Developmental changes in misinterpretation of garden-path wh-questions in French

Romy Lassotta, Akira Omaki & Julie Franck

Submit your article to this journal

Article views: 130

View related articles

View Crossmark data

Citing articles: 2 View citing articles

Developmental changes in misinterpretation of garden-path wh-questions in French

Romy Lassotta[1], Akira Omaki[2], and Julie Franck[3]

[1]Department Linguistik, Universität Potsdam, Potsdam, Germany
[2]Department of Cognitive Science, Johns Hopkins University, Baltimore, MD, USA
[3]Laboratoire de Psycholinguistique, Université de Genève, Geneva, Switzerland

This study explores (mis)interpretation of biclausal wh-questions by French-speaking adults and children, aiming to investigate cross-linguistic differences in sentence revision mechanisms. Following previous work in Japanese the ambiguity of wh-questions was manipulated: In ambiguous questions, the fronted wh-phrase could be associated with the first, main-clause verb or the second, embedded-clause verb, while in garden-path questions, an inserted filled-gap prepositional phrase (PP) blocked main-clause attachment. Importantly, French differs from Japanese in that the filled gap arises after the first verb—that is, after the wh-phrase has been interpreted within the main clause. Two story-based comprehension experiments were conducted to probe the effect of word order on revision performance. Adults and children frequently provided main-clause interpretations of ambiguous questions. In filled-gap questions, children displayed relatively acute sensitivity to the filled-gap in wh-argument questions (Experiment 2), but not in wh-adjunct questions (Experiment 1); adults showed surprisingly low sensitivity to it, frequently misinterpreting adjunct and argument questions. Acceptability ratings (Experiment 3) showed that adults systematically prefer in situ questions over wh-fronting questions. We conclude that timing of the error signal influences revision, and that whereas French-speaking children prioritize syntactic cues, adults prioritize distributional information about the optionality of wh-fronting in French.

Keywords: Wh-questions; Garden-path; Misinterpretation; Filled-gap dependency; Sentence revision.

Decades of psycholinguistic research have generated much evidence that the parser incrementally assigns accurate syntactic and semantic representations to the input in according with linguistic constraints (e.g., Marslen-Wilson, 1975). Nevertheless, it has also been reported that adult native speakers often misinterpret temporarily ambiguous structures, or leave ambiguities unresolved. For example,

Christianson, Hollingworth, Halliwell, and Ferreira (2001) reported evidence for misinterpretations of a sentence–question pair as shown in (1). In the garden-path sentence in (1a), *the baby that was cute and cuddly* could be temporarily analysed as the argument of the verb *dressed*, but this analysis is incompatible with the subsequent input *played* that requires a subject.

1. a. While Anna dressed the baby that was cute
      and cuddly played in the crib.
   b. Did Anna dress the baby?

Christianson et al. (2001) found that adult native speakers of English tended to answer "yes", even though the only grammatical interpretation is that Anna dressed herself and not the baby. Such misinterpretations may reflect an assignment of ill-formed structural representations that are syntactically "good enough" for the purpose of comprehension (Christianson et al., 2001; Ferreira, 2003; Ferreira & Patson, 2007; Tabor, Galantucci, & Richardson, 2004), or an incomplete revision of the initial syntactic analysis or semantic interpretation (Slattery, Sturt, Christianson, Yoshida, & Ferreira, 2013; Staub, 2007; Sturt, 2007; van Gompel, Pickering, Pearson, & Jacob, 2006).

Evidence for incomplete sentence revision has been reported in children's sentence comprehension as well, although these misinterpretations typically occur in structural environments where adults are able to successfully revise and inhibit their initial analyses (for reviews, see Omaki & Lidz, 2015; Snedeker, 2013; Trueswell & Gleitman, 2007). For example, Trueswell, Sekerina, Hill, and Logrip (1999) conducted a visual-world eye-tracking study to examine 5-year-old children's interpretation of garden-path sentences like (2). Here, the prepositional phrase (PP) *on the napkin* is temporarily ambiguous, as it could be a modifier of the noun phrase (NP) *the frog*, or the PP argument of the verb *put* that specifies the destination.

2. Put the frog on the napkin in the box.

When the act-out instruction in (2) was presented with an array of objects (e.g., a frog on a napkin, another empty napkin, an empty box, and a towel), children's initial eye movements showed adult-like patterns indicating that they immediately analysed *on the napkin* as the destination (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). However, subsequent eye movements revealed that children did not shift their gaze to the ultimate destination (the box) as much as adults did, which suggests that children did not revise their initial interpretation. Moreover, in the act-out performance

that reflects their ultimate interpretation of the sentence, children's actions often diverged from those of adults: They typically moved the frog to the empty napkin, even though the only grammatical interpretation of (2) given the array of objects is to move the frog to the empty box. Since the adult control group did not make such errors, these findings suggest that sentence revision abilities are still immature at age 5. Similar sentence revision failures in PP attachment ambiguity have been replicated in experiments that reduced the cognitive and linguistic demands in a variety of ways (Hurewitz, Brown-Schmidt, Thorpe, Gleitman, & Trueswell, 2000; Weighall, 2008; cf. Meroni & Crain, 2003), and similar findings have been reported in verb-final languages (Choi & Trueswell, 2010).

