22	55	0.40	58	43	1.34
	P*G	−16	49	−0.33	−49	77	−0.64	−32	47	−0.68	−49	56	−0.87
	F*P*G	47	52	0.90	42	43	0.97	50	41	1.22	53	44	1.20

Post-critical region (e.g., criminals)

No statistically significant effect or interactions were found, save for main effects of Group in first-pass, regression path, and total reading times.

Tail of the dependency (e.g., help)

At the region where the filler should finally be associated with the licit gap, both participant groups showed slower first-pass and rereading times for the implausible condition compared to the plausible one for the infinitival pair, while there was no real difference between the finite pair. For first-pass reading times we found main effects of Finiteness, Plausibility and Group, as well as a significant two-way interaction between Plausibility and Finiteness not modulated by Group. The same pattern was found for participants' regression path, rereading and total

https://doi.org/10.1017/S1366728915000942 Published online by Cambridge University Press

Table 4. *LME summary for first-pass reading time, group-wise analysis, at the post-verbal adjunct region (where (*) marks p < .06, and * marks p < .05).*

	Est.	SE	t-value
Native speakers			
Main effect of Finiteness	184	73	2.52*
Main effect of Plausibility	96	49	1.96(*)
Finiteness*Plausibility	133	63	2.09*
Non-native speakers			
Main effect of Finiteness	249	112	2.22*
Main effect of Plausibility	136	64	2.1*
Finiteness*Plausibility	49	48	1.02

Table 5. *LME summary for regression-path reading time, group-wise analysis, at the post-verbal adjunct region (where * marks p < .05, ** marks p < .01).*

	Est.	SE	t-value
Native speakers			
Main effect of Finiteness	137	55	2.49*
Main effect of Plausibility	111	43	2.58*
Finiteness*Plausibility	117	52	2.25*
Non-native speakers			
Main effect of Finiteness	253	74	3.41**
Main effect of Plausibility	201	78	2.57*
Finiteness*Plausibility	184	63	2.92**

reading times. Planned comparisons again showed that the difference between the infinitival pair was significant for first-pass ($t_1(43) = 2.55$, $p < .05$, $t_2(23) = 2.59$, $p < .05$), regression path ($t_1(43) = 3.51$, $p < .001$, $t_2(23) = 3.43$, $p < .01$), rereading ($t_1(43) = 2.71$, $p < .01$, $t_2(23) = 2.69$, $p = .01$) and total reading times ($t_1(43) = 4.93$, $p < .0001$, $t_2(23) = 4.58$, $p < .0001$), while there was no statistical difference in first-pass ($t_1(43) = 1.69$, $p > .05$, $t_2(23) = 1.54$, $p > .05$), regression path ($t_1(43) = 1.21$, $p > .05$, $t_2(23) = 1.19$, $p > .05$), rereading ($t_1(43) = 1.69$, $p > .05$, $t_2(23) = 1.54$, $p > 0.05$) or total reading times ($t_1(43) = 1.75$, $p > .05$, $t_2(23) = 1.71$, $p > .05$) between the finite pair.

Analysis of the control conditions

In order for us to examine potential effects of our finiteness manipulation independently from the presence of syntactic gaps, our stimulus materials included two further sentence types (16a,b) which did not contain any *wh*-gaps. Table 6 provides an overview of participants' reading times for these two conditions, and Table 7 summarises the statistical results.

We found no significant effects or interactions, other than main effects of Group reflecting the non-native speakers' generally slower reading times, for this sentence pair, however.

Summary

The analysis of the eye-movement data revealed that the L1 group showed the predicted Finiteness by Plausibility interaction during their initial reading of the adjunct (e.g., *massively and cleverly*) immediately following the critical verb. This interaction was found to be slightly delayed in the L2 group, who initially showed main effects of Finiteness and Plausibility only. However, both groups patterned together during their regression path, rereading and total reading time measures for this region, as well as during their reading of the verb at the tail of the dependency, where we also found significant Finiteness by Plausibility interactions that were not modulated by the factor Group. Our comparison of the two gapless control conditions yielded no significant main effect of Finiteness and no Finiteness x Group interactions.

