RESEARCH

Unacceptable but comprehensible: the facilitation effect of resumptive pronouns

Andrea Beltrama and Ming Xiang

University of Chicago, US

Corresponding author: Andrea Beltrama (andremormora@uchicago.edu)

It is often assumed in the theoretical syntax literature that intrusive resumptive pronouns can rescue island violations. However, recent experimental investigations did not provide strong evidence for such a rescuing effect. The current study examines intrusive resumption in Italian and English. In four experiments, we show that resumption indeed improves island violation to some degree, but such an effect is sensitive to task and contextual manipulations. In particular, the rescuing effect only surfaces with a comprehensibility but not a traditional acceptability task, and the effect is strongest when the antecedent of the resumptive pronoun is made salient through additional context. At the same time, however, the effect of resumption in longer embedded clauses (compared to shorter ones) is much weaker. We discuss these findings in terms of how resumptive pronouns, although ungrammatical in English, can facilitate parsing in particular yet principled ways.

1 Introduction

Resumptive pronouns (henceforth, RPs) have drawn considerable attention in theoretical and experimental syntax. Informally defined as pronouns which are found in a position "in which, under other circumstances, a gap would appear" (McCloskey 2006: 26), RPs are subject to well-documented patterns of cross-linguistic variation (McCloskey 2006). A distinction has been made between *grammatical* resumptives as found in Hebrew, Swedish, Irish, certain varieties of Arabic, etc. (Chao & Sells 1983; Engdahl 1985; Shlonsky 1992; McCloskey 2006) and *intrusive* resumptives found in English. In languages with grammatical resumption, RPs have been reported to be perfectly acceptable, both when they are obligatory (e.g. in direct object relatives in Palestinian Arabic, see Shlonsky 1992), or when they occur in free variation with gaps (e.g. in direct object relatives in Hebrew, see Shlonsky 1992). Intrusive resumption, instead, is reported to be ungrammatical and generally unacceptable, as the following contrast shows (example from Erteschik-Shir 1992: 89).

(1)   a.   *This is the girl that John likes **her**. (RP)
         b.   This is the girl that John likes __. (Gap)

This paper focuses on RPs of the intrusive kind. Two interesting and seemingly contrasting observations have emerged from previous studies on intrusive RPs. On the one hand, linguists have commonly assumed, based on introspective judgments, that resumptives aid processing of long distance dependencies in situations where the processing demand is high, such as syntactic islands and dependencies with multiple embeddings. On the other hand, controlled experiments have not been able to consistently find amelioration effects

in the acceptability of RPs over gaps in such environments (see Section 2.2), challenging the claims in the theoretical literature.

This paper aims to reconcile these seemingly paradoxical observations. Based on evidence from four experiments, we argue that intrusive RPs, while less consequential for *acceptability judgment*, do make the sentence more *comprehensible* in the presence of island violations. In particular, Experiments 1 and 2 provide evidence that RPs improve comprehensibility in both Italian and English. Experiments 3 and 4 suggest that the effect of RPs does not emerge with an acceptability task (Experiment 3) or without a sufficiently rich preceding context (Experiment 4). In all of these experiments, we also tested the effect of resumption with increased level of embedding. Consistent with previous findings, the effect of resumption in deeply embedded clauses is relatively subtle. We discuss these findings in terms of how resumptive pronouns, although ungrammatical in English, nevertheless could facilitate parsing in particular yet principled ways. Methodologically, our findings also raise questions about the role of task in eliciting linguistic judgments.

2 Background: The puzzle of intrusive resumptives
2.1 Resumptives as processing facilitators

Our discussion starts from two previous observations. First, intrusive RPs, though ungrammatical, are systematically found in spontaneous speech and laboratory-based speech production studies (Prince 1990; Creswell 2002; Ferreira & Swets 2005; Bennett 2008; also see Francis et al. 2015 on Cantonese Chinese). Second, linguists' introspective judgments suggest that RPs "sound better" in at least two environments: Island violations and long distance dependencies with multiple embeddings (Ross 1967; Kroch 1981; Sells 1984; Prince 1990; Erteschik-Shir 1992; Asudeh 2004; Asudeh 2011). For instance, in example (2), which contains a syntactic island, a resumptive pronoun is reported to improve the status of the sentence compared to a gap ((2a) vs. (2b)). It is important to note, however, that there is no consensus on what the best measure is to operationalize the amelioration effect. For the time being, we use the sign "$\gg$" to notate the yet-to-be specified meaning of "sounds better."

(2)   (from Asudeh 2004: 320)
    a. I'd like to meet the linguist that Peter knows a psychologist that works for **her**. (island with RP) $\gg$
    b. I'd like to meet the linguist that Peter knows a psychologist that works for _. (island with gap)

Sentences with multiple embeddings represent another environment in which resumption has been reported to make the sentence "sound better." Two separate contrasts are relevant here. RPs in sentences with multiple embeddings are reported to be better than RPs in simpler dependencies ((3a) vs. (3b)); they are also reported to be better than gaps with the same number of embeddings ((3c) vs. (3d)).

(3)   (from Erteschik-Shir 1992: 89)
    a. This is the girl that Peter said that John thinks that yesterday his mother had given some cakes to **her.** (complex dependency with RP) $\gg$
    b. This is the girl that Peter gave some cakes to **her** (simple dependency with RP)
    c. This is the girl that Peter said that John thinks that yesterday his mother had given some cakes to **her.** (complex dependency with RP) $\gg$
    d. This is the girl that Peter said that John thinks that yesterday his mother had given some cakes to _. (complex dependency with gap)

To explain the reported improvement of sentences like (2a), (3a) and (3c), and to account for the fact that RPs are relatively frequent in spontaneous speech, it has been suggested that intrusive resumptives can facilitate production and/or comprehension in unfavorable processing conditions (Kroch 1981; Prince 1990; Erteschik-Shir 1992; Asudeh 2004; 2011; 2012, among others). More specifically, both islands and multiple embeddings might overload the parser due to their structural complexity. In both of these cases, the presence of a resumptive pronoun could facilitate performance by alleviating the processing burden. As for the exact mechanism whereby they do this, various hypotheses have been put forward.

Kroch (1981) argues that resumption can be used to fix errors due to poor planning in production. If speakers begin to articulate an utterance before having a complete planning of the syntactic structure for the whole sentence, they might find themselves in trouble midway through the utterance. This happens, for example, when an island boundary is encountered during the production of a long-distance dependency. In this situation, the only way to deliver a coherent message without disrupting fluency is to insert an RP. According to Kroch, the end result of this process is a sentence that is ungrammatical, but, contrary to an island with a gap, such a sentence is at least somewhat *interpretable*. Prince (1990) proposes a similar account, arguing that RPs in English are "officially ungrammatical" (Prince 1990: 480), but are not uncommon in speech. Observing that nearly 70% of spontaneously produced RPs occur in islands, she proposed a processing account, suggesting that in these environments resumption serves the purpose of "making the best out of a bad job" (Prince 1990: 483). For the appearance of RPs in dependencies with multiple embeddings, similar processing-based explanations have been proposed. In particular, Erteschik-Shir (1992) suggests that RPs such as those in (2) – (3) are cognitively advantageous: they help the hearer to make sense of the extracted NP, which has been pushed out of short-term memory due to the relatively long temporal interval that has elapsed since the initial encounter with this NP. Finally, Asudeh (2004; 2011; 2012) provides a unified theory of intrusive resumption in islands and in complex dependencies with multiple embeddings. He argues that RPs, while ungrammatical, can nevertheless help the formation of a *locally well-formed* structure (see Section 2.3 for more details).

We now turn to discuss a set of controlled acceptability studies that appear to pose some challenges to the reported introspective judgments on RPs.

2.2 The empirical challenge from the acceptability judgment studies

As mentioned earlier, the two representative environments that have been reported to host intrusive RPs are syntactic islands, and sentences with multiple levels of embedding. To operationalize this intuition, a number of researchers have made the following two predictions, both of which hypothesized that RPs lead to improved acceptability. First, RPs in islands are hypothesized to be more acceptable than their gapped counterparts; second, RPs in dependencies with multiple embeddings should be more acceptable than both their gapped counterparts and RPs in simpler dependencies. Surprisingly, however, these predictions were not completely borne out in previous experimental investigations: RPs were not more acceptable than gaps in many of the previous experiments; neither was RPs' acceptability consistently ameliorated by increased levels of embedding. We review some of the experimental studies below.

In Ferreira and Swets (2005), participants were cued to produce target sentences containing a resumptive pronoun inside a wh-island, such as "*This is a donkey that I don't know where it lives*". Then the same participants were asked to rate the acceptability of the sentences they produced earlier. Regardless of whether the stimuli were presented in auditory or written format, sentences containing RPs turned out to be significantly less acceptable than the control sentences (e.g. "*This is a donkey that doesn't know where*

it lives"), showing that RPs in islands were not accepted even by the same speakers who produced them. A caveat of this result is that the study did not directly compare the acceptability of RPs with their gapped counterparts, and therefore is inconclusive as to whether intrusive RPs within islands are more acceptable than gaps. At the very least, however, the results strongly suggest that production of RPs does not automatically entail their full acceptability (also see Zukowski & Larsen 2004, for similar design and results).

