Applied Psycholinguistics (2025), **46**, e9, 1–34
doi:10.1017/S0142716425000074

CAMBRIDGE UNIVERSITY PRESS

ORIGINAL ARTICLE

Do readers anticipate *wh*-in-situ questions? Cross-linguistic reading time evidence from Mandarin Chinese and French

Leticia Pablos Robles[1], Yang Yang[2], Jenny Doetjes[1] and Lisa Lai-Shen Cheng[1]

[1]Leiden University, Leiden, Netherlands and [2]Center for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies, Guangzhou, China
Corresponding author: Leticia Pablos Robles; Email: l.pablos.robles@hum.leidenuniv.nl

(Received 21 November 2022; revised 13 November 2024; accepted 14 January 2025)

Abstract

The understanding of *wh*-in-situ questions relies naturally on contextual and prosodic information for their early discrimination from declarative sentences. However, there is scarce evidence on the parsing processes involved during the online incremental processing of these questions. In this study, we investigate the incremental reading of *wh*-in-situ sentences with no prosodic or contextual information available to aid the parser by comparing them to their declarative counterparts. We investigated two *wh*-in-situ languages: Mandarin Chinese (in-situ only) and French (optionally in situ). This comparison allows us to determine whether *wh*-in-situ questions are processed similarly across languages and whether the parsing process is related to language-specific question formation strategies. Results of four word-by-word self-paced reading experiments on two types of *wh*-in-situ phrases (simplex or complex) in Mandarin Chinese and French show an interpretation strategy in which the most frequent structure, declarative, is considered in both languages, independently of the available question formation strategy. Nevertheless, the timing of the online interpretation and the observed effects are affected by the nature of the *wh*-phrases (simplex or complex) and the definiteness of the noun phrases contained in the declaratives, which confirms that several processes occur concurrently introducing a limit on the capability to extract conclusions on the processes based solely on behavioral measures.

Keywords: Complex wh-questions; definiteness; French; In-situ wh-questions; Mandarin Chinese; question formation strategies

Introduction

How information-seeking questions (also known as *wh*-questions) are interpreted is an issue that has received quite a lot of attention in the sentence-processing

literature. Most of the studies focus on long-distance dependencies examining how fronted *wh*-phrases are interpreted at their canonical position and the effects that result from keeping the fronted *wh*-phrase in the parser's working memory until resolving the open dependency. Standard "filled-gap effects" (Crain & Fodor, 1985; Stowe, 1986; see Pablos, 2008 for overview) are associated with reading time evidence showing that readers expect the *wh*-phrase (the filler) to be discharged and interpreted at the first available grammatical position. The failure of discharging the *wh*-phrase results in longer reading times than their declarative counterparts. In addition, there is growing evidence that the interpretation of sentence meaning is achieved incrementally with comprehenders *predicting* upcoming information (including lexical and syntactic structure) based on available input (e.g. Altmann & Kamide, 1999; Levy, 2008). Under prediction accounts, projecting a *wh*-gap in fronted *wh*-questions occurs at the moment the *wh*-phrase is encountered.

The above scenario cannot be directly extended to *wh*-questions where the *wh*-phrase stays in its canonical position, known as *wh-in-situ* questions. The interpretation of these questions may result in a temporary syntactic ambiguity in comparison with their declarative counterparts that have a non-*wh*-word at the same site. Research shows that available contextual and prosodic information is used to predict the upcoming structure (Fodor, 2002; Déprez et al., 2013; Gryllia et al., 2016; Gryllia et al., 2020; Yang et al., 2019; Kawahara et al., 2022), practically resolving the ambiguity before encountering the *wh*-phrase. Nonetheless, this raises the question of how such *wh*-questions are parsed and to what extent readers anticipate an upcoming in-situ *wh*-question in the absence of other information (e.g., prosody, context, information structure) during the reading process. These questions form the focus of this study. More specifically, we examine how speakers of two *wh*-in-situ languages, Mandarin Chinese and French, proceed in the real-time reading of sentences presented without any preceding contextual or prosodic cues that could bias interpretations as questions or declarative statements. These two languages differ in the types of strategies they permit for *wh*-questions. Whereas Mandarin *wh*-questions are always in situ, French permits both in situ and fronted *wh*-questions. We investigate whether this variation in *wh*-in-situ strategies influences processing difficulty and predictability. In addition, we also examine an additional potential factor which might influence the parsing of *wh*-in-situ questions, namely, the complexity of the *wh*-phrase (i.e., simplex *wh*-phrase such as *who* or complex *wh*-phrase such as *which person*). Our second research question thus addresses whether there are processing differences between the complex and simplex *wh*-phrases in in-situ *wh*-questions.

Mandarin Chinese and French

Question formation strategies across languages

Languages differ in the number and type of strategies for forming *wh*-questions (see for instance Cheng, 1991) and can be categorized into three primary groups. The first type obligatorily fronts *wh*-words in *wh*-questions, as in English (1).[1,2]

https://doi.org/10.1017/S0142716425000074 Published online by Cambridge University Press

(1) **Who**$_i$ did you meet t_i at the art museum yesterday?

The second type consists of languages that always retain *wh*-words in their canonical position (i.e., in-situ) when formulating a *wh*-question, as in Mandarin Chinese (2).

(2) Nǐ zuótiān zài yìshù bówùguǎn yùjiàn le **shéi**?
you yesterday in art museum meet PERF who
'Who did you meet at the art museum yesterday?'

The third language type permits both fronting and in-situ *wh*-question formation, as in French illustrated in (3a) and (3b).[3,4]

(3) a. Tu as vu **qui** ?
you have seen who
'Who have you seen?'

b. **Qui**$_i$ tu as vu t_i ?
who you have seen
'Who have you seen?'

Adli (2015) examined the prevalence and distribution of *wh*-in-situ questions in relation to other variants of *wh*-question formation in French. He presented an assessment of spontaneous speech in French obtained from the *Sgs* database (with 10943 sentences) and showed that 56.2% of the total number of 1721 interrogative utterances (excluding echo-questions) are in-situ *wh*-questions. The study further found that the relative frequency of these in-situ questions is 0.62 for *wh*-adjuncts and 0.43 for *wh*-objects. A more recent study found an increase in the use of *wh*-in-situ in the last decade (Baunaz & Bonan, 2023).

These different strategies pose interesting questions regarding the processes used in the online comprehension of *wh*-in-situ constructions where the clause type of the sentence (question or declarative) is only obvious when the *wh*-word is encountered. If English were to permit in-situ *wh*-questions like Mandarin and French, a comparison of (4) and (5) illustrates that the difference between the *wh*-in-situ question (4) and the declarative sentence (5) is only revealed at the postverbal object position (see Note 3).

(4) (You said) Peter would like to meet **whom** tomorrow?
(5) (You said) Peter would like to meet **a friend** tomorrow.

Crucially, unless prosodic or contextual information is available, no distinction can be made between these two sentences by readers (up to the object position) as they proceed incrementally in the interpretation.

https://doi.org/10.1017/S0142716425000074 Published online by Cambridge University Press

The syntax and processing of in-situ wh-questions: previous studies

Syntactic studies of in-situ *wh*-questions analyze these questions as involving a covert dependency, such that the in-situ *wh*-phrase either is related to an interrogative operator or raised to the structurally higher operator position (SpecCP position) at the Logical Form (LF; for further discussion see Aoun and Li, 1993; Cheng, 1991, 2003, 2009; Huang, 1982; Tsai, 1994; and Bayer & Cheng, 2017 for an overview). The covert dependency is thus on par with overt dependencies in questions with overt *wh*-fronting in that it involves a syntactic representation where an (covert) operator is in the structurally higher position (i.e., SpecCP), which determines the clause type of the sentence.[5] This in turn raises an interesting question concerning the representation of in-situ *wh*-questions in the processing system. If the same processing mechanisms are used in processing dependencies, the abstract link between the *wh*-phrase and the SpecCP position in the case of *wh*-in-situ questions should manifest as a nonlocal dependency formation.

There has been to date limited research on the processing of in situ *wh*-questions. In French, for example, most of the research focused on the production of the prosodic features or the acceptability of in-situ *wh*-questions, but not on how these questions are interpreted incrementally (see Adli, 2004, 2006; Beyssade et al., 2007; Delattre, 1966; Deprez et al., 2013; Wunderli, 1983, 1984; Oiry, 2011; Tual, 2017a, 2017b from discussion in Glasbergen-Plas, 2021). Ueno and Kluender's (2009) study of Japanese *wh*-in-situ constructions showed an effect, manifested as a right-lateralized-anterior negativity (RLAN), on longer distance covert dependency formation. In Japanese, however, the question marker (also a scope marker) is morphologically overt. In Mandarin Chinese, studies by Xiang et al. (2013, 2015) examined the processing of in-situ complex *wh*-questions (i.e., *which x* questions) of different lengths (mono-clause vs. embedded clause) in comparison with declaratives (mono-clause vs. embedded clause) using the Speed-Accuracy Tradeoff (SAT) methodology. They looked at differences in *wh*-dependency length across their stimuli and found that length had an impact on processing accuracy but not on processing speed. Their results showed that questions such as (7b) had lower processing accuracy than those in (7a) but were equally slow in comparison to declaratives such as (6a) and (6b).

(6) *a. Declarative; Short*

市政府	严惩了	**那些官员** .
shì zhèngfǔ	yánchěng-le	**nàxiē guānyuán**.
city-council	*punish*	*those officials*

'The city council punished those officials.'

b. Declarative; Long

市长	命令	市政府	严惩了	**那些官员**.
Shìzhǎng	mìnglìng	shì-zhèngfǔ	yánchěng-le	**nàxiē guānyuán**.
mayor	*order*	*city-council*	*punish*	*those officials*

'The mayor ordered the city council to punish those officials.'

https://doi.org/10.1017/S0142716425000074 Published online by Cambridge University Press

(7) *a. Wh-Q; Short*

市政府	严惩了	**哪些官员?**
Shì zhèngfǔ	yánchěng-le	**nǎxiē guānyuán?**
city council	*punish*	*which officials*

'Which officials did the city council punish?'

b. Wh-Q; Long

市长	命令	市政府	严惩了	**哪些官员?**
Shìzhǎng	mìnglìng	shì zhèngfǔ	yánchěng-le	**nǎxiē guānyuán?**
mayor	*order*	*city-council*	*punish*	*which officials*

'Which officials did the mayor order the city council to punish?'

The increased processing time of *wh*-in-situ questions in (7a, b) was attributed to the effects of establishing a long-distance covert dependency. The effect of length on the accuracy, but not on the speed of processing of *wh*-in-situ, supports the notion of a covert dependency retrieved by a content-addressable memory process (McElree, 2000; McElree et al., 2003).

Predictions for parsing in-situ wh-questions

The evolution of the sentence comprehension models over the years has led to a consensus of an *incremental interpretation* process where the human parser interprets the available information incrementally building up a representation of the sentence meaning as the input unfolds, without delay. Still, there are a few aspects in which available models differ and that are relevant for the interpretation of observed processing difficulty. A growing amount of evidence points to the predictive nature of the comprehension processes (e.g. Levy, 2008; Altmann & Mirkovic, 2009), although there remains a debate on the interpretation of the concept of *prediction* and the difference with respect to *integration* processes (for a summary discussion see Pickering & Gambi, 2018 and Kuperberg & Jaeger, 2016, and counter-arguments in Huettig & Mani, 2016). In simple terms, prediction implies the activation of linguistic information *before* input is available. Predictive models can thus be understood in a probabilistic framework in which the parser updates continuously the projected structure and expected lexical content based on the information as it becomes available (Levy, 2008).

Considering the fact that declarative sentences are more frequent than questions in the world's languages and that they tend to be the most unmarked of the clause types (Ma et al., 2011), in the case of parsing a *wh*-in-situ question up to the *wh*-phrase (i.e., parsing the part of the sentence which is the same in both *wh*-questions and declaratives), the initial prediction made by the parser would be based on the most *frequent* structure. It should also be noted that Adli (2015) in his study of spontaneous speech in French also reported that questions (including *wh*-in-situ questions) constitute only 15.72% of utterances (1721 out of 10943 sentences), highlighting the dominance of declaratives in the dataset and thus reinforcing the expectation of declaratives over interrogatives. We therefore predict a processing slowdown at the *wh*-phrase when processing *wh*-in-situ questions as compared to the declarative counterpart.

https://doi.org/10.1017/S0142716425000074 Published online by Cambridge University Press

The nature of the observed processing difficulty, however, can have a different interpretation depending on the theoretical processing model considered. In "classical terms," it can be considered an indication of *re-analysis* to reconstruct the projected structure (Fodor & Ferreira, 1998), or the *activation* of the alternative structure. Further, the level of difficulty has been postulated to be quantified by measures such as the *surprisal* (Levy, 2008) and *entropy* (Linzen & Jaeger, 2016), which represent formalizations of the predictability of a word or structure in a certain context. These measures can be estimated from corpora or, traditionally, from Cloze probability procedures. These models are considered serial as only one interpretation is active at a given time, in comparison to models where multiple interpretations are concurrently active in parallel with different levels of activation. Under parallel activation, we can consider *activation-based retrieval* models (Van Dyke & Lewis, 2003; Lewis & Vasishth, 2005) or more recently, the proposed *parallel architecture model* (Huettig et al., 2022). In the *activation-based retrieval models*, processing difficulties at the *wh*-in-situ site reflect *reactivation* of the alternative structure combined with the *integration* of the covert dependency. In the *parallel architecture model* (Huetting et al., 2022), the potential structures, encoded as a lexicon (Jackendoff & Audring, 2020), are all active simultaneously as the first words are encountered (*within-item activation*) with different "resting activations", linked to their frequency.

