The pronoun which comprehenders who process it in islands derive a benefit

CHRISTOPHER HAMMERLY — *UMass Amherst, Department of Linguistics*

Preprint Draft 1.1, October 22nd, 2019

1 INTRODUCTION

Resumptive Pronouns (RPs) are well known to "rescue" island-violating structures in many languages of the world (Aoun, Choueiri, and Hornstein, 2001; McCloskey, 2006). However, the effect of RPs on islands in so-called *intrusive* resumption languages such as English is far less clear. Intuitive judgments within theoretical syntax have frequently suggested that RPs in English do in fact rescue, or at least facilitate the comprehension of, syntactic islands (Ross, 1967; Prince, 1990; Asudeh, 2004). For example, while still considered sub-standard English, most speakers report an increase in acceptability with the resumptive sentence in (1a) compared to its gapped counterpart (1b) (example from Ross, 1967, p. 433).

(1)   a. %<u>All the students</u> who the papers which <u>they</u> submitted were lousy I'm not going to allow to register next term.
        b. *<u>All the students</u> who the papers which ____ submitted were lousy I'm not going to allow to register next term.

Recent decades have seen a proliferation of attempts to see whether these effects can be observed in an experimental setting. In contrast to the reported judgments, initial studies in this vein found no evidence that RPs in English ameliorate islands, as measured by changes in acceptability with Likert scales (Heestand, Xiang, and Polinsky, 2011; Polinsky, Clemens, Morgan, Xiang, and Heestand, 2013) and magnitude estimation (Alexopoulou and Keller, 2007; Omaki and Nakao, 2010).

Taking these results at face value, a puzzle that emerges is why resumption is so commonly observed in production both in natural speech and in the lab (e.g. Prince, 1990; Ferreira and Swets, 2005; Morgan and Wagers, 2018) if it is considered ungrammatical by comprehenders. One leading idea is that resumption in English provides no detectable benefit to listeners, but does serve as an aid for speakers (Polinsky et al., 2013). This amounts to adopting a so-called selfish-speaker model, where speakers (unconsciously) optimize based on decreasing processing cost for themselves, rather than decreasing cost for the listener. Speakers, who are presumably aware of the coreference dependencies which they intend to convey, use resumption just in case these dependencies becomes difficulty to produce: i.e. in islands and dependencies that span multiple clauses.

Recent work has challenged the claim that RPs in English have no benefit to comprehenders. Beltrama and Xiang (2016) attribute the lack of acceptability to the framing of the task and context: ratings increase when participants are asked to judge the comprehensibility of a sentence rather than the acceptability, and more contextually salient antecedents for RPs are rated more highly. In a similar vein, Ackerman, Frazier, and Yoshida (2018) show that when participants explicitly compare gapped and resumptive counterparts facilitation increases, suggesting that previous failures to observe an ameliorating effect of resumption is the result of a task effect. Finally, Chacón (2019) reports that resumption becomes increasingly acceptable as memory load increases, supporting an account where the acceptability of resumption is tied to difficulty in a listeners ability to engage in active gap filling.

While the above results are highly suggestive, none provides direct evidence for an online benefit for RPs in the comprehension of islands. Existing work on the real-time processing of resumption in English has only examined the role of RPs in non-island violating structures (Dickey, 1996; Hofmeister and Norcliffe, 2013),

both of which show that RPs facilitate (i.e. speed-up) reading as dependency length grows longer. While these results are suggestive, at present, there has been no direct test of whether RPs in islands provide a similarly facilitatory effect for English comprehenders.

2 THE CURRENT EXPERIMENT

The current study examines the role of resumptive pronouns in the online comprehension of *wh*-islands. We ask two questions: (i) whether there is evidence of a real-time benefit for RPs in island violating structures as measured by facilitation in reading time, and (ii) whether such facilitation can be linked to better performance in questions targeting sentence comprehension. A full list of items and fillers including comprehension questions, as well as raw data, analysis scripts, and experimental scripts, can be accessed here: https://osf.io/g9q5a/

The experiment (N = 48) included 24 experimental items of the form in (2), where ISLANDHOOD (non-island/wh-island) and DEPENDENCY (gap/RP) were manipulated. Names stereotypically mismatching the gender of the RP were used to prevent a locally coherent parse of the embedded clause from creating facilitation (e.g. as in ...*Aaron announced that the agency had recruited him*...). Each sentence was split into chunks, denoted by slashes in (2), for moving-window self-paced reading (run in lab using Ibex Farm; Drummond, 2013). Participants read through the chunks of the sentence one-by-one, pressing the spacebar with their right pointer finger to move from one chunk to the next. Response time (RT) was measured by recording the time between button presses, providing an approximate measure of reading time for each chunk of the sentence. The critical region in-

cluded the embedded verb and the RP/gap. The spillover was a three-word PP. The spillover region was always followed by a final wrap-up region.