A later set of acceptability studies directly compared RPs with gaps and found either no facilitation effect or very limited improvement associated with RPs. Alexopoulou & Keller (2007) compared gaps and RPs in islands by testing wh-questions containing complement clauses in English, German, and Greek. In a magnitude estimation task, they showed that in all tested environments, RPs did turn out to be more acceptable than gaps. Heestand et al. (2011) tested the acceptability of RPs in relative clause islands and adjunct islands. In a Likert scale acceptability judgment task, the participants were explicitly instructed to "judge (the acceptability) based on their native-speaker intuition rather than any prescriptive rules, and to go with their first instinct rather than spending time pondering on their answers" (Heestand et al. 2011: 142). In addition to a regular offline acceptability task, a speeded presentation task was also employed to impose a certain amount of time pressure on the participants. In neither of these tasks were RPs rated more acceptable in islands than gaps, although there was a numerical trend that with the relative clause islands the acceptability judgment was made slightly faster on the RP conditions (Experiment 2). The acceptability rating results from Heestand et al. (2011) were later replicated on a set of auditorily-presented stimuli, generalizing the absence of the island-rescuing effect to a different modality (Clemens et al. 2012; Polinsky et al. 2013).

It is worth noting that the studies reviewed above all compared RPs with gaps in islands that involve object extraction. A number of other studies examined islands with both object and subject extractions (McDaniel & Cowart 1999; McKee & McDaniel 2001; Keffala & Goodall 2011; Keffala 2011; Han et al. 2012). These studies replicated the findings that RPs were not rated more acceptable than gaps in object-extracted islands; but crucially they also found higher ratings for RPs than gaps in subject-extracted islands. Yet, it is unclear whether such an improvement should be entirely due to a rescuing effect of RPs. As Keffala (2011) suggested, subject relative clause islands with gaps record extremely low acceptability judgments in that they evoke two different kinds of syntactic violations: island constraints and ECP effects. Whereas the presence of a resumptive pronoun does not rescue the island violation *per se*, it salvages the ECP effect[1], preventing a further degrading in acceptability.[2]

Concerning RPs in dependencies with multiple levels of embedding, resumption appears to have an effect on acceptability, but only under some situations. First of all, across different studies it was shown that increased level of embedding did not yield RPs more acceptable than gaps (Alexopoulou & Keller 2007; Keffala and Goodall 2011; Keffala 2011; Han et al. 2012; Hofmeister & Norcliffe 2013). When RPs in sentences with longer embedding are compared with RPs in shorter embedding, the strongest effect was recorded when zero-embedding was compared to one or more levels of embeddings (Alexopoulou & Keller 2007). The differences between higher numbers of embedding (e.g. two vs. three-levels)

---

[1] Cowart & McDaniel (1999), along similar lines, argue that resumptives cannot rescue violations of constraints on movement (e.g. subjacency). They are however able to improve violations of constraints on the distribution of traces in certain positions, which include the ECP. We thank an anonymous reviewer for pointing this out to us.

[2] The experimental literature on resumption and islands in English has primarily focused on wh-islands and relative clause islands. It should be noted that McKee & McDaniel (2001) also found that RPs were significantly more preferred than gaps when the extracted element was a genitive object (e.g. "This is the pirate whose Minnie Mouse buried treasure"). Francis et al. (2015) more systematically tested resumption in possessive phrases in Cantonese and also found higher acceptability ratings on RPs than gaps.

were very subtle – many of the studies above found that RPs in longer embedded clauses received the same, but not higher, acceptability ratings as those in shorter ones. However, since acceptability degradation associated with length was independently observed for sentences with gaps or declarative controls in these studies, one could also argue that the fact that RPs "neutralize" the negative length effect on acceptability is itself a demonstration of the amelioration effect of RPs.

2.3 Looking for the source of the processing facilitation

Although the experimental findings reviewed in the last section raise some questions on the facilitating effect of RPs, it would be too hasty to conclude that RPs' processing facilitation is illusory. In particular, we will suggest that the failure to find the facilitation effect of RPs is closely tied to the particular task employed in previous studies, i.e., that acceptability judgment is not necessarily the best measure to capture or operationalize the facilitation effect of RPs.

   Given that there is consensus that English resumption is of the "intrusive" type and is ungrammatical (e.g. Kroch 1981; Chao & Sells 1983; Prince 1990), we already have some initial reasons to ask whether acceptability judgment is the most appropriate index to quantify the facilitation effect of RPs (see more discussion in 4.1).[3] There are also proposals that explicitly argue that RPs impact the comprehension of a construction, as opposed to its acceptability. One of the most detailed accounts on how RPs can facilitate comprehension comes from Asudeh (2004; 2011; 2012). In his system, there are two distinct levels according to which the well-formedness of a sentence is evaluated: a *global* one, which concerns the sentence in its entirety, and a *local* one, which concerns the smaller segments that combine to form the sentence itself. Syntactic islands represent an example of *globally* ill-formed constructions – a filler (i.e., the extracted element) cannot be successfully interpreted as an argument of the verb, leading to ungrammaticality. The difference between the presence of a gap and the presence of a RP emerges at the local level. Sentences with a gap are locally ill-formed: given the impossibility of integrating the filler, the gap after the verb is perceived as an illicitly missing argument. By contrast, the presence of a RP ensures local well-formedness, as it supplies an argument to the verb. An example is given below (from Asudeh 2004: 320), with the underlined part representing the relevant local segment where the gap/RP is found.

(4)   a. *I'd like to meet the linguist that ***Peter knows a psychologist that works with __.**   LOCALLY: * GLOBALLY: *

        b. *I'd like to meet the linguist that ↻ **Peter knows a psychologist that works with her.**   LOCALLY: ↻ GLOBALLY: *

According to Asudeh, restoring local well-formedness in a globally ill-formed structure allows the speaker to produce a sentence that is consistent with the message plan. At the same time, it makes it possible for the listener to put together a coherent interpretation, extracting a meaningful message even if the structure is not grammatical. It follows from this account that resumption has little effect on the grammatical status of a sentence. Instead, the processing facilitation should be specifically related to the *comprehensibility* of a construction, which refers to how easily a speaker can construct a coherent interpretation out of an utterance. If this is true, the particular task adopted by previous experiments, i.e., the acceptability judgment task, may not be the most appropriate task to capture the facilitation effect of RPs.

---

[3] Grammatical resumption is not our main focus in this paper. But we note that not all studies on grammaticalized resumption have shown higher acceptability ratings of RPs than gapsin islands (e.g. Tucker et al. 2016, on Modern Standard Arabic).

This hypothesis gains some initial support from studies that did not measure the effect of RPs with an acceptability task. Hofmeister & Norcliffe (2013), besides collecting acceptability judgments, adopted a self-paced reading task paradigm to compare the processing difficulty of 2 and 3-embedding sentences with gaps and RPs in English. While, as discussed above, the authors did not find an improvement in acceptability, they did find an effect of resumption on reading times. Going from 2 to 3-embeddings, reading times on regions following the RP/gap increased significantly on the sentences with gaps, and decreased for sentences with RPs. Based on this finding, Hofmeister & Norcliffe argued that resumptive pronouns make comprehension easier than gapped sentences do in situations where processing pressure is high (e.g., dependencies with three embeddings). Similar results were also reported in Dickey (1996), in which RPs were found to speed up online reading times for sentences with multiple-embeddings. Experiments on other languages also obtained similar findings: Ning (2008) found that in Mandarin Chinese, RPs in more deeply embedded contexts (e.g., indirect object relative clauses) were read faster than those in simpler contexts (e.g., subject and direct object relative clauses).

Besides reading time measures, the benefit of RPs has also been shown in forced choice tasks. Ackerman et al. (2014) tested the effect of RPs using two forced choice tasks. In one task, participants were asked to choose the more acceptable option between two given sentences, one with a RP and the other one with a gap. In the other task, participants were given an incomplete sentence and asked to complete it by selecting either a segment containing a gap or a segment containing a resumptive. A range of constructions was tested, including wh-islands, adjunct islands, relative clause islands and also their non-island counterparts. Across all types of islands, RPs were found to be more preferred than gaps.

As a whole, these results show that the facilitation effect of RPs surfaces in controlled experiments once standard acceptability judgment is removed from the task. Hofmeister & Norcliffe (2013) argued for a direct link between comprehension and the facilitation effect of RPs, and suggested that the failure to find a facilitation effect of RPs in previous acceptability judgment studies is due to "the lack of measurements of comprehension difficulty." The reading time results reviewed above are certainly consistent with this hypothesis. Even for the force-choice task employed in Ackerman et al. (2014), it could also be argued that when participants were forced to choose between two given options (i.e. they were not given the option that "neither is acceptable"), they could resort to all possible dimensions of comparison, including comprehensibility, in order to make a response.

To further pin down the relationship between resumption and comprehensibility, the current paper explores whether a minimal change on the original acceptability judgment task, i.e., a comprehensibility task, is sufficient to bring out the facilitation effect of RPs. If participants' judgments can be modulated by whether they are asked to judge the "acceptability" or the "comprehensibility" of a sentence, it provides strong empirical support to the hypothesis that intrusive RPs, although ungrammatical, can indeed facilitate the comprehension process. We present a total of four experiments below, one in Italian (Experiment 1) and three in English (Experiments 2–4). Our investigation will focus on the two well-known "RP-friendly" environments: relative clause islands and sentences with multiple embeddings.

3 Experiments
3.1. Experiment 1

Experiment 1 investigated RPs in Italian. From a typological perspective, Italian patterns with English with respect to resumption: while RPs are not grammatical in regular dependencies (as in (5a)), they have been reported to improve the status of sentences containing island violations (see the different status of (5b) and (5c); examples and

judgments are from Belletti 2006). It has also been pointed out (Belletti 2006) that resumption in Italian, albeit ungrammatical, is particularly frequent in informal and colloquial registers.

