GLOSSA psycholinguistics

Cairncross, A., Vogelzang, M., & Tsimpli, I. (2024). Evaluating the Pseudorelative-First Hypothesis: Evidence from self-paced reading and persistence effects. *Glossa Psycholinguistics*, 3(1): 4, pp. 1–29. DOI: https://doi.org/10.5070/G6011225

eScholarship University of California

Evaluating the Pseudorelative-First Hypothesis: Evidence from self-paced reading and persistence effects

Alex Cairncross, University of Cambridge, UK, aac61@cam.ac.uk

Margreet Vogelzang, University of Cambridge, UK, mv498@cam.ac.uk

Ianthi Tsimpli, University of Cambridge, UK, imt20@cam.ac.uk

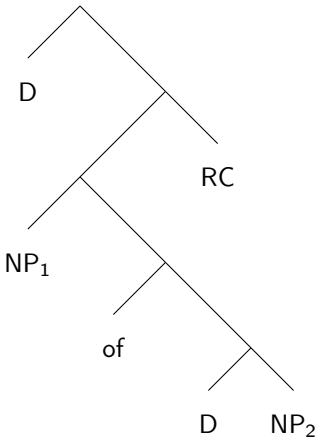
Within the psycholinguistic literature, there has been a longstanding debate regarding whether we resolve syntactic parsing ambiguities via universal or language-specific biases. The present study investigates attachment biases in the online parsing of 'relative clause' (RC) attachment in Italian with respect to pseudorelative (PR) availability. Following the PR account Grillo (2012), languages are assumed to universally prefer local attachment. When languages appear to prefer non-local attachment, this is due (at least partially) to the availability of PRs. Specifically, Grillo and Costa (2014) suggest that whenever a string is ambiguous between a PR and a RC, the parser will prefer the PR parse, resulting in apparent non-local attachment. Although there is growing evidence that PR availability indeed affects offline interpretations, few studies have explored this account from an online perspective. Hence, we conducted a self-paced reading task in Italian. In that task, we directly manipulated PR availability and attachment. Reading times for the critical and postcritical regions along with accuracy to comprehension questions were subjected to mixed-effect regressions. Consistent with the PR account, online results indicated a clear bias for local attachment with true RCs. When PRs were available, we observed a non-local bias. Additionally, the present study provides novel evidence in support of the PR-First Hypothesis, as results indicated that the initial preference for PRs may persist and affect the interpretation of even globally disambiguated items.

1. Introduction

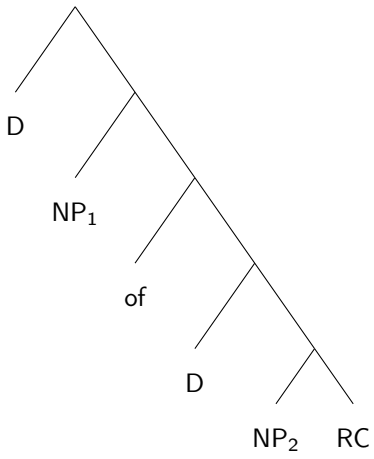
Upon encountering an incoming string, the hearer/reader must decode the physical signal and arrive at a syntactic representation. In order to achieve this task, many researchers have assumed that the parser relies on principles of computational efficiency, such as locality or minimal structure, to resolve ambiguity and reduce processing cost (e.g., De Vincenzi, 1991; Frazier, 1978; Kimball, 1973). This has led to considerable interest in the parsing of relative clause (RC) attachment. To see why, take (1). In that example, the bracketed RC may modify either NP_1 or NP_2. These two readings are known as high attachment (HA) and low attachment (LA), respectively, due to the relative height of the RC, as sketched in (2).

(1) a. Someone shot the $maid_1$ of the $actress_2$ [that was standing on the balcony].
b. Alguien disparó contra la $criada_1$ de la $actriz_2$ [que estaba en el balcón]. [Spanish]
(Cuetos & Mitchell, 1988)

(2) a. High Attachment



b. Low Attachment



This presents an example of an attachment ambiguity. Should there be a universal bias that favours local attachment, we would expect speakers of any language with RCs to prefer LA readings in such contexts. Since Cuetos and Mitchell (1988), however, it has been well known that speakers of different languages exhibit contrasting biases in their interpretation of sentences like (1). Whereas speakers of English have been shown to exhibit the expected LA bias, speakers of Spanish have been repeatedly shown to exhibit a HA bias (Carreiras & Clifton, 1993, 1999; Cuetos & Mitchell, 1988; Dussias, 2003; Fernández, 2003, i.a.). This has led to some researchers rejecting locality as a universal parser bias (e.g., Mitchell & Cuetos, 1991, et seq.) as well as a classification of languages as either LA languages (e.g., English: Cuetos & Mitchell, 1988; Fernández, 2003, i.a.; Chinese: Shen, 2006) or HA languages (e.g., Spanish; French: Zagar et al., 1997; Greek: Papadopoulou & Clahsen, 2003).

While subsequent research has demonstrated that various factors, such as implicit prosody (Fernández, 2003; Fodor, 1998), referentiality (Gilboy et al., 1995), and verbal semantics (Rohde et al., 2011), may influence RC attachment, these factors alone cannot fully account for the observed crosslinguistic variation. More recently, however, Grillo (2012) and Grillo and Costa (2014) have suggested that this crosslinguistic difference in parser biases may only be apparent. Rather, they point out that there is a hidden structural difference which must be taken into account. Namely, many of the languages that have been previously reported to exhibit a HA bias also admit pseudorelatives (PRs), whereas languages that exhibit a LA bias do not. PRs, despite being string-identical to true RCs, display a number of semantic and syntactic differences. Those authors also observed that in the contexts that have been previously studied in the psycholinguistic literature, PRs are only compatible with HA. Thus, they suggest that once this syntactic confound is accounted for, we can maintain locality as a universal parsing bias.

Whereas there is growing evidence that PR-availability affects the offline interpretation of sentences like (1) (Aguilar & Grillo, 2021; Grillo & Costa, 2014; Grillo et al., 2015; Pozniak et al., 2019, i.a.), we can also derive predictions regarding the online parsing of PR/RC ambiguities from the PR account. To date, these online predictions have received less attention, and the available results paint a mixed picture. In response, the present paper presents two new experiments (using a self-paced reading task and an eye-tracking-while-reading task) that investigate the resolution of PR-RC parsing ambiguities in Italian.

1.1 The PR account

Despite being string-identical to true RCs, PRs exhibit a number of syntactic and semantic differences (Casalicchio, 2013; Cinque, 1992; Radford, 1975, i.a.). On the semantic side, RCs denote properties of the NPs they modify, but PRs denote event situations (Moulton & Grillo,

2015). Therefore, the RC reading in (3a) tells us something about the relevant boy (i.e., that he ate ice cream) whereas the PR reading in (3b) tells us what Gianni saw (i.e., an event of eating in which a particular boy was the agent).

(3) a. Gianni ha visto [$_{DP}$ il [$_{NP}$ bambino [$_{CP}$ che mangiava il gelato]]]. [Italian]
'Gianni saw the boy that was eating the ice cream.'

b. Gianni ha visto [$_{PR}$ [$_{DP}$ il [$_{NP}$ bambino]] [$_{CP}$ che mangiava il gelato]].
'Gianni saw the boy eating ice cream.'

On the syntactic side, when a PR occupies the complement of a predicate of perception, as in (3b), the embedded CP is a sister to the DP. This means that when a string with a complex DP like the ones in (1) and (4) is parsed as containing a PR, only the structurally higher of the DPs may c-command into the embedded CP and be interpreted as its subject. That is to say, PR parses force HA in this context.[1] Conversely, RCs are adjuncts within the DP. As such, they may freely attach to either NP_1 or NP_2 in strings like (1) and (4).

(4) Gianni ha visto il figlio del medico che mangiava il gelato. [Italian]
'Gianni saw the son of the doctor (that was) eating the ice cream.'

With the above in mind, it is not enough for the parser to simply decide where to attach the embedded 'RC' when presented with such strings. Rather it must additionally decide whether the embedded CP should be parsed as a RC or a PR. To guide the resolution of this ambiguity, Grillo and Costa (2014) propose the PR-First Hypothesis in (5).

(5) PR-First Hypothesis: When PRs are available, everything else being equal, they will be preferred over RCs. (Grillo & Costa, 2014)

They suggest that the preference for PRs derives from the fact that PRs are structurally and presuppositionally simpler than true RCs. This is because PRs exhibit a number of structural restrictions relative to true RCs (e.g., tense and aspect restrictions; for more examples, see Grillo & Costa, 2014) which those authors take to suggest that RCs may have a "richer and more articulated functional domain" than PRs (p. 166). RCs also require the presupposition of a contrast set, whereas PRs do not. Therefore, for the structure in (3a) to be felicitous, we must presuppose the existence of at least one other boy who did not eat ice cream. The same does not hold of the structure in (3b).