(2)   *Non-island (a) and wh-island (b) example experimental stimulus. Slashes denote self-paced reading regions. Critical and spillover regions labeled.*

    a.   Mary trained / the spy / who Beth / announced that the agency /

       /$_{\text{crit}}$ had recruited {him, __ } /$_{\text{spill}}$ over the summer / for the program. /

    b.   Mary trained / the spy / who Beth / announced which agency /

       /$_{\text{crit}}$ had recruited {him, __ } /$_{\text{spill}}$ over the summer / for the program. /

Following the final chunk, a comprehension question with response options of the type exemplified in (3) was displayed. To prevent guessing and to gather a measure of comprehension difficulty, participants were specifically instructed to use the *confusion* response "I don't know" (3d) in case they felt they could not answer the question accurately. The other possible responses always corresponded to the *filler* (3a), the *non-local* gender mismatching distractor (3b), or the *local* distractor (3c). The order of responses was randomized for each trial, except for the confusion response which was always the fourth option. Participants made responses with their left hand using the number keys on the top of the keyboard.

(3)   *Example comprehension question for example item in (2)*

    Who was recruited over the summer?

    a.   The spy   *Filler*

    b.   Beth   *Non-local*

    c.   The agency   *Local*

    d.   I don't know   *Confusion*

A total of 60 grammatical filler sentences were randomly interspersed with the experimental items, making 86% of the sentences in the experiment grammatical. 24 of the fillers were items for a separate experiment on presupposition accommodation. Half of these fillers included a multi-clausal filler-gap dependency. Given that the comprehension questions for the critical experimental items always had the filler response as the target answer, we formulated questions that targeted different aspects of the sentence. The remaining 36 fillers were designed to further counterbalance the experimental items, including items that allowed for a valid resolution of the pronominal dependency in the embedded clause and items with superficial similarities to the island conditions. Again, comprehension questions were designed to target a variety of aspects of the sentence.

3 RESULTS: SPR RESPONSE TIME

Trials where RT for any given chunk was lower than 100ms or higher than 6s were discarded, resulting in a loss of 3.23% of trials overall. All reported values, including those for comprehension discussed in the following section, are based on the trimmed data set where these trials were rejected.

In the critical region, the key contrast between gaps and RPs was confounded by length (conditions with an RP were always 3 characters longer than those with a gap). Therefore the focus of the analysis is on the spillover region. Full region-by-region RT for each condition is given in Table 1, with RT for the spillover region alone shown in Figure 1. Mean RT in the spillover shows that when the dependency was tailed by a gap, RT was slower with *wh*-islands compared to non-islands. However, the slow-down associated with *wh*-islands in the gapped conditions was

attenuated by the presence of an RP, while RPs slowed reading relative to gaps in non-islands.

		Matrix	Filler	Name	Island	Critical	Spillover	Wrap-up
Non-island	Gap	1053 (62)	1025 (53)	1203 (75)	1635 (113)	854 (51)	768 (35)	819 (45)
	RP	1162 (71)	1118 (81)	1197 (78)	1617 (100)	1039 (63)	826 (41)	819 (47)
wh-island	Gap	1151 (74)	1013 (53)	1159 (65)	1537 (86)	884 (43)	902 (46)	875 (49)
	RP	1228 (80)	1051 (58)	1198 (81)	1671 (87)	1031 (56)	815 (43)	853 (50)

