

Language, Cognition and Neuroscience

ISSN: 2327-3798 (Print) 2327-3801 (Online) Journal homepage: www.tandfonline.com/journals/plcp21

Online processing of *wh*-questions in Italian: evidence from eye tracking

Elena Pagliarini, Mirta Vernice & Maria Teresa Guasti

To cite this article: Elena Pagliarini, Mirta Vernice & Maria Teresa Guasti (2025) Online processing of *wh*-questions in Italian: evidence from eye tracking, Language, Cognition and Neuroscience, 40:7, 1002-1014, DOI: 10.1080/23273798.2025.2501048

To link to this article: https://doi.org/10.1080/23273798.2025.2501048

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 12 May 2025.

Submit your article to this journal

Article views: 1124

View related articles

View Crossmark data

Citing articles: 1 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=plcp21

LANGUAGE, COGNITION AND NEUROSCIENCE
2025, VOL. 40, NO. 7, 1002–1014
https://doi.org/10.1080/23273798.2025.2501048

Routledge
Taylor & Francis Group

REGULAR ARTICLE

OPEN ACCESS Check for updates

Online processing of *wh*-questions in Italian: evidence from eye tracking

Elena Pagliarini [a], Mirta Vernice [b] and Maria Teresa Guasti [c]

[a]Centre de Lingüística Teòrica, Departament de Filologia Catalana, Universitat Autònoma de Barcelona, Belleterra, Spain; [b]Dipartimento di Studi Umanistici, Università di Urbino Carlo Bo, Urbino, Italy; [c]Dipartimento di Psicologia, Università degli Studi di Milano-Bicocca, Milan, Italy

ABSTRACT

This study investigates the processing of subject and object direct questions headed by *wh*-elements such as *chi* (i.e. *who*) and *quale-NP* (i.e. *which-NP*) in Italian, a pro-drop language that allows post-verbal subjects. In Italian, *chi* and *quale-NP* questions are ambiguous: they can be interpreted as either subject or object, in the absence of morphological or pragmatic disambiguation. Using eye tracking techniques, we explore how different types of *wh*-elements (*chi* vs. *quale-NP*) and extraction sites (subject vs. object) are processed when disambiguated by number verbal agreement. A distinct processing pattern emerges between *chi* and *quale-NP* questions: while *quale-NP* questions generally demand more cognitive effort when encountering the verb, *chi* questions necessitate increased processing during post-verbal NP integration. Additionally, a *wh*-subject extraction preference is observed, which reflects an active gap filling strategy across types of *wh*-elements. These findings are discussed in terms of the processing mechanisms underlying the elaboration of *wh*-questions.

ARTICLE HISTORY
Received 31 January 2024
Accepted 14 April 2025

KEYWORDS
Wh-questions; subject-object asymmetry; syntactic parsing; eye tracking; Italian

Introduction

In Italian, direct *wh*-questions introduced by *chi* (i.e. who) and *quale-NP* (i.e. which-NP) could be either a subject ((1a), (2a)) or an object ((1b), (2b)) question, in the absence of pragmatic or morphological disambiguation. Additionally, since Italian is a pro-drop language that allows post-verbal subjects, the post-verbal NP can function as either the sentence's direct object or as an inverted subject. Parsing *wh*-questions involves linking a particular *wh*-element back to its original position, in order to retrieve the correct interpretation. In the literature, the moved constituent is called filler whereas the empty base-position is called gap.

(1) Chi bacia il ragazzo?
Who kisses the boy?
a. Who$_i$ t$_i$ kisses the boy? (subject-extraction)
b. Who$_k$ does the boy kiss t$_k$? (object-extraction)

(2) Quale ragazza bacia il ragazzo?
Which girl kisses the boy?
a. Which girl$_i$ t$_i$ kisses the boy? (subject-extraction)
b. Which girl$_k$ does the boy kiss t$_k$? (object-extraction)

The processing of *wh*-question has been extensively studied across languages and populations (adults, children, agrammatical aphasics, children with language impairments). Various measures like reaction times, reading times, and event-related brain potentials revealed a common preference for *wh*-subject extraction, namely the displaced *wh*-element is initially preferentially taken to be the subject of the sentence (Contemori et al., 2018; Crain & Fodor, 1985; De Vincenzi, 1991a; Friedmann & Novogrodsky, 2011; Grimaldi et al., 2023; Guasti et al., 2012; Penolazzi et al., 2005; Schlesewsky et al., 2000; Stowe, 1986; Tanenhaus et al., 1989; Ueno & Kluender, 2009).

Different theoretical accounts have been proposed to explain the subject-bias. According to the *Minimal Chain Principle* (De Vincenzi, 1991b) and the *Active Filler Strategy* (Frazier, 1987; Frazier & Clifton, 1989; Frazier & D'Arcais, 1989), the syntactic processor assigns the filler to the first possible gap position within the sentence, which is the subject position. In terms of processing, since the parser has a structural preference for the *wh*-subject extraction, if the *wh*-element turns out to be the object of the sentence, the parser will have to revise the structure, because its initial analysis is incompatible with the overall structure of the sentence. The parser's tendency to favour the early completion of dependencies has also been observed in the processing of backward anaphora (Drummer & Felser, 2018; Giskes & Kush, 2021) and backward sluicing, a form of clausal ellipsis (Kim, Orth, et al., 2024).

CONTACT Elena Pagliarini elena.pagliarini@uab.cat Centre de Lingüística Teòrica, Departament de Filologia Catalana, Universitat Autònoma de Barcelona, Belleterra 08193, Spain

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

The preference for *wh*-subject extraction is also predicted by probabilistic or experience-based accounts, which propose that the distributional frequencies of subject questions determine their processing ease relative to object questions (Hale, 2001; Levy, 2008; Roland et al., 2007).

Here we focus on the processing of *chi* and *quale-NP* questions in Italian, as they share similarities but also differences. From a linguistic perspective, *chi* does not bear a referential index, but it enters into an operator-variable relation (Cinque, 1999). *Quale-NP* is a discourse linked phrase; it can enter into an antecedent-pronoun relation with an overt or covert resumptive pronoun due to its referential status (Cinque, 1999; Pesetsky, 1987). From a psycholinguistic perspective, *chi* is a filler that actively seeks for a gap and must be assigned to a gap position as soon as possible (which will be the subject position), whereas *quale-NP* is to be assigned a referential index and its interpretation is postponed until the lexical information about the verb and the following NP have been accessed (De Vincenzi, 1991a). In other words, in case of *a which-NP element*, the parser does not need to immediately assign it to the subject position. According to De Vincenzi (1991a), in syntactically and pragmatically ambiguous questions, this decision is taken at the post-verbal NP. However, recent findings challenge the distinction between D-linked and non-D-linked *wh*-elements. Kim, Orth, et al. (2024) tested the processing of D-linked *wh*-elements in simple filler-gap dependency constructions and backward sluicing constructions, showing that in both contexts, the parser actively searches for the antecedent as soon as possible, that is D-linked *wh*-elements involve a common active search mechanism.

Previous studies in Italian have confirmed that the parser prefers *wh*-subject extraction for *chi* questions, as showed by increased reading time and a P300 and a P600 at the verb region in case of *wh*-object extraction (De Vincenzi et al., 2004; Penolazzi et al., 2005). On the contrary, controversial results were found in the case of *quale-NP* questions. A self-paced reading study showed increased reading time at the verb region and at the post-verbal NP region of the *wh*-object extraction (Grimaldi et al., 2023), suggesting that the parser postpones its decision until it reaches the post-verbal NP. As a confirmation, in *quale-NP* questions, only a P600 emerged at the verb region for the *wh*-object extraction, confirming the absence of quick anomaly detection (no P300) (Grimaldi et al., 2023). However, another self-paced reading study didn't reveal any structural preference for *wh*-subject extraction for *quale-NP* questions (De Vincenzi, 1991a).