All the models described above would predict readers of Mandarin Chinese and French to have additional processing costs (observed as longer reading times) when encountering the *wh*-in-situ phrase, as compared to the non-*wh* noun phrase in the declarative counterpart. This processing cost could either be due to reanalysis, reactivation or covert dependency integration. The extent to which the parser anticipates upcoming structure when there is no other cue available might be modulated by the *likelihood* of encountering in-situ *wh*-phrases in each of the languages under study. In Mandarin, an in-situ question is the only option for *wh*-questions. In contrast, in French, as mentioned above, Adli (2015) showed that 56.2% of the produced interrogative utterances in the *Sgs* database were *wh*-in-situ.

The complexity and definiteness of (wh)-noun phrases

The processing study of Mandarin Chinese that we mentioned above by Xiang et al. (2013) and Xiang et al. (2015) used only complex *wh*-phrases (i.e., *which* x phrases). Nonetheless, there is experimental evidence showing differences in the processing of complex and simplex *wh*-questions for languages such as English, Dutch and Italian in that the complex *wh*-questions take longer to read than simplex *wh*-questions (De Vincenzi 1996; Kaan et al., 2000; Donkers et al., 2011). Other studies, however, provide opposite claims on the processing cost of complex *wh*-phrases, where these are facilitated (see Frazier & Clifton 2002; Clifton et al., 2006; Hofmeister et al., 2007; Hofmeister & Sag, 2010).

In addition, the syntactic and semantic literature has made different claims as to which type of noninterrogative noun phrase is more comparable to the type of *wh*-phrase, even though previous processing research on in-situ *wh*-questions primarily focused on comparisons with declaratives with definite noun phrases. Evidence from the theoretical syntax and semantics literature (Cheng, 1991, 1994)

https://doi.org/10.1017/S0142716425000074 Published online by Cambridge University Press

shows that in Mandarin Chinese, simplex *wh*-words are closer to indefinite noun phrases, whereas complex *wh*-phrases are more akin to definite noun phrases (Giannakidou & Cheng, 2006). Previous sentence processing studies showed differences in reading time depending on the referential nature of the noun phrase being tested (e.g., Warren & Gibson, 2002, 2005; Gordon, et al., 2004; Kaan & Vasić, 2004). These studies based their predictions on the *Accessibility* or *Givenness Hierarchy* (Gundel et al., 1993), which determines the accessibility of referents in the discourse and the relation between the type of noun phrase and the degree to which its antecedent is accessible, and they found that, in the absence of prior discourse, definite noun phrases take longer time to be read than their indefinite counterparts. This is because definite noun phrases require the reader to reconstruct their referent from zero, whereas indefinite noun phrases usually introduce new referents and do not require the reader to search for one.

Given the potential influence of noun phrase definiteness on parsing differences between *wh*-questions and declaratives, our experimental manipulation introduced two declarative types: one with definite and one with indefinite noun phrases in the *wh*-phrase position. To investigate the predictions outlined above and extend research on in-situ *wh*-question processing in Mandarin Chinese and the processing of complex and simplex *wh*-phrases, as well as studies on definite and indefinite noun phrases, we conducted four self-paced reading (SPR) experiments (see Jegerski, 2014 for a summary description). SPR's incremental processing methodology is well-suited for this investigation. The first two experiments in French compared the processing of in-situ questions with simplex object *wh*-phrases (*qui* "who") and complex *wh*-phrases (*quel N* "which N") with their declarative counterparts containing both definite and indefinite noun phrases. The second two experiments carried out the same comparisons in Mandarin Chinese. The next sections describe the experimental paradigm, design, and results.

Processing in-situ questions in French

Experiment 1: processing in-situ questions with simplex wh-phrases in French

As described above, research on the processing of in-situ *wh*-questions in French is scarce and researchers mainly concentrated on the prosodic characteristics of these questions or on their acceptability but not so much on their reading comprehension. The goal of Experiment 1 is to determine whether French in-situ questions with simplex *wh*-phrases (*qui* "who") incur predicted processing costs at the disambiguation point in the absence of prosodic and contextual cues, compared to their declarative counterparts.

Method

Participants

Participants ($n = 36$, mean age = 22 years, 18 females) were all native speakers of French. They were recruited in two groups: one from the University of Nantes

https://doi.org/10.1017/S0142716425000074 Published online by Cambridge University Press

(France) (n = 30, mean age = 20 years, 16 females) and one from the Expat French community in the Leiden area[6] (n = 6, mean age = 35 years, 2 females). Testing participants at different locations was done for practical reasons and to achieve the required statistical power. None of the participants suffered from dyslexia and all of them had normal or corrected vision. All participants provided informed consent and were monetarily compensated for their participation.

Materials

We compared object in-situ *wh*-questions with *qui* "who" in (8a), with indefinite noun phrases such as *quelqu'un* "someone" in (8b), and with monosyllabic (n = 9) and bisyllabic (n = 15) half masculine (n = 12) and half feminine (n = 12) proper names such as *Marie* in (8c).

An example of a stimuli set is given in (8).[7] The sentences were presented word-by-word incrementally from left-to-right.

(8) *a. In-situ question with a simplex* wh*-phrase*

Le	braqueur	de	banque	a	blessé	**qui**	dans	sa	fuite ?
The	*robber*	*of*	*bank*	*has*	*hurt*	*whom*	*on*	*his*	*escape?*

'Who did the bank robber hurt on his escape?'

b. Declarative with indefinite object noun phrase

Le	braqueur	de	banque	a	blessé	**quelqu'un**	dans	sa	fuite.
The	*robber*	*of*	*bank*	*has*	*hurt*	*someone*	*on*	*his*	*escape.*

'The bank robber hurt someone on his escape.'

c. Declarative with Proper Name object

Le	braqueur	de	banque	a	blessé	**Marie**	dans	sa	fuite.
The	*robber*	*of*	*bank*	*has*	*hurt*	*Marie*	*on*	*his*	*escape.*

'The bank robber hurt Marie on his escape.'

The experiment consisted of 24 sets of three sentences distributed across three lists in a Latin Square design, which were combined with 72 filler sentences of similar length. Half of the fillers were questions and the other half declaratives.

The modifier of the subject noun phrase *le braqueur* "the robber" varied minimally in its length between two and three words. Most of the items (i.e., 20/24) contained two-word modifiers for the subject, as *de banque* "of bank" in *le braqueur de banque* "the bank robber" (8). The region *dans sa fuite* "on his escape" following the critical position given in bold (i.e., *wh*-word *qui*, indefinite *quelqu'un,* or proper name *Marie*) also differed minimally in length across items, ranging between three (in 15/24 sentences) to four words (in 9/24 sentences). This variation was kept so that the stimuli would sound as natural as possible to French speakers. All materials were checked for grammaticality and naturalness by a French native speaker.

Procedure

Participants signed an informed consent form before the experiment in compliance with the Ethics Code for linguistic research in the Faculty of Humanities at Leiden

https://doi.org/10.1017/S0142716425000074 Published online by Cambridge University Press

University. A self-paced-reading, word-by-word moving window task (Just et al., 1982, Aaronson & Scarborough, 1976) was conducted on a MacBook Pro laptop running the software Linger (Rhode, 2003) in a quiet room at the University of Nantes and in Leiden University. Each trial began with a group of dashes that corresponded to each word in the sentence. Therefore, participants could see the length of the sentence but not the words behind the dashes. Participants were asked to press the space bar to read the sentence word-by-word and to reply to the comprehension question that appeared immediately afterwards on a different screen by pressing the "F" (YES) or "J" (NO) buttons. These responses were indicated with a sticker above the corresponding keys. As participants pressed the space bar to read the sentences, each word was revealed individually and the previously read word disappeared. The punctuation mark at the end of the sentence, which unambiguously determined the interrogative or declarative nature of the sentence, appeared together with the last word of the sentence. This meant that in principle readers of French could not determine whether they were reading a question or a declarative until they reached this sentence final position. Therefore, the reason to choose a word-by-word moving window was to check what the predictions with respect to upcoming material were per word. To keep participants attentive, each sentence was followed by a yes/no comprehension question. The experiment lasted approximately 30 minutes. An example question for item (8a) (repeated here) is shown below.[8]

(8) *a. In-situ question with a simplex* wh-*phrase*

Le	braqueur	de	banque	a	blessé	**qui**	dans	sa	fuite ?
The	*robber*	*of*	*bank*	*has*	*hurt*	*whom*	*on*	*his*	*escape?*

'Who did the bank robber hurt on his escape?'

Comprehension Question:

- Est-ce un braqueur de bijouterie qui a blessé quelqu'un dans sa fuite ?
- *Was it a jewelry store robber who injured someone on his escape?* (Answer : No)

Reading time data analysis

All trials (independently of whether the corresponding comprehension question was answered correctly or not) were included in the analysis. The average comprehension accuracy for the 36 participants was 96% (SD = 1.95%); thus, no participant was rejected on this basis. There was no significant difference in accuracy between declaratives (97.7%) and questions (96.9%), ($\chi^2(1, N = 859) = 0.277$, *Fischer's* $p = 0.49$).

The regions used for the analysis corresponded to single words, except for those cases where French clusters the determiner or preposition with the noun by means of an apostrophe (e.g., *l'infirmière* "the nurse," *d'une* "of one"). The collected reading time data was inspected and outlier data points with reading times smaller than 150 ms or larger than 2000 ms were removed. The total number of discarded data points represented ~1% of the complete data including both fillers and experimental sentences.

https://doi.org/10.1017/S0142716425000074 Published online by Cambridge University Press

There is experimental evidence that word length and frequency impact reading time both in eye tracking (e.g., Kliegl et al., 2004; Hyönä & Olson, 1995) and in self-paced-reding studies (e.g., Bultena et al., 2014), with low frequency and longer words both shown to display increased fixation or reading durations, which is associated with a higher processing cost. To avoid possible confounding effects unrelated to our experimental manipulation and research questions, we addressed the impact of word length and word frequency (when relevant) of the critical regions in the obtained reading times by conducting an ad hoc analysis. First, we tackled the relation of reading time with *word length* in two ways: by calculating length-corrected residual reading times (RSRT) (Ferreira & Clifton, 1986) and by considering individual experimental items as a random factor in the mixed effects model analysis (Barr et al., 2013). The reading time was residualized by computing a linear regression between the word length and reading time for each subject and then subtracting the predicted reading time from the observed reading time for each word. The resulting RSRT were used for all subsequent analyses. Second, to account for the possible effects of *word frequency*, the experimental dataset was expanded with the information contained in the *Lexique* database (New et al., 2004) for French. This was done by extracting a frequency of use for each critical word and matching it to the relevant syntactic category. For inflected words, we used the frequency of the lemma to account for possible effects of word familiarity, whereas in the clusters containing an apostrophe (e.g., *l'infirmière* 'the nurse', *d'une* "of one") we used the frequency of the noun (e.g., *infirmière* "nurse").

We analyzed differences in the RSRT at two regions, - the site of the disambiguation (*wh*-question or NP *qui/quelqu'n/Marie*) and the following word to account for possible spillover effects[9] (Vasishth, 2006), using Linear Mixed Effects Regression (LMER; Baayen et al., 2008) by means of the statistical computing language *R* (R Core Team, 2016) and the *lm4* package (Bates et al., 2015). The model included one fixed-effect factor, *Condition,* with three levels (*wh*-word, indefinite and Proper Name). In the region of the *wh*-question/NP (region 8 in Figure 1), as shown in (8), all the experimental items consist of the same pronouns *qui* or *quelqu'n* or a Proper Name, so we did not include in the model the word frequency for that region.[10] In the region immediately after the *wh*-site, in addition to *Condition,* a fixed effect for *Word Frequency* was considered based on the log-transformed, centered word frequency as extracted from *Lexique*. The maximal random effects structure justified by the model was considered (Barr et al. 2013): variance introduced by subjects and items was modeled as random intercepts. In addition, we considered random slopes by subject for the factor *Condition*.