(5) a. *L'   uomo   che   lo   arresteranno   se   continua   così
         The   man   that   him   arrest-FUT-3PL   if   continue-3SG so
         'The man that they will arrest if he goes on like that' (RP outside of island)

     b. *L'   uomo   che   temo   il   pericolo   che   arresteranno
         The   man   that   fear.1P the   danger   that arrest-FUT-3PL
         'The man that they will arrest if he goes on like that' (gap in island)

     c. (?)?L'   uomo   che   temo   il   pericolo   che   lo   arresteranno
         The   man   that   fear.1SG the   danger   that him   arrest-FUT-3PL
         'The man that they will arrest if he goes on like that' (RP in island)

Experiment 1 introduced two important design features of the current study. First, we explicitly asked subjects to focus on the *comprehensibility* of the target sentence, as opposed to its *acceptability*. If RPs can facilitate the construction of a more coherent semantic interpretation, we expect to see such a facilitation effect emerge in the comprehensibility rating. Second, whereas previous acceptability judgment studies often presented the target sentence in isolation, we embedded the target sentence in a short conversation between two partners, such that the target sentence was always preceded by a context sentence.

3.1.1 Material

In a 2x2x2 factorial design we created 8 conditions, resulting from crossing the following three factors: a) Island, b) Resumption, and c) Embedding. For the *Island* factor, the experimental sentence was either a grammatical definite NP relative clause or an ungrammatical NP relative clause with an island violation. For the *Resumption* factor, the experimental sentence contained either a gap or a resumptive pronoun. For *Embedding*, the experimental sentence was presented either with two levels of embedding (2-level) or with three levels of embedding (3-level). Each item consisted of two sentences. The first one described a context and was the same across all of the conditions. The second sentence was framed as a natural continuation of the first one and was manipulated according to the factors above. The example in (6) demonstrates the full paradigm for an item.

(6) **Context sentence (same across conditions, spoken by a female voice):**

Ieri   ci   sono   stati   disordini per strada e   alcune persone
Yesterday   there   were   been   riots   for street and some   people
sono   rimaste   ferite   dalla   polizia
were   remained wounded   from.the   police
'Yesterday there were riots in the street, and some people were wounded by the police.'

Critical sentence (spoken by a male voice):

a. Questo   è   il   ragazzo   che   il   poliziotto   che
    This   is   the   guy   that   the   cop   that
    guidava le   operazioni ha   picchiato __
    led.PST the   operations has   beaten   __
    'This is the guy that the cop who was leading the operation beat up.'
    (Non-island, 2-level embedding, Gap)

b. Questo   è   il   ragazzo che   il   poliziotto che
    This   is   the   guy   that   the   cop   that
    guidava   le   operazioni l'   ha   picchiato.
    led.PST   the   operations him   has   beaten
    'This is the guy that the cop who was leading the operation beat him up.'
    (Non-island, 2-level embedding, RP)

c. Questo   è   il   ragazzo che   il   giornale riporta   che
    This   is   the   guy   that   the   paper   reports   that
    il   poliziotto che   guidava le   operazioni ha   picchiato __.
    the   cop   that   led.PST the   operations has   beaten   __
    'This is the guy that the paper reports that the cop who was leading the
    operation beat up.' (Non-island, 3-level embedding, Gap)

d. Questo è   il   ragazzo che   il   giornale riporta che
    This   is   the   guy   that the   paper   reports that
    il   poliziotto che   guidava le   operazioni l'   ha   picchiato.
    the   cop   that   led.PST the   operations him   has   beaten.
    'This is the guy that the paper reports that the cop who was leading the
    operation beat him up.' (Non-island, 3-level embedding, RP)

e. Questo   è   il   ragazzo che   il   poliziotto che
    This   is   the   guy   that   the   cop   that
    ha   picchiato __   deve   essere sospeso.
    has   beaten   __   must   be   suspended.
    'This is the guy that the cop who beat up must be suspended." (Island,
    2-level embedding, Gap)

f. Questo   è   il   ragazzo che   il   poliziotto che
    This   is   the   guy   that   the   cop   that
    l'   ha   picchiato   deve   essere sospeso.
    him   has   beaten   must   be   suspended.
    'This is the guy that the cop who beat him up must be suspended." (Island,
    2-level embedding, RP)

g. Questo   è   il   ragazzo che   il   giornale riporta che
    This   is   the   guy   that the   paper   reports that
    il   poliziotto che   ha   picchiato __ deve   essere sospeso.
    the   cop   that   has   beaten   __ must   be   suspended.
    'This is the guy that the paper reports that the cop who beat up must be
    suspended.' (Island, 3-level embedding, Gap)

h. Questo è   il   ragazzo che il   giornale riporta   che
    This   is   the   guy   that the   paper   reports   that
    il   poliziotto che   l'   ha   picchiato   deve   essere   sospeso.
    the   cop   that   him   has beaten   must   be   suspended.
    'This is the guy that the paper reports that the cop who beat him up must
    be suspended. (Island, 3-level embedding, RP)'

Sixty-four sets of items were created. These items were distributed into eight lists with a
Latin Square design, so that every subject was tested on only one condition for a given item.
We also created 40 additional fillers, which consisted of two sentences: the first sentence
provided a context, and the second one introduced a relative clause. All filler sentences

were grammatical. Every subject was tested on 104 items total. In order to increase the naturalness of the task, we presented the stimuli auditorily, instead of in written form. All the items were first recorded by two native speakers of Italian (a man and a woman) from the same region as the subjects of the experiment.[4] In this way, the participants encountered an accent they were already fully familiar with, minimizing the disruption potentially generated by encountering accents associated with geographically distant areas of the country. In addition, to make the interaction as natural as possible, we explicitly asked our two speakers not to conceal their accents, and to read the sentences with a similar prosody to the one that they would use in an informal conversation with their peers. The context sentence was always read by the woman, while the target sentence was read by the man. In addition to the experimental items and the fillers, there were also six practice items.

3.1.2 Participants

Forty-three participants participated in the study. To ensure that the subject pool was dialectally homogenous, all the subjects were recruited from the northern Italian region of Lombardia. All subjects were between 18 and 40 years old and were either high school or college graduates.

3.1.3 Procedure and statistical analysis

At the beginning of the experiment, subjects read a paragraph (in Italian) on the monitor introducing the task of the experiment. They were told that they would listen to some short conversations between a girl named Cecilia and a man named Pietro, and their task was to judge the comprehensibility of the man's sentence after each conversation. For the rating task, the participants received the following instructions:

> "You will have to answer with a score ranging from 1 (the sentence is completely incomprehensible) to 7 (the sentence is perfectly comprehensible). We want you to judge these sentences based on how easy they are for you to understand".

After each trial, the participants received the following prompt:

How comprehensible is Pietro's sentence?

$$1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7$$

(– comprehensible   + comprehensible)

Each item (i.e., a mini-conversation) was presented auditorily, and participants could only listen to it once. Before the actual experimental session began, participants completed six practice trials with the same format. The test items and the fillers were presented in a randomized order.

For statistical analysis, we first z-transformed all the raw ratings of each individual subject, and then ran a mixed-effects model on the transformed data with the R statistical package lme4 (Bates et al. 2014).[5] Data analysis on all subsequent experiments also followed this procedure. The fixed effect predictors included Gap, Embedding, Island and their interactions, and the random effects included at least random intercepts for subjects and items. Random slopes were also included whenever the resulting model could converge.

---

[4] The speakers were from the town of Bergamo, while the participants were from the town of Sondrio. The two towns are located less than 50 miles apart from each other in the northern part of Lombardia, a region in northern Italy.

[5] Including or excluding filler trials in the calculation of z-scores did not change the results of the statistical analyses. The results reported here were based on ratings derived from only the experimental trials, excluding the filler trials.

All predictors were sum coded before the data analysis, with island, three-embedding, and gap coded as 1, and non-island, two-embedding and RP coded as -1.

3.1.4 Results

Raw rating results from Experiment 1 are presented in Figure 1. Mixed effects models show a significant main effect of Island ($\beta=-0.36$, se$=0.03$, p<0.0001), reflecting the fact that all island conditions are rated significantly lower than their non-island counterparts (gaps: 2-level embedding, $\beta=-1.45$, se$=0.1$, $p<0.0001$, 3-level embedding, $\beta=-0.67$, se$=0.1$, $p<0.0001$; RPs: 2-level embedding, $\beta=-0.45$, se$=0.09$, $p<0.0001$, 3-level embedding, $\beta=-0.34$, se$=0.09$, $p<0.001$). Since we are most interested in whether RPs can rescue island violations, we report the results for islands and non-islands conditions separately below. The effect of embedding is also assessed below for islands and non-islands separately.

3.1.4.1 Without island

When there are no island violations, the effect of Gap and Embedding are both significant (Gap: $\beta=0.18$, se$=0.04$, $p<0.001$; Embedding: $\beta=-0.17$, se$=0.03$, $p<0.0001$). But crucially the interaction between the two is also highly significant ($\beta=-0.19$, se$=0.03$, $p<0.0001$). Although gaps are (unsurprisingly) more comprehensible than RPs with shorter embeddings ($\beta=0.74$, se$=0.1$, $p<0.0001$), the difference disappears with 3-level embedding ($\beta=-0.02$, se$=0.09$, $p>0.8$) due to the fact that more embedding reduces the comprehensibility of gapped conditions ($\beta=-0.72$, se$=0.1$, $p<0.0001$), but RPs were not affected by more embeddings ($\beta=0.04$, se$=0.09$, $p>0.6$).