No prior research has directly contrasted how *chi* and *quale-NP* questions are processed. Our study aims to bridge this gap by using a silent reading task with eye tracking to analyse the impact of different *wh*-elements (*quale-NP* vs. *chi*) on questions featuring numerical morphology disambiguation on the verb. Eye tracking enables us to discern early and late processing effects, unveiling distinct stages in sentence elaboration. Three research questions were addressed: (i) Are there processing distinctions between these *wh*-elements? (ii) Is there a processing preference for *wh*-subject extraction in both *chi* and *quale-NP* questions? (iii) Do these *wh*-elements differ in how they integrate the post-verbal NP? We grounded our theoretical assumptions in the *Minimal Chain Principle* (De Vincenzi, 1991a, 1991b) and in the *Active Filler Strategy* (Frazier, 1987; Frazier & Clifton, 1989; Frazier & D'Arcais, 1989). Accordingly, we hypothesise prolonged reading times and more regressions at the verb region for both question types during *wh*-object extraction, due to reanalysis from a (preferred) *wh*-subject extraction to a *wh*-object extraction. Although no prior studies directly compare *chi* and *quale-NP* questions, following the prediction of De Vincenzi (1991a) and based on the results of Grimaldi et al. (2023), we hypothesise an early penalty for object-extraction with *chi*, which may exceed that for *quale-NP* if *chi* elicits an early commitment whereas the attachment choice for *quale-NP* is delayed. Building on this prediction, we expected higher integration costs at the post-verbal NP if the parser delays its decision until reaching the post-verbal NP in *quale-NP* questions, regardless of the site of *wh*-extraction. Additionally, extended processing costs at the post-verbal NP in *quale-NP* questions are also predicted under a discourse-processing-based distance metric (Warren & Gibson, 2002) which is grounded in the distance-based dependency locality theory (DLT) (Gibson, 1998, 2000). According to the DLT, the comprehension of a question requires at least two types of computational resources. First, it requires structural integration, which involves linking each new word to the current sentence structure. Second, it requires structural storage, which refers to the cost needed to keep track of unfinished dependencies as the sentence unfolds. Gibson suggested that when a new syntactic head (*h*) is introduced into a sentence, the parser must retrieve (and reactivate) the existing elements in the structure that *h* connects to syntactically. Assuming a reactivation-based model with a storage component, in both *chi* and *quale-NP* questions (i) the filler must be stored in memory and (ii) the filler must be retrieved/reactivated at the endpoint of the dependency. Nevertheless, other factors distinguishing the two types of *wh*-

elements can influence the storage-retrieval cost. In *chi* questions, the relative pronoun must be held in memory, and after encountering the verb, an additional NP needs to be integrated into the sentence structure. In *quale-NP* questions, the *wh*-element together with its pied-piped NP must be stored in memory, and later, another NP must be incorporated into the structure. Given the referentiality nature of definite descriptions (Warren & Gibson, 2002) and the similarity between the two stored elements which could lead to memory interference process (Gordon et al., 2001; Kim, Wellwood, et al., 2024), increased processing cost is expected at the post-verbal NP in case of *quale-NP* questions.

Method

Participants

Thirty-one Italian students from the University of Milano-Bicocca (mean age 26;24 years; SD: 4;15; 8 males) with normal or corrected-to-normal vision voluntarily participated in the study. Informed consent was collected prior to testing. The study was approved by the Ethics committee of the University of Milano-Bicocca (prot. 20974/13) according to the standards of the Helsinki Declaration (1964).

Materials

The experiment used 48 sets of sentences based on those used in Grimaldi et al. (2023) (available at OSF https://osf.io/edhbv/) and 48 filler sentences.

The experiment used a 2 × 2 factorial design manipulating the type of *wh*-element (*chi* vs. *quale-NP*) and the extraction site (subject vs. object). The sentences were designed to differ only in the regions containing the *wh*-element (region 2: *chi* vs. *quale-NP*) and the verb (region 3: singular vs. plural), as the verbal morphology signalled the extraction site (see Table 1). This means that when the *wh*-element is encountered, the sentence is still ambiguous, as the parser cannot yet determine whether it is a subject or an object question. Disambiguation does not occur until the verb is processed, since verbal morphology always matches the subject. Therefore, in (3a) and (3c) the *wh*-element must be interpreted as the subject, because it agrees with the verb. Conversely, in (3b) and (3d) the post-verbal NP agrees with the verb, indicating that it is the actual subject, which forces the *wh*-element to be interpreted as the object.

The four conditions of each item were distributed across four lists using Latin square counterbalancing. We created four lists, such that each list contained an equal number of different versions of the items, and such that each item appeared an equal number of times in all the versions of the experimental set, resulting in 12 trials per condition.

The sentences were presented in random order, interspersed with 24 yes/no comprehension questions, which were underlined to distinguish them from the experimental items. Subjects with an accuracy rate below the 75th percentile were excluded based on their performance on comprehension questions. A low accuracy rate is likely to reflect inattention rather than a difficulty in processing the task, as the comprehension questions required simply understanding the sentences. Therefore, we chose these criteria to ensure that only participants who were fully focused on the task were included in the analyses.

Data acquisition and analysis

Eye-movements were recorded with an Eyelink 1000 system (SR-Research, Ottawa, CA), tracking the dominant eye and using the desktop mount configuration. Each sentence was displayed on one line. Six participants were excluded for various reasons, including poor eye tracking data (2 participants), comprehension scores below 75% (2), or bilingualism (1). The final dataset included a total of 25 participants whose accuracy on the comprehension questions was 89% (SD = 3%).

Five canonical reading time measures (Rayner et al., 2005) were computed: (i) first pass reading time, i.e. sum of first pass fixation durations in an interest area excluding instances in which participants skipped the relevant region during the first pass; (ii) first pass regression out probability, i.e. whether a regression from a region to a previous one occurred before the region was exited in a forward direction; (iii) total reading time, i.e. the total duration of all fixations within a region; (iv) regression out full probability, i.e. whether a regression from a region to a previous one occurred, encompassing all regressions; (v) regression out full count, i.e. the number of times a region was left towards a previous region. Note that overall, these measures offer complementary insights into sentence processing. First-pass reading time and first-pass regression out probability are generally considered a measure of early processing (e.g. Falkauskas & Kuperman, 2015; Schmidtke et al., 2018), capturing early lexical and syntactic processing (Rayner, 1998). Total reading time, regression out full probability and regression out full count are generally considered an index of overall processing difficulty that includes both early and late processing (Clifton et al., 2007).

The sentences were divided into five regions, see (3a) – (3d) on Table 1. The analyses focused on regions 3 and 4, as these are the critical regions of interest: region 3 corresponds to the verb, which carries the morphological

Table 1. Sample experimental stimuli by condition.

	Region 1	Region 2	Region 3	Region 4	Region 5	Condition
(3) a.	Puoi dirmi	chi	saluta	i nonni	alla partenza in stazione?	*chi, subject-extraction*
	Can you tell me	*Who*	*greet-3SG*	*the grandparents*	*at the departure at the station?*	
(3) b.	Puoi dirmi	chi	salutano	i nonni	alla partenza in stazione?	*chi, object-extraction*
	Can you tell me	*Who*	*greet-3PL*	*the grandparents*	*at the departure at the station?*	
(3) c.	Puoi dirmi	quale nipote	saluta	i nonni	alla partenza in stazione?	*quale, subject-extraction*
	Can you tell me	*which grandchild*	*greet-3SG*	*the grandparents*	*at the departure at the station?*	
(3) d.	Puoi dirmi	quale nipote	salutano	i nonni	alla partenza in stazione?	*quale, object-extraction*
	Can you tell me	*which grandchild*	*greet-3PL*	*the grandparents*	*at the departure at the station?*	

information crucial for disambiguation, while region 4 contains the post-verbal NP. These regions were chosen to minimise the effect of regions' length. Note that region 3 varied in length by a maximum of only two characters across conditions, while region 4 was identical in length across all conditions. Given the minor length differences in these regions, they were considered suitable for analysis. In contrast, region 2 was excluded due to substantial differences in length between *chi* and *quale-NP* questions. This mismatch in length would have made it difficult to determine whether differences in reading time were caused by syntactic processing or simply by length variations.