The best model fitting the data was obtained by likelihood ratio test of models including and excluding the relevant effect and against a "*null*" model containing only an intercept parameter and the random effects structure. A follow-up analysis was performed when a significant effect of *Condition* was found to assess if a different behavior appears between the *wh*-in-situ question and the two types of declaratives.[11]

https://doi.org/10.1017/S0142716425000074 Published online by Cambridge University Press

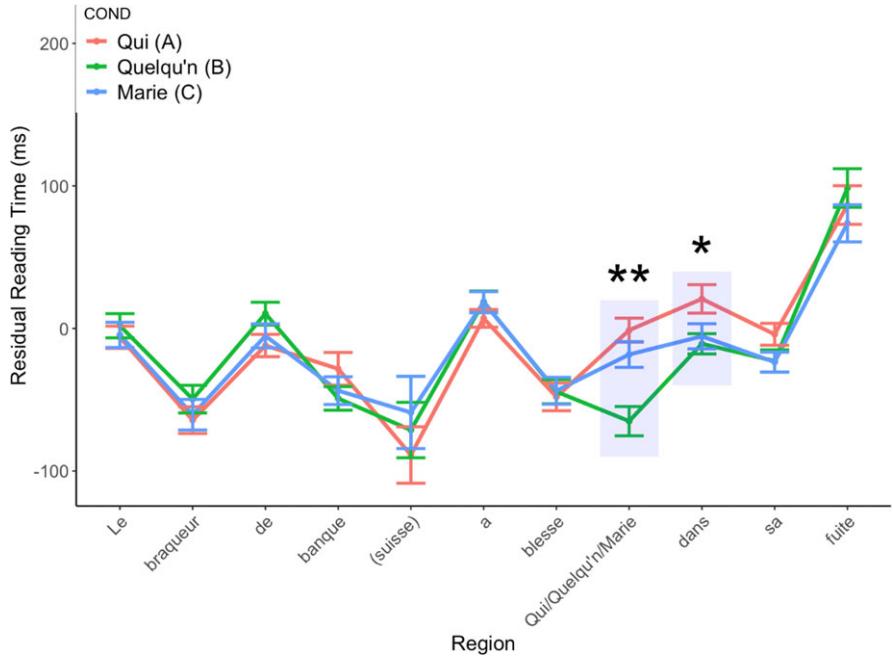

Figure 1. Mean RSRT per word for the comparison between in-situ questions with simplex wh-phrases (*Qui*), declaratives with indefinites (*Quelqu'un*) and proper names (*Marie*) in Experiment 1. Bars indicate the standard error per region.

Results

Figure 1 shows the average RSRT at the different regions of the experimental items against a sample sentence for reference. As shown in this figure, there are two regions that show significant effects. One is the critical region (i.e., *wh*-word "*Qui*"/ indefinite "*Quelqu'n*"/ Proper Name "*Marie*") and the other is the immediately following region (i.e., the preposition "dans"). In both regions, the in-situ *wh*-question condition in (8a) is read significantly slower than its indefinite declarative counterpart in (8b). The definite declarative with a proper name in (8c) is only read significantly slower than the indefinite declarative in (8b) at the critical word region. There is a difference observed between the definite declarative in (8c) and the *wh*-question condition in (8a) at the region following the critical region (i.e., "dans"), where the question appears to be read slower, but this difference did not reach statistical significance.

Post hoc analysis at the critical region (region 8: *Qui/Quelqu'n/Marie*), presented in Table 1, confirmed both in-situ questions with simplex *wh*-phrases (*qui*) ($D = 63.70$ ms, $\chi^2(1) = 12.37$, $p = 0.001$) and declaratives with proper names (*Marie)* ($D = 46.99$ ms, $\chi^2(1) = 8.43$, $p = 0.007$) were read significantly slower than declaratives that contain indefinites (*quelqu'un*). No significant difference was found between the reading time of in-situ questions (*qui*) and declaratives that contain proper names (*Marie*). At the region immediately after the critical region (region 9: *dans* in the example in (8)), we observe a significant increase in reading

https://doi.org/10.1017/S0142716425000074 Published online by Cambridge University Press

Table 1. Pairwise comparison for RSRT at the critical region "Qui/Quelqu'n/Marie" (region 8) and following word "dans" (region 9) in Experiment 1. P-values adjusted with the Holm method for multiple comparisons

	Region 8 (*qui/quelqu'n/Marie*)		Region 9 (*dans*)	
Comparison	*Difference*	χ^2	*Difference*	χ^2
wh-simplex—indefinite pronoun	**63.68****	12.375	**31.71***	6.11
wh-simplex—proper name	16.94	1.481	26.34	3.79
Indefinite pronoun—proper name	***−46.99****	*8.426*	*−5.38*	0.19

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

time on the interrogative condition when compared to the declarative with indefinite pronoun (D = 31.72 ms, $\chi^2(1) = 6.11$, $p = 0.04$), but no significant difference in reading time when comparing the interrogative condition with the declarative with a proper name.

Table 2 provides a summary of the maximal fitted model for the *wh/NP* disambiguation site (region 8: *Qui/Quelqu'n/Marie*) and the region after (region 9), respectively.

The results above show the expected increased effort in processing the in-situ questions with a simplex *wh*-phrase, such as *qui* "who," in French, when compared to declaratives with an indefinite NP. However, this effect is absent when contrasted with a declarative with a definite (proper name) NP. Furthermore, the same processing difficulty is observed between the two declaratives: *ProperName* conditions such as (8c) are also read slower than indefinite declaratives such as (8b). This difference between indefinite and proper names relative to in-situ questions with a simplex *wh*-phrase can be attributed to the greater integration difficulty of proper names compared to other definite or indefinite noun phrases (Ledoux et al., 2007; Camblin et al., 2007). This finding aligns with the *Accessibility Hierarchy* (Gundel et al., 1993), which posits that indefinites require minimal contextual information for interpretation, while definites and proper names necessitate prior knowledge of the referent.

Experiment 2: processing in-situ questions with complex wh-phrases in French

The sentence processing literature showed that complex *wh*-phrases presented in isolation produce longer reading times than their simplex *wh*-phrase counterparts (see De Vincenzi, 1996; Donkers et al., 2011). In Experiment 2, we compared, again in the absence of prosodic and contextual information, the processing of in-situ questions with complex *wh*-phrases (e.g., *quelle caissière* "which cashier," *quel garçon* "which boy") with declaratives with a definite or indefinite noun phrase at the *wh*-phrase. Our prediction again, following the hypothesis described earlier on a bias for a declarative interpretation, is that in-situ questions with complex *wh*-phrases will be read slower than both declarative definite and declarative indefinite sentences in French, as shown by Adli (2015).

https://doi.org/10.1017/S0142716425000074 Published online by Cambridge University Press

Table 2. Model summary for RSRT at the critical region "Qui/Quelqu'n/Marie" (region 8) and following word "dans" (Region 9) in Experiment 1

	Region 8 (*qui/quelqu'n/Marie*)		Region 9 (*dans*)	
Predictors	*Estimates*	*Conf. Int (95%)*	*Estimates*	*Conf. Int (95%)*
(Intercept)	−1.35	−23.95 – 21.25	11.87	−8.75 – 32.49
Word frequency			16.52	−4.55 – 37.59
Interrogative: *wh*-simplex (*Qui*)	*Reference*		*Reference*	
Declarative: indefinite pronoun (*Quelqu'n*)	−63.68 ***	−99.22 – −28.15	−31.71 *	−56.91 – −6.52
Declarative: proper name (*Marie*)	−16.69	−43.62 – 10.23	−26.34	−52.87 – 0.20
Random effects				
σ^2	21345.91		21102.91	
τ_{00}	658.89 $_{SUBJ}$		0.00 $_{SUBJ}$	
	938.03 $_{ITEM}$		72.55 $_{ITEM}$	
τ_{11}	6424.85 $_{SUBJ.CONDQuelqu'n\ (B)}$		627.17 $_{SUBJ.CONDQuelqu'n\ (B)}$	
	1371.19 $_{SUBJ.CONDMarie\ (C)}$		1285.60 $_{SUBJ.CONDMarie\ (C)}$	
ρ_{01}	−0.22 $_{SUBJ.CONDQuelqu'n\ (B)}$			
	−0.90 $_{SUBJ.CONDMarie\ (C)}$			
N	36 $_{SUBJ}$		36 $_{SUBJ}$	
	24 $_{ITEM}$		24 $_{ITEM}$	
Observations	857		861	
Marginal R^2/Conditional R^2	0.033/NA		0.012/NA	
AIC	11048.363		11033.268	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
p-values calculated based on conditional *F*-tests with Kenward-Roger approximation.
Marginal R2 based on Nakagawa et al., (2017).

The aim of this experiment was to examine: first, if complex *wh*-phrases lead to comparable processing cost as the processing of in-situ questions with simplex *wh*-phrases, as examined in Experiment 1 and secondly, whether the contrast between complex *wh*-phrases and their declarative definite/indefinite counterparts will show similar effects in terms of timing and effect size as those observed in Experiment 1. The second research question is motivated by the syntactic and semantic debate regarding the comparability of different noun phrase and of *wh*-phrase types (Giannakidou & Cheng, 2006), as discussed above.

Method

Participants

Participants ($n = 36$, mean age = 22 years, 25 females) were all native speakers of French. They were recruited in two groups: one from the University of Nantes ($n = 31$, mean age = 20 years, 22 females) and one from the Expat French community in the Leiden area[12] ($n = 5$, mean age = 35 years, 3 female) respectively. As in Experiment 1, participants were tested at different locations for practical reasons and to achieve the statistical power we needed. Participants in this experiment were different from those in Experiment 1 to avoid repetition of the content in the experimental stimuli that differed minimally at the critical region. None of the participants suffered from dyslexia and all of them had normal or corrected vision. They all provided informed consent and were monetarily compensated for their participation.

Materials

A sample set of stimuli for Experiment 2 is given in (9).[13] Here, we compared object in-situ *wh*-questions formed with complex *wh*-phrases such as *quelle caissière* "which cashier" in (9a) with indefinite noun phrases such as *une caissière* "a cashier" in (9b), and with definite noun phrases such as *la caissière* "the cashier" in (9c).

(9) *a. In-situ question with a complex wh-phrase*
Le braqueur de banque a blessé **quelle caissière** dans sa fuite ?
The robber of bank has hurt which cashier in his escape?
'Which cashier did the bank robber hurt on his escape?'

b. Declarative with indefinite object noun phrase
Le braqueur de banque a blessé **une caissière** dans sa fuite.
The robber of bank has hurt a cashier in his escape.
'The bank robber hurt a cashier on his escape.'

c. Declarative with definite object noun phrase
Le braqueur de banque a blessé **la caissière** dans sa fuite.
The robber of bank has hurt the cashier in his escape.
'The bank robber hurt the cashier on his escape.'

Experiment 2 consisted of 24 sets of three sentences distributed across three lists in a Latin Square design, which were combined with 72 filler sentences of similar

https://doi.org/10.1017/S0142716425000074 Published online by Cambridge University Press

length. Half of the fillers were questions and half were declaratives. The fillers of Experiment 2 were a bit different from those of Experiment 1 to match the variation at the critical *wh*-phrase/noun phrase region.

Procedure

The Ethics Protocol and experimental procedure were the same as in Experiment 1. As in Experiment 1, we chose to present the stimuli in a strict word-by-word manner, rather than as constituents (e.g., *which book*). Both methodological approaches were previously used to present stimuli consisting of complex *wh*-phrases. In research employing the SPR methodology, the majority of studies adopted the approach of presenting them as constituents, where the *wh*-determiner (*which)* of the complex *wh*-phrase was displayed to readers together with the noun with which it formed the complex *wh*-phrase (e.g., *quale ingegnere* "which engineer" in Italian (De Vincenzi, 1991, 1996), *welke bediende* "which servant" in Dutch (Donkers et al., 2011) and *nǎxiē guānyuán* "which$_{\text{plural}}$ officials" in Mandarin Chinese (Xiang et al., 2013; Xiang et al., 2015). Other studies, such as Kaan et al.'s (2000) ERP study, adopted the word-by-word approach where the *wh*-determiner was presented separately from the *wh*-phrase noun in complex *wh*-phrases (e.g., *which popstar)*. Since the main research question of our study was to run a direct comparison between the processing of declaratives with in-situ *wh*-questions, and we wanted to keep a close parallel between the way simplex *wh*-phrases and complex *wh*-phrases were presented and read by participants, we kept a word-by-word presentation for both in-situ questions with simplex and complex *wh*-phrases. This allowed the comparison between the incremental processing of the two kinds of questions with their declarative counterparts to be as closely matched as possible.