Table 1: Average RT (ms) and by-subjects SEM by region and condition. Regions denoted by slashes in (2). Note that the Island and Critical regions are confounded by number of characters across conditions.
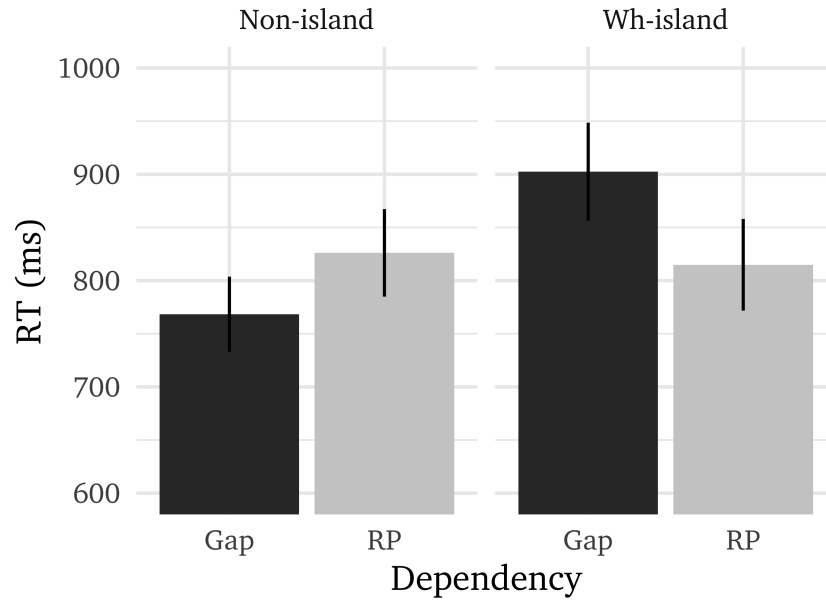

Figure 1: Average RT (ms) and by-subjects SEM for spillover region only. RPs slow RT in the non-island condition compared to gaps, but facilitate RT in the *wh*-island condition.

Raw RT for the spillover region was submitted to a maximal linear mixed effects model using the *lmerTest* package in R (Kuznetsova, Brockhoff, and Christensen, 2017) with fully crossed fixed effects of DEPENDENCY and ISLANDHOOD and subject and item random effects. The model revealed a significant main effect of ISLAND-

HOOD (Estimate (SE) = -54.96 (25.35), t = -2.16, p = 0.031), as well as a signifi-cant two-way interaction between DEPENDENCY and ISLANDHOOD (Estimate (SE) = 143.73 (50.71), t = 2.84, p = 0.005). In addition to a general slow-down in the *wh*-island conditions compared to non-islands, the results support the presence of a critical *crossover interaction*: RPs slowed reading relative to gaps in the non-island conditions, but produced a facilitatory effect compared to gaps in *wh*-islands. This supports the view that RPs facilitate the comprehension of island violating struc-tures. I leave further interpretation of this result to the discussion section.

4 RESULTS: COMPREHENSION QUESTIONS

The average response proportion for each of the four response options, broken down by the four conditions, is given in Table 2, with the proportion of filler re-sponses and confusion responses shown in Figure 2. For the analysis of compre-hension questions, I focus on the proportion of correct *filler* responses as well as the proportion of *confusion* responses. More filler responses were given in the non-island compared to the *wh*-island condition, and in turn more confusion responses were given in the *wh*-island condition compared to non-islands. The decrease in filler responses associated with *wh*-islands is slightly attenuated by the presence of an RP. Similarly, RPs are associated with a decrease in the proportion of confusion responses in the *wh*-island condition.

		Filler	Confusion	Non-local	Local
Non-island	Gap	0.80 (0.029)	0.10 (0.018)	0.07 (0.018)	0.04 (0.012)
	RP	0.79 (0.028)	0.10 (0.020)	0.07 (0.016)	0.03 (0.011)
wh-island	Gap	0.69 (0.031)	0.16 (0.023)	0.07 (0.016)	0.08 (0.018)
	RP	0.72 (0.038)	0.11 (0.023)	0.11 (0.026)	0.06 (0.012)