The parallel observations that adults and children often retain the initial interpretation of garden-path sentences suggest that developmental comparisons can shed light on the sentence revision mechanism in at least two ways. First, dramatic sentence revision difficulties in children can provide a useful window into mechanisms of sentence revision and potentially reveal the details of sentence revision processes that are not readily observable in adults with efficient comprehension mechanisms. Second, developmental changes in sentence revision performance can provide insights on cognitive and linguistic factors that constitute sentence revision mechanisms (e.g., Mazuka, Jincho, & Oishi, 2009; Novick, Hussey, Teubner-Rhodes, Harbison, & Bunting, 2013; Novick, Thompson-Schill, & Trueswell, 2005; Novick, Trueswell, Thompson-Schill, 2010). While studies that use developmental comparisons are starting to shed light on the nature of sentence revision mechanisms, at present the majority of empirical evidence for children's immature sentence revision comes from PP attachment ambiguity in English like (2) (for a related work on Mandarin passive sentences, see Huang, Zheng, Meng, & Snedeker, 2013). The reliance on a narrow range of syntactic environments and languages is potentially problematic, because the severe garden-path effects found in (2) may be specific to the particular linguistic property of the sentence, the language, or the task set-up. One of the main goals of this paper is thus to

extend developmental comparisons of sentence revision behaviours to a larger variety of structural environments and languages, and to investigate whether syntactic differences across languages cause differences in sentence revision behaviours.

The present study explores developmental changes in the misinterpretation of garden-path sentences with filler-gap dependencies in French, which shares the critical grammatical properties (e.g., wh-fronting and basic word order) with English. Filler-gap dependencies provide a useful testing ground for cross-linguistic developmental comparisons of incrementality and sentence revision. Research on adult sentence processing has reported much evidence that the parser associates a wh-phrase or a dislocated constituent (called *filler*) with the first available thematic position (called *gap*) in the sentence. For example, Stowe (1986) presented filler-gap dependency sentences in English like *My brother wanted to know **who** Ruth will bring us home to _____ at Christmas,* and observed that the reading time at the direct object *us* was greater than in a control condition in which the filler was substituted for an embedded *if*-clause without such a filler-gap dependency. This so-called *filled-gap effect* indicates that the parser had already posited the object gap before checking whether the object position was occupied. Evidence for active gap filling has been reported in a variety of experimental paradigms and across languages with different word orders, suggesting that this is a robust property of the adult parser (e.g., Aoshima, Phillips, & Weinberg, 2004; Crain & Fodor, 1985; Frazier & Flores D'Arcais, 1989; Garnsey, Tanenhaus, & Chapman, 1989; Omaki & Schulz, 2011; Sussman & Sedivy, 2003; Traxler & Pickering, 1996; Wagers & Phillips, 2014; for review, see Phillips & Wagers, 2007). Although none of these previous studies investigated filler-gap dependency processing in French, we predict that French-speakers should show an active gap-filling bias like English speakers when the wh-phrase is fronted. However, we also note that French differs from the other languages

that have been studied so far, in that French wh-phrases can optionally stay in situ in their original thematic position. This means that the relative frequency of wh-fronting in French is lower than that in English and other languages with obligatory wh-fronting, which may have consequences on the strength of the incrementality principle in completing filler-gap dependencies, and/or on the ease of revising completed filler-gap dependencies. As such, the present study on French wh-questions can shed light on whether mechanisms of filler-gap dependency processing are influenced by differences in grammatical and distributional properties of wh-question formation.

With respect to children's processing of filler-gap dependencies, Omaki, Davidson White, Goro, Lidz, and Phillips (2014) used a question-after-story design (De Villiers, Roeper, & Vainikka, 1990) in English and Japanese and reported evidence for active gap filling and filled-gap effects in 5-year-old children's ultimate interpretation of adjunct wh-questions (for cross-modal priming evidence for active gap filling, see Love, 2007; Roberts, Marinis, Felser, & Clahsen, 2007). This study explored offline interpretation preferences in English (3a) and Japanese wh-questions (3b).