Discussion

Even though our non-native speakers demonstrated native-like awareness of subject islands and the finiteness restriction on PGs in an offline judgement task (Experiment 1), we observed some subtle L1/L2 processing differences in the online reading task (Experiment 2). At the adjunct phrase immediately following the critical verb inside the island domain (e.g., *massively and cleverly*), only the L1 speakers showed selective sensitivity to acceptable PG environments in their first-pass reading times. This was reflected in a statistical interaction between the two factors we manipulated, Plausibility and Finiteness, with plausible fillers eliciting shorter reading times at potential PG sites than implausible ones in non-finite environments only.

This finding replicates and extends earlier findings by Phillips (2006) and suggests that during L1 processing, no gaps are postulated within finite subject islands. Our L2 participant group differed from the native controls in that they fleetingly postulated direct object gaps in both infinitival and finite island environments, thus indicating that island constraints do not necessarily prevent the processor from positing gaps within an island region. Another interesting finding is our observation that the attempt to create an implausible FGD at potential PG sites (e.g., *to inspire houseplants*) not only slowed down processing at this point, but also impeded filler integration at the terminal gap further downstream. This was the case for both the L1 and the L2 participants. These findings, and their possible implications, will be discussed in turn.

L2 speakers' knowledge of PGs inside subject islands

The results from Experiment 1 show that our L2 group patterned with the native speaker controls in selectively

https://doi.org/10.1017/S1366728915000942 Published online by Cambridge University Press

Table 6. *Native and non-native speakers' reading times at five regions of analysis in milliseconds (SDs in parentheses) for control conditions*

		Native speakers				Non-native speakers			
		First-pass reading time	Regression path	Rereading time	Total reading time	First-pass reading time	Regression path	Rereading time	Total reading time
Pre-critical region	Infinitival	211 (109)	423 (299)	363 (200)	574 (215)	241 (219)	454 (301)	441 (409)	682 (494)
	Finite	226 (129)	419 (326)	352 (217)	578 (204)	266 (194)	410 (315)	387 (384)	653 (451)
Subject verb	Infinitival	231 (136)	237 (121)	151 (127)	382 (358)	256 (162)	332 (159)	231 (112)	487 (305)
	Finite	256 (172)	218 (146)	143 (163)	399 (348)	241 (144)	311 (174)	214 (127)	455 (342)
Post-verbal adjunct	Infinitival	234 (142)	314 (168)	395 (173)	629 (392)	241 (192)	523 (327)	621 (201)	862 (352)
	Finite	252 (148)	364 (145)	404 (201)	656 (383)	267 (158)	582 (307)	611 (143)	878 (374)
Post-critical region	Infinitival	293 (142)	321 (339)	304 (257)	597 (371)	286 (222)	353 (182)	284 (367)	570 (292)
	Finite	274 (163)	372 (302)	335 (279)	609 (399)	291 (183)	373 (175)	312 (354)	603 (400)
Tail of dependency	Infinitival	213 (137)	259 (252)	244 (211)	457 (274)	263 (214)	397 (226)	283 (372)	546 (318)
	Finite	255 (241)	269 (218)	202 (234)	457 (328)	225 (273)	421 (183)	332 (364)	557 (364)

Table 7. *Summary of statistical analyses for four eye-movement measures at five regions of text in Experiment 2 (control conditions), (*) marks p < .06, * marks p < .05, ** marks p < .01* F = Finiteness, G = Group

		First pass reading time			Regression path			Rereading time			Total reading time		
		Est.	*SE*	*t*	*Est.*	*SE*	*t*	*Est.*	*SE*	*t*	*Est.*	*SE*	*t*
Pre-critical region	F	43	52	0.83	48	39	1.23	67	46	1.44	25	53	0.47
	G	70	51	1.37	21	34	0.62	113	87	1.30	183	118	1.55
	F*G	33	62	0.53	46	44	1.05	28	31	0.90	54	65	0.83
Subject verb	F	10	44	0.22	40	62	0.65	−25	42	−0.60	17	49	0.35
	G	12	23	0.52	183	81	2.26*	151	99	1.53	160	104	1.54
	F*G	19	54	0.35	59	74	0.80	76	64	1.19	28	56	0.50
Post-verbal adjunct	F	45	72	0.63	40	59	0.68	−1	39	−0.03	43	54	0.80
	G	22	37	0.59	427	112	3.81**	433	121	3.58**	455	185	2.45*
	F*G	12	42	0.29	82	100	0.82	28	46	0.61	83	74	1.12
Post-critical region	F	14	38	0.37	72	53	1.36	59	38	1.55	45	69	0.65
	G	10	17	0.59	33	64	0.52	−43	46	−0.93	−33	34	−0.97
	F*G	7	24	0.29	23	46	0.50	26	20	1.30	18	28	0.64
Tail of dependency	F	4	30	0.13	35	61	0.57	19	24	0.79	11	22	0.50
	G	20	49	0.41	284	96	2.95**	163	84	1.94(*)	189	110	1.72
	F*G	13	25	0.52	52	73	0.71	39	43	0.91	61	58	1.05