To account for the confounding effect of character length in regions 3 and 4, in first pass and total reading time, we first applied a log transformation to reduce skewness and approximate a normal distribution. We then extracted residuals from preliminary mixed-effects models that incorporated length (i.e. number of characters) as a predictor along with random effects (see Marelli & Baroni, 2015). The residuals partialled out the variance explained by length, allowing us to analyse our experimental factors more accurately (Chacón, 2022; Ferreira & Clifton, 1986; Miyamoto, 2003). Regarding regression out full count, we analysed the raw count data without any transformation. Note indeed that count measures might have a discrete and skewed distribution. However, given the relatively small range of values in our dataset, we avoid transformation to preserve a more direct interpretation of the data without introducing distortions due to data transformations. The remaining regression variables were treated as categorical.

Data and code used for data analyses are provided on the OSF. Data of regions 3 and 4 were analysed using mixed-effects models in R. Specifically, Linear Mixed Models (LMM) were used for continuous variables (residuals of first-pass and total reading time), and Generalised Linear Mixed Models (GLMM) were applied for categorical data (first-pass regression out and regression out full). As for regression out full count, as the data was characterised by overdispersion and zero inflation, we employed a Zero-Inflated Negative Binomial (ZINB) model (Green, 2021). This model allows us to account for the distribution of counts and the excess of zero observations. All the models were fitted using the lmerTest (for LMM and GLMM; Kuznetsova et al., 2017) and glmmTMB (for ZINB; Brooks et al., 2017) packages in R (4.3.3).

In all models, we included the same set of fixed factors and their interaction, namely type of *wh*-element (*chi* vs. *quale-NP*), extraction site (subject-extraction vs. object-extraction), and their interaction. Additionally, length (i.e. the number of characters in the region) was included only for categorical and count variables. All the models involved the fixed effects and their interaction, regardless of statistical significance. This choice was motivated by theoretical considerations, as it ensures that the models accurately capture the structure of the experimental design and the underlying hypotheses. Additionally, note that removing non-significant predictors can bias the interpretation of the remaining effects (Gelman & Hill, 2007), particularly in cases where interactions are theoretically expected based on previous research (cf. Grimaldi et al., 2023). In each model, to determine the best-fitting random structure, we started with a maximal model including all relevant random effects and random slopes. If convergence issues arose, we performed model comparisons to refine the structure. The parameter estimates for models fitted to continuous dependent variables are associated with *t* values, while those of models fitted to categorical or count data are associated with z values.

Figures 1(a,b), 2(a,b), 3(a,b) report the line graphs (a) the heatmaps (b) for first pass reading times, total reading times and first pass regression out probability in regions 3 and 4 as a function of extraction site and *wh*-element. Table 2 presents the descriptive statistics in regions 3 and 4 for all the variables considered. Finally, Table 3 summarises the parameters of the best-fitting models in the regions of interest.

Region 3 – verb

First-pass reading times were significantly longer in *chi* questions compared to *quale-NP* questions ($\beta = -0.20$,

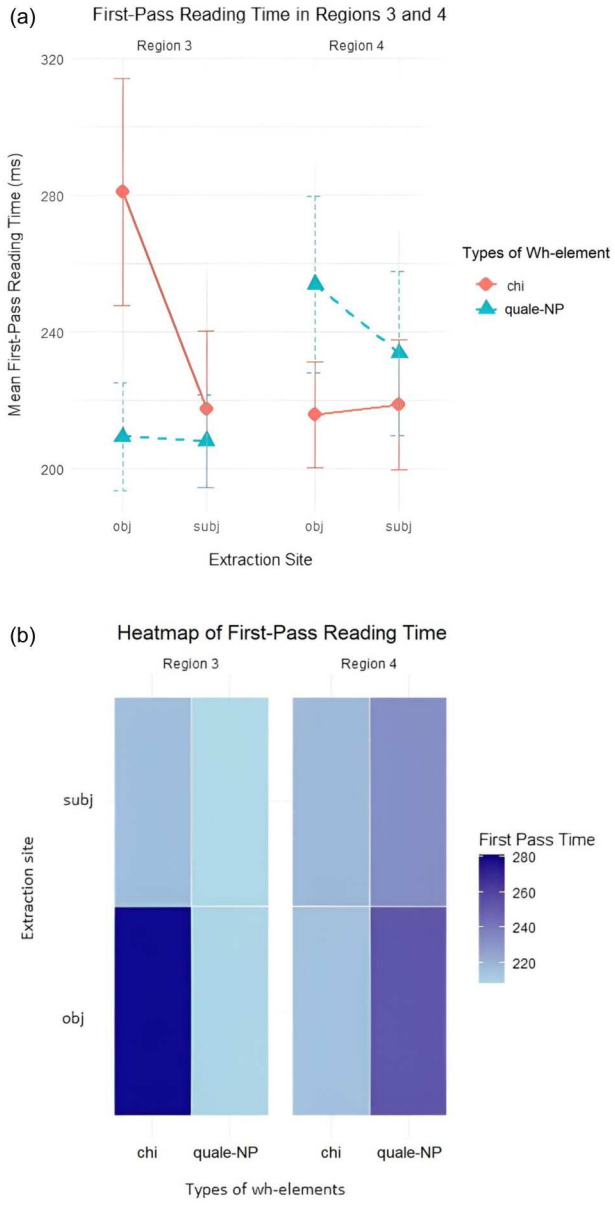

Figure 1. (a) The line graph represents the mean first-pass reading time as a function of extraction sites labelled as "subj" (subject extraction) and "obj" (object extraction) and types of *wh*-element (*chi* vs. *quale-NP)* in Regions 3 and 4. It shows the mean duration of first-pass reading for *chi* and *quale-NP* questions, with error bars indicating the standard error of the mean (SE). Solid lines correspond to *chi* questions, while dashed lines represent *quale-NP* questions. (b) The graph represents the heatmap of the mean first-pass reading time in regions 3 and 4. The heatmap shows the average first-pass reading times for types of *wh*-element (rows) and extraction site (columns) in regions 3 and 4. Darker shades of blue represent longer reading times, while lighter shades indicate shorter times.