3.1.4.2 With island

For conditions with islands, the most striking effect is a significant main effect of Gap ($\beta=-0.15$, se$=0.04$, $p<0.01$), reflecting that RPs are rated significantly higher than gaps across both 2 and 3-level embeddings (2-embedding, $\beta=0.26$, se$=0.1$, $p<0.01$; 3-embedding, $\beta=0.35$, se$=0.1$, $p<0.001$). We also observe that, the depth of embedding does not have any effect ($\beta=0.05$, se$=0.04$, $p>0.2$), nor is there an interaction between Gap and Embedding ($\beta=-0.02$, se$=0.03$, $p>0.4$), reflecting that longer embedding does not change the ratings for gaps or RPs within island. This is particularly interesting for the gap conditions in light
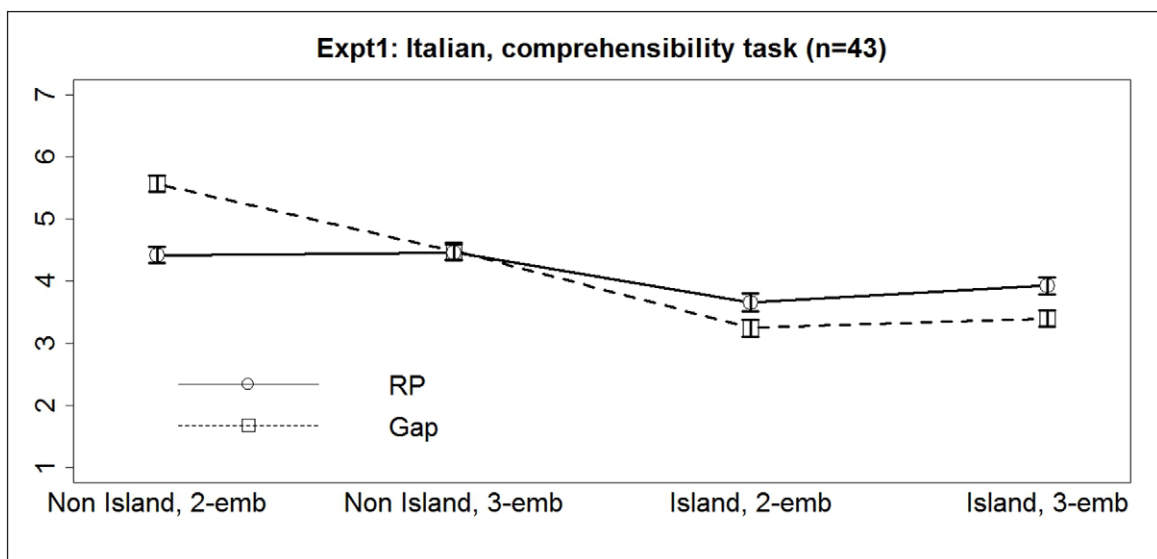
Figure 1: Comprehensibility judgments on a 1-7 scale for Experiment 1. The y-axis indicates average ratings. Error bars indicate standard errors.

of the finding that longer embedding does significantly reduce the comprehensibility of gaps without islands. We come back to this point in the Discussion section.

3.1.5 Discussion of Experiment 1

The most salient result from this first experiment is that, within islands, RPs are rated higher than gaps. Such an effect provides evidence that intrusive RPs do indeed help comprehension in cases where a syntactic violation complicates the overall processing of the sentence. This finding thus constitutes a crucial difference with respect to those of previous experiments, in which RPs were never rated better than gaps in islands.

With respect to embedding, grammatical dependencies with gaps receive reduced ratings with 3-level embeddings, but embedding does not seem to have an effect on the comprehensibility of resumptives – in complex 3-embedding dependencies, RPs are never rated higher than gaps, nor are they rated higher than 2-embedding sentences with RPs.

Overall, these results provide evidence that Italian RPs within islands do facilitate processing. This confirms our hypothesis that RPs could improve comprehensibility of island violations, in contrast to previously reported acceptability results on English RPs. However, this conclusion may be questioned on the ground that the observed effect may be due to some special properties of Italian RPs, rather than reflecting a more general rescuing effect of RPs. For example, Italian resumption, albeit ungrammatical in relative clauses, is a strategy that independently exists in the grammar. In particular, Italian requires the presence of resumptive clitic in Clitic Left Dislocation (CLLD: Cinque 1990; Belletti 2006) – a particular kind of unbounded dependency, as shown in (7).[6]

(7)   a.   Mario   lo   ho   visto   domenica.
               Mario   him   have   seen   Sunday.
               'Mario I saw him on Sunday'

        b. ?? Mario   lo   ho   visto   domenica.
               Mario   him   have   seen   Sunday.
               'Mario I saw him on Sunday'

It is possible that the presence of a resumptive structure in the grammar might lead Italian speakers to be less biased against sentences with RPs across the board, putting the experimental subjects in a position to more easily perceive the processing facilitation effect of RPs in islands. To assess whether the observed facilitation effect of RPs could be generalized, we performed the same task on English RPs in Experiment 2.

3.2 Experiment 2
3.2.1 Materials, design and procedure

The design was identical to Experiment 1. The procedure was largely the same as Experiment 1, but with two modifications. First, Experiment 2 was carried out on Amazon Mechanical Turk; and second, all materials were translated from the Italian stimuli in Experiment 1 and were presented in the written rather than auditory format. Participants first read a context sentence, then the target sentence, and finally were asked to judge how comprehensible the target sentence was on a scale from 1 to 7. Fifty-two self-reported native English speakers participated in the experiment (between 18–35 years old). Only

---

[6] Belletti (2006) analyzed this construction in terms of movement of a constituent to the left periphery of the sentence, with a clitic RP (*lo* in (7a)) that phonologically realizes the trace of the dislocated constituent. The only available reading of (7b) is one with contrastive focus on the dislocated constituent, such as "MARIO (and not Gianni), ho visto domenica", translated as 'MARIO (and not Gianni) I saw on Sunday'.

subjects with a US IP address were allowed to participate. For the comprehensibility rating task, we gave participants the same instruction as Experiment 1:

> "You will have to answer with a score ranging from 1 (the sentence is completely incomprehensible) to 7 (the sentence is perfectly comprehensible). We want you to judge these sentences based on how easy they are for you to understand."

An example (with all conditions) participants received is given in (8):

(8) An example trial:

(Context)
Have you heard?   Yesterday there were riots in the streets.   Some people were wounded. Look here, they're talking about it in the paper.

(Target sentence)
a. This is the boy that the cop who was leading the operation beat up. (Non island, 2-level embedding, Gap)
b. This is the boy that the cop who was leading the operation beat him up. (Non island, 2-level embedding, RP)
c. This is the boy that the newspaper reports that the cop who was leading the operation beat up. (Non island, 3-level embedding, Gap)
d. This is the boy that the newspaper reports that the cop who was leading the operation beat him up. (Non island, 3-level embedding, RP)
e. This is the boy that the cop who beat up was leading the operation. (Island, 2-level embedding, Gap)
f. This is the boy that the cop who beat him up was leading the operation. (Island, 2-level embedding, RP)
g. This is the boy that the newspaper reports that the cop who beat up was leading the operation. (Island, 3-level embedding, Gap)
h. This is the boy that the newspaper reports that the cop who beat him up was leading the operation. (Island, 3-level embedding, RP)

"How comprehensible is the last sentence?"

3.2.2 Results

The rating results are overall similar to the Italian results in Experiment 1 (Figure 2). We found a main effect of Island ($\beta = -0.12$, se $= 0.02$, $p < 0.0001$). However, the effect of Island is also modulated by the gap/RP difference. Although island conditions with gaps are rated significantly lower than their non-island counterparts (2-level embedding, $\beta = -0.79$, se $= 0.09$, $p < 0.0001$; 3-level embedding, $\beta = -0.42$, se $= 0.09$, $p < 0.0001$), ratings on RPs were not different between island and non-island conditions (2-level embedding, $\beta = 0.12$, se $= 0.09$, $p > 0.1$, 3-level embedding, $\beta = 0.13$, se $= 0.09$, $p > 0.1$). Like Experiment 1, we report below more detailed analyses for islands and non-islands conditions separately.