This principle then interacts with locality to derive the crosslinguistic pattern. In languages like Italian or Spanish, the PR-First Hypothesis straightforwardly accounts for the observed HA bias. Whenever a string is ambiguous between a RC and PR parse, the parser will prefer a PR

[1] Although there is no true high/low-attachment ambiguity in such cases, we will continue to call this a 'HA' reading, following Grillo and Costa (2014), because the subject of the embedded CP is interpreted as the structurally higher of the relevant NPs.

parse, resulting in HA. In languages like English, however, the PR-First Hypothesis can only apply vacuously, as PR parses are not licit.[2] In that case, the embedded CP must be parsed as a RC, resulting in a true attachment ambiguity. Principles of locality would then obtain, driving the observed LA bias. The interaction of these two principles is sketched in **Figure 1**, with preferred parses noted in boldface.[3]

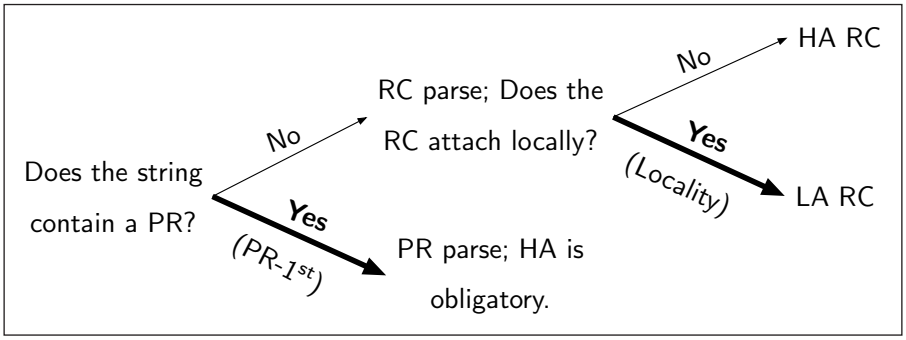


Figure 1: Summary of (relevant) universal PR/RC parse options.

1.2 Previous work on PRs

As the PR account derives the crosslinguistic pattern from this interaction of principles, we can also use it to make testable predictions language-internally. Namely, when PRs are locally blocked in an otherwise HA language, we would expect a LA bias to emerge with true RCs. To explore this prediction, Grillo and Costa (2014) conducted a sentence interpretation task in which they manipulated PR availability via the matrix predicates. Although PRs may occupy the complement of a verb of perception, such as *vedere* ('see'), they may not follow non-perceptual predicates, such as *sposare* ('marry'). This is shown in (6), as the proper name *Maria* blocks a (restrictive) RC reading of the embedded CP, forcing it to instead be interpreted as a PR, where grammatical. Changing the *Maria* to a common noun in (7), we can see that RCs are not similarly restricted by the selectional properties of the matrix predicate.

(6) a. Gianni ha visto Maria che correva. [Italian]
'Gianni saw Maria running.'
b. *Gianni ha sposato Maria che correva.

(7) Gianni ha sposato la donna che correva. [Italian]
'Gianni married the woman that was running.'

[2] For the scope of this paper, we set aside embedded CPs without an overt complementiser. See Grillo et al. (2015) regarding the PR-First Hypothesis and reduced relatives/small clause ambiguities in English.

[3] Although sketched serially, the PR-First Hypothesis is compatible with ranked parallel models.

Exploiting this restriction on PRs, Grillo and Costa (2014) constructed sentences like (8). Half of the time, items contained a perceptual matrix predicate and, half the time, they contained a non-perceptual one. They then presented these to native Italian speakers and asked them *who*-questions targeting the subject of the embedded CP. When items contained a matrix predicate of perception, the authors observed a HA bias (78.6% HA). However, consistent with the predictions of the PR account, participants preferentially selected LA (24.2% HA) when the same items contained non-perceptual predicates. Convergent results have since been reported in other HA languages (e.g., French: Pozniak et al., 2019; Spanish: Aguilar & Grillo, 2021).

(8) a. Gianni ha visto il figlio del medico che correva la maratona. [Italian]
'Gianni saw the son of the doctor (that was) running the marathon.'

b. Gianni vive con il figlio del medico che correva la maratona.
'Gianni lives with the son of the doctor that was running the marathon.'

Moreover, results from English, a LA language, suggest that the findings from HA languages cannot be reduced to the semantics of the matrix predicate, but rather should be interpreted as an effect of PR (un-)availability. Grillo et al. (2015) conducted a similar experiment with native English speakers which contained translation equivalents of the items in Grillo and Costa (2014), as in (9). Consistent with the PR account, however, the authors only report a small effect of semantic facilitation, not a wholesale shift from LA to HA between non-perceptual and perceptual predicates, as observed in Italian.

(9) a. Kelly heard the grandma of the girl that was screaming.

b. Kelly works with the grandma of the girl that was screaming.

In addition to offline interpretive biases, we can also derive testable predictions regarding online parsing from the PR account. The most straightforward of these is that whenever a string is ambiguous between a PR and a RC, disambiguation toward a RC should come with a processing cost (PR-First Hypothesis). To explore this prediction, Pozniak et al. (2019) conducted an eye-tracking-while-reading study in French. For this, they manipulated two factors: the type of matrix predicate (perceptual vs. non-perceptual) and its tense (match vs. mismatch with the embedded predicate). As discussed above, PRs are licit after perceptual predicates but not after non-perceptual ones. As an additional restriction, the tensed element in a PR must match in tense with the matrix predicate. This is demonstrated in (10), with a proper name to block a RC parse. Again, no such restriction holds for RCs (11).

(10) a. Gianni ha visto$_{\text{PST}}$ Maria che correva$_{\text{PST}}$. [Italian]
'Gianni saw Maria running.'

b. *Gianni ha visto$_{\text{PST}}$ Maria che corre$_{\text{PRS}}$.

(11) Gianni ha visto$_{PST}$ la donna che corre$_{PRS}$. [Italian]
'Gianni saw the woman that is running.'

Therefore in items like (12a, b), the PR-FIRST HYPOTHESIS would predict an initial PR bias, because the matrix predicate is a verb of perception. Upon encountering the embedded predicate, however, a PR parse becomes incompatible with the string in (12b), due to the tense mismatch. Therefore, the account would predict the embedded predicate in (12b) to be processed more slowly than in (12a). No such difference is expected between the tense match and mismatch items in (12c, d) because the matrix predicates are non-perceptual. As such, we would predict an interaction of *predicate type* and *tense match* in the eye-tracking measures.

(12) a. Jean a vu$_{PST}$ la fille qui poussait$_{PST}$ la femme. [French]
'Jean saw the girl (that pushed / pushing) the woman.'

b. Jean voit$_{PRS}$ la fille qui poussait$_{PST}$ la femme.
'Jean sees the girl that pushed the woman.'

c. Jean était$_{PST}$ marié à la fille qui poussait$_{PST}$ la femme.
'Jean was married to the girl that pushed the woman.'

d. Jean est$_{PRS}$ marié à la fille qui poussait$_{PST}$ la femme.
'Jean is married to the girl that pushed the woman.'

Consistent with the prediction above, the authors observed evidence for an interaction of *predicate type* and *tense match* in regression path duration (or 'go-past') times for the embedded predicate as well as the proportion of regressions out of the region. In both measures, there was greater processing cost for tense mismatch items when the matrix predicate was perceptual but not when it was non-perceptual. However, this effect was only present in the first half of the experiment. Nonetheless, when the authors compared the French results to results from native English speakers with translation equivalents, they observed three-way interactions with *language*, indicating the advantage for tense match under predicates of perception is restricted to French, the language that admits PRs. Thus, on the one hand, this study provides initial online evidence in support of the PR-FIRST HYPOTHESIS. On the other, however, the effect disappeared by the second half of the experiment, suggesting that it is subtle and can be obfuscated by adaptation to the experimental context. In this particular experiment, this may have been due to the highly reliable nature of tense (as well as the relatively low rate of filler, only 48% of the French items). As the embedded predicates were always in a past tense form, upon encountering a matrix perceptual predicate in the present, it is plausible that participants could eventually anticipate that the embedded CP could only be parsed as a RC.

Although Pozniak et al. (2019) provide some initial evidence for an online preference for PRs over RCs in simple contexts, they did not investigate how this affects attachment preferences. Therefore,

building off this, Aguilar, Ferré, Gavilán, Hinojosa, and Demestre (2021) explored how PR-firstness interacts with attachment in the online parsing of Spanish. Following the PR account, they expected contrasting biases when PRs are, and are not, locally available. When PRs are available, items disambiguated toward LA are expected to be harder to process than items disambiguated toward HA. This is because PR parses are assumed to be preferred over RC parses (PR-FIRST HYPOTHESIS) and PRs are incompatible with LA in the relevant contexts. However, when PRs are unavailable and the string must be parsed as containing a RC, LA disambiguated items are predicted to be easier to process than HA items, due to principles of locality (e.g., LATE CLOSURE, Frazier, 1978).