accepting PGs in non-finite subject island environments. Numerically, the L2 group judged the 'both gaps, infinitival' condition even more highly than did the L1 group. Note that, even though German permits extraction from nominal subjects under certain conditions (see note 7), there is no equivalent in Standard German of PGs of the kind that were examined in the current study. The fact that our L2 group showed sensitivity to the finiteness restriction on PGs in the offline task is thus unlikely to be due to any direct positive transfer from their L1.

As we noted earlier, L2 speakers' sensitivity to island constraints is compatible, in principle, both with grammar-based and processing-based approaches to islands, and we do not claim that our results can differentiate between these. Our finding that L1 German-speaking learners of English show native-like sensitivity to PGs is surprising, however, from the point of view of emergentist or exposure-based approaches to L2 acquisition and processing, given that PGs inside subject islands are extremely rare in the input and are also not normally explicitly taught. Instead, our finding that L2 learners who lack PGs inside subject islands in their native language can nevertheless acquire native-like knowledge of restrictions on their acceptability points towards possible universal constraints being involved here that do not need to be learnt.

Processing PGs in real time

The native English speakers' selective sensitivity to our plausibility manipulation during the processing of infinitival subject islands suggests that in L1 processing, the initial gap search is restricted to domains where PGs are considered acceptable. Even though we already saw a similar numerical pattern at the embedded verb itself, the predicted Finiteness by Plausibility interaction only reached statistical significance at the post-verbal adjunct. This is unsurprising given that the function of the post-verbal adjunct was to increase the distance between the embedded verb (e.g., *inspire*) and its direct object (e.g., *the sensible criminals*) so as to lengthen the time for which a temporary PG could be sustained.

The L2 speakers also showed the predicted interaction, but only for regression path, rereading and total viewing times, which are thought to reflect somewhat later points in the processing time-course than the first-pass measure, for which they showed a different pattern from the L1 group. During their initial reading of the adjunct region, our L2 speakers appeared to be sensitive to the different clause types, processing infinitival subject islands faster than finite ones, whilst also showing sensitivity to the plausibility of the filler-verb relationship irrespective of clause type. That is, the L2 group briefly seemed to violate the subject island constraint by also postulating gaps in environments in which they are normally deemed unacceptable.

The L2 participants' overly liberal positing of PGs can hardly reflect a lack of awareness of the finiteness restriction on PGs, as shown by their native-like performance in Experiment 1, and given that their sensitivity to this restriction was not very much delayed. As our L2 group tended to read the stimulus sentences more slowly than the native controls (as demonstrated by the main effects of Group that were present in several eye-movement measures and sentence regions), it is conceivable that the observed L1/L2 difference reflects a difference in processing speed. That is, the L2 groups' sensitivity to illicit PG environments might have been delayed because they processed the information that induces islandhood here (notably, finiteness and/or the presence of a covert *wh*-operator) more slowly than the native controls. An explanation in terms of reduced sensitivity to islandhood cues associated with our finiteness manipulation during early processing stages would be consistent with the Shallow Structure Hypothesis.

Alternatively, the L2 processing pattern observed in the current study is consistent with L2 comprehenders employing a strong version of the AFS according to which gaps are initially postulated wherever they are locally licensed, and illicit gaps filtered out subsequently (e.g., Clifton & Frazier, 1989; Freedman & Forster, 1985). Our L2 group's reading times began to converge with those of the L1 controls only a little later though, demonstrating sensitivity to the finiteness restriction on PGs during real-time processing. This is in line with previous findings of L2 comprehenders showing sensitivity to other types of island in real-time processing tasks (e.g., Felser et al., 2012; Omaki & Schulz, 2011).