3. a. Where did Lizzie tell someone that she was gonna catch butterflies?
   b. Doko-de Yukiko-chan-wa choucho-o tsuka-maeru-to itteta-no?
      where-at Yukiko-Dim-Top pro butterfly-Acc catch-Comp was telling-Q[1]
      "Where was Yukiko telling someone that she will catch butterflies?"
   c. Doko-de Yukiko-chan-wa **kouen-de** choucho-o tsukamaeru-to itteta-no?
      where-at Yukiko-Dim-Top park-at pro butterfly-Acc catch- Comp was telling-Q
      "Where was Yukiko telling someone that she will catch butterflies in the park?"

The biclausal wh-questions in (3a) and (3b) contain a global ambiguity, as the fronted wh-phrase can be associated with either the

---

[1]The gloss abbreviations are as follows: Acc = accusative case marker, Comp = complementizer, Dim = diminutive marker, Gen = genitive case marker, Q = question particle, Top = topic marker.

main-clause verb (*tell someone*) or the embedded-clause verb (*catch butterflies*). A series of question-after-story tasks with adults and children found that English-speaking adults and 5-year-olds systematically preferred the main-clause association (i.e., answering the location for the telling event) in sentences like (3a) (cf. De Villiers, Roeper, Bland-Stewart, & Pearson, 2008; De Villiers et al., 1990), whereas Japanese-speaking adults and 5-year-olds showed a systematic preference for the embedded-clause association in (3b) (i.e., answering the location for the butterfly-catching event). This contrasting behaviour provides evidence for first verb association preference: The main clause is completed first in English, while the embedded-clause verb is completed first in Japanese, due to its verb-final word order. This first verb association preference is compatible with active gap filling, because the first verb in the sentence is the earliest position for completing the filler-gap dependency.

Moreover, a stronger argument for active gap filling comes from the filled-gap condition in Japanese, which added a filled-gap PP that specified the location of the embedded-clause event (3c). Here, adults systematically adopted the main clause (i.e., second verb) interpretation, because the overt locative PP modifier already specifies the location of the embedded-clause event and forces the main-clause association. In other words, the filled-gap served as an error signal—that is, a revision cue for the initial embedded-clause attachment of the wh-phrase. However, 5-year-olds provided the embedded-clause (i.e., first verb) interpretation as often as they did in the ambiguous condition, suggesting that they struggled to inhibit the strong bias to associate the fronted wh-phrase with the embedded-clause verb, despite the presence of the overt PP that blocks an embedded-clause interpretation. This finding resembles the misinterpretation pattern observed in the PP attachment study (e.g., *Put the frog on the napkin in the box*) by Trueswell et al. (1999). Here, the incremental destination analysis of the ambiguous PP was disconfirmed by subsequent input, but children nevertheless preserved the initial, incremental analysis in their ultimate interpretation. In other words, the child's

initial interpretation perseverates very robustly (for a review of relevant findings, see Omaki & Lidz, 2015).

While the adults in the Japanese study showed no sign of incomplete sentence revision, this may be partly due to the relative position of the filled-gap PP and the verb: In a head-final language like Japanese, the filled-gap PP arrives before the verb and could thus be used to prevent an assignment of interpretation to the filler-gap dependency. It is important to note that evidence for incomplete sentence revision or good-enough representations in adults has come from research on English garden-path sentences in which the error signal arrives after the initial syntactic analysis, and interpretation has been fully assigned [see (1)]. In this sense, if we present translations of filled-gap sentences like (3c) in a head-initial language like English or French, where the filled gap arrives after the verb, it is plausible that even adults may show signs of incomplete sentence revision. The interest of using French rather than English is to explore a particularity of this head-initial language allowing both fronted and in situ wh-phrase. A cross-linguistic approach can thus enrich evidence about head-initial languages in general and their specific syntactic properties in particular.

In summary, head position critically determines the relative position of the error signal (i.e., the filled gap) and the first verb in the sentence: The filled-gap precedes the first verb in head-final languages like Japanese, whereas it follows it in head-initial languages like English or French. The filled gap in Japanese may thus serve as a *preventive* signal, but it can only serve as a *corrective* signal in English or French, after active gap filling has taken place, and the filler-gap dependency has been assigned an interpretation. The present study uses three experiments to explore the effect of postverbal filled gaps in French garden-path wh-questions and tests whether there are developmental changes in (mis)interpretation of filler-gap dependencies. If corrective error signals are weaker cues than preventive signals, French-speaking children and possibly even adults are expected to show difficulties in parsing garden-path wh-questions. Experiment 1 aims to

replicate the Japanese finding in French, using adjunct wh-questions (i.e., *where*) and the same methodology as that used by Omaki et al. (2014). In order to test whether revision difficulties generalize to another type of filler-gap dependency, Experiment 2 explores how French-speaking children and adults interpret garden-path questions with argument wh-phrases (i.e., *to whom*), which may syntactically facilitate the revision of initial filler-gap dependency interpretation. Experiment 3 uses an acceptability judgement task to ensure that the garden-path sentences used in Experiments 1 and 2 are acceptable sentences of French. The study demonstrates severe garden-path effects in both French-speaking children and adults, unlike in the previous Japanese study, and provides novel evidence that the success of sentence revision mechanisms in children is influenced by the grammatical properties of the language.