For their eye-tracking-while-reading experiment, they created items as in (13). In these items, they manipulated the type of matrix predicate (perceptual vs. non-perceptual) to affect PR availability, as in Pozniak et al. (2019). Rather than a tense manipulation to force RC readings in otherwise PR-compatible contexts, however, Aguilar et al. (2021) exploited gender marking on secondary depictive predicates within the embedded CPs to force attachment. As the authors counterbalanced both the gender of NP_1 and the gender marking of the secondary depictive predicate, this study lacked a reliable cue like the one in Pozniak et al. (2019). Moreover, fillers also included a number of unambiguous PRs as in (14), which has been observed to mitigate adaptation effects (Fernandes et al., 2018).

(13) a. Juan vio al entrenador$_M$ de la tenista$_F$ que lloraba amargad(-o$_M$ / -a$_F$) por la derrota. [Spanish]
'Juan saw the coach of the tennis player (that wept / weeping) bitterly for the defeat.'

b. Juan conoció al entrenador$_M$ de la tenista$_F$ que lloraba amargad (-o$_M$ / -a$_F$) por la derrota.
'Juan has met the coach of the tenis player that wept bitterly for the defeat.'

(14) a. El técnico de laboratorio observó a Rosa que estaba escribiendo las fórmulas en la pizarra. [Spanish]
'The lab technician observed Rosa writing the formulas on the board.'

At the disambiguating secondary predicate, the authors reported a significant interaction of *predicate type* and *attachment* in early processing measures (first fixation and gaze or 'first-pass' duration) which was driven by a LA bias in the non-perceptual items. No HA bias was observed in perceptual ones. In the later total time measure, however, this pattern reversed. There, the interaction was driven by a HA bias in perceptual items with no LA bias in the non-perceptual ones. As such, this experiment provides two pieces of evidence in support of the PR account. First, despite previous studies repeatedly categorising Spanish as a HA language, once PRs are controlled for, the authors find evidence for an online LA bias with true RCs. Second, at least in total times, there was also online evidence for a HA bias when the matrix predicates were

compatible with PRs. However, this later interaction was further modulated by the order of presentation, with the HA bias only appearing in the first half of items. Thus, despite the study lacking a reliable cue for PR/RC disambiguation, PR-firstness appeared susceptible to adaptation, as in Pozniak et al. (2019). The authors suggest that this may have been caused by the fact that RCs outnumbered PRs even after the inclusion of PR filler items.

More recently, Lee and De Santo (2022) conducted a self-paced reading study with native Italian speakers in which participants were presented sentences like the ones in (15). In these items, the authors manipulated the type of matrix predicate (perceptual vs. non-perceptual) and the number agreement on the embedded predicate to force attachment high or low. After each item, participants were asked a *who*-question targeting the subject of the embedded CP. Offline results were non-significant. Online results, however, indicated a significant interaction of *predicate type* with *attachment* on the window containing the embedded verb. This surfaced in the non-perceptual items as longer reading times on the embedded verb when the RC was forced to attach high (although see also Foppolo & Abbondanza, 2021, who did not find online evidence of a LA bias in their self-paced reading task focusing on RCs in Italian). In the perceptual items, the effect of *attachment* was non-significant. However, De Santo and Lee (2022) presented an additional analysis of that data in which they investigated the immediately following word (*la* in (15)). There, they observed a significant interaction of *predicate* and *attachment* which was driven by significantly longer reading times for LA perceptual items. That is to say, they observed an online effect of PR availability, just not at their disambiguating region. How to reconcile the Italian results with those in Aguilar et al. (2021) is unclear, as the method and manipulation in the two studies are dissimilar. Moreover, Lee and De Santo (2022) do not mention item order effects, which might obfuscate PR-firstness as discussed above.

(15) a. Gianni vide il figlio$_{SG}$ dei medici$_{PL}$ che correv (-a$_{SG}$ / -ano$_{PL}$) la maratona. [Italian]
'Gianni saw the son of the doctors (that ran / running) the marathon.'

b. Gianni viveva con il figlio$_{SG}$ dei medici$_{PL}$ che correv (-a$_{SG}$ / -ano$_{PL}$) la maratona.
'Gianni lived with the son of the doctors that ran the marathon.'

Thus in summary, whereas there is growing evidence that PR availability modulates interpretive biases for 'RC' attachment ambiguities, the available online results are somewhat mixed. On the one hand, there is evidence in support of the PR-First Hypothesis in simple contexts which abstract away from attachment. On the other hand, of the two experiments that have previously looked at PR-firstness and attachment, only one reports a HA bias at the critical window in PR-compatible contexts. Moreover, those studies that have reported online evidence for PR-firstness indicate that the effect is subtle and susceptible to adaptation.

2. Research questions

Given the inconsistency between the results of Aguilar et al. (2021) and Lee and De Santo (2022), we set out to investigate the PR-First Hypothesis in the online processing of Italian. Specifically, we formulated the two following sub-questions:

I. Can we observe a locality effect in the processing of true Italian RCs?

II. Can we observe a PR-firstness effect in the processing of items that are PR-compatible?

In response to these questions, we conducted a self-paced reading task (Experiment 1) and an eye-tracking-while-reading task (Experiment 2) with the same materials.

3. Experiment 1

3.1 Participants

For our self-paced reading experiment, participants consisted of 66 native Italian speakers (Female = 34, Male = 32) who (i) had lived in Italy from birth until at least the age of 16, (ii) had no diagnosed language-related disorder, and (iii) reported growing up monolingually.[4] At the time of testing, all participants were living in Italy, and the average age of the group was 43.47 years (SD = 7.39).

3.2 Stimuli and design

Critical items consisted of 32 temporarily ambiguous sentences modelled after Aguilar et al. (2021). These all contained a complex NP (NP_1 of NP_2) followed by a finite embedded CP. Within these items, we manipulated two binary factors: PR availability and attachment. To manipulate PR availability, we exploited the selectional properties of the matrix predicate (counterbalanced *within* items); half of the time, items appeared with a matrix predicate of perception, and half the time, they appeared with a non-perception predicate. To force attachment, we used gender marking on a secondary predicate within the embedded CP ($-o_M$, $-a_F$). To that end, the two NPs in the complex NP always differed in gender (with the gender of NP_1 counterbalanced *across* items). We opted for gender disambiguation over a number manipulation, as it allowed us to maintain a uniform length of the critical window across conditions.[5] For consistency, all NPs were animate and singular. For NP_1, we additionally ensured that we did not use any kinship terms that may carry a presupposition of uniqueness (e.g., *marito* 'husband'), as that may have

[4] When specifically asked about potential exposure to 'dialect' as a child (for example, with one's grandparents), 14 reported some exposure.

[5] We did not consider a tense manipulation, as in Pozniak et al. (2019), because this would not have allowed us to force attachment.

artificially biased RC attachment away from NP_1 (as noted above, restrictive RCs presuppose a contrast set, whereas PRs do not). An example of an item in all 4 conditions is provided in (16).

(16) a. Perceptual HA [Italian]
Gianni | ha visto | il collega$_M$ | della biologa$_F$ | che correva | sporco$_M$ | di fango.

b. Perceptual LA
Gianni | ha visto | il collega$_M$ | della biologa$_F$ | che correva | sporco$_F$ | di fango.
'Gianni saw the colleague of the biologist (that was) running covered in mud.'

c. Non-perceptual HA
Gianni | vive | con il collega$_M$ | della biologa$_F$ | che correva | sporco$_M$ | di fango.

d. Non-perceptual LA
Gianni | vive | con il collega$_M$ | della biologa$_F$ | che correva | sporco$_F$ | di fango.
'Gianni lives with the colleague of the biologist that was running covered in mud.'

These sentences were distributed across 4 lists such that participants saw each item only once, with 8 sentences per condition. The experiment also included pronominal distractor items (N = 32) as well as unrelated, unambiguous fillers (N = 40).

3.2.1 Norming of the stimuli

To account for any potential semantic biases in the embedded CP (in particular, biases relating to our gender disambiguation), we initially constructed 50% more items (N = 48) than we intended to use in the main experiments. From each of those items, we then formed 2 simple, unambiguous sentences corresponding to the HA and LA readings of the original item (17). These were then distributed across 2 lists such that they contained 24 sentences corresponding to a HA interpretation (17a) and 24 corresponding to a LA interpretation (17b). To these lists, we appended another 62 unrelated sentences.