Although we attempted to control for potential complexity differences between our infinitival and finite conditions in Experiment 2 by including two additional control conditions, it is still possible that finite clauses are more difficult to extract from than non-finite ones (e.g., Kluender, 2004; Michel & Goodall, 2013). Based on the results from acceptability judgement experiments, Michel and Goodall argue that finiteness is not problematic in itself but may become a problem when it intervenes in certain dependencies, notably extractions from island domains. Note that in our materials, extraction from finite subject islands presumably involves crossing another (null) *wh*-operator in the embedded RC, which should then give rise to intervention effects (Rizzi, 2013). The primary reason preventing readers from postulating PGs inside finite subject islands may thus not be finiteness *per se*, but rather the need for establishing a second *wh*-dependency within the subject phrase in our finite conditions, which may also require creating additional functional structure at the left clausal periphery.

The L2 participants in the current study might have taken a little longer than the native speakers to compute a full structural representation for finite RCs that lack an overt *wh*-pronoun, and thus carried on searching for a gap for the original filler during their first reading of the island region. Note that even universal constraints on FGD formation (regardless of whether these are grammaticalised or not) can only apply once a sufficiently detailed sentence representation has been computed. Even though our L2 comprehenders may initially have failed to notice the relevant island-inducing cues, they nevertheless caught up with the native controls fairly quickly, and before reading the remainder of the sentence.

Note that our reading-time data fail to support expectation-based accounts of subject islands and PGs such as that proposed by Chaves (2013). This particular account would predict that, due to PGs inside subject islands being extremely rare in English, they should initially be overlooked but then cause garden-path effects later on during the sentence, forcing backtracking and reprocessing of the island region. In the current study, however, potential parasitic gaps were already postulated during both the L1 and L2 participants' initial reading of the island region, and even more liberally so by our L2 speakers.

In summary, while processing resource limitations may indeed play a crucial role in bringing about island effects, sensitivity to (covert) linguistic detail is also sometimes vital for island effects to show up in real-time processing tasks. As our L2 data illustrate, the failure to compute sufficiently detailed representations in real time (Clahsen & Felser, 2006) may result in island effects being delayed.

Effects at the ultimate gap

Recall that at the tail of the dependency we saw longer reading times for the infinitival implausible condition as compared to the others, where the PG was plausible and/or was prevented by a finite subject island. It looks like the earlier positing of an implausible PG, then, also impedes filler integration at the licit gap further downstream. Where the initial attempt to link the filler to a potential subcategoriser yielded an implausible semantic association, as was the case in our infinitival implausible condition (e.g., *to inspire houseplants*), this association appeared to persist despite the fact that, in our materials, all putative PGs were subsequently disconfirmed by the appearance of an overt direct object (e.g., *to inspire . . . the sensible criminals*).

Our observation that implausible fillers for PGs also cause processing difficulty at the terminal gap indicates that the processor tries to build sentence representations that are connected. Conceivably, we may be looking at a special case of 'semantic persistence' here, with structural reanalysis not fully dissolving an initially formed interpretative link (compare e.g., Sturt, 2007). That is to say, previous interpretations of the filler as a PG may be maintained even after the PG is syntactically disconfirmed by the presence of an overt object noun phrase. If that initial interpretation was implausible, it may then inhibit the filler's integration with its true semantic licenser at the terminal gap site.

Concluding remarks

Our results show that L1 German-speaking, proficient learners of English fleetingly posited gaps within both infinitival and finite subject islands during real-time processing. Native English speakers, in contrast, do not postulate parasitic gaps during their processing of finite subject islands. The observed L1/L2 processing difference is rather subtle, however, and can be accounted for by assuming that the L2 participants took slightly longer than native comprehenders to process island-inducing cues such as finiteness and/or the presence of a covert *wh*-operator.

The fact that our L2 participants patterned with the L1 control group in the offline judgement task – as well as at later processing stages – shows that even learners who lack equivalent PG structures in their L1 are aware of the licensing conditions on PGs inside English subject islands, despite the lack of explicit instruction and the fact that such PGs are very rare. This is in line with previous findings suggesting that L2 learners are sensitive to extraction islands in processing tasks, and furthermore points towards possible universal constraints being involved in determining islandhood.