Table 2: Average response proportion and by-subjects SEM by response type and condition. Rows may not add up to 1 due to rounding error.
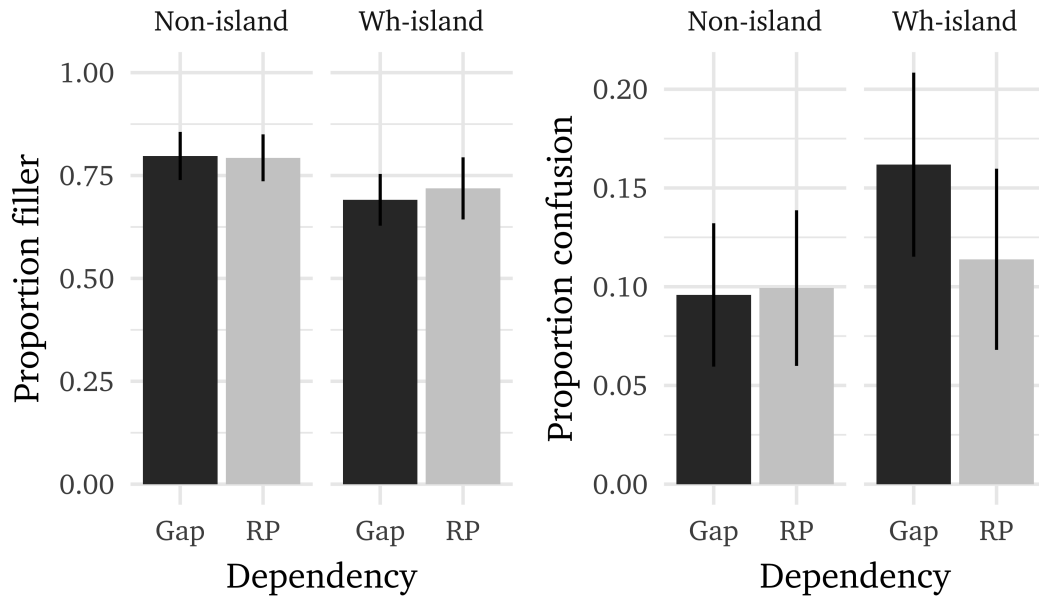
Figure 2: Average response proportion and by-subjects 95% confidence intervals for filler (left) and confusion (right) response options. Non-islands show increased filler responses and decreased confusion compared to *wh*-islands. RPs numerically increase filler responses and decrease confusion in *wh*-islands, but the result is not statistically significant in the reported models.

Maximal logistic mixed effects models with the same effects structure as in the RT analysis were fit separately on the response data for the filler and confusion response types. Both models revealed only a significant main effect of ISLANDHOOD (*Filler model*: Estimate (SE) = 0.61 (0.17), $z = 3.62$, $p < 0.001$; *Confusion model*: Estimate (SE) = -0.41 (0.20), $z = -2.07$, $p = 0.039$). In neither case did the interaction of ISLANDHOOD and DEPENDENCY approach significance. These results fail to support the presence of a facilitatory effect of RPs in island-violating structures for these measures.

5 DISCUSSION

The present experiment provides the first evidence for a real-time benefit of resumptive pronouns in island-violating structures. The results further support a view where RPs in non-island violating structures disrupt processing. Previous studies have shown that RPs facilitate processing in longer (i.e. multi-clausal) non-island structures (Alexopoulou and Keller, 2007; Dickey, 1996; Hofmeister and Norcliffe, 2013). In this study, the ameliorating effect of RPs in *wh*-islands cannot be attributed to length (as measured by number of intervening clauses between the filler and the tail of the dependency), as all conditions were matched in this respect. Furthermore, the presence of a *disruptive* effect of RPs in the non-island condition cannot be explained by length: if dependency length were the determining factor, then facilitation should have been observed regardless of islandhood.

The pattern of results can most readily be understood in light of past findings in filler-gap processing. Considering first the *wh*-island conditions, previous work has shown that islands generally block a comprehender's ability to assign a filler into a gap (e.g. Traxler and Pickering, 1996; Wagers and Phillips, 2009). This amounts to a suspension of the *active filler strategy* (Frazier and Flores d'Arcais, 1989) within islands. Under normal conditions, when a filler is encountered, the parser engages in a search for a gap. This leads the parser to attempt dependency formation between the filler and the first encountered gap site (i.e. the first unfilled argument position). Islands block active dependency formation of this sort.