3.2.2.1 Without island

In the absence of island violations, gaps are more comprehensible than RPs, as reflected by a main effect of Gap ($\beta = 0.2$, se $= 0.05$, $p < 0.0001$). Paired comparisons between gaps and RPs confirm that gaps are more comprehensible than RPs for both the shorter and the longer embeddings (2-level embedding: $\beta = 0.52$, se $= 0.09$, $p < 0.0001$; 3-level embedding: $\beta = 0.33$, se $= 0.09$, $p < 0.001$). In addition, there is a main effect of Embedding ($\beta = -0.12$, se $= 0.03$,
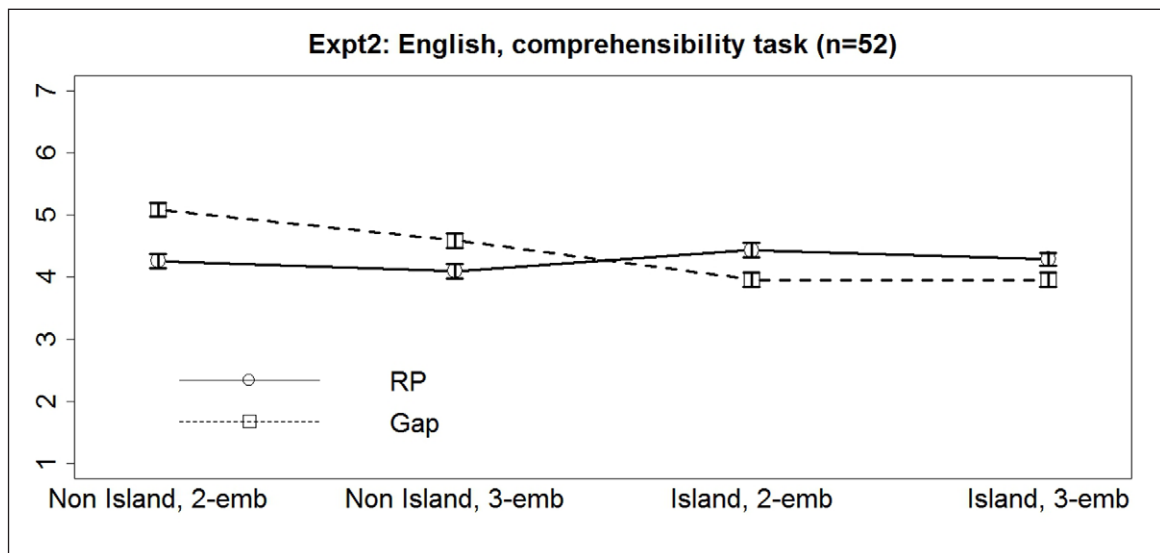
Figure 2: Comprehensibility judgments for Experiment 2. Error bars indicate standard errors.

$p < 0.01$). Longer embedding reduced the comprehensibility of gaps ($\beta = -0.35$, se $= 0.09$, $p < 0.001$), and there is a similar but only marginal effect on RPs ($\beta = -0.16$, se $= 0.09$, $p < 0.09$). There is no interaction between Gap and Embedding ($\beta = -0.04$, se $= 0.03$, $p > 0.1$).

3.2.2.2 With island

Within islands, RPs are once again more comprehensible than gaps across both 2 and 3-level embeddings (2-embedding, $\beta = 0.38$, se $= 0.09$, $p < 0.0001$; 3-embedding, $\beta = 0.22$, se $= 0.09$, $p < 0.05$). We also observe that within islands, the depth of embedding does not have any effect ($\beta = -0.03$, se $= 0.03$, $p > 0.2$), nor is there an interaction between Gap and Embedding ($\beta = 0.04$, se $= 0.03$, $p > 0.2$).

3.2.3 Discussion of Experiment 2

Experiment 2 replicated the "comprehensibility rescuing effect" of RPs on syntactic islands in English with a comprehensibility judgment task. With respect to embedding, the pattern is also similar to Experiment 1. First, gaps in grammatical dependencies (i.e., without islands) are considerably less comprehensible with three embeddings; but the effect of embedding disappears for gaps in ungrammatical dependencies (i.e., within islands). Second, embedding only has a very weak effect on the comprehensibility of RPs, with three embeddings showing marginally lower ratings than two embeddings.

Overall, the observations above suggest that the findings of Experiment 1 were not due to language specificities of Italian, but to more general properties of intrusive resumption and the particular comprehensibility judgment task we employed. For both languages, we showed that when participants' attention is focused on assessing the comprehensibility difficulty, and when the test sentences are preceded by a context sentence, RPs can indeed facilitate processing. In the next two experiments, we aim to assess the impact of each of the two factors separately: Experiment 3 replaces the comprehensibility judgment with an acceptability judgment task, while Experiment 4 tests for comprehensibility without a context sentence.

3.3 Experiment 3
3.3.1 Materials, design and procedure

The design, stimuli, and procedure of Experiment 3 were identical to Experiment 2, except that acceptability judgments were elicited rather than comprehensibility judgments. All

target sentences were still preceded by a context sentence. Thirty-six subjects participated in the experiment. We present below the specific instruction to the participants prior to the whole experiment:

  Instruction to the participants:

  "You will be given 104 short paragraphs, each of which contains 2–3 sentences. After each paragraph, you will have to answer the following question:

  How acceptable is the last sentence of each paragraph?

  You have to answer this question with a score ranging from 1 (= the sentence is completely unacceptable) to 7 (the sentence is perfectly acceptable). Please make your judgments based on how good the last sentence sounds in English given the context it is in."

3.3.2 Results

The results are presented in Figure 3. The effect of Island is significant ($\beta = 0.28$, se $= 0.08$, $p < 0.001$). But this effect is again modulated by the gap/RP difference. Similar to Experiment 2, although island conditions with gaps are rated significantly lower than their non-island counterparts (2-level embedding, $\beta = -1.01$, se $= 0.1$, $p < 0.0001$; 3-level embedding, $\beta = -0.28$, se $= 0.09$, $p < 0.001$), ratings on RPs were not different between island and non-island conditions (2-level embedding, $\beta = 0.09$, se $= 0.08$, $p > 0.2$; 3-level embedding, $\beta = 0.08$, se $= 0.08$, $p > 0.3$). Separate analyses for islands and non-islands conditions are presented below.

3.3.2.1 Without island

Outside of islands, gaps are more acceptable than RPs with both shorter ($\beta = 0.98$, se $= 0.09$, $p < 0.00001$) and longer embedding ($\beta = 0.34$, se $= 0.09$, $p > 0.7$, $p < 0.001$). Furthermore, there is an interaction between Gap and Embedding ($\beta = -0.16$, se $= 0.02$, $p < 0.0001$). Longer embedding reduced the acceptability of gaps ($\beta = -0.60$, se $= 0.07$, $p < 0.00001$), but not of RPs ($\beta = 0.04$, se $= 0.07$, $p > 0.6$).

3.3.2.2 With island

Crucially different from Experiment 1 and 2, within islands, there is no difference between RPs and gaps, across both 2 and 3-level embeddings (2-level embedding: $\beta = 0.03$, se $= 0.09$,
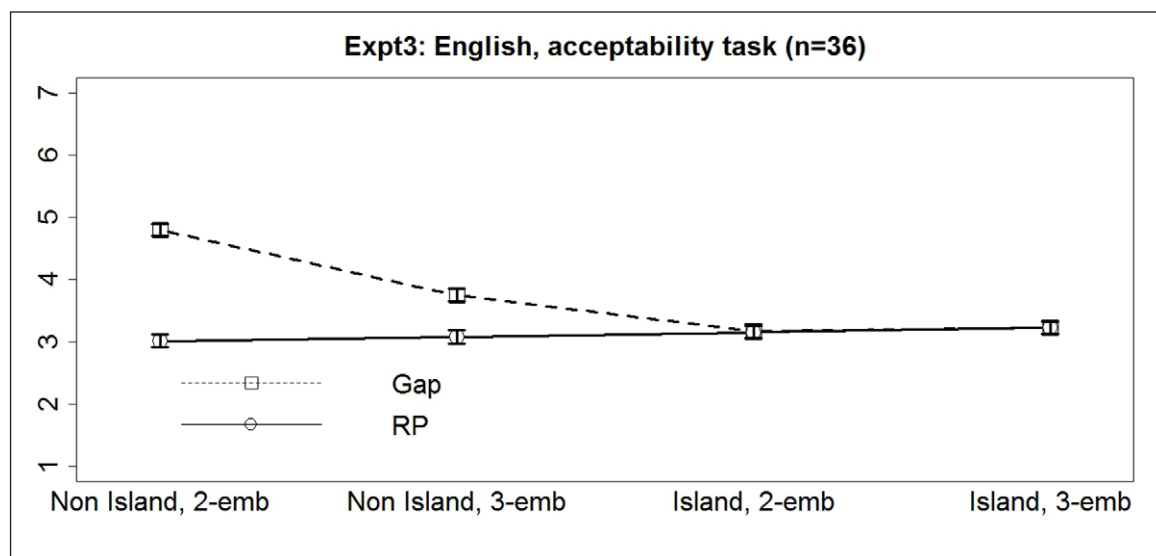
Figure 3: Acceptability judgments for Experiment 3. Error bars indicate standard errors.

$p > 0.7$; 3-level: $\beta = 0.03$, se $= 0.09$, $p > 0.7$). In addition, we once again observe that the depth of embedding does not have any effect ($\beta = 0.01$, se $= 0.03$, $p > 0.5$), nor is there an interaction between Gap and Embedding ($\beta = -0.003$, se $= 0.02$, $p > 0.9$).

3.3.3 Discussion of Experiment 3

The crucial observation from Experiment 3 is that, when the experimental task is changed to acceptability judgments, RPs in islands are no longer rated higher than gaps. This result is in line with previous acceptability judgment studies in the literature (see Section 2.2). This also constitutes a crucial difference from the findings of Experiments 1–2, in which the ratings of RPs in islands were better than those of gaps under a comprehensibility task.

3.4 Experiment 4
3.4.1 Materials, design and procedure

In order to evaluate the importance of the context sentence, in the final study, we restored comprehensibility instructions and presented the target sentence in isolation, with no context sentence introducing it. The design and procedure were otherwise identical to Experiment 2. Thirty-six participants participated in the study.

3.4.2 Results

The results are presented in Figure 4. There is an effect of Island ($\beta = 0.59$, se $= 0.08$, $p < 0.0001$), and this effect is once again modulated by the gap/RP difference. Island conditions with gaps are rated significantly lower than their non-island counterparts (2-level embedding, $\beta = -1.18$, se $= 0.08$, $p < 0.0001$; 3-level embedding, $\beta = -0.59$, se $= 0.08$, $p < 0.0001$), but ratings on RPs were not different between island and non-island conditions (2-level embedding, $\beta = 0.01$, se $= 0.07$, $p > 0.8$; 3-level embedding, $\beta = 0.008$, se $= 0.07$, $p > 0.9$). Separate analyses for island and non-island conditions are presented below.