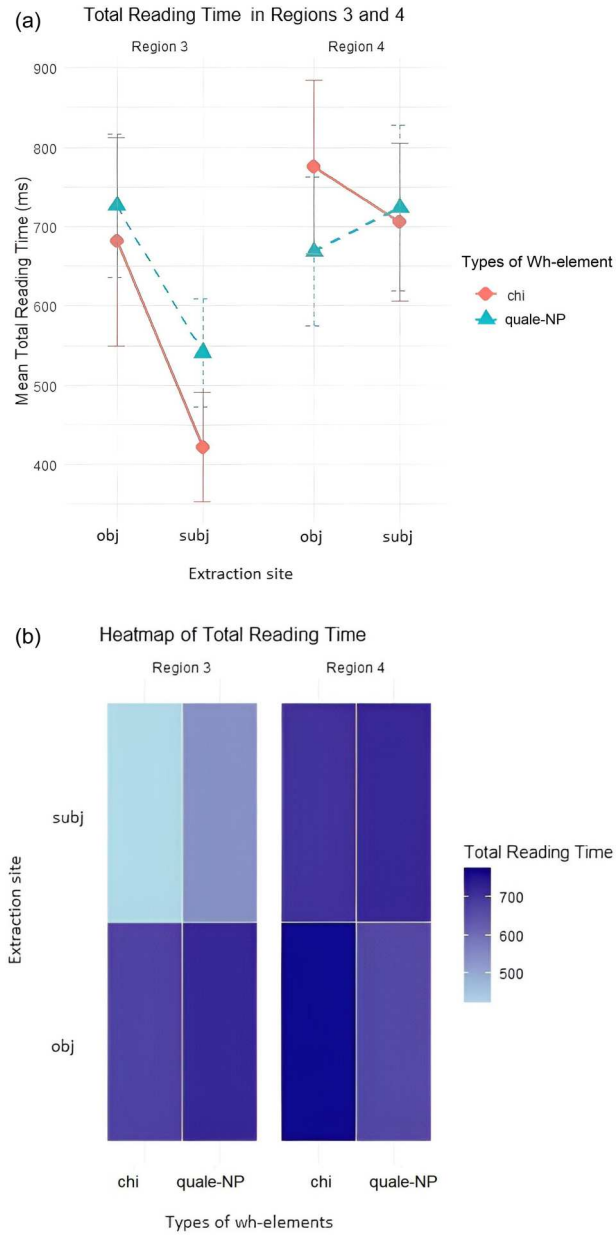

Figure 2. (a) The line graph represents the mean total reading time as a function of extraction sites labelled as "subj" (subject extraction) and "obj" (object extraction) and types of *wh*-element (*chi* vs. *quale-NP)* in Regions 3 and 4. It shows the mean duration of total reading time for *chi* and *quale-NP* questions, with error bars indicating the standard error of the mean (SE). Solid lines correspond to *chi* questions, while dashed lines represent *quale-NP* questions. (b) The graph represents the heatmap of mean total reading time in regions 3 and 4. The heatmap shows the average total reading times for types of *wh*-elements (rows) and extraction site (columns) in regions 3 and 4. Darker shades of blue represent longer reading times, while lighter shades indicate shorter times.

SE = 0.06, $t = -3.56$, $p < .001$) and in object-extraction compared to subject-extraction questions ($\beta = -0.22$, SE = 0.07, $t = -3.14$, $p = .002$). The significant interaction ($\beta = 0.22$, SE = 0.08, $t = 2.80$, $p = .005$; Figure 1(a)) was unpacked by means of a series of pairwise comparisons performed using the *emmeans* package (Lenth, 2024). Object-extracted *chi* questions led to significantly slower first-pass than subject-extracted *chi* ($\beta = 0.22$,

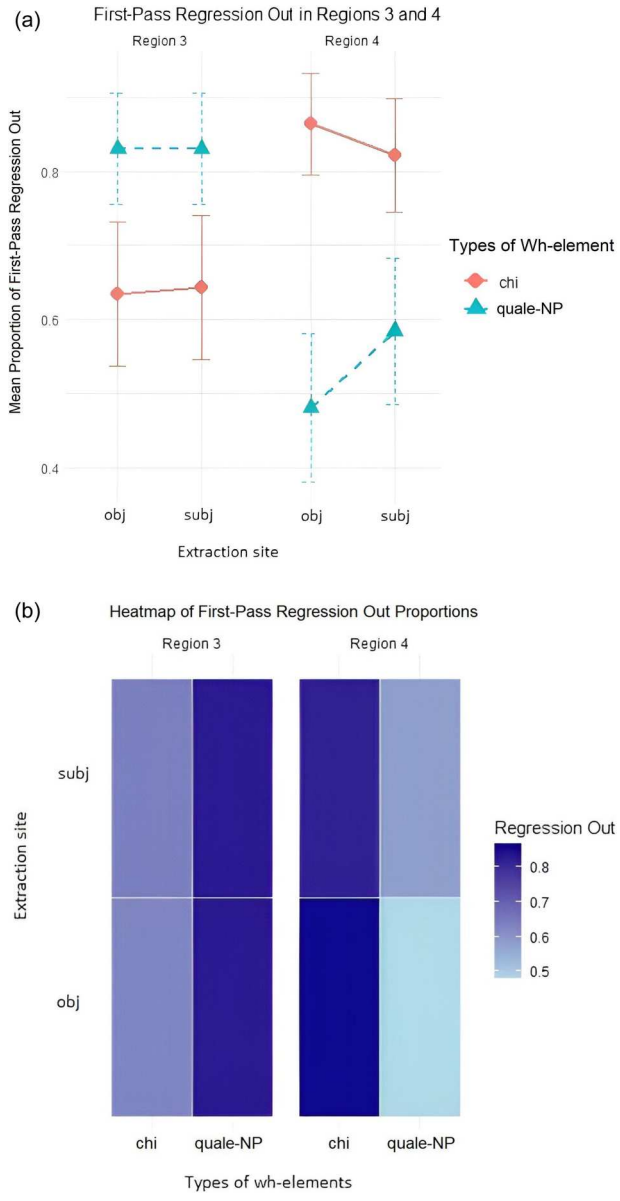

Figure 3. (a) The line graph represents the proportion first-pass regression out as a function of extraction sites labelled as "subj" (subject extraction) and "obj" (object extraction) and wh-element (*chi* vs. *quale*-NP) in Regions 3 and 4. It shows the mean probability of regressions out of the interest area for *chi* and *quale*-NP questions, with error bars indicating the standard error of the proportion (SE). Solid lines correspond to *chi* questions, while dashed lines represent *quale-NP* questions. (b) Heatmap of first-pass regression out proportions in regions 3 and 4. The heatmap shows the mean proportions of regression out occurrences for types of *wh*-elements (rows) and extraction site (columns) in regions 3 and 4. Darker shades of blue represent higher proportions, while lighter shades indicate lower proportions.

SE = 0.07, $t = 3.11$, $p = .012$). Additionally, object-extracted *chi* questions were read more slowly than both object-extracted *quale-NP* ($\beta = 0.20$, SE = 0.06, $t = 3.49$, $p = .003$) and subject-extracted *quale-NP* ($\beta = 0.20$, SE = 0.06, $t = 3.36$, $p = .005$). By contrast, no significant differences emerged between subject-extracted *chi* and object-extracted *quale-NP* ($\beta = -0.03$, SE = 0.06, $t = -0.46$, $p = 1$), nor between subject-extracted *chi* and subject-extracted *quale-NP* ($\beta = -0.03$, SE = 0.06, $t = -0.43$, $p = 1$). Similarly, object-extracted *quale-NP* and subject-extracted *quale-NP* did not differ significantly ($\beta = 0.002$, SE = 0.04, $t = 0.01$, $p = 1$). Overall, these results suggest that object-extracted *chi* leads to increased processing costs compared to both subject-extracted *chi* and object-extracted *quale-NP*. However, extraction site does not seem to modulate the reading times for *quale-NP* sentences, potentially indicating that the presence of an NP influences sentence processing differently than the bare *wh*-word *chi*. Moving to regressions, we found that regressions during the first pass of the verb region were more likely in *quale-NP* questions (first-pass regression out: $\beta = 3.44$, SE = 0.61, $z = 5.60$, $p < .001$) (see Figure 2(b); Table 4) .

Considering late processing measures, *quale-NP* questions resulted in significantly longer total reading times ($\beta = 0.23$, SE = 0.12, $t = 1.93$, $p = .05$). The significant interaction ($\beta = 0.32$, SE = 0.17, $t = 1.93$, $p = .05$), that can be visually inspected in Figure 2(a), was unpacked by means of pairwise comparisons performed using the

Table 2. Means and standard deviations for reading times and regression measures for regions 3 and 4 as a function of *wh*-element (*chi* vs. *quale-NP*) and extraction site (subject-extraction vs. object-extraction).

		3		4	
		M	SD	*M*	SD
First pass Reading Time	*chi* object	281.10	166.79	216.00	77.70
	chi subject	217.52	115.19	218.78	95.72
	quale-NP object	209.38	79.04	254.05	129.03
	quale-NP subject	208.14	67.90	233.85	119.79
(First pass) Regression out	*chi* object	0.63	0.49	0.86	0.34
	chi subject	0.64	0.49	0.82	0.38
	quale-NP object	0.83	0.38	0.48	0.50
	quale-NP subject	0.83	0.38	0.58	0.49
Total reading time	*chi* object	680.98	654.84	775.33	550.08
	chi subject	422.02	344.67	706.27	497.74
	quale-NP object	726.19	449.80	668.17	469.09
	quale-NP subject	541.00	339.60	723.48	518.64
Regression out full probability	*chi* object	0.68	0.47	0.88	0.33
	chi subject	0.64	0.49	0.84	0.37
	quale-NP object	0.85	0.35	0.57	0.50
	quale-NP subject	0.83	0.38	0.66	0.47
Regression out full count	*chi* object	0.90	0.83	1.36	0.89
	chi subject	0.95	0.85	1.16	0.80
	quale-NP object	1.23	0.84	0.84	0.89
	quale-NP subject	1.12	0.71	0.94	0.85

Table 3. Schematic summary of the main findings for regions 3 and 4.