EXPERIMENT 1: COMPREHENSION OF ADJUNCT WH-QUESTIONS

Experiment 1 adopts the same experimental design as that of Omaki et al. (2014) to investigate the comprehension of ambiguous and filled-gap wh-questions like (4) in French-speaking adults and children.

4. a. Où est-ce qu'Aline a expliqué qu'elle allait attraper des papillons?
   where Q Aline has explained that she went catch some butterflies
   "Where did Aline explain that she was going to catch butterflies?"

b. Où est-ce qu'Aline a expliqué **dans le salon** qu'elle allai attraper des papillons?
   where Q Aline has explained in the room that she was going to catch some butterflies
   "Where did Aline explain in the living room that she was going to catch butterflies?"

In the globally ambiguous condition (4a), the wh-phrase can be associated with either the main-clause verb (*expliquer* "explain") or the embedded-clause verb (*attraper* "catch"). On the other hand, in garden-path sentences like (4b),

the filled-gap PP (*dans le salon* "in the living room") should block the association of the wh-phrase with the main verb, and the embedded-verb interpretation should be adopted. In order to explore how effectively the filled-gap PP changes the interpretation bias, we explore how much the rate of embedded-clause interpretation increases from the ambiguous to the filled-gap condition.

If French-speaking adults and children have a bias to incrementally complete filler-gap dependencies, as observed in English and Japanese, they should demonstrate a main-clause attachment preference in the globally ambiguous condition (4a). In the filled-gap condition, however, interpretation patterns may diverge between adults and children. Based on the Japanese findings and on previous reports of children's general struggle with the inhibition of initial analyses, French-speaking children are expected to frequently fail to inhibit the main-clause interpretation. Japanese adults were shown to be at ceiling in revising their initial parse, but if the timing of the error signal arrival influences the ease with which the mature adult parser revises its initial analysis, French-speaking adults may also demonstrate some difficulty in inhibiting the main-clause (first verb) interpretation bias.

Given the goal of Experiment 1 to identify potential cross-linguistic differences between French and Japanese, the same adjunct wh-questions as those of Omaki et al. (2014) were adopted here. However, French offers an alternative parse for these filled-gap constructions that is absent in Japanese. In French, the adjunct wh-phrase can potentially be associated with the main-clause verb by embedding it within the filled-gap PP itself (e.g., *where in the living room?*), which results in a question about a sublocation inside the larger location. In the Japanese example in (3c), this sublocation question strategy is syntactically unavailable (Omaki et al., 2014). Nevertheless, the sublocation interpretation was made infelicitous in our experiment by the use of story contexts that failed to provide any information about a sublocation (see the Materials section). Thus, although the lack of a context licensing the attachment of the wh-phrase to the main clause was expected to prevent the parser from adopting

that interpretation, one cannot exclude that possibility. We return to this point in the Discussion.

Method

Participants

Thirty French-speaking children (12 female) aged 5;1–8;2 years (mean age 6;10 years) took part in this experiment. Participants were recruited at two elementary schools in Geneva. Based on the bilingual exposure questionnaire developed by Unsworth (2013), which assessed the amount of exposure to two (or more) languages in a child's environment, we found that 17 (57%) of the children were bilingual, although in all these bilingual children French was their first and dominant language. Data from eight additional children were excluded from the analyses, as they were too distracted to concentrate on the experiment.

Furthermore, 30 French-speaking adults (25 female) aged 20–39 years (mean age 25 years) also participated in this experiment. They were recruited from the student community of the University of Geneva and received course credit for their participation. Eight (27%) of them were bilingual according to self-reports, but French was the dominant native language that was primarily used at the university and in their daily communication.

Materials

The experiment used a question-after-story design (De Villiers et al., 1990) in which a child-friendly story is presented first, followed by a question about the content of the story. Each story is accompanied by a cartoon that consists of three scenes: a main-clause scene, an embedded-clause scene, and a distractor scene (see *Story Design* below). All of the stories and questions were prerecorded and were combined with the visual stimuli in an animated movie format.