(17) Gianni (ha visto / vive con) il collega$_M$ della biologa$_F$ che correva sporco$_M$ di fango.
a. Il collega$_M$ della biologa$_F$ correva sporco$_M$ di fango.
b. La biologa$_F$ correva sporca$_F$ di fango.

These sentences were then presented as part of a naturalness judgment task to 30 native Italian speakers (Female = 16; Male = 14)[6] who did not take part in either of our main experiments. During the naturalness judgment task, participants read the sentences in isolation. After each sentence, they were asked to rate how natural the sentence sounded to them from 1 (very unnatural) to 5 (very natural). Prior to testing, 'very natural' was explicitly defined to the norming participants as meaning that 'they might hear such a sentence in a normal conversation

[6] Due to issues with data logging, the judgments for one participant were lost.

with other native speakers of Italian.' The ratings for the a and b versions of each potential item were then compared using *t*-tests. Any potential item for which the difference between the ratings for the two versions was found to approach significance (i.e., $p \leq 0.1$) was rejected. We then selected the 32 items with the highest global naturalness ratings (final mean = 4.32/5) such that we were able to maintain counterbalancing of the gender of NP_1.

3.3 Procedure

Participants were recruited through Prolific Academic, and the experiment was run via PCIbex (Zehr & Schwarz, 2018). All participants provided informed consent, and prior to testing, we obtained ethics approval for the study from the Ethics Committee of the Faculty of Modern and Medieval Languages and Linguistics at the University of Cambridge.

For the self-paced reading task, items were always presented alone on the participant's screen. They initially appeared as a series of underscores corresponding to the windows (boundaries indicated by |s in (16)). In order to view the first window, participants pressed the space bar. When they had read this section, they pressed the space bar again. This caused the current window to disappear and the next window to become visible. After reading the final window for each item, participants were presented with a polar comprehension question. For the critical items, this always asked about the subject of the non-matrix CP. In half of the items, the question asked about NP_1 and in half of the items, it asked about NP_2 (counterbalanced *within* items). To respond, participants pressed 'F' or 'J.' An example is presented in (18). To familiarise themselves with the procedure, participants were presented with practice items (N = 7) prior to testing.

(18) La biologa correva?
'Was the biologist$_F$ running?'
F: Sì
'Yes'
J: No
'No'

We opted to include these comprehension questions, as previous studies on different types of garden-path sentences have found that readers' initial/preferred parse may persist even after reanalysis/re-ranking (e.g., Christianson et al., 2001; Slattery et al., 2013; Sturt, 2007). This can be exploited to derive a secondary and indirect way of tapping parser biases, namely, by comparing response accuracy after items in which we expect reanalysis and after items in which we do not. If accuracy is lower for the items that are expected to require reanalysis, this would support the idea that reanalysis has indeed taken place.

To ensure that sufficient attention was paid to the stimuli, we also included polar comprehension questions after the unambiguous filler items. These were counterbalanced such

that the correct answer was 'F' in half of the items. All participants in the present sample were required to achieve a minimum accuracy of 85% on those items.

3.4 Data cleaning and analysis

In the following, we present two planned analyses as well as one unplanned analysis. For the planned analyses, we consider the reading times of the disambiguating window (which consisted of the secondary depictive predicate) as well as the accuracy to the comprehension questions. For the unplanned analysis, we also consider the reading times of the post-critical window (which consisted of the sentence-final PPs). We opted to present this additional analysis given De Santo and Lee's (2022) additional analysis of the data reported in Lee and De Santo (2022). Despite learning of their reanalysis after beginning data collection, all of our items included sentence-final PPs to push the critical region off the sentence edge, allowing us to also perform the additional analysis.

Due to a coding error, we lost 3 items for 12 participants. To clean the reading time data for the remaining trials, we first coded as missing any items for which the analysed window, or any window prior to it, had an implausibly fast reading time (<200 ms). This affected 1.83% of the data in both the critical and post-critical windows. For both windows, we then calculated the interquartile range (IQR) for each of our 4 conditions. For each condition, we coded as missing any values that lay 1.5 IQR above the third quartile for that window. This affected a further 6.79% of trials in the critical window (for a total data loss of 8.62%) and 7.27% of trials in the post-critical window (total data loss: 9.10%). The remaining data was equally distributed across the 4 conditions (critical: $\chi^2(1,1897) = 0.23; p = 0.93$; post-critical: $\chi^2(1,1887) = 0.98; p = 0.81$). For the analysis of the comprehension questions, we only consider those items included in the planned analysis of the critical window.

To analyse the data, we used the *lme4* package (Bates et al., 2015) in R (R Core Team, 2022). We started by identifying the best random-effects structure. To that end, we conducted a family of intercept-only models in which we considered the various random-effects structures of our theoretically relevant fixed predictors (i.e., random intercepts by *item* and *participant* as well as random slopes by *attachment* and *predicate*). From these, we selected the best-fitting random effects structure (Matuschek et al., 2017) that contained, at a minimum, random intercepts by *item* and *participant*, using the Akaike Information Criterion (AIC). We then fit a base model with only the fixed effects that we were theoretically interested in using sum coding (i.e., –0.5, 0.5). For *predicate*, we coded non-perceptual as the negative level. For *attachment*, LA was coded as the negative level. As previous work has suggested that trial order may obfuscate one of the effects we are interested in (i.e., PR-firstness) due to adaptation to the experimental context, we then conducted a further two models which included *order* (centred over the experiment) as either a simple effect or potential interaction term. We compared these additional models against the

base model, using the AIC to assess whether the added terms improved the fit. If complicating the model reduced the AIC by 2 or more, that extra factor was kept. Otherwise, we selected the simpler model.

For the analysis of response accuracy, once we had identified the best fitting model with regard to *order*, we ran two further models which included *response target* (sum coded, negative level: yes) as either a simple predictor or interaction term. This was done to account for any potential effects of the acquiescence bias (Holbrook, 2008). We then selected the best fitting model in the same manner.

3.5 Hypotheses and predictions

From the above, the following hypotheses were formulated:

(H1) In the RC-Only items, participants will exhibit a LA bias.

(H2) In the PR/RC items, participants will exhibit a HA bias.

For reading times, then, we predict that LA items should be read faster than HA items when the matrix predicate is non-perceptual, due to locality (e.g., LATE CLOSURE). When the matrix predicate is perceptual, however, we expect the opposite. HA items should be read faster than LA items due to the PR-FIRST HYPOTHESIS. This should result in a two-way interaction of *attachment* and *predicate*. If the best-fitting model also includes a three-way interaction with *order*, this would be expected to surface as a PR-firstness effect in the first half of the experiment that disappears in the second half (adaptation).

As we take response accuracy in this context to be an indirect measure of processing biases, we expect the same pattern of results. When the matrix predicate was non-perceptual, we expected greater accuracy for LA items than HA ones. Conversely, when the matrix predicate is perceptual, there should be greater accuracy with HA items than LA ones, resulting in an interaction of *attachment* and *predicate*. Should there be a three-way interaction, this is expected to be due to a reduced PR-firstness effect as the experiment progresses.

3.6 Results

3.6.1 Reading times for the critical window

Figure 2 presents the average reading times for the critical window, broken down by the matrix predicate and attachment. After non-perceptual matrix predicates, LA items were read noticeably faster than HA items (LA: 700.47 ms; HA: 811.79 ms). After perceptual predicates, however, participants read LA disambiguated items (mean = 763.24 ms) slower than HA disambiguated items (mean = 718.97 ms).

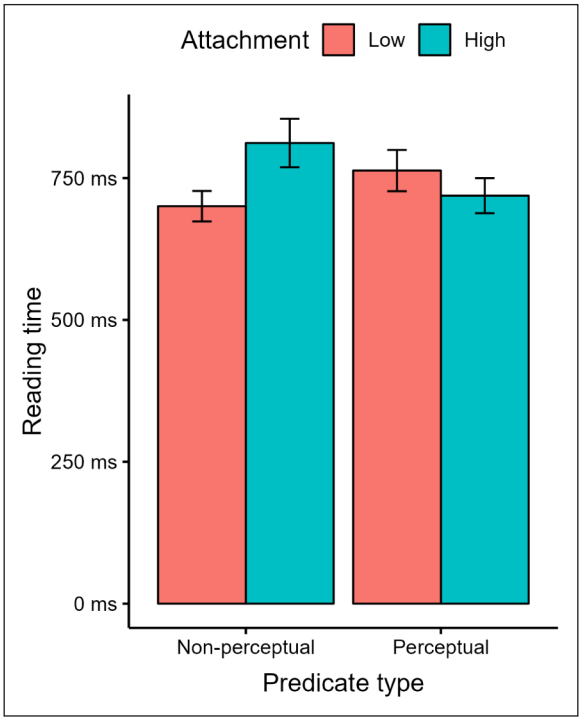


Figure 2: Average reading times for the critical window with 95% confidence intervals.