Returning to the present experiment, we observed an increase in RT at the site of an unfilled argument position within an island (i.e. in the gapped *wh*-island condition). The disruption can be attributed to the confusion caused by the presence of an unfilled argument position, with no immediately available filler to form a de-

pendency into this site. This supports the view that the activity of the filler, and therefore the active search for a gap, is suspended in islands. However, the disruptive effect caused by the lack of active filler is attenuated when the argument position is filled by an RP. The RP facilitates the formation of a dependency between the displaced filler and the argument position within the island in lieu of the active-filler strategy.

On the other hand, the disruptive effect of RPs compared to gaps in the non-island condition can be attributed to a *filled-gap effect* (Crain and Fodor, 1985; Stowe, 1986). For example, in sentences containing a filler-gap dependency like those in (4a), reading time is disrupted at *students* compared to a non-filler-gap counterpart (4b). This is directly due to active-filler strategy: when the verb *like* is encountered and there is an active filler, the parser immediately posits a gap to place the filler into. When this gap turns out to already be filled by *students*, processing is disrupted.

(4)   a.   Kaan forgot <u>which font</u> the professor **likes** students to use __

        b.   Kaan forgot whether the professor **likes** students to use a specific font

In the present experiment, the RP in the non-island condition filled the gap site, leading to a processing disruption in these conditions compared to conditions where this position was unfilled. The results from the comprehension questions show that ultimately this pronoun was interpreted as coreferent with the filler, but not without causing an initial disruption.

The general picture is one in which a parser's ability to actively maintain a filler is the critical determinant of whether RPs in intrusive resumption languages like English result in facilitation. In short, RPs become useful when fillers are difficult or

impossible to keep active. This aligns with findings that increases on memory-load due to factors such as dependency length or task demands increase the facilitatory effect of RPs (Hofmeister and Norcliffe, 2013; Chacón, 2019). They do so by aiding in the formation of a dependency, allowing an otherwise unfillable (or difficult to fill) argument position to be resolved. Island-violating structures fall naturally into this explanation. As reviewed above, islands are independently known to suspend the active filler strategy, resulting in an inactive filler that cannot be used to immediately associate with the argument position. The processing of the otherwise unfilled argument position within an island is facilitated by the presence of an RP, allowing a dependency to be formed.

The view advocated for here is related to the one recently put forward by Chacón (2019), with one key conceptual difference. Chacón argues that fillers trigger the predictive formation of a gapped structure, and that increased memory load leads to degradation of this prediction. The current account argues that it is instead a decrease in the activity of the *filler* under memory duress that leads to a suspension of the active filler strategy. The reason for preferring a filler-driven theory over a gap-driven theory in the context of the current study is the disruptive effect of the open argument position in islands. The parser has no trouble identifying open argument positions (i.e. potential gap sites), but does have trouble finding something to interpret in that position. This supports a filler-driven theory of the active filler strategy, and the view that filler degradation is responsible for processing difficulty in islands.

One question that remains is why the RT measure showed evidence of the facilitatory effect of RPs in islands, but the comprehension questions did not. There are a number of reasonable possibilities (which are not mutually exclusive). The first relates to the relatively small effect size. While the numerical trends in the averages

aligned with the RT patterns, it is possible that the experiment did not provide sufficient power to observe the effect of interest. The second concerns the distribution of responses. Filler responses (which can be considered "correct" responses) generally appeared to be close to ceiling given the task. Similarly, the overall proportion of confusion responses was relatively low. It is possible that a more difficult set of comprehension questions would lead to increased variability in responses and allow the effect to be observed. Finally, it is possible that by the time the comprehension question was encountered, participants were generally able to reconstruct the gist of the sentence regardless of online difficulty. That is, comprehenders benefitted (or were disrupted) by RPs online, but by the end of the sentence, participants could recover the meaning anyway.

While the results do not support the view that RPs in English are part of the grammar per se, it is clear that they serve a purpose that goes beyond what the speaker-oriented model advocated by Polinsky et al. (2013) would suggest. It is still possible that speakers produce resumptive pronouns solely to help themselves form difficult dependencies, but the current study shows that comprehenders derive a benefit anyway. This benefit is not, so far, reflected strongly in acceptability, but has been evident in comprehensibility ratings (Beltrama and Xiang, 2016). To the extent that acceptability reflects the state of the grammar, and comprehensibility the difficulty with which a sentence is understood, resumption in English continues to live in a gray zone: not grammatical, but still beneficial under the various pressures of real-time processing.