REGION 3	Types of *wh*-elements (*chi* vs. *quale-NP*)	Extraction sites (subject vs. object)	Types of *wh*-elements*Extraction sites
First Pass	*** *quale-NP* < *chi*	**obj > subj	**
Total Reading Time	**quale-NP* > *chi*	ns	*
First Pass Regression Out	****quale-NP* > *chi*	ns	ns
Regression Out Full Probability	** *quale-NP* > *chi*	*obj > subj	ns
Regression Out Full Count	** *quale-NP* > *chi*	***obj > subj	ns
REGION 4	**Types of wh-elements (*chi* vs. *quale-NP*)**	**Extraction sites (subject vs. object)**	**Types of wh-elements*Extraction sites**
First Pass Reading Time	*** *quale-NP* > *chi*	ns	ns
Total Reading Time	ns	ns	ns
First Pass Regression Out	****quale-NP* < *chi*	ns	*
Regression Out Full Probability	****quale-NP* < *chi*	ns	ns
Regression Out Full Count	****quale-NP* < *chi*	ns	ns

Notes: The symbols "<" and ">" should be interpreted as follows: in continuous variables, the caption "subject < object" means significantly lower reading times in the subject-extraction condition than in the object-extraction condition, conversely "subject > object" indicates longer reading times in the subject-extraction condition. In categorical variables, "subject < object" indicates a reduced probability of making regressions in the subject-extraction condition. A similar logic applies to "*quale-NP* < *chi*" and "*quale-NP* > *chi*". *Ns* refers to non-significant.

emmeans package (Lenth, 2024), revealing significant differences in total reading times across conditions. Specifically, subject-extracted *chi* questions were read significantly faster than subject-extracted *quale-NP* ($\beta = -0.55$, SE = 0.12, $t = -4.58$, $p = .001$), suggesting that lexical processing caused additional elaboration in subject-extraction. Additionally, we observed two distinct effects: subject-extracted *quale-NP* questions were read significantly faster than object-extracted *chi* ($\beta = 0.37$, SE = 0.12, $t = 3.06$, $p = .0201$) and subject-extracted *chi* questions were read significantly faster than object-extracted *quale-NP* ($\beta = -0.41$, SE = 0.12, $t = -3.47$, $p = .006$), indicating a cumulative effect of both site of extraction and lexical processing. By contrast, no significant differences were observed between object- and subject-extracted *quale-NP* questions ($\beta = -0.14$, SE = 0.12, $t = -1.15$, $p = 1$). Similarly, there was no significant difference in reading times between object- and subject-extracted *chi* questions ($\beta = 0.19$, SE = 0.12, $t = 1.55$, $p = .74$), nor between sentences where *chi* and *quale-NP* involved object-extraction ($\beta = -0.23$, SE = 0.12, $t = -1.90$, $p = .37$).

Overall, these results indicate that subject-extraction led to faster reading times as compared to object-extractions in *chi* questions, while in *quale-NP* there were additional processing costs, leading to increased reading times even in subject-extraction. The lack of significant differences in total reading times between *chi* and *quale-NP* in object-extraction suggests that the overall processing cost of object-extraction might mask the additional difficulty of integrating the NP with the verb.

Regarding regressions, participants were significantly more likely to regress to previous regions in *quale-NP* questions (regression out full probability: $\beta = 0.75$, SE = 0.23, z = 3.25, $p = .001$; regression out full count: $\beta = 0.33$, SE = 0.10, z = 3.47, $p < .001$). Additionally, regressions were more frequent in object-extraction than in subject-extraction questions (regression out full probability: $\beta = -0.57$, SE = 0.27, z = −2.12, $p = .034$), though this effect was only marginal for regression out full count ($\beta = -0.25$, SE = 0.14, z = −1.86, $p = .06$).

Taken together, these results suggest that upon encountering the verb, *quale-NP* questions, especially in object-extraction, trigger more regressions, indicating a greater need for reanalysis. Specifically, we propose that in the object-extraction condition, the parser initially commits to a subject interpretation. Upon encountering the verb, verbal agreement disconfirms this initial analysis, requiring a revision towards an object interpretation. The increased regressions in *quale-NP* object-extraction structures thus indicate greater processing difficulty associated with this reanalysis.

To summarise, the subject-object asymmetry appears to manifest differently across the two types of *wh*-element, emerging in early measurements for *chi* and mainly in late measurements for *quale-NP*.

Region 4 – post-verbal NP

During the first pass of the sentence, *quale-NP* elicited significantly longer first-pass reading times ($\beta = 0.10$, SE = 0.03, $t = 3.43$, $p < .001$). In contrast, *chi* questions led to more first-pass regressions out to previous regions ($\beta = -2.54$, SE = 0.26, z = −9.63, $p < .001$) and to a significant interaction ($\beta = .82$, SE = 0.35, z = 2.37, $p < .02$) (see Figure 3(b)). The results of pairwise comparisons (Lenth, 2024) indicate that the likelihood of making a regression was significantly higher with *chi* questions in object-extraction. Specifically, the probability of making a regression was significantly higher in object-extracted *chi* questions as compared to both object-extracted *quale-NP* ($\beta = 2.54$, SE = 0.26, z = 9.63, $p < .001$) and subject-extracted *quale-NP* ($\beta = 2.026$, SE = 0.25, z = 8.08, $p < .001$). However, the difference

Table 4. Summary of the best-fitting models in the different regions of interest.

Dependent Variable	Effect	Estimate	SE	t/z-value	*p*-value
Region 3					
First Pass Reading Time (residuals)	(Intercept)	0.18	0.05	3.58	0.001***
	Types of wh-elements	−0.20	0.06	−3.56	0.001***
	Extraction sites	−0.22	0.07	−3.14	0.01**
	Interaction	0.22	0.08	2.80	0.01**
Total Reading Time (residuals)	(Intercept)	−0.10	0.08	−1.22	0.22
	Types of wh-elements	0.23	0.12	1.93	0.05*
	Extraction sites	−0.18	0.12	−1.57	0.12
	Interaction	0.32	0.17	1.93	0.05*
Regression Out Full Count	(Intercept)	−1.26	0.40	−3.15	0.00
	Types of wh-elements	0.33	0.10	3.47	0.00*
	Extraction sites	−0.25	0.14	−1.86	0.06.
	Interaction	0.15	0.15	1.01	0.31
	Length (characters)	0.08	0.03	2.14	0.03*
First Pass Regression Out Probability	(Intercept)	−6.57	2.03	−3.24	0.001***
	Types of wh-elements	3.44	0.61	5.60	0.001***
	Extraction sites	−0.20	0.56	−0.36	0.72
	Interaction	−0.42	0.53	−0.79	0.43
	Length (characters)	0.39	0.19	1.98	0.04*
Regression Out Full Probability	(Intercept)	−1.66	0.95	−1.74	0.08.
	Types of wh-elements	0.75	0.23	3.25	0.001***
	Extraction sites	−0.57	0.27	−2.12	0.03*
	Interaction	0.42	0.30	1.41	0.16
	Length (characters)	0.18	0.09	1.94	0.05*
Region 4					
First Pass Reading Time (residuals)	(Intercept)	−0.04	0.02	−1.74	0.08.
	Types of wh-elements	0.10	0.03	3.43	0.00*
	Extraction sites	−0.00	0.03	−0.08	0.94
	Interaction	−0.05	0.04	−1.24	0.21
Total Reading Time (residuals)	(Intercept)	0.06	0.07	0.82	0.41
	Types of wh-elements	−0.13	0.11	−1.16	0.25
	Extraction sites	0.04	0.09	0.47	0.64
	Interaction	−0.06	0.12	−0.47	0.64
Regression Out Full Count	(Intercept)	−0.39	0.28	−1.39	0.16
	Types of wh-elements	−0.48	0.09	−5.64	0.001***
	Extraction sites	−0.11	0.08	−1.43	0.15
	Interaction	0.22	0.12	1.77	0.07.
	Length (characters)	0.05	0.02	2.24	0.02
First Pass Regression Out Probability	(Intercept)	−0.54	1.06	−0.51	0.61
	Types of wh-elements	−2.54	0.26	−9.64	0.001***
	Extraction sites	−0.30	0.26	−1.18	0.24
	Interaction	0.82	0.35	2.37	0.02**
	Length (characters)	0.24	0.09	2.71	0.01**
Regression Out Full Probability	(Intercept)	0.44	0.91	0.48	0.63
	Types of wh-elements	−1.97	0.25	−7.87	0.00*
	Extraction sites	−0.17	0.26	−0.66	0.51
	Interaction	0.61	0.34	1.78	0.07.
	Length (characters)	0.15	0.07	2.02	0.04*