Reading times were logged and subjected to a family of mixed-effect regression, as described above. Model comparison indicated that the best-fitting model included *order* as a simple effect (improvement to AIC $>$ 2) but not as a potential interaction term. The output from that model is reported in **Table 1**.

Table 1: Model output for reading times in the critical region.

	Estimate	**Std. Error**	***t*-value**	***p*-value**	
Intercept	6.52	0.04	165.19	$<$**0.001**	***
Predicate	–0.01	0.02	–0.86	0.39	
Attachment	0.03	0.02	1.71	0.09	.
Order	–0.001	0.0003	–4.60	$<$**0.001**	***
Predicate:Attachment	–0.15	0.03	–4.56	$<$**0.001**	***

That model indicated a significant effect of *order* ($\hat{\beta}$ = –0.001; t = –4.60; $p < 0.001$), with participants reading more quickly as the experiment progressed. The model did not indicate significant effects of *predicate* or *attachment*, but there was a significant interaction between the two ($\hat{\beta}$ = –0.15; t = –4.56; $p < 0.001$). To follow up on this interaction, we ran pairwise

comparisons of *attachment* within *predicate* using the *emmeans* package (Lenth, 2022). For these comparisons, we present Holm-Bonferroni corrected *p*-values. Results indicated a significant effect of *attachment* within the non-perceptual items ($\hat{\beta} = 0.10$; $t = 4.42$; $p < 0.001$) with LA disambiguated items being read more quickly. Results also indicated a significant effect within the perceptual items ($\hat{\beta} = -0.05$; $t = -2.02$; $p = 0.04$). In those items, however, this surfaced as faster reading times for HA disambiguations.

3.6.2 Reading times for the postcritical window

Average reading times by attachment and matrix predicate for the postcritical region are presented in **Figure 3**. For the non-perceptual items, participants read LA disambiguated items (mean = 777.49 ms) faster than HA disambiguated ones (mean = 971.94 ms). For perceptual items, participants read LA disambiguated items (mean = 963.56 ms) slower than HA disambiguated items (mean = 854.33 ms).

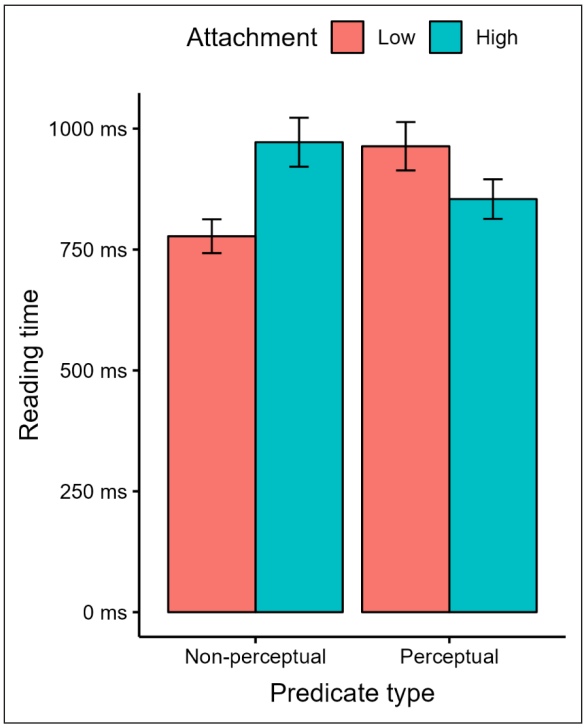


Figure 3: Average reading times for the postcritical window with 95% confidence intervals.

The best fitting model contained *order* as a simple effect. The output for that model is presented in **Table 2**. The model indicated a significant effect of *attachment* ($\hat{\beta} = 0.05$; $t = 2.66$; $p < 0.01$), which surfaced as longer reading times for HA items. There was also a significant effect of *order* ($\hat{\beta} = -0.003$; $t = -9.72$; $p < 0.001$), with participants reading more quickly as the experiment

Table 2: Model output for reading times in the postcritical region.

	Estimate	**Std. Error**	***t*-value**	***p*-value**	
Intercept	6.67	0.04	167.84	$<$**0.001**	***
Predicate	0.02	0.02	1.17	0.24	
Attachment	0.05	0.02	2.66	$<$**0.01**	**
Order	–0.003	0.0003	–9.72	$<$**0.001**	***
Predicate:Attachment	–0.30	0.04	–8.13	$<$**0.001**	***

progressed. The effect of *predicate* was not significant but its interaction with *attachment* was ($\hat{\beta}$ = –0.30; t = –8.13; $p < 0.001$). To follow up on that interaction, we again conducted pairwise comparisons of *attachment* within *predicate* using the *emmeans* package (Lenth, 2022). For these, we report Holm-Bonferroni corrected *p*-values. The pairwise comparisons indicated a significant effect of *attachment* within the non-perceptual items ($\hat{\beta}$ = 0.20; t = 7.55; $p < 0.001$) with LA disambiguated items being read more quickly. There was also a significant effect of *attachment* within the perceptual items ($\hat{\beta}$ = –0.10 t = –3.90; $p < 0.001$), with LA disambiguated items being read more slowly.

3.6.3 Response accuracy

Figure 4 presents the average accuracy (by participant) for the comprehension questions, broken down by the matrix predicate and attachment. In the non-perceptual items, participants gave the correct answer at similar rates for both LA (mean = 91.98%) and HA disambiguated items (mean = 89.16%). In the perceptual items, however, participants' responses were noticeably more accurate after HA disambiguated items (mean = 95.45%) than LA disambiguated ones (mean = 83.76%).

Responses (coded as ± correct) were entered into a family of mixed effect logistic regression, as described above. Model comparison indicated that neither of the models with *order* nor the model with *response target* as a potential interaction term fit the data better than the simple model (improvements to AIC < 2). However, the inclusion of *response target* as a simple predictor did improve the model fit (improvement to AIC > 2). The output for that model is presented in **Table 3**.

The model indicated a significant effect of *response target* ($\hat{\beta}$ = –0.46; z = –2.67; $p < 0.01$), with participants giving more accurate responses when the target response was *sì* ('yes'). The effets of *predicate* and *antecedent* were not significant, but their interaction was ($\hat{\beta}$ = 2.04; z = 5.61; $p < 0.001$). To follow up on that interaction, we conducted pairwise comparisons of *attachment* within *predicate* and report Holm-Bonferroni corrected *p*-values, as above. That indicated that the effect of *attachment* was non-significant within the non-perceptual items

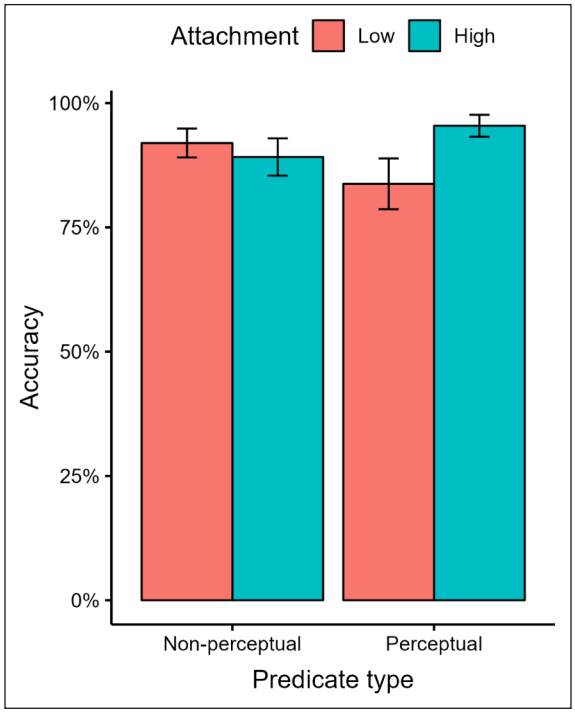


Figure 4: Average response accuracy (by participant) with 95% confidence intervals.

Table 3: Model output for response accuracy.

	Estimate	**Std. Error**	***z*-value**	***p*-value**	
Intercept	2.94	0.21	13.89	**<0.001**	***
Predicate	0.06	0.18	0.35	0.73	
Attachment	0.61	0.37	1.66	0.10	.
Response target	–0.46	0.17	–2.67	**<0.01**	**
Predicate:Attachment	2.04	0.36	5.61	**<0.001**	***

($\hat{\beta} = -0.41; z = -1.02; p = 0.31$). Within the perceptual items, however, the effect of *attachment* was significant ($\hat{\beta} = 1.63$; $z = 3.88$; $p < 0.001$), with participants giving more accurate responses after HA items.