Reading time data analysis

The analysis procedure was as in Experiment 1 with the following exceptions: i) mixed-effect model considered the same predictors, interactions, and random effects but in this case *Word Frequency* was considered in all regions, ii) analysis was performed in three regions, the two corresponding to the complex *wh*-phrase, and the immediately posterior region "*dans*". The average comprehension accuracy for all 36 participants was 95.3%, without significant difference in accuracy between declaratives (95.3%) and questions (95.5%), ($\chi^2(1, N = 856) = 0.0001$, *Fischer's* $p = 1.0$), so no participant was excluded from the analysis. As in Experiment 1, each word was considered a region, except for those words that contained a determiner (or preposition), plus noun clusters connected via an apostrophe. The same exclusion criteria for outliers were used as in Experiment 1 resulting in 1.5% of data discarded, including both filler and experimental items.[14]

Results

Figure 2 shows the average RSRT per region for the three conditions described in (9) with a sample sentence. As it can be seen in the figure, there are two main effects. One occurs at the noun within the critical *wh*-phrase and noun phrase, where the

https://doi.org/10.1017/S0142716425000074 Published online by Cambridge University Press

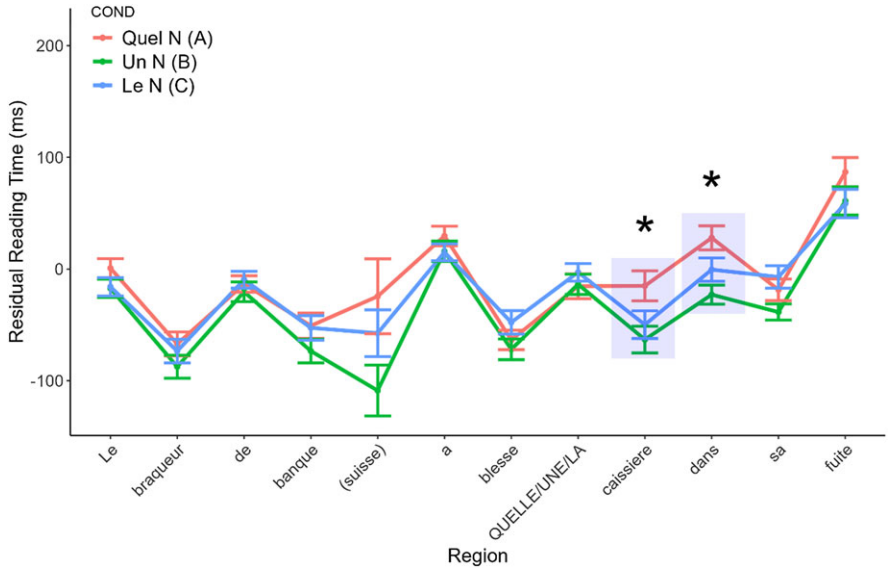

Figure 2. Mean RSRT per word region for the comparison in Experiment 2 between in-situ questions with complex wh-phrases ("*quelle caissière*") and declaratives with indefinites ("*une caissière*") and definites ("*la caissière*").

in-situ questions with complex wh-phrases ("*quelle caissière*") in (9a) are read significantly slower than the declaratives with indefinites ("*une caissière*") in (9b) at the noun "*caissière" (cashier)*. This effect is also present at the following region of the preposition "dans". The declaratives with definites ("*la caissière*") in (9c) are read faster at the noun "*caissière" (cashier)* and at the preposition *"dans"* than the interrogatives in (9c) but this difference is not significant.

Table 3 contains the mixed-effect model for each of the three regions of interest. The LMER model fitted included random intercepts for both *Subject* and experimental *Item* and random slopes by *Condition* for *Subject.* A significant effect of the experimental manipulation was observed in regions 9 and 10 (*caissiere* "cashier" and *dans* "in" in example (9)). *Word Frequency* did not affect the observed results, and introducing the effect did not provide an improvement of the model fit for the observed reading time.

As in Experiment 1, a follow-up analysis of the regions with significant differences (regions 9 and 10 = *caissiere* "cashier" and *dans* "*in*") was performed. Post hoc pairwise comparisons (see Table 4 and Figure 3) confirmed that complex *wh-in-situ* questions (*quelle caissière*) are read significantly slower than indefinites (*une caissière*) in the noun part of the *wh*-phrase *(caissière)* ($D = 47.30$ *ms,* $\chi^2(1) = 5.80$, $Pr(>\chi^2) = 0.048$). However, there is no significant difference between complex *wh-in-situ* questions and declarative sentences with definite noun phrases (*la caissière*) ($D = 30.97$, $\chi^2(1) = 2.10$, $Pr(>\chi^2) = 0.29$). In addition, the effect appears to continue in the region after the *wh*-phrase (preposition "*dans*" in (9)), where again, only the indefinite declarative shows a significant difference from the interrogative condition ($D = 55.85$, $\chi^2(1) = 12.84$,

https://doi.org/10.1017/S0142716425000074 Published online by Cambridge University Press

Table 3. Model summary for RSRT at the critical regions 8, 9 and 10 in (9) ("*quelle*", "*caissière*," and "*dans*") in Experiment 2

	Region 8 (*quelle/une/la*)		Region 9 (*caissiere*)		Region 10 (*dans*)	
Predictors	*Estimates*	*Conf. Int (95%)*	*Estimates*	*Conf. Int (95%)*	*Estimates*	*Conf. Int (95%)*
(Intercept)	40.56	−34.81 – 115.93	2.74	−42.35 – 47.84	23.01	−3.75 – 49.76
Word frequency	210.64	−61.95 – 483.23	17.19	−10.78 – 45.17	10.20	−14.95 – 35.35
Interrogative: wh-complex (Quel N)	*Reference*		*Reference*		*Reference*	
Declarative: indefinite (Un N)	−244.35	−563.99 – 75.30	**−47.38** *	−85.85 – −8.92	**−55.85** **	−86.44 – −25.25
Declarative: definite(Le N)	−272.62	−642.52 – 97.29	−30.92	−72.73 – 10.89	−26.29	−58.58 – 6.00
Random effects						
σ^2	24445.80		41094.28		26406.63	
τ_{00}	370.95 $_{\text{SUBJ}}$		3449.16 $_{\text{SUBJ}}$		752.13 $_{\text{SUBJ}}$	
	439.93 $_{\text{ITEM}}$		260.43 $_{\text{ITEM}}$		740.92 $_{\text{ITEM}}$	
τ_{11}	952.51 $_{\text{SUBJ.CONDUn N (B)}}$		2850.44 $_{\text{SUBJ.CONDUn N (B)}}$		2349.80 $_{\text{SUBJ.CONDUn N (B)}}$	
	359.32 $_{\text{SUBJ.CONDLe N (C)}}$		5370.24 $_{\text{SUBJ.CONDLe N (C)}}$		3346.25 $_{\text{SUBJ.CONDLe N (C)}}$	
ρ_{01}	−1.00 $_{\text{SUBJ.CONDUn N (B)}}$		−0.05 $_{\text{SUBJ.CONDUn N (B)}}$		−0.62 $_{\text{SUBJ.CONDUn N (B)}}$	
	−1.00 $_{\text{SUBJ.CONDLe N (C)}}$		−0.48 $_{\text{SUBJ.CONDLe N (C)}}$		−0.56 $_{\text{SUBJ.CONDLe N (C)}}$	
ICC			0.11		0.08	
N	36 $_{\text{SUBJ}}$		36 $_{\text{SUBJ}}$		36 $_{\text{SUBJ}}$	
	24 $_{\text{ITEM}}$		23 $_{\text{ITEM}}$		24 $_{\text{ITEM}}$	
Observations	855		810		893	
Marginal R^2/Conditional R^2	0.004/NA		0.010/0.117		0.019/0.096	
AIC	11071.539		10957.805		11671.156	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
p-values calculated based on conditional F-tests with Kenward-Roger approximation.
Marginal R2 based on Nakagawa et al., (2017).

Table 4. Pairwise comparison for RSRT at the noun region in the *wh*-complex phrase (region 9: "caissiere" in (9)) and following word (Region 10: *dans* in (9)) in Experiment 2

	Region 9 (*caissiere*)		Region 10 (*dans*)	
Comparison	*Difference*	χ^2	*Difference*	χ^2
wh-complex - indefinite noun phrase	**47.31***	5.80	**55.85****	12.84
wh-simplex - definite noun phrase	30.97	2.10	26.29	2.55
Indefinite noun phrase–Definite noun phrase	*–16.34*	*0.62*	***–29.56 *** ($p = 0.055$)	4.84

° $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
p-values adjusted with Holm method for multiple comparisons.

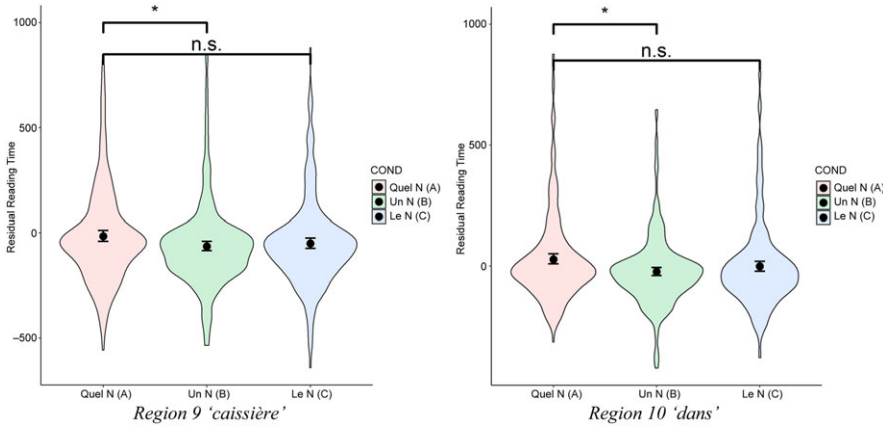

Figure 3. Mean RSRT in Experiment 2 at the *wh*-phrase word "*caissiere*" (left) and at the word after the *wh*-phrase "*dans*" *(right).* Black dot indicates the mean value, and bars indicate the 95% bootstrapped confidence interval of the mean.

$Pr(>\chi^2) = 0.001$). Declaratives with definite names (as in 9c) do not show a significant effect compared with the interrogative (9a), but show, as in the case of questions with simplex *wh*-phrases, a larger reading time compared with indefinite declaratives (9b), although marginally significant ($D = 29.56$ *ms,* $\chi^2(1) = 4.84$, $Pr(>\chi^2) = 0.055$).

The results show a similar pattern as in-situ questions with simplex *wh*-phrases and are consistent with the expected increased effort in the processing of in-situ questions with *wh*-phrases in French compared to declaratives. Again, this is only the case with indefinite declaratives. In the case of complex *wh*-phrases, the effect is observed not when the *wh*-word is encountered, at the determiner position, but rather at the *wh*-phrase boundary (noun). Further, it extends to the word immediately after the *wh*-phrase, in an apparent spillover effect.

https://doi.org/10.1017/S0142716425000074 Published online by Cambridge University Press

Qualitative comparison of results from processing French in-situ wh-questions

In general, the results of Experiments 1 and 2 support the hypothesis that, in the absence of prosody and contextual information, in-situ *wh*-questions in French are not anticipated during parsing. Both in-situ questions with a simplex *wh*-phrase and in-situ questions with a complex *wh*-phrase are processed significantly slower than indefinite declaratives.

The finding that indefinites (e.g., *quelqu'un* "someone", *une caissière* "a cashier") in declaratives are read faster than in-situ wh-questions aligns with previous research (Warren & Gibson, 2002; Warren & Gibson, 2005, Gordon et al, 2004, Kaan & Vasić, 2004) resulting in longer reading times for definite noun phrases in the absence of previous discourse. This pattern supports the *Accessibility Hypothesis* (Gundel et al., 1993), which posits that indefinites require less processing effort due to their lower referential demands. In our case, experimental sentences were presented in isolation and no prior discourse was introduced; therefore, the processing of a definite noun phrase or proper name (Ledoux et al., 2007; Camblin et al., 2007) can be as costly (and therefore result in longer reading times) as the processing of an in-situ *wh*-phrase.