Notes: All factors were coded using treatment coding, with the *quale-NP* and subject-extraction conditions serving as reference levels for the *wh*-element and extraction site factors, respectively. Analyses were performed separately for critical regions 3 and 4. Significant effects at a $p \leq .05$ level are marked with a *. For the random effects, we included both subjects and items intercepts in all models, with the following exceptions: (i) total reading time (residuals) included a random slope for the types of *wh*-elements by subjects and the site of extraction by items in both region 3 and region 4; (ii) regression out included a random slope for the types of *wh*-elements by subjects and items in region 3; (iii) regression out full included a random slope for the types of *wh*-elements by subjects in region 3. We started with full mixed-effects models which involved random intercepts and slopes for subjects and items. However, we simplified the random effects structure due to problems of convergence (e.g. boundary fits and singular variance-covariance matrices). Specifically, random slopes that did not contribute significantly to model fit were removed, while random intercepts for subjects and items were kept.

between *chi* in object-extraction and *chi* in subject-extraction was not significant ($\beta = 0.304$, SE = 0.257, z = 1.18, $p = 1$). Additionally, object-extracted *quale-NP* questions were associated with a significantly lower probability of regression compared to subject-extracted *chi* ($\beta = -2.24$, SE = 0.25, z = −9.23, $p < .001$), but not with subject-extracted *quale-NP* ($\beta = -.51$, SE = 0.21, z = −2.44, $p = .08$). Finally, the difference between subject-extracted *chi* and subject-extracted *quale-NP* was significant ($\beta = 1.72$, SE = 0.25, z = 6.95, $p < .001$). Overall, these findings suggest that object-extraction in *chi* questions led to a significantly higher probability of regressions, indicating increased processing difficulty compared with all other conditions. Furthermore, even in subject-extraction *chi* was associated with higher probability of regressions as compared to *quale-NP*.

In total reading time, no significant main effects or interactions were observed, suggesting that processing costs in this region were more evident in early measures rather than in later ones (all *ps* > .25). As for regression out full probability, there was a significantly higher probability of regressions to a previous region in *chi* questions ($\beta = -1.97$, SE = 0.25, z = −7.87, $p < .001$). Similarly, regression out full count analysis showed that *chi* questions led to more regressions overall ($\beta = -0.48$, SE = 0.09, z = −5.64, $p = .001$).

In general, when integrating the post-verbal NP, object-extraction questions led to a higher probability of first-pass regressions and more total regressions, but this was particularly true for *chi* questions. Furthermore, the significant interaction in first-pass regression out data underlies that object-extraction imposes different processing demands depending on the *wh*-question, particularly when encountering the post-verbal NP.

Discussion

In this paper we compared the processing of *chi* and *quale-NP* questions in Italian by means of eye tracking methodology. Two are the main findings of our study: a subject-object asymmetry for both types of *wh*-elements and a different processing pattern for *chi* and *quale-NP* questions.

Subject-extraction preference for both types of wh-elements

Overall, early processing measures suggest different strategies for processing the verb region between *chi* and *quale-NP* questions. On the one hand, the longer first-pass reading times in *chi* questions indicate increased initial processing effort in the object-extraction condition, due to the parser's preference for a subject extraction. On the other hand, the higher likelihood of first-pass regressions in *quale-NP* questions suggests a greater need for reanalysis upon encountering the verb. When integrating the post-verbal NP, *quale-NP* questions require extra processing in early measures, while *chi* questions, especially in object-extraction condition, lead to more reanalysis, as evidenced by increased first-pass regressions.

The overall pattern of results suggests an active gap-filling strategy across different types of *wh*-elements, with object extraction leading to extended reading times and regressions compared to subject extraction. Therefore, we argue that the *Minimal Chain Principle* (De Vincenzi, 1991b) or alternatively – the *Active Filler Strategy* (Frazier, 1987; Frazier & D'Arcais, 1989) – apply to both *quale-NP* questions and bare quantifiers like *chi*. Both in the case of *chi* questions and *quale-NP* questions, once the processor has identified a filler, it assigns it to the first possible position, that in both cases is the "subject" position. Hence, when the processor reaches the verb region and realises that the filler should be analysed as an object, the first analysis must be revised. Our results on the parsing of *quale-NP* questions are in line with previous findings by Contemori et al. (2018) on adults' and children's processing that tested the online comprehension of *which*-questions in English using visual-word paradigm (see also Pontikas et al., 2023 for similar evidence from bilingual children). Contemori et al. (2018) found that both groups fixated significantly more on the competitor picture (compatible with a *wh*-subject extraction) when processing an object question than when processing a subject question, suggesting a strong preference for a subject interpretation of the *wh*-element in *which*-questions. Preference for attaching the fronted *wh*-element to the first available gap has also been found in biclausal *wh*-questions (Lassotta et al., 2024). Our results are also in line with recent evidence by Kim, Orth, et al. (2024) showing that the parser actively searches for a resolution to complete the filler-gap dependencies as soon as possible for D-linked *wh*-elements.

The greater difficulty of object extraction compared to subject extraction in both *chi*- and *quale-NP* questions contrasts with evidence from other languages, where this asymmetry appears to be present only in *which-NP* questions, although these studies used offline measure, which may be less sensible than online ones to subtle differences. Notably, the studies supporting this cross-linguistic pattern (Avrutin, 2000; Friedmann et al., 2009; Hickok & Avrutin, 1996) are based on data from children or patients with Broca's aphasia. For instance, using a picture-pointing task, Avrutin (2000) found that English-speaking children aged 3;5 to 5;2 (mean age = 4;3) demonstrated lower comprehension of *which*-object questions compared to *which*-subject questions (48% vs. 86% correct responses). In contrast, no such asymmetry was observed for *who* questions, where accuracy remained consistent at 80% for both subject and object questions.

The preference for *wh*-subject extraction also aligns with probabilistic and experience-based accounts, which propose that the parser, being sensitive to the higher frequency of subject questions, anticipates a subject gap when processing an object question (Hale, 2001; Levy, 2008; Roland et al., 2007). In other words, participants accept the more frequent subject-extraction structure in both types of questions, and when this expectation is violated, they experience processing difficulty.

Difference in the processing costs for the two types of wh*-elements*

The comparison between early and late measures shows that the parser uses different strategies in *chi* and *quale-NP* question in the reanalysis in case of object extraction. In *chi* questions, the diagnosis of the anomaly is fast because *chi* can be only singular and the verb triggers the agreement whereas in *quale-NP* the cue is not that clear because *quale-NP* can carry either singular or plural features so the parser does not know whether to revise the *quale-NP* or the verb features. The presence of an alternate solution for the revision of the interpretation causes the persistence of the garden path until the reaching of the post-verbal NP (Fodor & Inoue, 1994).