Given the relatively small number of incorrect responses (N = 194, 10.23% of responses) as well as relatively small differences in error rates for the comprehension questions (about 12 percentage points in the perceptual items), we additionally inspected which items participants got wrong. This was done to see if the trend reported above could have been driven by a small set of potentially bad items. Inspection of the items, however, indicated that errors were fairly well distributed, with all items receiving some wrong responses (mean accuracy by item = 89.77%;

SD = 4.06%). To look for potential outliers, we calculated the IQR for accuracy by item. For no items did accuracy fall 1.5 IQR below/above the 1st/3rd quartile.

3.7 Interim discussion

Above, we have presented a self-paced reading task in which we were interested in two potential effects: first, a LA bias with true RCs when PRs are blocked, and second, a HA bias when PRs are locally available.

For items with a non-perceptual matrix predicate, reading times for the disambiguating region were significantly longer when the RC was forced to attach high than when it was forced to attach low. Thus, despite Italian being a so-called "HA language," a LA bias emerges in the online processing of true RCs once PRs are locally blocked. This is convergent with Italian and Spanish results from Lee and De Santo (2022) and Aguilar et al. (2021), respectively (although see also Foppolo & Abbondanza, 2021, who did not find online evidence of a LA bias for Italian RCs using a mixture of disambiguating features).

When the matrix predicate was perceptual, however, reading times for both the critical and post-critical windows were significantly longer when the embedded CP was forced to attach low. As LA is incompatible with PRs in our items, we interpret this as an online bias for PRs where locally available. Again, this is convergent with the previously reported results, although this effect was not restricted to the post-critical window (cf. De Santo & Lee, 2022) nor we did find evidence for a loss of PR-firstness due to adaptation to the experimental context (cf. Aguilar et al., 2021).

Interestingly, the online bias for PRs also appears to have affected offline comprehension responses. That is to say, we observed greater accuracy after HA perceptual items than LA ones. This is what we would expect if (i) participants first projected a PR parse in those items upon encountering the embedded CP and (ii) that intermediate parse (which must be discarded upon encountering a LA disambiguation) may persist in short term memory and interfere when participants are responding to the comprehension questions. However, we may want to be sceptical of this finding and interpretation, given that we did not find a persistence effect in the non-perceptual items, despite a clear LA bias in the reading times of both windows. As such, before trying to interpret this asymmetry, it is worth checking if the result is replicable. To that end, we now turn to Experiment 2.

4. Experiment 2

4.1 Participants

For this experiment, we recruited 27 native Italian speakers (Female = 18, Male = 9) who had not participated in Experiment 1. As with the self-paced reading experiment, all participants (i) had lived in Italy from birth until at least the age of 16, (ii) had no diagnosed language-related disorder,

and (iii) did not report speaking any other languages at home as a child.[7] At the time of testing, all participants were living in Italy, and the average age of the group was 41.37 years (SD = 13.34).

Although Experiment 2 was initially planned as a sister experiment to our self-paced reading experiment, in order to control for any potential task effects that may have led to the differences in the results reported by Aguilar et al. (2021) and Lee and De Santo (2022), our sample for Experiment 2 is relatively small for an eye-tracking study, as pointed out to us by the anonymous reviewers and editor.[8] As such, we will refrain from analysing or basing conclusions on the eye-movement measures. Rather, we will focus on the behavioural data to investigate whether the pattern observed in offline comprehension responses is replicable in a new sample. Nonetheless, for transparency, all the eye-movement data that we planned to analyse but did not include in the final manuscript is available on the associated OSF page along with a description of the measures and the data filtering process.

4.2 Stimuli and design

Items in Experiment 2 were identical to those in the self-paced reading experiment. To reiterate, critical items consisted of 32 temporarily ambiguous sentences which contained a complex NP (NP_1 of NP_2) followed by a finite embedded CP. Within these items we manipulated PR-availability and attachment. To manipulate PR-availability, we exploited the selectional properties of the matrix predicate (counterbalanced *within* items); half of the time, items appeared with a matrix predicate of perception, and half the time, they appeared with a non-perceptual predicate. To force attachment, we used gender marking on a secondary predicate within the embedded CP ($-o_M$, $-a_F$). For that reason, the two NPs in the complex NP always differed in gender (with the gender of NP_1 counterbalanced *across* items). An example item in all 4 conditions is provided in (19).

(19) a. Perceptual HA
Gianni ha visto il collega$_M$ della biologa$_F$ che correva | sporco$_M$ | di fango.

b. Perceptual LA
Gianni ha visto il collega$_M$ della biologa$_F$ che correva | sporca$_F$ | di fango.
'Gianni saw the colleague of the biologist that was running covered in mud.'

c. Non-perceptual HA
Gianni vive con il collega$_M$ della biologa$_F$ che correva | sporco$_M$ | di fango.

d. Non-perceptual LA
Gianni vive con il collega$_M$ della biologa$_F$ che correva | sporca$_F$ | di fango.
'Gianni lives with the colleague of the biologist that was running covered in mud.'

[7] When specifically asked about potential exposure to 'dialect' as a child (for example, with one's grandparents), 7 reported some exposure.

[8] For comparison, Pozniak et al.'s (2019) eye-tracking study reported data from 62 French-speaking participants, while Aguilar et al.'s (2021) reported data from 44 Spanish-speaking participants.

Sentences were distributed across 4 lists, such that participants saw each item only once, with 8 sentences per condition. The experiment also included pronominal distractor items (N = 32) and unambiguous fillers (N = 40).

4.3 Procedure

After providing informed consent, participants read the instructions and performed an initial calibration.[9] They were then presented with a practice phase to familiarise themselves with the eye-tracking procedure. Before each item, they were required to fixate on a cross on the left-hand side of the screen in order to perform a drift check. If this were successful, the item would then appear immediately to the right of the fixation cross. If the drift check were unsuccessful, it would initiate a re-calibration sequence. After reading the item, participants pressed the space bar to view the polar comprehension question. The two possible answers were always labelled 'F' and 'J.' To respond, participants pressed the corresponding key. As with the self-paced reading task, we also included polar comprehension questions after the unambiguous filler items. All participants were required to achieve a minimum accuracy of 85% on those items to ensure that sufficient attention was paid to the stimuli.

Prior to testing, ethics approval for the study was obtained from the Ethics Committee of the Faculty of Modern and Medieval Languages and Linguistics at the University of Cambridge.

4.4 Data cleaning and analysis

In the following, we only present our planned analysis of the accuracy to the comprehension questions. For this, we coded as missing any trials in which the critical region itself was not fixated. This affected 1.16% of the data. We then excluded missing responses. This affected a further 0.93% of trials, for a total exclusion of 2.08%. The remaining observations were equally distributed across the 4 conditions ($\chi^2(1, 846) = 0.06; p = 1.00$).

To analyse the comprehension data, we again conducted families of intercept-only models to identify the best-fitting random-effect structure. We then fit a base model with only the fixed effects that we were theoretically interested in, using sum coding (i.e., –0.5, 0.5). For *attachment*, we coded LA as the negative level. For *predicate*, non-perceptual was coded as the negative level. To account for potential adaptation effects, we then conducted two more models which included (trial) *order* (centred over the experiment) as either a simple effect or potential interaction term. If complicating the model improved the AIC by 2 or more, the more complicated fixed effect structure was retained. Once we had identified the best fitting model with regard to *order*, we repeated this process with *response target*.

[9] For details on the eye-tracking procedure for the unanalysed eye-movement measures, please consult the OSF page.

4.5 Hypotheses and predictions

Our hypotheses were the same as in the self-paced reading task:

(H1) In the non-perceptual items, participants will exhibit a LA bias.

(H2) In the perceptual items, participants will exhibit a HA bias.

For response accuracy analysis, we expected greater accuracy for LA items than HA items when the matrix predicate is non-perceptual. When it is perceptual, however, we expected the opposite pattern. This was expected to result in interaction of *attachment* and *predicate*. If the best-fitting models were to include a three-way interaction with *order*, we would expect this to surface as a reduction of the PR-firstness effect over time.

4.6 Results

Figure 5 presents the average accuracy (by participant) for the comprehension questions, broken down by the matrix predicate type and attachment. In the non-perceptual condition, responses to LA and HA items had a similar accuracy (LA: 96.63%; HA: 96.76%). In the perceptual condition, however, the accuracy for LA items was noticeably lower than for HA items (LA: 90.74%; HA: 99.54%).