Last but not least, we found that where a PG was initially postulated, thus allowing for the filler to be thematically interpreted, a lack of a plausible fit at this point also later impeded comprehenders' ability to integrate the filler at the terminal gap site. This suggests that comprehenders create sentence representations in which the two gaps are linked at some level of representation. Among the questions that remain are the empirical questions of whether the present findings generalise to other types of PG configurations, and to other languages or language combinations.

Supplementary Material

For supplementary material accompanying this paper, visit http://dx.doi.org/10.1017/S1366728915000942

References

Aldwayan, S., Fiorentino, R., & Gabriele, A. (2010). Evidence of syntactic constraints in the processing of *wh*-movement: A study of Najdi Arabic learners of English. In B. VanPatten & J. Jegerski (eds.). *Research in second language processing and parsing*. Amsterdam: John Benjamins, 65–86.

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

Barr, D., Levy, R., Scheepers, C., & Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68,* 255–278.

Bošković, Ž. (1996). Selection and the categorial status of infinitival complements. *Natural Language and Linguistic Theory, 14,* 269–304.

Chaves, R. (2013). An expectation-based account of subject islands and parasitism. *Journal of Linguistics, 49,* 285–327.

Chomsky, N. (1962). The logical basis of linguistic theory. In M. Halle (ed.) *Preprints of the papers for the ninth International Congress of Linguists,* 509–74.

Chomsky, N. (1973). Conditions on transformations. In S. Anderson and P. Kiparsky (eds.). *A festschrift for Morris Halle*. New York: Holt, Rinchart & Winston.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

Clahsen, H., & Felser, C. (2006). How native-like is non-native language processing? *Trends in Cognitive Sciences, 10* (*12*), 564–570.

Clifton, C., & Frazier, L. (1989). Comprehending sentences with long-distance dependencies. In G. Carlson & M. Tanenhaus (eds.). *Linguistic structure in language processing*. Dordrecht: Kluwer.

Culicover, P. (2001). Parasitic gaps: A history. In P. Culicover & P. Postal (eds.). *Parasitic gaps*. Cambridge, MA: MIT Press, 3–68.

Cunnings, I., Batterham, C., Felser, C., & Clahsen, H. (2010). Constraints on L2 learners' processing of wh-dependencies: Evidence from eye movements. In B. VanPatten & J. Jegerski (eds.). *Research in second language processing and parsing*. Amsterdam: John Benjamins, 87–110.

Dallas, A., & Kaan, E. (2008). Second language processing of filler-gap dependencies by late learners. *Language and Linguistics Compass* 2, 372–388.

Deane, P. (1991). Limits to attention: A cognitive theory of island phenomena. *Cognitive Linguistics, 2,* 1–63.

Ellis, N. (1998). Emergentism, connectionism and language learning. *Language Learning 48,* 631–64.

Engdahl, E. (1983). Parasitic gaps. *Linguistics and Philosophy, 6,* 5–34.

Felix, S. (1985). Parasitic gaps in German. In W. Abraham (ed.). *Erklärende Syntax des Deutschen*. Tübingen, 173–200.

Felix, S. (1988). UG-generated knowledge in adult second language acquisition. In S. Flynn & W. O'Neil (eds.) *Linguistic theory in second language acquisition*. Dordrecht: Kluwer, 277–294.

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

Felser, C., Cunnings, I., Batterham, C., & Clahsen, H. (2012). The timing of island effects in nonnative sentence processing. *Studies in Second Language Acquisition, 34,* 67–98.

Freedman, S., & Forster, K. (1985). The psychological status of overgenerated sentences. *Cognition, 24,* 171–186.

Gibson, E., & Warren, T. (2004). Reading time evidence for intermediate linguistic structure in long-distance dependencies. *Syntax, 7,* 55–78.

Haegeman, L., Jimenez-Fernandez, A. L., & Radford, A. (2014). Deconstructing the Subject Condition in terms of cumulative constraint violation. *The Linguistic Review, 31,* 73–150.

Haider, H. (1993). *Deutsche Syntax - generativ*. Tübingen: Narr.

Hofmeister, P., & Sag, I. (2010). Cognitive constraints and island effects. *Language, 86,* 366–415.