The different processing costs outlined for the two types support De Vincenzi's (1991a) claim, which distinguishes between *chi* and *quale-NP* elements. According to De Vincenzi, *chi* functions as a filler that actively searches for a gap and must be assigned to a gap position as soon as possible (typically the subject position), whereas *quale-NP* is assigned a referential index, delaying its interpretation until lexical information about the verb and the following NP has been processed. Further discussion of this phenomenon is provided below.

The integration of the post-verbal NP

When post-verbal NP integration occurred, the results showed a varied pattern: *quale-NP* questions caused a longer first pass, while in all other measures there was a consistent pattern indicating that *chi* questions caused longer reading times and a higher tendency to regress to a previous region. This effect was also consistent at the verb region, with early measure denoting more demands for *chi* and late measure indicating a greater cost for *quale-NP*. To account for such findings, we propose an analysis following the *Minimal Chain Principle*: in *chi* questions, the parser's preference for assigning the subject role to the *wh*-element causes an immediate disruption at the verb region in the case of object-extraction. The matching between the verb and the subject feature causes a clash between the singular-feature *chi* and the plural-feature verb which leads immediately to the detection of the incongruence, resulting in longer first pass reading times in the verb region. We propose that, in order to repair the garden path, the parser postulates a *pro* subject between the *wh*-element and the plural verb; it deletes the subject trace previously presupposed and assigns to the *wh*-element the object role. This local arrangement of the sentence leads to a OSV order sentence, with S being the null subject pro. The consistency of the pattern of late measures found in the post-verbal region suggests that the integration of the post-verbal NP was significantly more costly in *chi* questions, in line with the hypothesis that the parser does not expect to find an NP, since the garden path has already been recovered in the verb region by building an OSV order sentence. In *quale-NP* questions, since the *quale-NP* can carry either singular or plural features (singular: *quale-NP*; plural*: quali-NPs)*, the parser is uncertain whether to revise the *wh*-element or the verb features. However, the post-verbal NP is easily integrated into the structure, suggesting that direct memory interference between the two NPs in the question is minimal. At the same time, the referential nature of *quale-NP* questions does not appear to directly affect the integration of the post-verbal NP (Warren & Gibson, 2002). Gordon et al. (2001) examined the comprehension of embedded clauses (object-extracted vs. subject-extracted) and found that sentences were harder to understand when they contained NPs of the same type. In their experiment, comprehension difficulties with object-extracted NPs were reduced by mixing names or indexical pronouns with definite descriptions as the NPs in a sentence. Since *quale-NP* questions require storing two NPs for later retrieval, extended reading times at the post-verbal NP could have been expected due to increased memory load (Gordon et al., 2001).

Comparing the processing of chi *vs* quale-NP

When comparing the overall processing patterns of the two *wh*-elements, *quale-NP* questions appear more costly to process than *chi* questions, both at the early and later stages of reading. This increased processing cost is consistently observed at the verb region across both early and late measures, with the exception of the first-pass reading time for *chi* object-extraction at the verb region. In this context, we cannot rule out the possibility that the higher processing cost associated with *quale-NP* questions reflects memory-related difficulties, potentially due to similarity-based interference (Gibson, 1998), or that it may be explained by the D-linking and referential nature of *quale-NP* questions (Pesetsky, 1987; Warren & Gibson, 2002). Nevertheless, our results suggest that this difficulty does not manifest itself as direct interference during the integration of the post-verbal NP. Instead, the increased processing cost in *quale-NP* questions may be due to higher retrieval demands at the verb combined with greater structural complexity.

Limitations of our study and future directions

The results of late measures suggest that participants recover from an initial misinterpretation of object

questions as subjects once they encounter disambiguating information at the verb region. However, our study does not provide insight into potential lingering effects, as our task was not designed to assess this phenomenon (Christianson et al., 2001, 2006; Ferreira et al., 2001; Ferreira & Henderson, 1991; Malyutina & den Ouden, 2016). The comprehension questions included in our study served only to ensure participants' attention, not to evaluate whether reanalysis was fully achieved. To address this, future research should employ a task where comprehension questions specifically target reanalysis.

Another limitation of our study regards the fact that in the experiment, the questions were not presented within a discourse context but rather in isolation. This may not fully capture the ecological nature of sentence processing.

Conclusions

Our findings indicate that the parser employs distinct strategies when processing *chi* and *quale-NP* questions concerning object extraction. In *chi* questions, rapid anomaly diagnosis occurs due to the singular nature of *chi*, along with the verb triggering agreement. However, in *quale-NP* questions, the presence of an alternate solution for the revision of the interpretation might cause the persistence of the garden path until the reaching of the post-verbal NP. Yet, the post-verbal NP does not require extra-processing in *quale-NP* questions. In conclusion, when considering multiple variables concurrently, the comprehension of *quale-NP* questions appears to pose a greater challenge compared to *chi* questions.

Acknowledgements

We thank all the participants who participated in the experiment. We are particularly grateful to Masaya Yoshida for providing valuable feedback on previous versions of the manuscript and for his insightful advice on statistical analyses. We also thank the two anonymous reviewers and the editor for their valuable comments.

Authors' contributions

M.T.G., E.P. and M.V. conceived the project; E.P. designed the experiment in consultation with M.V, E.P. collected the data; M.V. analysed the data; E.P. and M.V. prepared the initial manuscript; all the authors revised the manuscript and approved its final version.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data availability statement

Anonymized data have been deposited in the OpenScienceFramework database (https://osf.io/edhbv/).

Funding

E.P. acknowledges the financial support of Grant PID2022-138413NB-I00, funded by MCIN/AEI/10.13039/501100011033 and by the European Union; of Grant 2021SGR00787 funded by the Generalitat de Catalunya; of Grant RYC2021-033969-I (Ramón y Cajal), funded by MCIN/AEI/10.13039/501100011033 and by European Union Next-GenerationEU/PRTR. M.V. disclosed receipt of the following financial support for publication of this article: a contribution from the Department of Humanities – DISTUM of the University of Urbino Carlo Bo (Department of Excellence 2023-27).

ORCID

Elena Pagliarini http://orcid.org/0000-0002-8644-0984
Mirta Vernice http://orcid.org/0000-0001-9911-2755
Maria Teresa Guasti http://orcid.org/0000-0002-4470-4187

References

Avrutin, S. (2000). Comprehension of discourse-linked and non-discourse-linked questions by children and Broca's aphasics. In Y. Grodzinsky, L. P. Shapiro, & D. Swinney (Eds.), *Language and the brain* (pp. 295–313). Academic Press. https://doi.org/10.1016/B978-012304260-6/50017-7

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Machler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, *9*(2), 378–400. https://doi.org/10.32614/RJ-2017-066

Chacón, D. A. (2022). Default is different: Relations and representations in agreement processing. *Language, Cognition and Neuroscience*, *37*(6), 785–804. https://doi.org/10.1080/23273798.2021.2022172

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

Christianson, K., Williams, C. C., Zacks, R. T., & Ferreira, F. (2006). Younger and older adults' "good-enough" interpretations of garden-path sentences. *Discourse Processes*, *42*(2), 205–238. https://doi.org/10.1207/s15326950dp4202_6

Cinque, G. (1999). *Adverbs and functional heads: A cross-linguistic perspective*. Oxford University Press.

Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in Reading words and sentences. In R. P. G. Van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements* (pp. 341–371). Elsevier. https://doi.org/10.1016/B978-008044980-7/50017-3

Contemori, C., Carlson, M., & Marinis, T. (2018). On-line processing of English which-questions by children and adults: A visual world paradigm study. *Journal of Child Language*, *45*(2), 415–441. https://doi.org/10.1017/S0305000917000277

Crain, S., & Fodor, J. D. (1985). How can grammars help parsers? In A. M. Zwicky, D. R. Dowty, & L. Karttunen (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives* (pp. 94–128). Cambridge University Press. https://doi.org/10.1017/CBO9780511597855.004

De Vincenzi, M. (1991a). Filler-gap dependencies in a null subject language: Referential and nonreferential WHs. *Journal of Psycholinguistic Research*, *20*(3), 197–213. https://doi.org/10.1007/BF01067215

De Vincenzi, M. (1991b). *Syntactic parsing strategies in Italian: The Minimal Chain Principle*. Kluwer Academic Publishers.