REFERENCES

Ackerman, Lauren, Michael Frazier, and Masaya Yoshida. 2018. Resumptive pronouns can ameliorate illicit island extractions. *Linguistic Inquiry* 49:847–859.

Alexopoulou, Theodora, and Frank Keller. 2007. Locality, cyclicity, and resumption: At the interface between the grammar and the human sentence processor. *Language* 83:110–160.

Aoun, Joseph, Lina Choueiri, and Norbert Hornstein. 2001. Resumption, movement, and derivational economy. *Linguistic Inquiry* 32:371–403.

Asudeh, Ash. 2004. Resumption as resource management. Doctoral Dissertation, Stanford University.

Beltrama, Andrea, and Ming Xiang. 2016. Unacceptable but comprehensible: the facilitation effect of resumptive pronouns. *Glossa* 1.

Chacón, Dustin A. 2019. Minding the gap?: Mechanisms underlying resumption in English. *Glossa* 4.

Crain, Stephen, and Janet Dean Fodor. 1985. How can grammars help parsers? In *Natural language parsing*, ed. David Dowty, Lauri Karttunen, and Arnold Zwicky, 94–128. Cambridge University Press.

Dickey, Michael Walsh. 1996. Constraints on the sentence processor and the distribution of resumptive pronouns. *Linguistics in the Laboratory* 19:157–192.

Drummond, Alex. 2013. Ibex farm. *Online server: http://spellout. net/ibexfarm* .

Ferreira, Fernanda, and Benjamin Swets. 2005. The production and comprehension of resumptive pronouns in relative clause "island" contexts. In *Twenty-first century psycholinguistics: Four cornerstones*, ed. Anne Cutler, 263–278. Mahwah, NJ: Lawrence Erlbaum.

Frazier, Lyn, and Giovanni B Flores d'Arcais. 1989. Filler driven parsing: A study of gap filling in dutch. *Journal of memory and language* 28:331–344.

Heestand, Dustin, Ming Xiang, and Maria Polinsky. 2011. Resumption still does not rescue islands. *Linguistic Inquiry* 42:138–152.

Hofmeister, Philip, and Elisabeth Norcliffe. 2013. Does resumption facilitate sentence comprehension? In *The core and the periphery: Data-driven perspectives on syntax inspired by Ivan A. Sag*, 225–246. CSLI Publications.

Kuznetsova, Alexandra, Per B Brockhoff, and Rune Haubo Bojesen Christensen. 2017. lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software* 82.

McCloskey, James. 2006. Resumption. *The Blackwell companion to syntax* 94–117.

Morgan, Adam Milton, and Matthew W Wagers. 2018. English resumptive pronouns are more common where gaps are less acceptable. *Linguistic Inquiry* 49:861–876.

Omaki, Akira, and Chizuru Nakao. 2010. Does English resumption really help to repair island violations. *Snippets* 21:11–12.

Polinsky, Maria, Lauren Eby Clemens, Adam Milton Morgan, Ming Xiang, and Dustin Heestand. 2013. Resumption in English. In *Experimental syntax and island effects*, ed. Jon Sprouse and Norbert Hornstein, 341–359. New York: Cambridge University Press.

Prince, Ellen F. 1990. Syntax and discourse: A look at resumptive pronouns. In *Proceedings of the Sixteenth Annual Meeting of the Berkeley Linguistics Society: Parasession on the Legacy of Grice*, ed. Kira Hall, Jean-Pierre Koenig, Michael Meacham, Sondra Reinman, and Laural A. Sutton, 482–497. Berkeley: University of California, Berkeley Linguistics Society.

Ross, John Robert. 1967. Constraints on variables in syntax. Doctoral Dissertation, Massachusetts Institute of Technology.

Stowe, Laurie A. 1986. Parsing *wh*-constructions: Evidence for on-line gap location. *Language and cognitive processes* 1:227–245.

Traxler, Matthew J, and Martin J Pickering. 1996. Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language* 35:454–475.

Wagers, Matthew W, and Colin Phillips. 2009. Multiple dependencies and the role of the grammar in real-time comprehension. *Journal of Linguistics* 45:395–433.