Juffs, A., & Harrington, M. (1995). Parsing effects in second language sentence processing: Subject and object asymmetries in *wh*-extraction. *Studies in Second Language Acquisition, 17,* 483–516.

Jurka, J., Nakao, C., & Omaki, A. (2011). It's not the end of the CED as we know it: Revisiting German and Japanese subject islands. *Proceedings of the 28th West Coast Conference on Formal Linguistics*. Somerville, MA: Cascadilla Proceedings Project.

Kathol, A. (2001). On the nonexistence of true parasitic gaps in standard German. In P. Culicover & P. Postal (eds.). *Parasitic gaps*. Cambridge, MA: MIT Press, 315–338.

Kayne, R. (1983). Connectedness. *Linguistic Inquiry, 14,* 223–49.

Kluender, R. (1992). Deriving island constraints from principles of predication. In H. Goodluck & M. Rochemont (eds.). *Island constraints: Theory, acquisition and processing*. Dordrecht: Kluwer, 223–258.

Kluender, R. (2004). Are subject islands subject to a processing account? In B. Schmeiser, V. Chand, A. Kelleher, A. Rodriguez (eds.). *Proceedings of the 23rd West Coast Conference on Formal Linguistics (WCCFL 23)*. Somerville, MA: Cascadilla Press, 101–125.

Kurtzman, H., & Crawford, L. (1991). Processing parasitic gaps. In T. Sherer (ed.). *Proceedings of the 21st annual meeting of the North Eastern Linguistics Society*. Amherst, MA: LSA Publications, 217–231.

Marinis, T, Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition, 27,* 53–78.

Michel, D., & Goodall, G. (2013). Finiteness and the nature of island constraints. In N. Goto, K. Otaki, A. Sato & K. Takita (eds.). *Proceedings of GLOW in Asia IX*: The Main Session. Mie University, Japan, 187–197.

Norusis, M. (2011). *IBM SPSS statistics 19 advanced statistical procedures companion*. Pearson.

Omaki, A., & Schulz, B. (2011). Filler-gap dependencies and island constraints in second language sentence processing. *Studies in Second Language Acquisition, 33,* 563–588.

Parker, A. (1999). Parasitic gaps in the Germanic languages. Master's dissertation, National University of Ireland, Dublin.

Phillips, C. (2006). The real-time status of island phenomena. *Language 82,* 795–823.

Phillips, C. (2013). On the nature of island constraints. I: Language processing and reductionist accounts. In J. Sprouse & N. Hornstein (eds.). *Experimental syntax and island effects*. Cambridge, UK: Cambridge University Press.

Rizzi, L. (2013). Locality. *Lingua, 130,* 169–186.

Ross, J. (1967). Constraints on variables in syntax. PhD dissertation, MIT, Cambridge, MA.

Schwartz, B., & Sprouse, R. (2000). When syntactic theories evolve: Consequences for L2 acquisition research. In J. Archibald (ed.). *Second language acquisition and linguistic theory*. Malden, MA: Blackwell, 156–186.

Staub, A., & Rayner, K. (2007). Eye movements and online comprehension processes. In G. Gaskell (ed.). *The Oxford handbook of psycholinguistics*. Oxford: Oxford University Press.

Stowe, L. (1986). Parsing *wh*-constructions: Evidence for on-line gap location. *Language and Cognitive Processes, 1,* 227–245.

Sturt, P. (2007). Semantic reinterpretation and garden-path recovery. *Cognition, 105,* 477–88.

https://doi.org/10.1017/S1366728915000942 Published online by Cambridge University Press

Szabolcsi, A. (2006). Strong vs. weak islands. In M. Everaert & H. van Riemsdijk (eds.), *The Blackwell companion to syntax, vol. 4*, pp. 479–531. Malden, MA: Blackwell Publishing.

Traxler, M., & Pickering, M. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language, 35,* 454–475.

Wagers, M., & Phillips, C. (2009). Multiple dependencies and the role of the grammar in real-time comprehension. *Journal of Linguistics, 45,* 395–433.

Williams, J., Möbius, P., & Kim, C. (2001). Native and non-native processing of English *wh*-questions: Parsing strategies and plausibility constraints. *Applied Psycholinguistics, 22,* 509–540.