Question design. A total of eight test questions was constructed, each of them appearing in the two conditions (ambiguous vs. filled-gap) shown in (4). These sentences all consisted of biclausal wh-questions. The ambiguous condition (4a) was designed in such a way that the wh-phrase could be associated with either the main-clause verb or the embedded-clause verb. The filled-gap condition, on the other hand, included a PP that specified the location of the main-clause event, such that the embedded-clause association would be enforced. The main-clause verb was either *expliquer* "explain" in half of the target items, and the other half was *raconter* "tell". An additional set of 10 filler questions was built to introduce some variability in the experimental structures. They consisted of four wh-island questions (e.g., *Où est-ce qu'Aline a expliqué où elle allait attraper des papillons?* "Where did Aline explain where she was going to catch butterflies?") and six simple wh-in situ questions (e.g., *Aline a parlé à ses parents où?* "Aline talked to her parents where?"). Each of the 18 questions (8 test and 10 fillers) was followed by a second question about the distractor location to ensure that participants were paying attention to the task. The second question was asked immediately after the participant's response to the first question.

The audio stimuli were prerecorded by a female native speaker of French and were carefully controlled to create a natural prosody that is compatible with either interpretation. The ambiguous condition contained a short pause of 39 ms on average before the embedded clause, while this pause was slightly longer in the filled-gap condition (86 ms on average) in order to promote the attachment of the filled-gap PP to the main verb. Also, the filled-gap condition contained hardly any pause between the main verb and its following PP to avoid a possible prosodic bias for positing a gap immediately after the main-clause verb. In both conditions, we implemented further prosodic adjustments such that the embedded-clause attachment of the wh-element was supported by prosody: The pitch of the main verb (*raconter* or *expliquer*) rose only 7 Hz on average from its second to its last syllable (+8 Hz in ambiguous; +5 Hz in filled gap); the syllable length of the main verb increased from 59 to 73 ms per phoneme only from the second to the last syllable in the ambiguous condition, and from 55–67 ms per phoneme in the filled-gap condition. Consequently, the prosody in our stimuli did not create a bias

towards an attachment of the wh-phrase to the main clause since there was no prosodic boundary after the main verb in terms of pitch rise or lengthening of the word-final syllable.

Materials were part of a block design such that ambiguous questions were presented in the first half of the experiment followed by filled-gap questions. This order was important because the ambiguous condition was designed to elicit the default attachment preference in ambiguous biclausal wh-questions, and we wanted to make sure that exposure to the unambiguous filled-gap questions did not interfere with the ambiguous question interpretation. Filler sentences were randomly spread across the whole session.

Story design. Stories were presented as short cartoons (Figure 1), in which a main character visited three different locations. Each location was associated with the main-clause event, the embedded-clause event, or a distractor event. In the story that preceded the question in (4), for example, Aline was first in her bedroom (distractor location; top right in Figure 1), thinking of what to do on a nice sunny day. She decided to go outside, so she went to the living room (bottom of Figure 1) and said to her sister (main-clause event) that she was going to catch butterflies in the park. Finally, Aline visits the park and catches butterflies (embedded-clause event). The events in each location were animated in order to make it easier for participants to recognize what happened. The first location was always a distractor event, and the order of main-clause and embedded-clause events was balanced across items. It is important to note that the story did not mention the presence of a sublocation in any of the three locations. This manipulation was critical for providing contextual cues that block the main-clause interpretation and favour embedded-verb attachment of the wh-element.

The materials were split in two experimental lists such that each participant only saw one version of the eight test questions (ambiguous or filled gap) and the stories that illustrated them. A list contained a total of 18 stories paired with a

Figure 1. *A sample scene presented during a story stimulus in Experiment 1. To view this figure in colour, please visit the online version of this Journal.*

question: eight stories paired with a test question (four ambiguous followed by four filled gap) and 10 paired with a filler question.

Procedure

Participants were instructed to carefully listen to the story and watch the animation. Children were asked to provide their response to the question by touching the corresponding picture on a touchscreen monitor, such that no verbal response was required, and the task demands were lowered. Adults provided their responses by answering the questions out loud. Prior to the testing session, each participant completed one practice trial in order to become familiar with the materials and with the testing procedure. A session lasted approximately 20 minutes.

Data coding and analysis

Offline responses were coded as (a) main-verb responses, (b) embedded-verb responses, (c) sublocation responses, or (d) other responses. Sublocation responses consisted of answers about a specific location inside a larger location described by the overt PP. For example, if the PP was *dans le salon* "in the living room", a potential sublocation answer would be "by the couch in the living room". Sublocation responses were only observed in adults and consisted in creative descriptions of sublocations based on objects displayed in the pictures, even though the story narrative never explicitly mentioned such sublocations. Other responses in children included pointing at the distractor location (5%), and similar errors were found in adults as well (4%).

Given that other responses represented a very small percentage of the overall observations, these responses were removed from the analyses. Performance to filler items and second questions was above 95%, showing that participants were attentive to the task.