With respect to the comparison between how in-situ simplex and complex *wh*-phrases behave, results show a timing difference: for simplex *wh*-phrases, effects occur already at the position of the *wh*-phrase *qui* "who", and at the word immediately after, whereas in complex *wh*-phrases, effects are apparent at the *wh*-phrase boundary (noun) and at the word immediately after this noun. In complex *wh*-phrases, readers might wait for the *wh*-phrase boundary completion to build an interpretation, considering that the determiner *quel(le)* "which" in French is not informative on its own. The observed "spill-over effect" to the next word can reflect an open processing of the *wh*-phrase as well, since complex *wh*-phrases can be followed by a postnominal modifier (e.g., *quelle caissière de supermarché* "which supermarket cashier" (lit. "which cashier of supermarket") which is not the case for simplex *wh*-phrases (**qui de supermarché* "*who of supermarket").

Experiment 2 on French data indicates that parsing comparisons between *wh*-questions and declaratives require a consideration of semantic factors beyond in-situ *wh*-question processing, such as the nature of the referential elements, which may influence the processing of declaratives.

Processing in-situ questions in Mandarin Chinese

Next, we present two experiments on in-situ questions containing simplex and complex *wh*-phrases in Mandarin Chinese, a language that always applies the in-situ question strategy formation, with a similar setup of Experiments 1 and 2 in French (considering what the cross-linguistic differences allowed) to compare qualitatively the results of French in-situ *wh*-questions processing with those of Mandarin Chinese.

https://doi.org/10.1017/S0142716425000074 Published online by Cambridge University Press

Experiment 3: Processing in-situ questions with simplex wh-phrases in Mandarin Chinese

In Experiment 3, we compared in-situ questions with simplex *wh*-phrases (*shéi* "who") with two types of declaratives, one containing a definite object, and the other an indefinite object.

Method

Participants

Participants (n = 36, mean age = 20 years, 16 females) were all native speakers of Mandarin Chinese and were recruited from Tsinghua University in Beijing, China. None of the participants suffered from dyslexia and all of them had normal or corrected vision. All participants provided informed consent and were compensated monetarily for their participation according to the local standards.

Materials

In this experiment, we compared in-situ *wh*-questions with the simplex *wh*-phrase *shéi* "who" in object position, as in (10a), with indefinite noun phrases such as *rén* "person/someone" in (10b), and with disyllabic proper names such as *Xiǎozhāng*[15] in (10c). We constructed sentences that contained an intensional verb and without perfective marker *–le* to constrain the reading of bare nouns such as *rén* "person/someone" to a nonspecific indefinite reading.[16] Moreover, the intensional context further allowed two regions following the in-situ *wh*-phrase to grant observation of effects that could potentially occur after the critical region. An example of a set of stimuli is given in (10).[17]

(10) a. *In-situ question with a simplex phrase*

那个	男生	想要	求	**谁**	解决	问题?
Nàgè	nánshēng	xiǎngyào	qiú	**shéi**	jiějué	wèntí?
That/the	boy	want	beg	who	solve	problem

'Who does the boy want to beg to solve the problem?'

b. *Declarative with indefinite object noun phrase*

那个	男生	想要	求	人	解决	问题.
Nàgè	nánshēng	xiǎngyào	qiú	**rén**	jiějué	wèntí.
That/the	boy	want	beg	person	solve	problem

'The boy wants to beg someone to solve the problem.'

c. *Declarative with Proper Name object*

那个	男生	想要	求	小张	解决	问题.
Nàgè	nánshēng	xiǎngyào	qiú	**Xiǎozhāng**	jiějué	wèntí.
That/the	boy	want	beg	Xiaozhang	solve	problem

'The boy wants to beg Xiaozhang to solve the problem.'

Experiment 3 consisted of 24 sets of three sentences distributed across three lists in a Latin Square design, which were combined with 72 filler sentences of similar length. Half of the fillers were questions and half were declaratives.

https://doi.org/10.1017/S0142716425000074 Published online by Cambridge University Press

Procedure

The Ethics Protocol was as in Experiment 1. A self-paced-reading, word-by-word (where each word consisted of one to three characters) moving window task (Just et al, 1982; Aaronson & Scarborough, 1976) was conducted on a MacBook Pro laptop running the software Linger (Rhode, 2003) in a quiet room in Tsinghua University. Each trial began with a group of dashes that corresponded to each word in the sentence. Therefore, participants could see the length of the sentence but not the words behind the dashes. Participants were asked to press the space bar to read the sentence word-by-word and to reply to the comprehension question that appeared immediately afterwards in a different screen by pressing the "F" (YES) or "J" (NO) buttons. These responses were indicated with a sticker above the corresponding keys. As participants pressed the space bar to read the sentences, each word was revealed individually and the previously read word disappeared. The punctuation mark at the end of the sentence, which unambiguously determined the interrogative or declarative nature of the sentence, appeared together with the last word of the sentence. As it is standard in Chinese, no spaces were provided in between characters, and readers read only one-to-two-character words at a time incrementally on the screen as they pressed the space bar. To keep participants attentive, each sentence was followed by a yes/no comprehension question. The experiment lasted approximately 30 minutes.

Reading time analysis

The average comprehension accuracy for the 36 participants was 96.3%, with no difference in accuracy between questions (97.6%) and declaratives (95.7%) ($\chi^2(1, N = 864) = 1.464$, *Fischer's* $p = 0.23$). Therefore, all trials were included in the analysis, regardless of comprehension question accuracy. All regions used for the analysis corresponded to single words, which ranged from one to two characters.[18]

Following previous studies on Mandarin (Wu et al., 2012; Xiang, et al. 2015; Li et al., 2019) and other East Asian languages (e.g., Kwon & Sturt, 2013 for Korean; Witzel & Witzel, 2016 for Japanese), we log-transformed the individual raw reading times (RWRTs) in each region to correct for the skewness of the distribution.[19] As in Experiments 1 and 2 for French, data points with RWRT shorter than 100 ms or larger than 2000 ms were excluded from the analysis, affecting <1% data. The resulting log-RTs at each region were analyzed using linear mixed effects model analysis (LMER; Baayen et al, 2008) with *Sentence Type (declarative vs question)* as fixed-effect factor. In addition, random intercepts for both *Item* and *Subject* were included as well as slopes by *Condition* for *Subject*. The effect of *Word Frequency* on the reading time was accounted for by introducing in the model fit a fixed factor *logW* which contained the log-frequency of the specific word as extracted from the Chinese database on film subtitles SUBTLEX-CH (Cai & Brysbaert, 2010). The variable introduced represents the logarithm of the count of appearances of the selected word in a database. As for the French *wh*-simplex experiment, in the critical region of the *wh*-question for Mandarin, all items have the same pronouns (谁/*shéi* in 10a, 人/*rén* in 10b) or proper names (as in (10c)), the latter not found in the database, therefore a *Word Frequency*

https://doi.org/10.1017/S0142716425000074 Published online by Cambridge University Press

factor is not included in the statistical analysis for that region. The predictor *logW* was centered and included in the LMER model as a fixed factor in the analysis for other regions with a variation of words across items. Analyses were performed as in the French experiments at two sites: the disambiguation region of the *wh*-question (region 5: in (10) 谁/*shéi*, 人/*rén*, 小张/ *Xiǎozhāng*) and in the immediately posterior region (region 5: in (10) 解决/ *jiějué*). As in Experiments 1 and 2, analyses were performed using the statistical computing language *R* (R Core Team, 2016) and the *lm4* package (Bates et al., 2015).[20]

Results

LMER model in Table 5 shows the final best-fitting model. A fixed effect of *Condition* was retained in the final model and random intercepts by Subject and Item. Random slopes did not significantly improve the fit.

Table 5. Model summary for RSRT at the critical region 5 (*shie/ren/Xiaozhang*) and the immediately posterior region 6 (*jiejue*) in (10) in Experiment 3

	Region 5 (*shie/ren/Xiaozhang*)		Region 6 (*jiejue*)	
Predictors	*Estimates*	*Conf. Int (95%)*	*Estimates*	*Conf. Int (95%)*
(Intercept)	2.53 ***	2.50 – 2.57	2.58 ***	2.55 – 2.61
Word frequency			0.01	−0.00 – 0.02
Interrogative: simplex question	*Reference*		*Reference*	
Declarative: indefinite	0.02 *	0.00 – 0.03	−0.03 **	−0.05 – −0.01
Declarative: proper name	0.02 **	0.01 – 0.04	−0.00	−0.02 – 0.02
Random effects				
σ^2	0.01		0.01	
τ_{00}	0.01 $_{SUBJ}$		0.01 $_{SUBJ}$	
	0.00 $_{ITEM}$		0.00 $_{ITEM}$	
τ_{11}	0.00 $_{SUBJ.CONDDeclarative\ Indefinite\ (B)}$			
	0.00 $_{SUBJ.CONDDeclarative\ Proper\ Name\ (C)}$			
ρ_{01}	−0.07 $_{SUBJ.CONDDeclarative\ Indefinite\ (B)}$			
	0.70 $_{SUBJ.CONDDeclarative\ Proper\ Name\ (C)}$			
ICC	0.46		0.38	
N	36 $_{SUBJ}$		36 $_{SUBJ}$	
	24 $_{ITEM}$		24 $_{ITEM}$	
Observations	863		864	
Marginal R^2/Conditional R^2	0.005/0.468		0.012/0.386	
AIC	−1360.600		−1149.053	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

https://doi.org/10.1017/S0142716425000074 Published online by Cambridge University Press

Table 6. Pairwise comparison for RSRT at the noun region in the *wh*-simplex phrase (region 5: "*shie/ren/Xiaozhang*" in (10)) and following word (Region 6: *jiejue* in (10)) in Experiment 3

	Region 5 (*shie/ren/Xiaozhang*)		Region 6 (*jiejue*)	
Comparison	*Difference*	χ^2	*Difference*	χ^2
wh-simplex - indefinite pronoun	–0.018	4.21	**0.031****	10.27
wh-simplex – proper name	**–0.023***	7.00	0.002	0.04
Indefinite pronoun – proper name	*–0.005*	*0.33*	***–0.029****	9.07

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
p-values adjusted with Holm method for multiple comparisons.

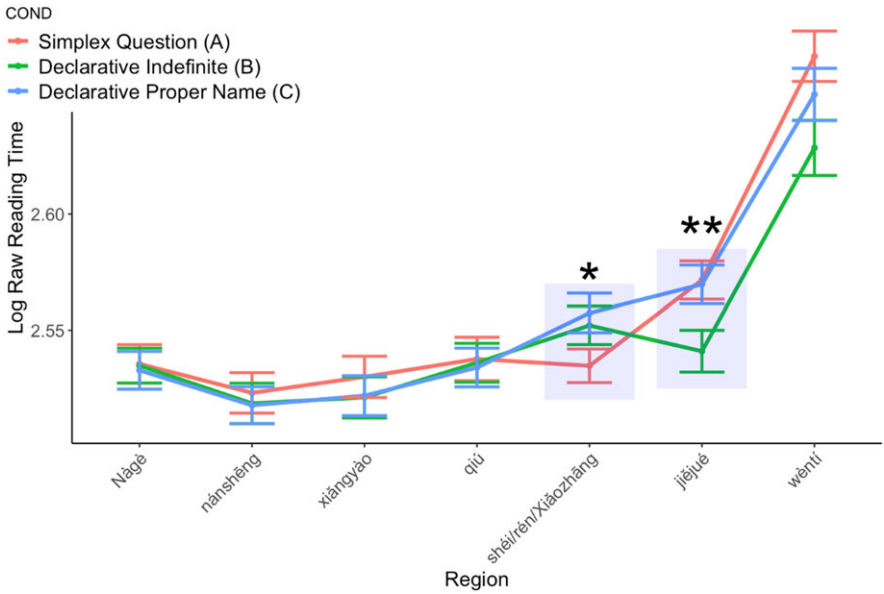

Figure 4. Mean Log Reading times per word/region for the comparison between in-situ questions with simplex *wh*-phrases with *shéi* and declaratives in Experiment 3.

Figure 4 shows the average LogRT at each region for all three experimental conditions in (10) against a sample sentence for reference. As shown in this figure, and corroborated by the post hoc analysis in Table 6, there are two regions that show significant effects. One is the critical region (i.e., 谁/*shei*, 人/*ren*, 小张/ *Xiǎozhāng*) and the other is the immediately following region (i.e., 解决/ *jiějué* "to solve"). In the region of the in-situ *wh*-question condition in (10a), the in-situ *wh*-phrase (谁/*shei*) is read significantly faster than the definite declarative counterpart ($D = -0.023$, $\chi^2(1) = 7.00$, $Pr(>\chi^2) = 0.02$) with a proper name in (10c). No other comparisons reached significance at that region. In the immediately following region (i.e., 解决/ *jiějué* "to solve") however, both the in-situ question with the simplex *wh*-phrase

https://doi.org/10.1017/S0142716425000074 Published online by Cambridge University Press

(D = 0.031, $\chi^2(1) = 10.27$, $Pr(>\chi^2) = 0.004$) and the definite declarative with a proper name (D = 0.029, $\chi^2(1) = 9.07$, $Pr(>\chi^2) = 0.005$) are read significantly slower than the indefinite declarative.