De Vincenzi, M., Di Matteo, R., & Job, R. (2004). Comparing who-question interpretation in Italian sentences. In *Proceedings of the 45th Annual Meeting of the Psychonomic Society*. Psychonomic Society Publications.

Drummer, J.-D., & Felser, C. (2018). Cataphoric pronoun resolution in native and non-native sentence comprehension. *Journal of Memory and Language*, *101*, 97–113. https://doi.org/10.1016/j.jml.2018.04.001

Falkauskas, K., & Kuperman, V. (2015). When experience meets language statistics: Individual variability in processing English compound words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(6), 1607. https://doi.org/10.1037/xlm0000132

Ferreira, F., Christianson, K., & Hollingworth, A. (2001). Misinterpretations of garden-path sentences: Implications for models of sentence processing and reanalysis. *Journal of Psycholinguistic Research*, *30*(1), 3–20. https://doi.org/10.1023/A:1005290706460

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*(3), 348–368. https://doi.org/10.1016/0749-596X(86)90006-9

Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, *30*(6), 725–745. https://doi.org/10.1016/0749-596X(91)90034-H

Fodor, J. D., & Inoue, A. (1994). The diagnosis and cure of garden paths. *Journal of Psycholinguistic Research*, *23*, 407–434.

Frazier, L. (1987). Syntactic processing: Evidence from Dutch. *Natural Language & Linguistic Theory*, *5*(4), 519–559. https://doi.org/10.1007/BF00138988

Frazier, L., & Clifton, C. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes*, *4*(2), 93–126. https://doi.org/10.1080/01690968908406359

Frazier, L., & D'Arcais, F. (1989). Filler driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language*, *28*(3), 331–344. https://doi.org/10.1016/0749-596X(89)90037-5

Friedmann, N., Belletti, A., & Rizzi, L. (2009). Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua*, *119*(1), 67–88. https://doi.org/10.1016/j.lingua.2008.09.002

Friedmann, N., & Novogrodsky, R. (2011). Which questions are most difficult to understand? The comprehension of Wh questions in three subtypes of SLI. *Lingua*, *121*(3), 367–382. https://doi.org/10.1016/j.lingua.2010.10.004

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1–76. https://doi.org/10.1016/S0010-0277(98)00034-1

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain* (pp. 95–126). MIT Press.

Giskes, A., & Kush, D. (2021). Processing cataphors: Active antecedent search is persistent. *Memory & Cognition*, *49*(7), 1370–1386. https://doi.org/10.3758/s13421-021-01176-z

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory inference during language processing. *Journal of Experimental Psychology: Learning Memory and Cognition*, *27*(6), 1411. https://doi.org/10.1037/0278-7393.27.6.1411

Green, J. A. (2021). Too many zeros and/or highly skewed? A tutorial on modelling health behaviour as count data with Poisson and negative binomial regression. *Health Psychology and Behavioral Medicine*, *9*(1), 436–455. https://doi.org/10.1080/21642850.2021.1920416

Grimaldi, M., De Vincenzi, M., Lorusso, P., Di Russo, F., Di Matteo, R., Rizzi, L., & Guasti, M. T. (2023). The processing of Which interrogative sentences: A behavioral and ERP study. *Journal of Neurolinguistics*, *68*, Article 101154. https://doi.org/10.1016/j.jneuroling.2023.101154

Guasti, M. T., Branchini, C., & Arosio, F. (2012). Interference in the production of Italian subject and object wh-questions. *Applied Psycholinguistics*, *33*(1), 185–223. https://doi.org/10.1017/S0142716411000324

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *2nd Meeting of the North American Chapter of the Association for Computational Linguistics, NAACL 2001*. https://doi.org/10.3115/1073336.1073357

Hickok, G., & Avrutin, S. (1996). Comprehension of wh-questions in two Broca's aphasics. *Brain and Language*, *52*(2), 314–327. https://doi.org/10.1006/brln.1996.0014

Kim, H., Orth, W. J., & Yoshida, M. (2024). Incremental structure building in the processing of ellipsis. *Quarterly Journal of Experimental Psychology*, *0*(0), 17470218241280567. https://doi.org/10.1177/17470218241280567

Kim, N., Wellwood, A., & Yoshida, M. (2024). Processing wh-filler-gap dependencies. *Quarterly Journal of Experimental Psychology*, *77*(12), 2391–2417. https://doi.org/10.1177/17470218241231872

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). Lmertest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lassotta, R., Panizza, D., Omaki, A., & Franck, J. (2024). Syntactic revision in wh-questions: Developmental trajectory and the role of cognitive control. *Language, Cognition and Neuroscience*, *39*(6), 713–734. https://doi.org/10.1080/23273798.2024.2355218

Lenth, R. V. (2024). emmeans: Estimated marginal means, aka least-squares means. *R Package Version 1.10.6*.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Malyutina, S., & den Ouden, D.-B. (2016). What is it that lingers? Garden-path (mis)interpretations in younger and older adults. *Quarterly Journal of Experimental Psychology*, *69*(5), 880–906. https://doi.org/10.1080/17470218.2015.1045530

Marelli, M., & Baroni, M. (2015). Affixation in semantic space: Modeling morpheme meanings with compositional

distributional semantics. *Psychological Review*, *122*(3), 485–515. https://doi.org/10.1037/a0039267

Miyamoto, E. T. (2003). Reanalysis of clause boundaries in Japanese as a constraint-driven process. *Language and Speech*, *46*(1), 23–50. https://doi.org/10.1177/00238309030460010301

Penolazzi, B., De Vincenzi, M., Angrilli, A., & Job, R. (2005). Processing of temporary syntactic ambiguity in Italian "who"-questions: A study with event-related potentials. *Neuroscience Letters*, *377*(2), 91–96. https://doi.org/10.1016/j.neulet.2004.11.074

Pesetsky, D. (1987). Wh-in-situ: Movement and unselective binding. In E. Reuland & A. ter Meulen (Eds.), *The representation of (in)definiteness* (pp. 98–129). MIT Press.

Pontikas, G., Cunnings, I., & Marinins, T. (2023). Online processing of which-questions in bilingual children: Evidence from eye-tracking. *Journal of Child Language*, *50*(5), 1082–1118. https://doi.org/10.1017/s0305000922000253

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. https://doi.org/10.1037/0033-2909.124.3.372

Rayner, K., Juhasz, B. J., & Pollatsek, A. (2005). Eye movements during reading. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 79–97). Blackwell Publishing. https://doi.org/10.1002/9780470757642.ch5

Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, *57*(3), 348–379. https://doi.org/10.1016/j.jml.2007.03.002

Schlesewsky, M., Fanselow, G., Kliegl, R., & Krems, J. (2000). The subject preference in the processing of locally ambiguous WH-questions in German. In B. Hemforth & L. Konieczny (Eds.), *German sentence processing* (pp. 65–93). Springer. https://doi.org/10.1007/978-94-015-9618-3_3

Schmidtke, D., Van Dyke, J. A., & Kuperman, V. (2018). Individual variability in the semantic processing of English compound words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(3), 421. https://doi.org/10.1037/xlm0000442

Stowe, L. A. (1986). Parsing wh-constructions: Evidence for online gap location. *Language and Cognitive Processes*, *1*(3), 227–245. https://doi.org/10.1080/01690968608407062

Tanenhaus, M. K., Stowe, L. A., & Carlson, G. (1989). Lexical expectation and pragmatics in parsing filler-gap constructions. In W. D. Marlsen-Wilson (Ed.), *Lexical representation and process* (pp. 311–331). MIT Press.

Ueno, M., & Kluender, R. (2009). On the processing of Japanese wh-questions: An ERP study. *Brain Research*, *1290*, 63–90. https://doi.org/10.1016/j.brainres.2009.05.084

Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition*, *85*(1), 79–112. https://doi.org/10.1016/S0010-0277(02)00087-2