Offline responses were analysed by constructing a generalized linear mixed-effects model with question type (ambiguous vs. filled gap) and age group (children vs. adults) as fixed effects and random intercepts for subjects and items (see Baayen, Davidson, & Bates, 2008; Jaeger, 2008), as the fully crossed model with random slopes (Barr, Levy, Scheepers, & Tily, 2013) failed to converge. All statistical analyses reported in this paper were conducted in the R statistical environment (R Development Core Team, 2013). All variables were dichotomous, so sum contrasts were used with one level of the factor coded as −0.5 and the other as 0.5. Estimates (β), standard errors (SE), and z- and p-values of the model were obtained via the glmer function of the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2014).

Results

In the ambiguous condition, children and adults both show a clear main-verb preference, as illustrated in Table 1. For children, only 15% of the responses showed the embedded-verb interpretation, while they represent 12% of adults' responses. In the filled-gap condition, on the other hand, the rate of embedded-verb responses was modulated to different degrees in adults and

Table 1. *Distribution of the offline responses in Experiment 1*

	Ambiguous questions			Filled-gap questions		
	Main-clause interpretations		Embedded-clause interpretations	Main-clause interpretations		Embedded-clause interpretations
Age group	Main verb	Sublocation	Embedded verb	Main verb	Sublocation	Embedded verb
Children	85 (3)	NA	15 (3)	88 (3)	NA	12 (3)
Adults	88 (3)	NA	12 (3)	49 (5)	18 (4)	33 (5)

Note: Offline responses in percentages. Standard errors in parentheses.

children. Children provided embedded-verb responses 12% of the time, similarly to the ambiguous condition. Adult participants provided the embedded-verb response 33% of the time, representing a 21% increase compared to the ambiguous condition. In 49% of their responses, adults provided the main-verb association by simply stating the location mentioned in the filled gap of the test question (*dans le salon* "in the living room"). In the remaining 18% of their responses, they provided the sublocation answer, reflecting the attachment of the wh-element to the main-clause PP (e.g., "by the couch in the living room"). Table 1 provides detailed descriptive statistics for the offline responses.

The mixed-effects model analysis of embedded-verb responses (see Table 2) revealed an effect of question type—that is, a higher proportion of embedded-verb responses in the filled-gap condition than in the ambiguous condition ($p = .037$) and an effect of age group ($p = .033$). The model also showed a significant interaction between question type and age group ($p = .001$), indicating that the increase of embedded-verb responses from the ambiguous to the filled-gap condition was greater in adults than in children. Planned pairwise comparisons by age group confirmed a higher proportion of embedded-verb responses in the filled-gap condition than in the ambiguous condition in adults ($\beta = -1.437$, $SE = 0.346$, $z = -4.151$, $p < .001$), but not in children ($\beta = 0.148$, $SE = 0.206$, $z = 0.718$, $p = .473$).

The offline response preference in the filled-gap condition varied substantially, not only between age groups, but also between subjects. In the child

group, we found one participant preferring embedded verb over main-verb responses, 26 preferring main-verb over embedded-verb responses, and three opting for embedded-verb responses as often as for main-verb responses. By contrast, in the adult group, 10 participants preferred embedded verb over main-verb responses, 14 preferred main-verb over embedded-verb responses, and six showed no preference. Thus, the majority of participants (90% in children and 80% in adults) showed a systematic interpretation pattern, whereas only a few selected the embedded-verb response at chance in the filled-gap condition.

Discussion

Results from Experiment 1 indicate that French-speaking children and adults preferentially attach the wh-phrase to the first verb in their comprehension of ambiguous biclausal wh-questions, which provides evidence for a main-clause interpretation bias in French. This pattern replicates the findings in Japanese by Omaki et al. (2014) and is consistent with the active gap-filling account that listeners incrementally associate the wh-phrase with the first verb in the sentence. Evidence for the first verb association bias as well as for misinterpretation was observed in the filled-gap condition in both age groups. French-speaking children up to age 8 demonstrated little sensitivity to the filled gap: They provided the main-clause response as often as they did in the ambiguous condition, mirroring the sentence revision difficulties observed in Japanese 5-year-olds. Adults showed a higher rate of embedded-verb interpretation than children in the filled-gap condition, which nevertheless remained extremely low (33% embedded-verb responses, representing a 21% increase from the ambiguous condition) compared to the Japanese finding in sentences with the same filled-gap manipulation where adults were nearly at ceiling in inhibiting their bias to associate the wh-phrase with the first verb.