In other words, for in-situ questions with a simplex *wh*-phrase (*shéi* "who") in Mandarin Chinese, the expected increased reading time of questions compared to declaratives is only observed with respect to definite declaratives conditions (the proper name *Xiǎozhāng*) at the region after the *wh*-word. The unexpected rapid processing of the interrogative compared with the other conditions at the *wh*-phrase position could be attributed to the difference in length between the conditions. Unlike the region that immediately follows the *wh*-phrase (i.e., 解决/ *jiějué* "to solve"), where all items have equal duration, the *wh*-question region presents variability: proper names consist of two characters (i.e., 小张/ *Xiǎozhāng*), while *wh*-phrase and indefinite pronouns comprise only one (i.e., 谁/*shei*, 人/*ren*). Although a small variation, word length in characters in Chinese has been shown to affect the reading time in measures of fixation duration in eye-tracking studies (Zang et al., 2018).

Experiment 4: Processing in-situ questions with complex wh-phrases in Mandarin Chinese

In Experiment 4, we compare in-situ questions with complex *wh*-phrases in Mandarin Chinese (*nǎgè tóngxué* "which classmate") with declaratives that contained a non-interrogative noun phrase (e.g., definite: *nàgè tóngxué* "the classmate"; indefinite: *yígè tóngxué* "a classmate").

Method

Participants

Participants ($n = 54$, mean age = 27 years, 31 females) were all native speakers of Mandarin Chinese and were recruited from the pool of MA and PhD students from China studying at Leiden University. Participants were recruited locally instead of in China for practical reasons. None of the participants suffered from dyslexia and all of them had normal or corrected vision. All participants provided informed consent and were monetarily compensated according to the local standards.

Materials

In this experiment, we compared object in-situ *wh*-questions with *wh*-phrases such as *nǎgè tóngxué* "which classmate" in (11a), with declaratives that contained indefinite nouns phrases such as *yígè tóngxué* "a classmate" in (11b), and declaratives that contained definite noun phrases such as *nàgè tóngxué* "the classmate" in (11c). As in Experiment 3 on in-situ questions simplex *wh*-phrases in Mandarin Chinese, we tested sentences that had an intensional verb and no perfective marker *-le*. (11) provides a set of sample stimuli.[21]

https://doi.org/10.1017/S0142716425000074 Published online by Cambridge University Press

(11) a. *In-situ question with a complex phrase*

那个	男生	想要	求	**哪个**	**同学**	解决	问题?
Nàgè	nánshēng	xiǎngyào	qiú	***nǎgè***	***tóngxué***	jiějué	wèntí?
That/the	boy	want	beg	which	classmate	solve	problem

'Which classmate does the boy want to beg to solve the problem?'

b. *Declarative with indefinite object noun phrase*

那个	男生	想要	求	一个	同学	解决	问题?
Nàgè	nánshēng	xiǎngyào	qiú	***yígè***	***tóngxué***	jiějué	wèntí?
That/the	boy	want	beg	a	classmate	solve	problem

'The boy wants to beg a classmate to solve the problem.'

c. *Declarative with definite object noun phrase*

那个	男生	想要	求	那个	同学	解决	问题?
Nàgè	nánshēng	xiǎngyào	qiú	***nàgè***	***tóngxué***	jiějué	wèntí?
That/the	boy	want	beg	the	classmate	solve	problem

'The boy wants to beg the classmate to solve the problem.'

Experiment 4 consisted of 24 sets of three sentences distributed across three lists in a Latin Square design, which were combined with 72 filler sentences of similar length. Half of the fillers were questions and half were declaratives.

Procedure

The Ethics Protocol and procedure were like in Experiment 3, except that it was tested at Leiden University in the Netherlands.

Reading time analysis

The analysis followed the same as Experiment 3. Starting from a maximal model, the simplest best-fitting model explaining the data was retained. This model contained the same predictors, interactions, and random effects as those used for Experiment 3. The average comprehension accuracy for the 54 participants was 95.4%, with a similar high rating on both questions (94.1%) and declaratives (96.7%), although with a significantly slightly higher accuracy level in declaratives ($\chi^2(1, N = 1280) = 4.228$, *Fischer's* $p = 0.04$). As in Experiment 2, we analyzed the regions corresponding to the *wh*-phrase/NP (region 5, *nǎgè* "which"/*yígè* 'a'/*nàgè* "the/that" and 6, *tóngxué* "classmate") and the immediately posterior word (region 7, *jiějué* "to solve").[22]

Results

Figure 5 shows the average LogRT at each region for all three experimental conditions in (11) against a sample sentence for reference. As shown in this figure and corroborated by the statistical analysis (Table 7), there are two regions that show significant effects. One is the determiner in the *wh*-phrase/NP in the critical region (i.e., *nǎgè* "which"/*yígè* "a"/*nàgè* "the/that") and the other is the immediately following noun region (i.e., *tóngxué* "classmate"). The word immediately following the *wh*-phrase/NP (*jiějué* "*to solve*") does not show any differences. In the region at

https://doi.org/10.1017/S0142716425000074 Published online by Cambridge University Press

Table 7. Pairwise comparison for RSRT at the *wh*-complex phrase regions (region 5: *nǎgè/yígè/nàgè* and region 6 *tóngxué*: in (11)) and following word (Region 7: *jiějué* in (11)) in Experiment 4

	Region 5 (nǎgè/yígè/nàgè)		Region 6 (*tóngxué*)		Region 7 (*jiějué*)	
Comparison	*Difference*	χ^2	*Difference*	χ^2	*Difference*	χ^2
wh-complex – indefinite	**0.023***	7.31	**0.055*****	30.11	0.016	1.89
wh-simplex – definite	0.018	3.70	**0.019***	4.47	–0.001	0.02
Indefinite– definite	*–0.005*	*0.32*	***–0.037****	17.47	–0.017	2.91

* $p < 0.05$, **, $p < 0.01$, *** $p < 0.001$.
p-values adjusted with Holm method for multiple comparisons.

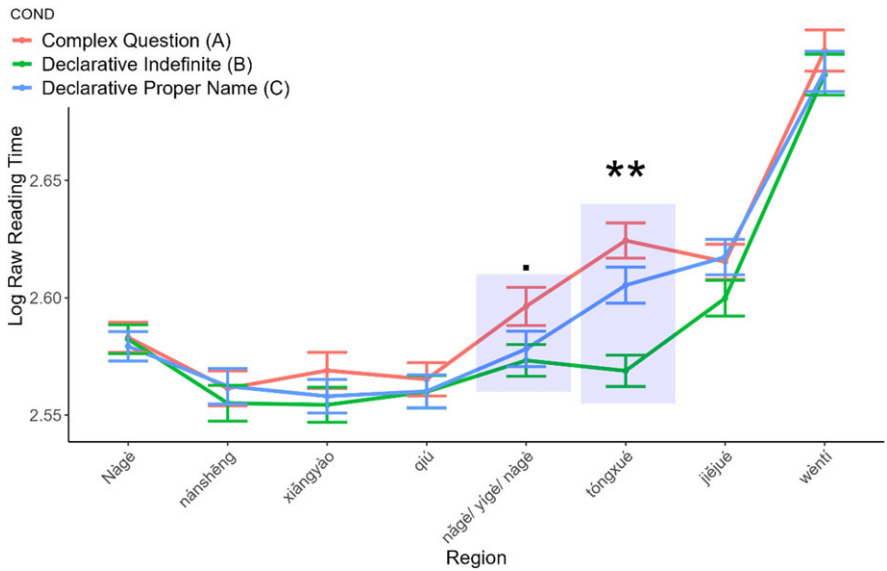

Figure 5. Mean log reading times per word/region for the comparison between in-situ questions with complex *wh*-phrases (nǎgè tóngxué "which classmate"), declaratives with an indefinite phrase (yígè tóngxué "a classmate") and declaratives with a definite phrase (nàgè tóngxué "that classmate") in Experiment 4.

the start of the *wh*-phrase/NP, the *wh*-in-situ condition (i.e., *nǎgè* "which") in (11a) is read significantly slower than the indefinite declarative (*yígè* "a") in (11b) *($D = 0.023$, $\chi^2(1) = 7.31$, $Pr(>\chi^2) < 0.05$)*. In the immediately following region (i.e., *tóngxué* "classmate"), the *wh*-in-situ condition (i.e., *nǎgè* "which") in (11a) is again read significantly slower than the indefinite declarative in (11b) ($D = 0.055$, *$\chi^2(1) = 30.11$, $Pr(>\chi^2) < 0.001$*) and the definite declarative (*nàgè* "the/that") in (11c) ($D = 0.019$, *$\chi^2(1) = 4.47$, $Pr(>\chi^2) < 0.05$*). Further, the indefinite declaratives in (11b) were read significantly faster than the definite declaratives in (11c) ($D = -0.037$, *$\chi^2(1) = 17.47$, $Pr(>\chi^2) < 0.01$*) at this noun position.

Table 8 shows the best-fitting LMER model with a retained fixed effect of *Condition*.

https://doi.org/10.1017/S0142716425000074 Published online by Cambridge University Press

Table 8. Model summary for RSRT at the *wh*-complex phrase regions (region 5: *nǎgè/yígè/nàgè* and region 6 *tóngxué*: in (11)) and following word (Region 7: *jiějué* in (11)) in Experiment 4

	Region 5 (nǎgè/yígè/nàgè)		Region 6 (*tóngxué*)		Region 7 (*jiějué*)	
Predictors	*Estimates*	*Conf. Int (95%)*	*Estimates*	*Conf. Int (95%)*	*Estimates*	*Conf. Int (95%)*
(Intercept)	2.60 ***	2.57 – 2.63	2.63 ***	2.60 – 2.65	2.61 ***	2.58 – 2.64
Word Frequency			0.00	−0.00 – 0.01	−0.01	−0.02 – 0.00
Interrogative: Complex Question	*Reference*		*Reference*		*Reference*	
Declarative: Indefinite	−0.02 **	−0.04 – −0.01	−0.06 ***	−0.08 – −0.04	−0.02	−0.04 – 0.01
Declarative: definite noun	−0.02	−0.04 – 0.00	−0.02 *	−0.04 – −0.00	0.00	−0.02 – 0.02
Random Effects						
σ^2	0.01		0.01		0.01	
τ_{00}	0.01 $_{SUBJ}$		0.01 $_{SUBJ}$		0.01 $_{SUBJ}$	
	0.00 $_{ITEM}$		0.00 $_{ITEM}$		0.00 $_{ITEM}$	
τ_{11}	0.00 $_{SUBJ.CONDDeclarative\ Indefinite\ (B)}$		0.00 $_{SUBJ.CONDDeclarative\ Indefinite\ (B)}$		0.00 $_{SUBJ.CONDDeclarative\ Indefinite\ (B)}$	
	0.00 $_{SUBJ.CONDDeclarative\ Proper\ Name\ (C)}$		0.00 $_{SUBJ.CONDDeclarative\ Proper\ Name\ (C)}$		0.00 $_{SUBJ.CONDDeclarative\ Proper\ Name\ (C)}$	
ρ_{01}	−0.94 $_{SUBJ.CONDDeclarative\ Indefinite\ (B)}$		−0.24 $_{SUBJ.CONDDeclarative\ Indefinite\ (B)}$		−0.07 $_{SUBJ.CONDDeclarative\ Indefinite\ (B)}$	
	−0.12 $_{SUBJ.CONDDeclarative\ Proper\ Name\ (C)}$		0.17 $_{SUBJ.CONDDeclarative\ Proper\ Name\ (C)}$		0.07 $_{SUBJ.CONDDeclarative\ Proper\ Name\ (C)}$	
ICC			0.37		0.40	
N	54 $_{SUBJ}$		54 $_{SUBJ}$		54 $_{SUBJ}$	
	24 $_{ITEM}$		24 $_{ITEM}$		23 $_{ITEM}$	
Observations	1293		1292		1240	
Marginal R^2/Conditional R^2	0.007/NA		0.023/0.384		0.004/0.400	
AIC	−1599.368		−1582.015		−1446.788	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

The results, therefore, show that in-situ questions with a complex *wh*-phrase in Mandarin Chinese are read significantly slower than their indefinite declarative counterparts already at the *wh*-determiner position with the slowdown of questions with respect to indefinite declaratives carrying over to the *wh*-phrase/NP boundary, the noun (i.e., *tóngxué* "classmate" in (11)).