3.4.2.1 Without island

Outside of islands, gaps are more comprehensible than RPs with both shorter ($\beta = 1.09$, se $= 0.08$, $p < 0.0001$) and longer embedding ($\beta = 0.48$, se $= 0.08$, $p < 0.0001$). In addition, there is an interaction between Gap and Embedding ($\beta = -0.16$, se $= 0.03$, $p < 0.0001$), reflecting the fact that longer embedding reduced the comprehensibility of gaps ($\beta = -0.63$, se $= 0.08$, $p < 0.0001$), but not of RPs ($\beta = 0.02$, se $= 0.08$, $p > 0.8$).
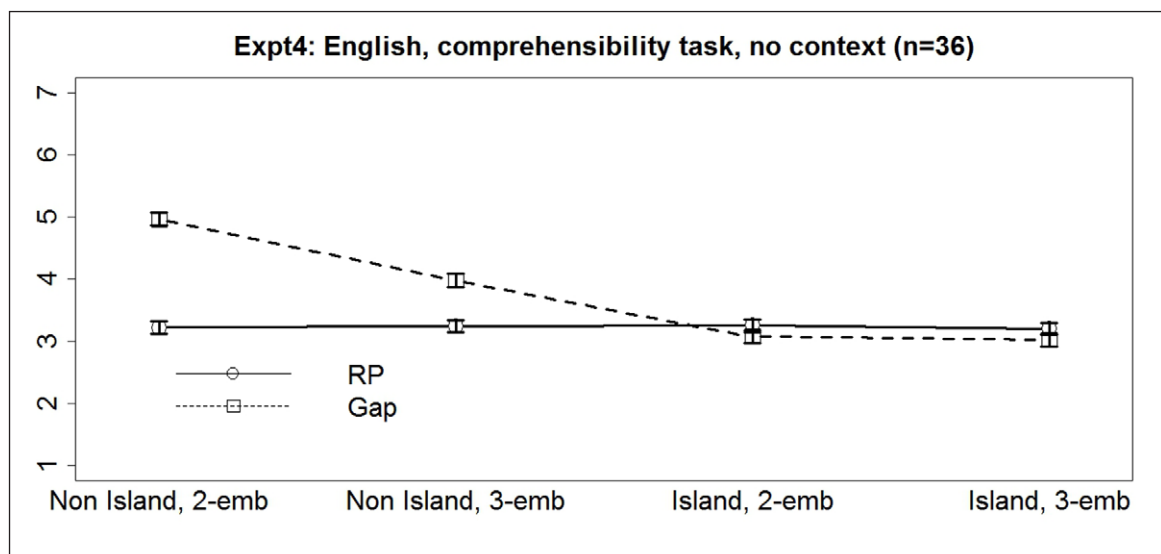
Figure 4: Comprehensibility judgments for Experiment 4. Error bars indicate standard errors.

3.4.2.2 With island

Although a trend can be observed in which RPs are more comprehensible than gaps with both 2 and 3-level embeddings, the difference is not statistically significant (2-level embedding: $\beta = 0.09$, se $= 0.07$, $p > 0.2$; 3-level embedding, $\beta = 0.12$, se $= 0.08$, p > 0.1). We also observe that within islands, the depth of embedding does not have any effect ($\beta = -0.01$, se $= 0.03$, p > 0.8), nor is there an interaction between Embedding and Gap ($\beta = -0.007$, se $= 0.03$, p > 0.8).

3.4.3 Discussion of Experiment 4

In Experiment 4, we restored the comprehensibility task while eliminating the initial context sentence. As in Experiment 3, no significant difference between RPs and gaps in islands was found, although there is a trend for RPs to be rated higher than gaps. Experiment 3 and 4 together show that both a comprehensibility task and a context sentence are important in order for the facilitation effect of RPs to emerge in the presence of islands. Regarding embedding, once again the depth of the dependency only had an effect on sentences with gaps, and only when no island violation was present.

3.5 Interim summary

In four experiments, we compared gaps and resumptive pronouns with respect to two linguistic manipulations: the presence/absence of islands, and the number of embedded clauses. We also compared two different experimental tasks: a comprehensibility task and a more traditional acceptability task. Two main observations emerged from our results. First, resumptive pronouns did turn out to rescue islands to some degree, confirming the previously reported introspective judgments. Crucially, however, such effects only emerged with a comprehensibility task and with the presence of a context sentence that we predicted would facilitate the retrieval of an antecedent. Second, the number of embeddings, which modulates the processing difficulty of long distance dependencies, only affected grammatical dependencies with gaps (i.e. with no islands and no RPs), in terms of both acceptability and comprehensibility judgments. However, it did not show a significant influence on sentences with islands or RPs. In the discussion below we assess the implications of these findings.

4 General discussion
4.1 Acceptability, comprehensibility, and introspective judgments

The first important finding from our results is that comprehensibility ratings better quantified the previously reported introspective judgments from professional linguists than acceptability ratings. This raises an important question about the effect of tasks in obtaining metalinguistic judgments. The traditional acceptability judgment task focuses a speaker's attention to the overall naturalness of a sentence, an important component of which is the syntactic well-formedness (i.e., grammaticality). Acceptability ratings, therefore, are largely determined by the syntactic form of a sentence (Sprouse and Almeida 2012), although it is also well documented that other factors, such as processing complexity, could influence the outcome of these judgments (e.g., Chomsky & Miller 1963; Kluender 1992; Hofmeister, Staum Casasanto and Sag 2014). It has long been recognized by linguists that intrusive resumptive pronouns are not grammatical in English (Kroch 1981; Prince 1990; Erteschik-Shir 1992; Asudeh 2004; 2011); therefore it should not be surprising that the amelioration effects of RPs reported in previous introspective judgments could not be detected by the acceptability judgment task. The comprehensibility task we adopted, on the other hand, shifted the speaker's attention from judging the overall naturalness of a sentence to a narrower focus of assessing whether and how easily a given sentence is interpretable (see more qualification below). This task is better suited to capture the amelioration effects of RPs.

Before we discuss further how resumptive pronouns help with the comprehension process, it is important to note that our results do not imply that just about any ungrammatical sentence can be perceived to have an improved status as long as speakers can somehow make sense of it. Particularly pertinent to this discussion are the experimental results from Maclay & Sleator (1960). In that study, participants gave judgments to sentences under three different types of task instructions, and we discuss two of them here, which are most relevant for the current purpose. Under one task, participants judged whether a given string of words formed a "grammatical" English sentence; in a different task, participants judged whether the same string of words formed a "meaningful" English sentence. Under both tasks, participants gave gradient judgments to different kinds of stimuli that were constructed based on syntactic well-formedness and semantic meaningfulness, and there was also a task effect on some of the stimuli types. Since the sentence stimuli in that study were not parallel to the ones we used in the current study, a direct comparison is not possible, but it is crucial to note that ungrammatical sentences like "Yesterday I the child the dog gave" received almost identical ratings both in terms of "grammaticality" and "meaningfulness" judgments (both at 26% "Yes" responses, see Table IV in Maclay & Sleator 1960). This suggests that the mere possibility of constructing a sensible interpretation out of an ungrammatical sentence, which in the case above is based on speakers' real world knowledge, is not sufficient to boost its comprehensibility rating (assuming the "meaningfulness" judgment is similar to the comprehensibility judgment in the current study). Given these considerations, we want to emphasize that the improved comprehensibility ratings of RPs (over gaps) observed in the current study, and the amelioration effects of RPs reported previously by trained linguists, were not reflecting just *any* kind of sensicality or plausibility judgments. We instead argue that RPs, being anaphoric, aid parsing in very particular and yet principled ways. More specifically, they help to construct a locally coherent parse, and they also help to retrieve the left-hand side of a non-local dependency. Both of these effects fit into a larger picture of standard parsing procedures. We elaborate on them in the sections below.

4.2 The processing facilitation effect of resumptive pronouns in islands

In this section we discuss the facilitation effect observed for RPs within islands. One possibility, as argued by Asudeh (2004; 2011; 2012; see Section 2.3), is that RPs can facilitate the comprehension of an utterance through assuring that the sentence is locally well-formed. Crucially, Asudeh's model separates the parsing benefits of RPs from the grammaticality of the construction, suggesting that RPs, while unable to render the structure grammatical, can at least facilitate the construction of a well-formed local parse. We note that the local coherence effect is not limited to Asudeh's model and the phenomenon of resumption. It is well known, for instance, that a coherent local parse can sometimes affect performance independent of the global parse. Tabor et al. (2004) showed that speakers were distracted by the presence of a locally coherent string when interpreting a globally difficult structure. In the sentence "The coach smiled at the player tossed a frisbee", for instance, the local structure "the player tossed a frisbee" should be parsed as a reduced relative clause (i.e. "the player <who was> tossed a frisbee"). However, it was shown that participants tend to parse the local string "the player tossed…" as a subject-verb structure, possibly because of the overwhelming parsing complexity at the global level.[7]

In addition to helping with the local parse, RPs can also aid the dependency formation between a "filler" – which is looking for a gap – and the argument position that the RP occupies. First of all, in a complex syntactic structure, such as syntactic islands, it may

---

[7] However, Tabor et al. (2004) cautiously pointed out that it was unclear whether the locally coherent parse is the cause or the consequence of the global parsing complexity.

be relatively difficult to identify where the tail of a dependency is. But a resumptive pronoun provides a clear perceptual cue for that. Second, compared to gaps, resumptive pronouns also provide explicit morphological cues, such as information about the animacy, gender, number, and person features of the antecedent, which can guide the retrieval of the appropriate antecedent more. Such cue-based retrieval mechanisms in pronouns fit into a more general memory retrieval architecture that accounts for a number of other phenomena in sentence processing (e.g. Van Dyke & Lewis 2003; Lewis & Vasishth 2005; Wagers et al. 2009). Gaps, on the other hand, provide little information beyond the verb subcategorization cues to help identify the appropriate antecedents. The processing difference between gaps and pronouns discussed here may also underlie some intuitions suggested in the previous proposals pertaining to the linguistic difference between gaps and RPs. For example, many previous proposals have suggested that while gaps are bound variables, intrusive RPs are anaphorically linked to their antecedents (Chao & Sells 1983; Prince 1990; Erteschik-Shir 1992; Alexopoulou & Keller 2007; Clemens et al. 2012; Han et al. 2012), and this difference is somewhat responsible for the fact that gaps are more sensitive to syntactic islands, whereas RPs can find their contextually salient discourse antecedents despite the intervening island boundaries (e.g. Clemens et al. 2012).