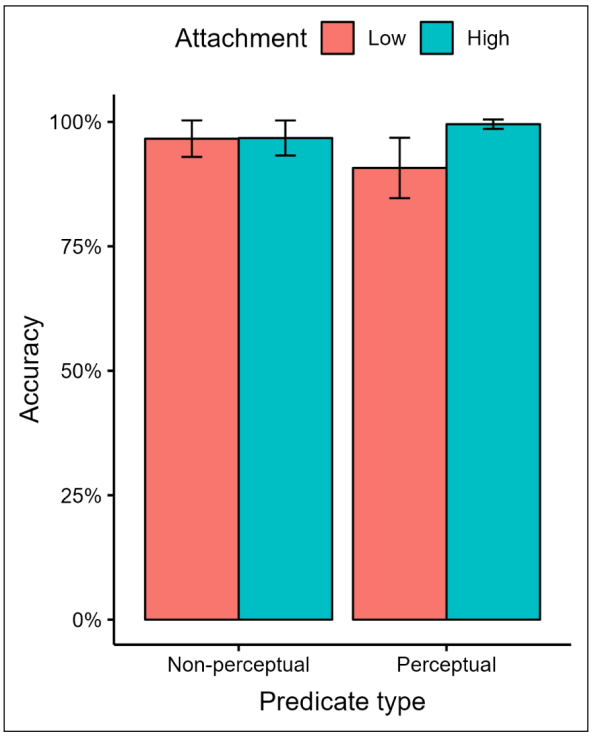


Figure 5: Average response accuracy (by participant) with 95% confidence intervals.

To analyse the accuracy for comprehension questions (coded as ± correct), we again conducted a family of logistic models. Model comparison indicated that the inclusion of *order* as a simple effect or potential interaction term did not improve the model fit (improvements to AIC < 2). However, the inclusion of *response target* as a simple predictor did improve the model fit (improvement to AIC > 2), whereas its inclusion as an interaction term did not (improvement to AIC < 2). The output for the best fitting model is presented in **Table 4**.

Table 4: Model output for response accuracy.

	Estimate	**Std. Error**	***z*-value**	***p*-value**	
Intercept	5.47	1.01	5.41	**<0.001**	***
Predicate	0.47	0.63	0.74	0.46	
Attachment	2.29	1.58	1.45	0.15	
Response Target	1.19	0.45	2.66	**<0.01**	**
Predicate:Attachment	3.73	1.30	2.87	**<0.01**	**

This indicated a significant effect of *response target* ($\hat{\beta} = 1.19$; $z = 2.66$; $p < 0.01$), with participants giving more accurate responses when the target response was 'no.' No significant effects of *predicate* or *attachment* were observed. There was, however, a significant interaction between the two ($\hat{\beta} = 3.73$; $z = 2.87$; $p < 0.01$). To follow up on this significant interaction, we ran pairwise comparisons of *attachment* within *predicate*. For these comparisons, we present Holm-Bonferroni corrected p-values. Pairwise comparison revealed a significant effect of *attachment* within the perceptual items ($\hat{\beta} = 4.16$; $z = 2.19$; $p = 0.03$), which surfaced as greater accuracy for HA items. However, the effect of *attachment* within the non-perceptual items was non-significant ($\hat{\beta} = 0.43$; $z = 0.29$; $p = 0.77$).

As the global number of errors was, again, relatively small (N = 35, 4.26% of responses), we then inspected how the errors were distributed, to investigate if the offline results could have been due to potentially problematic items, as discussed above. We found that 22 of the 32 items received at least one wrong response (mean accuracy by item = 95.84%; SD = 3.84%). We then calculated the IQR for accuracy by item. For no items did accuracy fall 1.5 IQR below/above the $1^{st}/3^{rd}$ quartile.

4.7 Discussion

In the preceding, we have presented the behavioural portion of Experiment 2 to further investigate how PR availability affects attachment biases. Specifically, we were interested in whether we could observe evidence of a HA bias when PRs were locally available, but a LA bias with true RCs when PRs were blocked in the final interpretation of our fully disambiguated items.

The results from the comprehension questions revealed that after items with perceptual matrix predicates, participants were significantly more accurate in their response to HA items than LA ones. This is the pattern we would expect if there were an interfering PR-firstness effect. Even though the global error rates in our data were relatively small, the fact that the same effect surfaced with a similar magnitude in both the eye-tracking and self-paced reading experiments, despite the differences in methodology and participants, suggests that the pattern is unlikely to be spurious. Moreover, in neither experiment was this effect attributable to a subset of potentially problematic items. As such, the present study presents a new type of evidence in support of the PR account; not only is a PR-firstness effect observable in the acceptability judgments of fully disambiguated items (Fernandes et al., 2018; Pozniak et al., 2019), interference from PR-firstness is also observable even in the final interpretation of such items.

In a similar vein, the replicability of the offline results also raises the question of whether we should interpret the lack of a LA bias in the comprehension of non-perceptual items in both experiments to be non-accidental. That interpretation would be interesting, as it would suggest that garden-pathing participants is not sufficient for the observation of persistence effects. Rather, we would need to say something about the structure being reanalysed. For example, we might point out that PRs may occupy the DO position of the matrix predicate, whereas RCs may not. As such, we might suggest that the reanalysis of a PR to a LA RC is somehow more difficult than the reanalysis of a LA RC to a HA RC. Nonetheless, without additional replication, we should be very cautious of interpreting this potential asymmetry between our perceptual and non-perceptual items, given that it rests on a null result.

The discussion has so far not touched on two asymmetries with the previous work on PR parsing: (i) the lack of adaptation effects in the present study and (ii) the lack of interference effects with comprehension questions in Lee and De Santo (2022). Let us explore each of these in turn.

Starting with the lack of adaptation effects in either of our experiments, this may relate to the quantity or quality of non-critical items. The present study contained a noticeably higher percentage of non-critical items than in Pozniak et al. (2019) (48% vs. 69%) and, unlike the non-critical items in Aguilar et al. (2021), we included distractor items. These accounted for 46% of our non-critical items and contained a different type of ambiguity (i.e., pronominal ambiguity). Our task was also noticeably shorter than that in Fernandes et al. (2018) (32 vs. 60 critical items). However, as Pozniak et al. (2019) and Fernandes et al. (2018) employed a very different manipulation from the gender manipulation in the present study and Aguilar et al. (2021), teasing these factors apart is not straightforward and is beyond the scope of the present paper.

Regarding the fact that the present study observed an offline effect, whereas Lee and De Santo (2022) did not, we suggest that this may be due to the manipulations used to force attachment. Whereas Lee and De Santo (2022) exploited a number manipulation, the present

paper used gender marking to force attachment. Although we are not aware of any paper that has compared the effect of the two manipulations on attachment in Italian, Slioussar, Antropova, and Chernova (2022) recently presented some evidence that, at least in Russian, interference from an initial/preferred parse is more evident with a gender manipulation than a number one. Namely, when they forced participial clause attachment with or against their HA bias, the authors noted an effect of attachment within the gender-disambiguated items (HA: 89% accuracy; LA 69%). However, the same was not true for items disambiguated by number (HA: 83%; LA 82%). This asymmetry is also in line with the findings in Carminati (2005) for a different type of processing bias. She observed that when null pronouns in Italian are forced to violate their antecedent bias, participants were more readily able to overcome the structurally preferred antecedent when items were disambiguated by number than when they were disambiguated by gender.

5. Conclusion

The aim of the present study was to investigate the role of PR availability in the online parsing of 'RC' attachment ambiguities in Italian. Specifically, we were interested in whether we could observe a locality bias with true RCs and an anti-locality bias when PRs were possible. Regarding our first question, the results clearly indicate that native Italian speakers exhibit an online locality bias in the processing of true RC attachment ambiguities, as would be expected if locality were a universal processing bias. Regarding our second question, results from our self-paced reading task also indicate that Italian speakers, like the Spanish participants in Aguilar et al. (2021), exhibit an online preference for PRs over RCs when PRs are locally available, which surfaces as an apparent non-local bias. Moreover, the present results provide a novel type of evidence in support of the PR account. Namely, when strings were temporarily compatible with a PR parse, we observed that the initial preference for PRs over RCs sometimes persisted and influenced the final interpretation, despite the strings being globally disambiguated.

Abbreviations

AIC – Akaike Information Criterion

F – feminine

HA – high attachment

LA – low attachment

M – masculine

PL – plural

PR – pseudorelative

PRS – present

PST – past

RC – relative clause

SG – singular

Data accessibility statement

The critical stimuli, data, and R script are accessible at: https://osf.io/9wuc2/?view_only=33c5efd73161483486eee1cfc99d7ca2

Ethics and consent

Ethical approval was obtained from the Ethics Committee of the Faculty of Modern and Medieval Languages and Linguistics at the University of Cambridge. Prior to testing, all participants provided informed consent.

Acknowledgements

We would like to thank Pietro Baggio for his linguistic judgements when we were constructing our initial stimuli. We are also grateful to Anna Cardinaletti and members of the BemboLab for their useful comments as well as their help in making the in-person data collection for Experiment 2 possible. Thanks also to the audiences at AMLaP 2022 and the Tsimpli research seminars for their feedback.