Qualitative comparison of results on processing Mandarin Chinese in-situ wh-questions

Mandarin in-situ *wh*-questions, both simplex (*shéi* "who") and complex (*nǎgè tóngxué* "which classmate"), exhibit slower processing than declaratives. However, the timing of these processing costs differs between *wh*-phrase types. Post hoc analyses revealed significantly faster reading times for declaratives with indefinite noun phrases (i.e., *rén* "person" and *yígè tóngxué* "a classmate") compared to those with definite noun phrases (i.e., *Xiǎozhāng* and *nàgè tóngxué* "the/that classmate") when contrasted with *wh*-questions.

As outlined in the initial predictions in the introduction section, question clause-type interpretation is triggered upon encountering a *wh*-phrase, not prior. This is mainly observed in Mandarin Chinese for complex *wh*-questions. The evidence is not so clear for simplex *wh*-questions, where we observe no significant effect at the *wh*-word. Instead, we observe a delayed effect one word later.

When the definiteness/indefiniteness of declaratives is considered, the slowdown is mainly observed between in-situ *wh*-questions and declaratives containing indefinite noun phrases. Declaratives containing definite noun phrases exhibit a distinct processing pattern.

Qualitative comparison of results of French and Mandarin Chinese

Both French and Mandarin Chinese show similar patterns where in-situ *wh*-questions are processed slower than indefinite declarative counterparts. The additional processing effort on the *wh*-in-situ sentences confirms the hypothesis that, in the absence of overt cues, the parser commits to a particular interpretation of the sentence (i.e., declarative) and only considers the interrogative interpretation when there is overt evidence for it (i.e., the *wh*-word). As outlined in the introduction section, the slowdown in reading time observed could be attributed to a syntactic integration process when an alternative structure needs to be either reactivated in *parallel processing accounts* (Jackendoff & Audring, 2020; Huettig et al., 2022), retrieved in *activation-based retrieval* accounts (Van Dijk & Lewis, 2003; Lewis & Vasishth, 2005), or re-analyzed as in "classic" processing accounts reflecting the reanalysis (Fodor & Ferreira, 1998) of the clause type of the sentence from a declarative to an interrogative interpretation, with the extra process of integrating the scope position (Spec CP) to associate it with the *wh*-word (see the work by Xiang et al., 2015 and Lo & Brennan, 2021 addressing specifically this process).

The present findings, demonstrating processing costs for in-situ *wh*-questions in both French and Mandarin Chinese, suggest that the availability of different *wh*-movement strategies is not a primary determinant of in-situ processing difficulty, as optional *wh*-in-situ languages like French also show processing

https://doi.org/10.1017/S0142716425000074 Published online by Cambridge University Press

difficulties. A direct comparison of the relative size of the effects is not possible cross-linguistically however, so to investigate if a modulation of the effect strength is present due to the language *wh*-movement strategy might not be possible using behavioral methods.

Although the processing pattern is common in both languages, French and Mandarin Chinese differ in the timing of the processing of in-situ *wh*-questions, depending on the nature of the *wh*-phrase. For simplex *wh*-questions, while both French and Mandaring speakers exhibit slower reading times relative to indefinite declaratives, the onset of this processing difficulty differs: it emerges at the *wh*-phrase position in French but one word later in Mandarin. As discussed earlier, it is not clear why at the *wh*-phrase position Mandarin Chinese readers show a facilitation effect of interrogatives relative to definite declaratives and answering this would require further research. For complex *wh*-phrases, Mandarin speakers show slowdown effects as soon as the *wh*-determiner is encountered whereas French speakers do not show these effects until the noun within the *wh*-phrase has been processed. Complex *wh*-phrases in French exhibit delayed processing relative to simplex *wh*-phrases, with effects emerging in the first post-*wh*-phrase region. The explanation we provide above for this delay is related to the possibility of having postnominal modification in French, which is not available in Mandarin Chinese. Further, the nature of the *wh*-determiner "which" (i.e., *quel(le)* vs. *nǎgè*) in each language may allow for more ambiguity in French than in Mandarin Chinese, where differences are observed at the onset of the *wh*-phrase.

Finally, the most pronounced difference between *wh*-in-situ questions and declaratives emerged in conditions involving indefinite noun phrases. Declaratives containing definite elements, such as proper names or definite noun phrases, appear to incur a processing load similar to that of *wh*-questions in both languages. Consistent with previous research (Warren & Gibson, 2002; Yen, 2007), proper names and definite noun phrases, lacking antecedent context, are generally more demanding to integrate than other referential noun phrases. This integration difficulty likely contributes to the observed processing costs for declaratives containing these elements, aligning them with the processing challenges of *wh*-questions.

One limitation of the presented work is that we did not have a within-subject configuration for the experiments conducted in each of the languages, therefore we could not compare statistically the simplex and complex in-situ *wh*-questions. This was done to avoid the discussion often observed in the literature of how valid it is to compare a simplex and a complex *wh*-phrase directly when the number of words read is different (e.g., one for "who" vs two for "which woman"). We discussed how differences in syllable length of our critical words for the four experiments complicated the comparison, therefore, including one more difference would have only complicated the interpretation of the results further.

Conclusion

In this study, we examined the word-by-word reading of in-situ questions that contained both simplex and complex *wh*-phrases in French and Mandarin and

https://doi.org/10.1017/S0142716425000074 Published online by Cambridge University Press

compared them with declaratives that contained definite/indefinite noun phrases. Our results showed that the parser assumes a declarative interpretation when reading these sentences incrementally, independently of the question formation strategy/strategies that the language has. Nevertheless, the nature of the *wh*-phrase (simplex or complex) and the declaratives' definiteness must be considered in the interpretation of the processing difficulty of *wh-in-situ* in the absence of contextual or prosodic information, as several cognitive processes are acting concurrently, to avoid incorrect conclusions on the processes observed.

Replication package. The supplementary materials including stimuli, data, and analysis code are available in https://osf.io/2cwqn/.

Acknowledgments. This work was supported by The Netherlands Organisation for Scientific Research (NWO) [project number 360-70-480] to the first author and by the Chinese Scholarship Council to the second author. We would like to thank Hamida Demirdache, Xiaolu Yang and Xose Lorenzo for providing us with the venues to test participants in France, China and The Netherlands and our research assistants Sylvie Cuchet, Lucas Tual and Juliette Angot for the help provided in conducting the French studies in France and in constructing the French stimuli. Finally, we would like to thank the editor and two anonymous reviewers for their insightful comments during the revision process.

Author contribution statement. Leticia Pablos Robles: Conceptualization, Methodology, Data Acquisition, Formal analysis, Data curation, Writing—original draft. **Yang Yang:** Conceptualization, Methodology, Data Acquisition, Formal analysis, Writing—review & editing. **Jenny Doetjes:** Conceptualization, Writing—review & editing, Funding acquisition. **Lisa Cheng:** Conceptualization, Writing—review & editing, Funding acquisition.

Notes

1 The position where the question/*wh*-word comes from is indicated by *t*, with an index associated with the *wh*-word.

2 It should be noted that English allows *wh*-in-situ in echo-questions (not regular information seeking questions) and multiple *wh*-questions (e.g., *Who bought what?*). Further, in some special cases (see, Pires & Taylor, 2007), given an extra-linguistic context, it is possible to invoke an in-situ option in English: if speaker A says: "I made desserts", speaker B can then say: "You made 'what kind of desserts?" Note that in such cases, there is also a special intonation involved. In this paper, we focus on single *wh*-questions without extra-linguistic contexts.

3 It should be noted that in embedded questions (i.e., indirect questions), *wh*-phrases in French must move (see an overview discussion in Glasbergen-Plas, 2021):

(i) a. *Je me demande que tu as vu qui.
I REFL *ask that you have seen who*
'I wonder that you have seen who'

b. Je me demande qui tu as vu.
I REFL *ask who you have seen*
'I wonder who have you seen'

4 Note that both (3a, b) are used in informal French only. In more formal registers, fronting is combined with subject-verb inversion or insertion of the question particle *est-ce que* (see, for instance, Boucher (2010) for an overview). Instead of (3b), one could thus have *Qui as-tu vu?* or *Qui est-ce que tu as vu?* The different strategies that are available in the case of fronting do not play a role in this paper.

5 Clause types refer to sentence types such as declaratives, interrogatives, imperatives, and exclamatives.

6 French speakers were all educated and grew up in France. They read and speak daily in French even though most of them spent more than ten years out of France.

https://doi.org/10.1017/S0142716425000074 Published online by Cambridge University Press

7 For a complete list of materials see: https://osf.io/2cwqn/
8 For a complete list of comprehension questions see: https://osf.io/2cwqn/
9 Spill-over effects are common in the processing literature on studies that examined different reading-time on different techniques such as eye-tracking or self-paced reading (see Just and Carpenter, 1978; or Kush et al., 2017, for a discussion on this effect). They are effects that emerge one or two words after the critical region and are considered a delayed response to the expected effect.
10 Frequency for proper names is, with few exceptions, not reflected in corpus databases. Nevertheless, we selected the proper names following the criteria that they should be common in daily French, by checking it with a French native speaker consultant.
11 For analysis script and further data processing details see: https://osf.io/2cwqn/
12 These speakers have the same language background as those in Experiment 1.
13 For the list of materials see: https://osf.io/2cwqn/
14 For analysis script and further data processing details see: https://osf.io/2cwqn/
15 Family names in China (e.g., Zhāng) tend to be monosyllabic. In addressing people, it is common to add adjectives such as *xiǎo/lǎo* 'young/old' to family name making the sequence disyllabic as in "Xiǎozhāng" in (10c).
16 When the verb carries the perfective marker *–le,* bare nouns such as *ren* 'person/someone' can have a definite or specific indefinite interpretation (see the discussion in Cheng and Sybesma 1999).
17 For the complete list of materials see: https://osf.io/2cwqn/
18 The critical *wh*-word region differed in having one to two characters. *Rén* 'someone' (人) and *shéi 'who'*(谁) consisted of one character, whereas all proper names consisted of two characters (e.g., *Xiǎozhāng* 小张).
19 Residual reading time as calculated as the residuals of a linear regression between the word/region length and reading time is not used in general for languages with logographic script. For languages that use logographic rather than alphabetic scripts, there is no clear way of calculating the length of the word/region unless the number of strokes per character or the character complexity are considered within the calculation. Log transformed reading time are normally distributed and adequate for usage in linear mixed models regression.
20 For analysis script and further data processing details see: https://osf.io/2cwqn/
21 For list of materials see: https://osf.io/2cwqn/
22 For analysis script and further data processing details see: https://osf.io/2cwqn/

References

Aaronson, D., & Scarborough, H. S. (1976). Performance theories for sentence coding: Some quantitative evidence. *Journal of Experimental Psychology: Human Perception and Performance*, **2**(1), 56–70.
Adli, A. (2004). Y at-il des morphèmes intonatifs impliqués dans la syntaxe interrogative du français? Le cas du qu-in-situ. In T. Meisenburg & M. Selig (Eds.), *Nouveaux Départs en Phonologie: Les Conceptions Sub- et Suprasegmentales* (p. 199215). Tübingen: Narr.
Adli, A. (2006). French wh-in-situ questions and syntactic optionality: Evidence from three data types. *Zeitschrift für Sprachwissentschaft*, **25**(2), 163–203.
Adli, A. (2015). What you like is not what you do: acceptability and frequency in syntactic variation. In A. Adli, M. García García, & G. Kaufmann (Eds.), *Variation in Language: System- and Usage-based Approaches* (pp. 173–200). Berlin, München, Boston: De Gruyter.
Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, **73**(3), 247–264.
Altmann, G. T., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, **33**(4), 583–609.
Aoun, J., & Li, A. Y.-H. (1993). Wh-elements in-situ: syntax or LF? *Linguistic Inquiry*, **24**(2), 199–238.
Baayen, R.H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modelling with crossed random effects for subjects and items. *Journal of Memory and Language*, **59**(4), 390–412.
Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language*, **68**(3), 255–278.
Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**(1), 1–48.
Baunaz, L., & Bonan, C. (2023). Activation levels: a fresh perspective on French *wh* in-situ. *Isogloss. Open Journal of Romance Linguistics*, **9**(1), 1–43.

https://doi.org/10.1017/S0142716425000074 Published online by Cambridge University Press

Bayer, J., & Cheng, L. L. S. (2017). *Wh*-in-situ. In M. Everaert, & H. van Riemsdijk (Eds.), *The Blackwell Companion to Syntax* (2nd ed., pp. 1–44). Hoboken, NJ: Wiley-Blackwell Publishing.