The anaphoric status of RPs also explains why the presence of a context sentence in our experiments had a significant impact on the comprehensibility ratings of RPs. In about one quarter of the experimental stimuli, the antecedent is directly mentioned in the context sentence. Two examples are given in (9) and (10). In examples like this, the context sentence serves to boost the salience of the relevant antecedent, making it more accessible for an anaphoric expression (Ariel 1990; Erteschik Shir 1992; Gundel et al. 1993; Roberts 2010).

(9)   Context: In the high school where I graduated, a **janitor** suddenly decided that he wanted a better education.
        Sentence: This is the **janitor** that the teacher who tutors **him** is really nice.
(10)   Context: The newly graduated mechanical **engineers** have gone through a series of job interviews, and some of them already received good news.
        Sentence: This the **engineer** that the manager who hired **him** has shown to trust young people.

For the majority of the experimental items, the context sentence did not directly mention the antecedent, as shown in the examples (11) and (12) (with (11) reproduced from (6)):

(11)   Context: Yesterday there were riots in the street, and some people were wounded by the police.
        Sentence: This is the **guy** that the cop who beat **him** up must be suspended.
(12)   Context: In track and field, someone always tries to cheat.
        Sentence: This is the **runner** that the umpire who disqualified **him** behaved very professionally.

In these examples, even though the antecedent for the pronoun wasn't explicitly mentioned in the context, the context sentence nevertheless sets up a situation model for the listener, which aids the memory maintenance of the antecedent-pronoun relationship. This is in line with Ariel's (1990) claim that the task of retrieving an antecedent is easier whenever the pronoun and the antecedent are part of a frame that is known and well-defined. It is also possible that an explicit background context can facilitate the memory encoding of

an antecedent, boosting its degree of discourse familiarity. A number of researchers (e.g. Ariel 1990; Roberts 2010) have argued that the more familiar the antecedent is, the easier it is to retrieve the antecedent later.

4.3 Resumptives under longer dependencies

Although the processing facilitation effect of RPs in syntactic islands is relatively clear in our results, their facilitation effect in longer dependencies (e.g., multiple embeddings) is not very robust. This finding is by and large consistent with previous studies that have experimentally examined the effect of RPs in structures with multiple embeddings (see Section 2.2). In this section we discuss some possible reasons for this result.

Generally speaking, it is well-established that dependency length has an effect on processing complexity: Longer dependencies are generally more difficult to process than shorter ones, as reflected in both offline and online measures of processing complexity (Gibson 1998; Warren & Gibson 2002; Van Dyke & Lewis 2003; Lewis & Vasishth 2005; Lewis et al. 2006). There are a number of possible underlying sources for the length effect. The memory representation of the retrieval target (e.g., the filler) could have decayed over a long period of time; more linguistic material introduced by the longer dependency could increase the likelihood of similarity-based interference (e.g., the features on the retrieval target are shared by some other entities in working memory), or semantic integration over longer distance and more linguistic material could be more costly than the integration of a simpler dependency. All of these possibilities could overload the parser and result in higher processing difficulty, making the construction of a coherent message difficult.

If resumptives can aid the processing of complex dependencies via facilitating comprehensibility, one may expect that RPs in longer dependencies would result in higher comprehensibility than gaps, or RPs in longer dependencies should receive higher comprehensibility rating than RPs in shorter dependencies. However, neither prediction was completely borne out in the current results: in grammatical dependencies, RPs in longer dependencies did not receive higher comprehensibility ratings than gaps; and RPs in sentences with three-embeddings received the same, but not higher, comprehensibility ratings as RPs in sentences with two-level embeddings. We discuss below a number of possibilities that could explain these results.

The first consideration concerns our stimuli. The current design only compares sentences with 3-level and 2-level embeddings. Since these two conditions only differ for one level of embedding, the facilitation effect of RPs may not be detectable. It is possible that the benefit of RPs on comprehensibility becomes observable only when there is a larger difference in embedding, as shown by the original example in Erteschik-Shir (1992) (see the example in (3)). The results from Alexopoulou and Keller (2007) also showed that the largest effect of embedding on resumption was observed when zero-embedding was compared to other levels of embedding.

It should also be pointed out that while the longer dependencies with RPs were not rated more comprehensible than the shorter ones, the lack of improvement can be reinterpreted as evidence for the processing facilitation effect of RPs. In three out of the four experiments (i.e. Experiment 1, 3, 4, but not Experiment 2), for the no-island proportion of the conditions, there is a robust interaction between Gap/RP and the level of embedding – while longer dependencies with gaps received lower ratings than short dependencies with gaps, the same degradation was not observed with RPs, suggesting a neutralization of the negative effect of embedding. This is in line with previous findings (e.g. Alexopoulou & Keller 2007; Han et al. 2012; Hofmeister & Norcliffe 2013) This interpretation, however,

should also be treated with some caution. It is possible that the stable low ratings of RPs across different number of embeddings may simply reflect a "floor effect" – that is, resumption in two and three-level embeddings are so bad that they are already at the bottom of participants' judgment scale.

Finally, rather than broadly stating that longer dependencies are always more costly than shorter ones, it is worth considering the exact source of complexity associated with dependency length. One of the major candidates discussed in the literature is the increasing likelihood of similarity-based interference when more material is introduced by longer dependencies (Lewis & Vasishth 2005; Lewis et al. 2006). Under this account, memory retrieval of a target is guided by a set of retrieval cues/features (e.g., an animate subject is being cued as a retrieval target). If the retrieval target shares features with other representations in the working memory, feature similarity among different representations will prevent the correct target from being retrieved accurately due to cue overload. Van Dyke & Lewis (2003) showed that sentences with similar length but different degrees of retrieval interference led to different processing complexity, suggesting that, rather than length per se, it is cue overload that caused difficulty for the parser. Coming back to the current study, it is possible that in some of our stimuli, the short and long conditions, though different in length, may not be different in overlapping cues. We illustrate this by the example in (6) above, repeated here as (13).

(13)   a. 2-embedding, non-island, RP
            This is the boy that the cop who was leading the operation beat *him* up.
        b. 3-embedding, non-island, RP
            This is the boy that the paper reports that the cop who was leading the operation beat *him* up.

At the RP "him", the parser is looking for a [+singular, +masculine, +animate] noun as the retrieval target. In the short condition (13a), in addition to the correct target "the guy", it also contains one interfering NP "the cop", which shares all the features with the right target and therefore is a serious competitor. The long condition (13b) contains two additional NPs "the cop" and "the paper," but "the paper" is at best a very weak competitor since it does not share many features with the retrieval target. In other words, although (13b) is longer than (13a), one does not necessarily expect processing complexity difference between the two. This could explain why we found no comprehensibility rating differences between long and short sentences with RPs.

5 Conclusion

To conclude, this paper provides novel empirical evidence that intrusive resumptive pronouns indeed can "rescue" syntactic islands, confirming previously reported introspective judgments by trained linguists. Yet, the rescuing effect is crucially not at the level of grammaticality or acceptability, but at the level of sentence comprehension/comprehensibility. We also argued that the facilitation effect of RPs on comprehension follows from the general parsing mechanisms that subserve sentence comprehension. Methodologically speaking, our findings contribute to a more nuanced understanding of the nature of different types of metalinguistic judgments. Behavioral judgments, such as acceptability judgments or truth value judgments, form the primary empirical base for linguistic theories. It is therefore of crucial interest for future research to be able to more precisely characterize the specific linguistic properties that each type of judgment task targets and what factors may influence these judgments.

Acknowledgments

Competing interests

The authors declare that they have no competing interests.

References

Ackerman, Lauren, Frazier Michael and Masaya Yoshida. 2014. Resumptive pronouns salvage island violations in forced-choice tasks. Poster presented at the CUNY Sentence Processing Conference.

Alexopoulou, Theodora & Frank Keller. 2007. Locality, cyclicity and resumption: At the interface between the grammar and the human sentence processor. *Language* 83. 110–160. DOI: http://dx.doi.org/10.1353/lan.2007.0001

Ariel, Miriam. 1990. *Accessing noun phrase antecedents*. London: Routledge.

Asudeh, Ash. 2004. Resumption as resource management. Palo Alto: Stanford University dissertation.