Funding

This work was supported by funding from the Philological Society, the Cambridge Trust, the Faculty of Modern and Medieval Languages and Linguistics, and the Section of Theoretical and Applied Linguistics.

Competing interests

The authors have no competing interests to declare.

Author contributions

Alex Cairncross: conceptualisation, methodology, funding acquisition, investigation, formal analysis, writing

Margreet Vogelzang: conceptualisation, supervision, review and editing

Ianthi Tsimpli: conceptualisation, supervision, review and editing

References

Aguilar, M., Ferré, P., Gavilán, J. M., Hinojosa, J. A., & Demestre, J. (2021). The actress was on the balcony, after all: Eye-tracking locality and PR-availability effects in Spanish. *Cognition*, *211*. DOI: https://doi.org/10.1016/j.cognition.2021.104624

Aguilar, M., & Grillo, N. (2021). Spanish is *not* different: On the universality of minimal structure and locality principles. *Glossa: a journal of general linguistics*, *6*. DOI: https://doi.org/10.5334/gjgl.1251

Bates, D., MÃd'chler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). DOI: https://doi.org/10.18637/jss.v067.i01

Carminati, M. N. (2005). Processing reflexes of the Feature Hierarchy (Person > Number > Gender) and implications for linguistic theory. *Lingua*, *115*(3), 259–285. DOI: https://doi.org/10.1016/j.lingua.2003.10.006

Carreiras, M., & Clifton, C. (1993). Relative clause interpretation preferences in Spanish and English. *Language and Speech*, *36*(4), 353–372. DOI: https://doi.org/10.1177/002383099303600401

Carreiras, M., & Clifton, C. (1999). Another word on parsing relative clauses: Eyetracking evidence from Spanish and English. *Memory & Cognition*, *27*(5), 826–833. DOI: https://doi.org/10.3758/BF03198535

Casalicchio, J. (2013). *Pseudorelative, gerundi e infiniti nelle varietà romanze: Affinità (solo) superficiali e corrispondenze strutturali* (Unpublished doctoral dissertation). Università degli Studi di Padova.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*(4), 368–407. DOI: https://doi.org/10.1006/cogp.2001.0752

Cinque, G. (1992). The pseudo-relative and ACC-ing constructions after verbs of perception. *University of Venice Working Papers in Linguistics*.

Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition*, *30*(1), 73–105. DOI: https://doi.org/10.1016/0010-0277(88)90004-2

De Santo, A., & Lee, S. Y. (2022). *Pseudo-relative clause effects on the online processing of Italian relative clause attachment*. Retrieved from https://oxford-abstracts.s3.amazonaws.com/dde5ba24-7264-4a77-9451-a24f0eea8913.pdf (presented at the 35th Conference on Human Sentence Processing)

De Vincenzi, M. (1991). *Syntactic parsing strategies in Italian: The Minimal Chain principle*. Springer Science & Business Media. DOI: https://doi.org/10.1007/978-94-011-3184-1

Dussias, P. E. (2003). Syntactic ambiguity resolution in L2 learners. *Studies in Second Language Acquisition*, *25*(4), 529–557. DOI: https://doi.org/10.1017/S0272263103000238

Fernandes, B., Alexiadou, A., Chow, W.-Y., Santi, A., & Grillo, N. (2018). *Adaptation to complex cues in sentence comprehension*. Retrieved from https://eprints.whiterose.ac.uk/132816/1/CUNY_2018_abstract_final.pdf (presented at the 31st Conference on Human Sentence Processing). DOI: https://doi.org/10.31234/osf.io/8uya9

Fernández, E. M. (2003). *Bilingual sentence processing: Relative clause attachment in English and Spanish*. John Benjamins Publishing Company. DOI: https://doi.org/10.1075/lald.29

Fodor, J. D. (1998). Learning to Parse? *Journal of Psycholinguistic Research*, *27*(2), 285–319. DOI: https://doi.org/10.1023/A:1023258301588

Foppolo, F., & Abbondanza, M. (2021). *Number, gender and animcay effects on RC-attachment preferences in Italian*. Retrieved from https://amlap2021.github.io/program/113.pdf (presented at Architectures and Mechanisms for Language Processing 2021)

Frazier, L. (1978). *On comprehending sentences: Syntactic parsing strategies* (Unpublished doctoral dissertation). University of Connecticut.

Gilboy, E., Sopena, J.-M., Clifton, C., & Frazier, L. (1995). Argument structure and association preferences in Spanish and English complex NPs. *Cognition*, *54*(2), 131–167. DOI: https://doi.org/10.1016/0010-0277(94)00636-Y

Grillo, N. (2012). Local and universal. In V. Bianchi & C. Chesi (Eds.), *Enjoy linguistics! Papers offered to Luigi Rizzi on the occasion of his 60th birthday* (pp. 234–245). CISCL Press. Retrieved from http://eprints.whiterose.ac.uk/110188/1/Grillo_2012_Local_and_Universal.pdf

Grillo, N., & Costa, J. (2014). A novel argument for the universality of parsing principles. *Cognition*, *133*(1), 156–187. DOI: https://doi.org/10.1016/j.cognition.2014.05.019

Grillo, N., Costa, J., Fernandes, B., & Santi, A. (2015). Highs and lows in English attachment. *Cognition*, *144*, 116–122. DOI: https://doi.org/10.1016/j.cognition.2015.07.018

Holbrook, A. (2008). Acquiescence response bias. In P. Lavrakas (Ed.), *Encyclopedia of survey research methods*. Sage Publications, Inc. DOI: https://doi.org/10.4135/9781412963947

Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition*, *2*(1), 15–47. DOI: https://doi.org/10.1016/0010-0277(72)90028-5

Lee, S. Y., & De Santo, A. (2022). *Online processing effects of pseudo relative clasue availability in Italian attachment ambiguity*. Retrieved from https://www.ling.upenn.edu/Events/PLC/plc46/PLC46Abstracts.pdf (presented at the 46th Penn Linguistics Conference).

Lenth, R. V. (2022). emmeans: Estimated marginal means, aka least-squares means [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=emmeans (R package version 1.7.5).

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. DOI: https://doi.org/10.1016/j.jml.2017.01.001

Mitchell, D. C., & Cuetos, F. (1991). The origins of parsing strategies. *Current issues in natural language processing,* 1–12. Retrieved from https://www.researchgate.net/publication/239666294_The_origins_of_parsing_strategies

Moulton, K., & Grillo, N. (2015). Pseudo relatives: Big and direct. In T. Bui & D. Ozyildiz (Eds.), *Proceedings of 45th annual meeting of the North Eastern Linguistic Society* (pp. 193–202). Retrieved from https://www.researchgate.net/publication/303034807_Pseudo-Relatives_Big_and_Direct

Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing. *Studies in Second Language Acquisition*, *25*(4), 501–528. DOI: https://doi.org/10.1017/S0272263103000214

Pozniak, C., Hemforth, B., Haendler, Y., Santi, A., & Grillo, N. (2019). Seeing events vs. entities: The processing advantage of pseudo relatives over relative clauses. *Journal of Memory and Language*, *107*, 128–151. DOI: https://doi.org/10.1016/j.jml.2019.04.001

R Core Team. (2022). R: A language and environment for statistical computing [Computer software manual]. Retrieved from https://www.R-project.org/

Radford, A. (1975). Pseudo-relatives and the unity of subject-raising. *Archivum Linguisticum*, *6*, 32–64.

Rohde, H., Levy, R., & Kehler, A. (2011). Anticipating explanations in relative clause processing. *Cognition*, *118*(3), 339–358. DOI: https://doi.org/10.1016/j.cognition.2010.10.016

Shen, X. (2006). *Late assignment of syntax theory: Evidence from Chinese and English* (Unpublished doctoral dissertation). University of Exeter.

Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language*, *69*(2), 104–120. DOI: https://doi.org/10.1016/j.jml.2013.04.001

Slioussar, N., Antropova, D., & Chernova, D. (2022). *To forget or not to forget: processing gender, number and case features.* (Presented at AMLaP 2022).

Sturt, P. (2007). Semantic re-interpretation and garden path recovery. *Cognition*, *105*(2), 477–488. DOI: https://doi.org/10.1016/j.cognition.2006.10.009

Zagar, D., Pynte, J., & Rativeau, S. (1997). Evidence for early closure attachment on first pass reading times in French. *The Quarterly Journal of Experimental Psychology Section A*, *50*(2), 421–438. DOI: https://doi.org/10.1080/713755715

Zehr, J., & Schwarz, F. (2018). PennController for internet based experiments (IBEX) [Computer software manual]. Open Science Framework. DOI: https://doi.org/10.17605/OSF.IO/MD832