The evidence for incremental comprehension mechanisms in both French-speaking children and adults is consistent with what has been found in previous developmental work on garden-path

Table 2. *Model of embedded-verb responses in Experiment 1*

Effect	β	SE	z	p
Intercept	−2.187	0.319	−6.864	***
Question type	−0.313	0.150	−2.086	*
Age group	0.371	0.175	2.127	*
Question Type × Age Group	−0.498	0.151	−3.289	**

*$p < .05$. **$p < .01$. ***$p < .001$.

sentences (Choi & Trueswell, 2010; Omaki et al., 2014; Trueswell et al., 1999). However, our results differ from these previous studies in that we found novel evidence for severe revision difficulties in adults. A comparison with the Japanese study of Omaki et al. (2014) on the same structures suggests that the timing of the error signal plays a crucial role in adults' ability to revise. In the Japanese sentences used in Omaki et al. (2014), the filled gap precedes the embedded-clause verb (i.e., the first verb) due to the verb-final word order in that language, but in French, the filled gap follows the main-clause verb. The Japanese–French contrast may thus be taken as evidence for the role of the timing with which the error signal occurs in determining revision difficulty. We return to this point in the General Discussion.

Nevertheless, one cannot exclude the possibility that the contrast between the two languages may be due to the fact that the French translation of the Japanese stimuli allowed the wh-phrase to attach to the filled-gap PP itself: *où* ("where") could be analysed as the head of a larger PP (e.g., *Où dans le salon* "where in the living room") that asks about a specific sublocation inside the larger location described by the overt PP. In our stimuli, this interpretation was made implausible, as the story stimuli mentioned no sublocation inside the relevant locations. However, some participants still ignored pragmatic felicity conditions and opted for a parse that does not involve a revision of the gap position. It is unclear to what extent this additional interpretive possibility affected the main-clause response rate. Children provided their response using a touch screen, and the picture selection data are not precise enough to distinguish the sublocation interpretation from a main-verb interpretation. As for adults who provided answers out loud, they were able to provide answers that reflected a sublocation interpretation. However, the lack of a contextual support for sublocation interpretations may have confused adults and led to an increase of main-verb interpretation.

In summary, the comparison of the previous Japanese study and the current French replication study provides preliminary evidence that the timing of filled-gap signals plays an important role in modulating the misinterpretations of garden-path sentences. Nevertheless, the data are also compatible with an alternative explanation that attributes the cross-linguistic contrast to the presence of an alternative sublocation interpretation in French. This possibility is addressed in Experiment 2, which uses a filler-gap dependency with an argument wh-phrase that eliminates this sublocation interpretation possibility.

EXPERIMENT 2: COMPREHENSION OF ARGUMENT WH-QUESTIONS

Experiment 2 tested whether sentence revision is observed in wh-argument questions like (5), in which the filled gap provides an error signal that syntactically blocks main-verb attachment or sublocation analysis and forces embedded-verb attachment.

5. a. A qui Marie a raconté qu'elle avait distribué des bonbons?
     to whom Marie has told that she had distributed some candy
     "To whom did Marie tell that she had given some candy?"
   b. A qui Marie a raconté **à son papa** qu'elle avait distribué des bonbons?
     to whom Marie has told to her Dad that she had distributed some candy
     "To whom did Marie tell her Dad that she had given some candy?"

The critical difference between Experiments 1 and 2 is that the wh-phrases differ in their syntactic relation to the verb: *Où* ("where") in Experiment 1 was an adjunct modifier that specified the location of an event, but *à qui* ("to whom") in Experiment 2, as shown in (5), is a potential argument for either of the verbs (*raconter* "tell" or *distribuer* "distribute"), as these verbs typically require a recipient of the direct object. Most importantly, the number of arguments for a verb is restricted based on their syntactic and semantic properties (e.g., Levin & Rappaport-Hovav, 1995; Perlmutter, 1978). In the present experiment, both the main-clause and embedded-clause verbs could take a theme and a

recipient as their internal arguments: The main-clause verb (either *expliquer* "explain" or *raconter* "tell") can take the content of speech (e.g., the content of the embedded clause) and its recipient (e.g., someone who hears the content) as their internal arguments, while the embedded-clause verbs were typically "transfer of object" verbs like *distribuer* "distribute" or *donner* "give", which could take a theme (the object given) and an optional recipient (someone who receives the object) as its internal arguments. Since the recipient argument is not obligatory for these verbs, sentences involving those verbs without recipients are grammatical.

Critically, the impossibility to attach the wh-phrase to the argument PP within the main clause prevents attachment of the argument wh-phrase to the argument PP in (5b). In this sense, the argument filled gap in (5b) syntactically prohibits the main-verb interpretation. Given the ungrammaticality of the main-verb attachment analysis, the embedded-verb response rate in this experiment was predicted to be greater than that in Experiment 1.