Asudeh, Ash. 2011. Local grammaticality in syntactic production. In Emily Bender and Jennifer Arnold (eds.), *Language from a cognitive perspective: Grammar, usage & processing. Studies in honor of Thomas Wasow.* Stanford: CSLI Publications.

Asudeh, Ash. 2012. *The logic of pronominal resumption.* Oxford: Oxford University Press. DOI: http://dx.doi.org/10.1093/acprof:oso/9780199206421.001.0001

Bates, Douglas, Martin Maechler, Ben Bolker and Steve Walker. 2014. lme4: Linear mixed-effects models using Eigen and S4. R package version.1–7.

Belletti, Adriana. 2006. Extending doubling to non-local domains: Complete vs. partial copying + deletion and related reconstruction issues. Unpublished Manuscript.

Bennett, Robert. 2008. English resumptive pronouns and the highest subject restriction: A corpus study. Paper presented at the Trilateral (TREND) Linguistics Weekend.

Chao, Wynn & Peter Sells. 1983. On the interpretation of resumptive pronouns. In Peter Sells et al. (eds.), *Proceedings of the 13th NELS Conference.*

Chomsky, Noam & George A. Miller. 1963. Introduction to the formal analysis of natural languages. In Robert R. Bush et al. (eds.), *Handbook of mathematical psychology*, 269–321. New York: Wiley.

Clemens, Lauren, Gregory Scontras & Maria Polinsky. 2012. English does not have resumptive: A cross-sentential account of English resumption. Talk presented at the 87th Meeting of the Linguistics Society of America.

Creswell, Cassandre. 2002. Resumptive pronouns, wh-island violations & sentence production. In *Proceedings of the Sixth International Workshop on Tree-Adjoining Grammar and Related Frameworks (TAG + 6),* 101–109. Venice: Università di Venezia.

Dickey, Michael Walsh. 1996. Constraints on the sentence processor and the distribution of RPs. In Michael Walsch Dickey and Susanne Tunstall (eds.), *Linguistics in the laboratory* 19, 157–192. Amherst, MA: GLSA.

Engdahl, Elizabeth. 1985. Parasitic gaps, resumptive pronouns and subject extraction. *Linguistics* 23. 23–44. DOI: http://dx.doi.org/10.1515/ling.1985.23.1.3

Erteschik-Shir, Naomi. 1992. Resumptive pronouns in islands. In Helen Goodluck and Michael Rochemont (eds.), *Island constraints. Theory, acquisition and processing*, 89–109.

Dordrecht: Kluwer Academic Publishers. DOI: http://dx.doi.org/10.1007/978-94-017-1980-3_4

Ferreira, Fernanda & Benjamin Swets. 2005. The production and comprehension of resumptive pronouns in relative clause island contexts. In Anne Cutler (ed.), *Twenty-first century psycholinguistics: Four cornerstones*, 101–122. Malway, NJ: Lawrence Erlbaum Associates.

Francis, Elaine J., Charles Lam, Carol Chun Zheng, John Hitz & Stephen Matthews. 2015. Resumptive pronouns, structural complexity & the elusive distinction between grammar and performance: evidence from Cantonese. *Lingua* 162. 56–81. DOI: http://dx.doi.org/10.1016/j.lingua.2015.04.006

Gibson, Edward. 1998. Linguistic complexity: locality of syntactic dependencies. *Cognition* 68. 100–176. DOI: http://dx.doi.org/10.1016/S0010-0277(98)00034-1

Gundel, Jeanette, Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69. 274–307. DOI: http://dx.doi.org/10.2307/416535

Han, Chung-hye, Noureddine Elouazizi, Christina Galeano, Emrah Gorgulu, Nancy Helberg, Jennifer Hinnell, Meghan Jeffrey, Kyeong-min Kim & Susannah Kirby. 2012. Processing strategies and resumptive pronouns in English. In Nathan Arnett and Ryan Bennett (eds.), *Proceedings of the 30th West Coast Conference on Formal Linguistics*, 153–161. Somerville, MA: Cascadilla Press.

Heestand, David, Ming Xiang & Maria Polinsky. 2011. Resumption still does not rescue islands. *Linguistic Inquiry 42*. 138–152. DOI: http://dx.doi.org/10.1162/LING_a_00032

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

Hofmeister, Philip, Laura Staum Casasanto and Ivan A. Sag. 2014. Processing effects in linguistic judgment data: (Super-)additivity and reading span scores. *Language and Cognition* 6. 111–145. DOI: http://dx.doi.org/10.1017/langcog.2013.7

Keffala, Bethany. 2011. Resumption and gaps in English relative clauses: Relative acceptability creates an illusion of 'saving'. In Chundra Cathcart et al. (eds.), *Proceedings of the 37th Annual Meeting of the Berkeley Linguistics Society* 140–154. DOI: http://dx.doi.org/10.3765/bls.v37i1.846

Keffala, Bethany and Grant Goodall. 2011. Do resumptive pronouns ever rescue illicit gaps in English? Poster presented at CUNY 2011 Conference on Human Sentence Processing, March 24–26, 2011, Stanford University

Kluender, Robert. 1992. Deriving islands constraints from principles of predication. In Helen Goodluck and Michael Rochemont (eds.), *Island constraints: Theory, acquisition and processing*, 223–258. Dordrecht: Kluwer. DOI: http://dx.doi.org/10.1007/978-94-017-1980-3_8

Kroch, Anthony. 1981. On the role of resumptive pronouns in amnestying island constraint violations. In Eric Williams et al. (eds.), *Proceedings of the 16th Chicago Linguistic Society (CLS 16)*, 125–135. Somerville, MA: Cascadilla Press.

Lewis, Richard L. & Shravan Vasishth. 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science* 29. 375–419. DOI: http://dx.doi.org/10.1207/s15516709cog0000_25

Lewis, Richard L., Shravan Vasishth & Julie A. van Dyke. 2006. Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science* 10. 447–454. DOI: http://dx.doi.org/10.1016/j.tics.2006.08.007

Maclay, Howard & Mary D. Sleator. 1960. Responses to language: Judgments of grammaticalness. *International Journal of Applied Linguistics* 26. 275–282. DOI: http://dx.doi.org/10.1086/464587

McCloskey, James. 2006. Resumption. In Martin Everaert and Henk van Riemsdijk (eds.), *Blackwell companion to syntax*, 94–115. Oxford: Blackwell.

McDaniel, Dana, & Wayne Cowart. 1999. Experimental evidence for a minimalist account of English resumptive pronouns. *Cognition* 70. B15–B24. DOI: http://dx.doi.org/10.1016/S0010-0277(99)00006-2

McKee, Cecile, & Dana McDaniel. 2001. Resumptive pronouns in English relative clauses. *Language Acquisition* 9. 113–156. DOI: http://dx.doi.org/10.1207/S15327817LA0902_01

Ning, Li Hsin, 2008. The grammar and processing of resumptive pronouns in Chinese relative clauses. Tapei: National Chengchi University Master Thesis.

Polinsky, Maria, Lauren Eby Clemens, Adam M. Morgan, Ming Xiang & David Heestand. 2013. Resumption in English. In Jon Sprouse and Norbert Hornstein (eds.), *Experimental syntax and island effects,* 341–360. Cambridge University Press. DOI: http://dx.doi.org/10.1017/cbo9781139035309.017

Prince, Ellen. 1990. Syntax and discourse: a look at resumptive pronouns. In Kira Hall et al. (eds.), *Proceedings of the 16th Berkeley Linguistics Society (BLS 16)*, 482–498. Somerville, MA: Cascadilla Press. DOI: http://dx.doi.org/10.3765/bls.v16i0.1719

Roberts, Craige. 2010. *Retrievability and incomplete descriptions*. Ms. The Ohio State University.

Ross, Haj. 1967. *Constraints on variables in syntax*. Cambridge: MIT dissertation.

Sells, Peter. 1984. *Syntax and semantics of resumptive pronouns.* Ahmerst: University of Massachusetts dissertation.

Shlonsky, Ur. 1992. Resumptive pronouns as a last resort. *Linguistic Inquiry* 23. 443–468.

Sprouse, Jon & Diogo Almeida. 2012. Assessing the reliability of textbook data in syntax: Adger's core syntax 1. *Journal of Linguistics* 48. 609–652. DOI: http://dx.doi.org/10.1017/S0022226712000011

Tabor, William, Bruno Galantucci & Daniel Richardson. 2004. Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language* 50. 355–370. DOI: http://dx.doi.org/10.1016/j.jml.2004.01.001

Tucker, Matthew, Ali Idrissi, Jon Sprouse & Diogo Almeida. 2016. Components of grammaticalized resumption as Island repair in Modern Standard Arabic. Poster presented at the 2016 Annual Meeting of the Linguistic Society of America. Washington, DC.

Van Dyke, Julie A. & Richard L. Lewis. 2003. Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language* 49. 285–316. DOI: http://dx.doi.org/10.1016/S0749-596X(03)00081-0

Wagers, Matt, Ellen Lau & Colin Phillips. 2009. Agreement attraction in comprehension: representations and processes. *Journal of Memory and Language* 61. 206–237. DOI: http://dx.doi.org/10.1016/j.jml.2009.04.002

Warren, Tessa & Edward Gibson. 2002. The influence of referential processing on sentence complexity. *Cognition* 85. 79–112. DOI: http://dx.doi.org/10.1016/S0010-0277(02)00087-2

Zukowski Andrea & Jaiva Larsen. 2004. The production of sentences that we fill their gaps. Poster presented at the CUNY Sentence Processing Conference, University of Maryland.