Beyssade, C., Delais-Roussarie, E., & Marandin, J.-M. (2007). The prosody of French interrogatives. *Noveaux Cahiers de Linguistique Française*, **28**, 163–175.

Boucher, P. (2010). Wh-questions in French and English. In E. Göbbel, & C. Breul (Eds.), *Comparative and Contrastive Studies of Information Structure* (pp. 101–138). Amsterdam: John Benjamins Publishing Company.

Bultena, S., Dijkstra, T., & van Hell, J. G. (2014). Cognate effects in sentence context depend on word class, L2 proficiency, and task. *Quarterly Journal of Experimental Psychology*, **67**(6), 1214–1241.

Cai, Q., & Brysbaert, M. (2010) SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PloS One*, **5**(6), e10729.

Camblin, C. C., Ledoux, K., Boudewyn, M., Gordon, P. C., & Swaab, T. Y. (2007). Processing new and repeated names: effects of coreference on repetition priming with speech and fast RSVP. *Brain Research*, **1146**, 172–184.

Cheng, L. L. S. (1991). *On the Typology of Wh-questions*. New York City: Garland Publishing Inc.

Cheng, L. L. S. (1994). *Wh*-words as polarity items. *Chinese Languages and Linguistics*, **2**, 615–650.

Cheng L. L. S. (2003). *Wh*-in-situ. *Glot International*, **7**(5), 129–137.

Cheng L. L. S. (2009). *Wh*-in-situ from the 1980s to now. *Language and Linguistics Compass*, **3**(3), 767–791.

Cheng, L. L. S., & Sybesma, R. (1999). Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry*, **30**(4), 509–542.

Clifton, C. Jr., Fanselow, G., & Frazier, L. (2006) Amnestying superiority violations: processing multiple questions. *Linguistic Inquiry*, **37**(1), 51–68.

Crain, S., & Fodor, J. D. (1985). How can grammars help parsers? In D. R. Dowty, L. Karttunen, & A. M. Zwicky (Eds.), *Natural Language Parsing: Psychological, Computational, and Theoretical Perspectives* (pp. 94–128). Cambridge: Cambridge University Press.

De Vincenzi, M. (1991). *Syntactic parsing strategies in Italian: The minimal chain principle* (Vol. **12**). Springer Science & Business Media.

De Vincenzi, M. (1996). Syntactic analysis in sentence comprehension: effects of dependency types and grammatical constraints. *Journal of Psycholinguistic Research*, **25**, 117–133.

Delattre, P. (1966). Les Dix intonations de base du français. *The French Review*, **40**(1), 1–14.

Déprez, V., Syrett, K., & Kawahara, S. (2013). The interaction of syntax, prosody, and discourse in licensing French wh-in-situ questions. *Lingua*, **124**, 4–19.

Donkers, J., Hoeks, J. C. J., & Stowe, L. A. (2011). D-linking or set-restriction? *Processing which-questions in Dutch. Language and Cognitive Processes*, **28**(1–2), 9–28.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, **25**(3), 348–368.

Fodor, J., & Ferreira, F. (1998). *Reanalysis in Sentence Processing (Vol. 21)*. Dordrecht: Springer Science & Business Media/ Kluwer.

Fodor, J. D. (2002). Psycholinguistics cannot escape prosody. *Proceedings of Speech Prosody*, **2002**, 83–90.

Frazier, L., & Clifton Jr., C. (2002). Processing "d-linked" phrases. *Journal of Psycholinguistic Research*, **31**, 633–660.

Giannakidou, A., & Cheng, L. L. S. (2006). (In)definiteness, polarity, and the role of wh-morphology in Free Choice. *Journal of Semantics*, **23**(2), 135–183.

Glasbergen-Plas, A. (2021). Questions in context: the case of French wh-in-situ. [Doctoral Dissertation, Leiden University]. LOT dissertation series. https://hdl.handle.net/1887/3239072.

Gordon, P. C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language*, **51**(1), 97–114.

Gryllia, S., Cheng, L. L.-S., & Doetjes, J. (2016) On the intonation of French wh-in-situ questions: What happens before the wh-word is reached? *Speech Prosody*, **2016**, 611–614.

Gryllia S., Doetjes J. S., Yang Y., & Cheng, L. L. S. (2020), Prosody, clause typing, and wh-in-situ: evidence from Mandarin, *Laboratory Phonology*, **11**(1), 19.

Gundel, J., Hedberg, H., & Zacharski, R. (1993). Referring expressions in discourse. *Language*, **69**(2), 274–307.

Hofmeister, P., Jaeger, T. F., Sag, I. A., Arnon, I., & Snider, N. (2007). Locality and accessibility in wh-questions. In S. Featherston, & W. Sternefeld (Eds.), *Roots: Linguistics in Search of its Evidential Base* (pp. 185–206). Berlin, New York: De Gruyter Mouton.

Hofmeister, P., & Sag, I. A. (2010). Cognitive constraints and island effects. *Language*, **86**(2), 366–415.
Huang, J. C.-T. (1982). *Logical Relations in Chinese and the Theory of Grammar*. New York: Taylor & Francis.
Huettig, F., Audring, J., & Jackendoff, R. (2022). A parallel architecture perspective on pre-activation and prediction in language processing. *Cognition*, **224**, 105050.
Huettig, F., & Mani, N. (2016). Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience*, **31**(1), 19–31.
Hyönä, J., & Olson, R. K. (1995). Eye fixation patterns among dyslexic and normal readers: effects of word length and word frequency. *Journal of Experimental Psychology. Learning, Memory and Cognition*, **21**(6), 1430–1440.
Jackendoff, R., & Audring, J. (2020). *The Texture of the Lexicon: Relational morphology and the Parallel Architecture*. Oxford: Oxford University Press.
Jegerski, J. (2014). Self-paced reading. In J. Jegerski, & B. VanPatten (Eds.), *Research Methods in Second Language Psycholinguistics* (pp. 20–49). New York, NY, USA: Routledge.
Just, M. A., & Carpenter, P. A. (1978). Inference processes during reading: Reflections from eye fixations. In J. W. Senders, D. F. Fisher, & R. A. Monty (Eds.), *Eye Movements and the Higher Psychological Functions* (pp. 157–174). London: Routledge.
Just, M.A., Carpenter, P.A., & Wooley, J.D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, **111**(2), 228–238.
Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language & Cognitive Processes*, **15**(2), 159–201.
Kaan, E., & Vasić, N. (2004). Cross-serial dependencies in Dutch: testing the influence of NP type on processing load. *Memory & Cognition*, **32**, 175–184.
Kawahara, S., Shaw, J. A., & Ishihara, S. (2022). Assessing the prosodic licensing of wh-in-situ in Japanese: a computational-experimental approach. *Natural Language & Linguistic Theory*, **40**, 103–122.
Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye-movements in reading. *European Journal of Cognitive Psychology*, **16** (1–2), 262–284.
Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, **31**(1), 32–59.
Kush, D., Lidz, J., & Phillips, C. (2017). Looking forwards and backwards: the real-time processing of strong and weak crossover. *Glossa: A Journal of General Linguistics*, **2**(1), 70.
Kwon, N., & Sturt, P. (2013). Null pronominal (pro) resolution in Korean, a discourse-oriented language. *Language and Cognitive Processes*, **28**(3), 377–387.
Ledoux, K., Gordon, P. C., Camblin, C. C., & Swaab, T. Y. (2007). Coreference and lexical repetition: mechanisms of discourse integration. *Memory & Cognition*, **35**(4), 801–815.
Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, **106**(3), 1126–1177.
Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, **29**, 1–45.
Li, L., Li, S., Xie, F., Chang, M., McGowan, V. A., Wang, J., & Paterson, K. B. (2019). Establishing a role for the visual complexity of linguistic stimuli in age-related reading difficulty: evidence from eye movements during Chinese reading. *Attention Perception & Psychophysics*, **81**, 2626–2634.
Linzen, T., & Jaeger, T. F. (2016). Uncertainty and expectation in sentence processing: evidence from subcategorization distributions. *Cognitive Science*, **40**(6), 1382–1411.
Lo, C. W., & Brennan, J. R. (2021) EEG correlates of long-distance dependency formation in Mandarin Wh-questions. *Frontiers in Human Neuroscience*, **15**, 591613.
Ma, J. K.-Y., Ciocca, V., & Whitehill, T. L. (2011). The perception of intonation questions and statements in Cantonese. *Journal Acoustical Society of America*, **129**(2), 1012–1023.
McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research*, **29**, 111–123.
McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, **48**(1), 67–91.
Nakagawa, S., Johnson, P., & Schielzeth, H. (2017) The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, **14**, 20170213.

New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: a new French lexical database. *Behavior Research Methods, Instruments, & Computers*, **36**(3), 516–524.

Oiry, M. (2011). A case of true optionality: Wh-in-situ patterns like long movement in French. *Linguistic Analysis*, **37**(1–2), 115–142.

Pablos, L. (2008). Filled-gap effects in sentence processing: different accounts compared. *Anuario del Seminario de Filología Vasca Julio de Urquijo*, **42**(1), 309–332.

Pickering, M. J., & Gambi, C. (2018), Predicting while comprehending language: a theory and review, *Psychological Bulletin*, **144**(10), 1002–1044.

Pires, A., & Taylor, H. (2007). The syntax of *wh*-in-situ and common ground. *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, **43**(2), 201–215.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rhode, D. (2003). *Linger* (Version 2.94) [Computer software]. MIT. http://tedlab.mit.edu/~dr/Linger/

Stowe, L. A. (1986). Parsing *wh*-constructions: evidence for on-line gap location. *Language and Cognitive Processes*, **1**(3), 227–245.

Tsai, W.-T. D. (1994). *On Economizing the Theory of A-Bar Dependencies*. London: Routledge.

Tual, L. (2017a). Long-distance *wh*-in-situ in French: an experimental study. *Paper Presented at Romance Interrogatives*, **17**.

Tual, L. (2017b). Prosodic and syntactic properties of French interrogative sentences. *Paper Presented at the 43rd Incontro Di Grammatica Generativa, Pavia*, **17**.

Ueno, M., & Kluender, R. (2009). On the processing of Japanese *wh*-questions: An ERP study. *Brain Research*, **1290**, 63–90.

Van Dyke, J., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: a cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, **49**(3), 285–316.

Vasishth, S. (2006). On the proper treatment of spillover in real-time reading studies: Consequences for psycholinguistic theories. In *Proceedings of the International Conference on Linguistic Evidence*, Tübingen, Germany, 96–100.

Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition*, **85**(1), 79–112.

Warren, T., & Gibson, E. (2005). Effects of NP type in reading cleft sentences in English. *Language and Cognitive Processes*, **20**(6), 751–767.

Witzel J, & Witzel, N. (2016). Incremental sentence processing in Japanese: a maze investigation into scrambled and control sentences. *Journal of Psycholinguistic Research*, **45**(3), 475–505.

Wu, F., Kaiser, E., & Andersen, E. (2012). Animacy effects in Chinese relative clause processing. *Language and Cognitive Processes*, **27**(10), 1489–1524.

Wunderli, P. (1983). L'intonation des interrogatives du type, 'Il est né en quelle année? *Romanica Gandensia*, **20**, 169–181.

Wunderli, P. (1984). L'intonation des questions sans marquée segmentale. *Travaux de Linguistique et de Littérature*, **22**(1), 203–250.

Xiang, M., Dillon, B., Wagers, M., Liu, F. Q., & Guo, T. M. (2013) Processing covert dependencies: an SAT study on Mandarin wh-in-situ questions. *Journal of East Asian Linguistics*, **23**(2), 207–232.

Xiang, M., Wang, S. P., & Cui, Y. L. (2015) Constructing covert dependencies—the case of wh-in-situ processing. *Journal of Memory and Language*, **84**, 139–166.

Yang, Y., Gryllia, S., Pablos, L., & Cheng, L. (2019). Clause type anticipation based on prosody in Mandarin. *International Journal of Chinese Linguistics*, **6**(1), 1–26.

Yen, H.-L. (2007). Processing of Proper Names in Mandarin Chinese: A Behavioral and Neuroimaging Study. [Doctoral Dissertation, University of Bielefeld]. https://core.ac.uk/download/pdf/15958155.pdf.

Zang, C., Fu, Y., Bai, X., Yan, G., & Liversedge, S. P. (2018). Investigating word length effects in Chinese reading. *Journal of Experimental Psychology. Human Perception and Performance*, **44**(12), 1831–1841.

Cite this article: Pablos Robles, L., Yang, Y., Doetjes, J., & Cheng, L. L. S. (2025). Do readers anticipate *wh*-in-situ questions? Cross-linguistic reading time evidence from Mandarin Chinese and French. *Applied Psycholinguistics*. https://doi.org/10.1017/S0142716425000074