Method

Participants
Thirty French-speaking children (12 female) aged 5;9–8;3 years (mean age 6;11 years) took part in this experiment. They were recruited at three elementary schools of Geneva. The bilingual status questionnaire (Unsworth, 2013) revealed that 19 (63%) of them were bilingual, but all had French as first and dominant language. Data from eight additional children were excluded from the analyses as they were too distracted to concentrate on the experiment ($N = 4$) or showed extremely long response times ($N = 3$), or due to an experimenter error ($N = 1$).

In addition, 30 French-speaking adults (24 female) aged 19–24 years (mean age 21 years) also participated in this experiment. They were recruited from the student community of the University of Geneva and received course credit for their participation. Fourteen (47%) of them were bilingual according to self-reports, but

French was always the dominant native language that was primarily used at the university and in their daily communication.

Materials
Question design. We constructed eight question sets with two conditions (ambiguous vs. filled gap) shown in (5). The filled-gap condition was created by adding an overt PP recipient to the main-clause verb, which syntactically blocked the attachment of the wh-phrase to the main-clause verb. As in Experiment 1, the main-clause verb was either *expliquer* "explain" or *raconter* "tell". In the embedded clause, we used eight different ditransitive verbs: *acheter* "buy", *distribuer* "distribute", *donner* "give", *envoyer* "send", *montrer* "show", *offrir* "offer", *prêter* "lend", and *servir* "serve". The eight question sets were counterbalanced across two lists and were combined with 10 filler questions. As in Experiment 1, in order to prevent a potential interference between conditions, we used a block design and presented the ambiguous condition in the first block and the filled-gap condition in the second block.

As in Experiment 1, the audio stimuli were prerecorded by a female native speaker of French and were carefully controlled to create a natural prosody that is compatible with either interpretation. The ambiguous condition contained a short pause of 83 ms on average before the embedded clause, but this clause was again longer in the filled-gap condition (117 ms on average), where we aimed at promoting the attachment of the filled-gap PP to the main verb. Also, the filled-gap condition contained hardly any pause between the main verb and its following PP in order to prosodically group these constituents. In both conditions, we implemented further prosodic adjustments such that the embedded-clause attachment of the wh-element was supported by prosody: The pitch of the main verb (*raconter* or *expliquer*) rose only 9 Hz on average from its second to its last syllable (ambiguous condition: +3 Hz; filled-gap condition: +15 Hz); the syllable length of the main verb increased from 76 to 96 ms per phoneme from the second to the last syllable in

the ambiguous condition and decreased from 68 to 48 ms per phoneme in the filled-gap condition.

Story design. The basic design of the stories was identical to that of Experiment 1. A main character undergoes three different events, starting with a distractor event, and followed by a main-clause event and an embedded-clause event. The only difference was that in this experiment, no specific location was described for the events in the stories.

Procedure
The experimental procedure was identical to that of Experiment 1, except that this time adult participants also provided their answers using a touchscreen monitor. This allowed us to record their response time, which may provide an implicit, finer measure of their sentence revision difficulty during the processing of the target filler-gap dependencies.

Data coding and analysis
As in Experiment 1, the offline responses were coded as (a) main-verb responses, (b) embedded-verb responses, and (c) other responses. Other responses included the selection of the distractor location (2% in children and 3% in adults) or null responses (<1% in children and 2% in adults) and were removed from the dataset for analyses. Performance in filler items and second questions was above 95%, indicating that participants were attentive to the task. Offline responses were analysed using a linear mixed-effects regression model with question type (ambiguous vs. filled gap) as fixed effect and random intercepts for subjects and items. The model was conducted on empirical logit transformed embedded-verb responses, a transformation that was required because adults provided zero embedded-verb responses in the ambiguous condition. For the response time analysis, responses with latencies over 2 standard deviations over the subject's mean were discarded (5% in children and 5% in adults). Response times were log-transformed and analysed by way of linear mixed-effects regression models with question type (ambiguous vs. filled gap), age group (children vs. adults), and type of offline

response (main verb vs. embedded verb) as fixed effects, and random intercepts for subjects and items. Finally, in order to compare the rate of embedded-clause responses in Experiments 1 and 2, a linear mixed-effects model analysis was conducted on empirical logit transformed embedded-verb responses with question type (ambiguous vs. filled-gap), age group (child vs. adult), and wh-phrase type (adjunct in Experiment 1 vs. argument in Experiment 2) as fixed effects. For all analyses, sum contrasts were used with one level of factor coded as 0.5 and the other as −0.5. Fully crossed models containing random slopes failed to converge. Estimates (β), standard errors (*SE*), degrees of freedom (*df*), and *t*- and *p*-values of the models were obtained via the lmer function of the lmerTest package (Kuznetsova et al., 2014).

Results

Overall, a strong main-verb preference was observed in the ambiguous condition for adults and children alike (see Table 3): Main-verb responses represent 95% of children's responses and 100% of adults' responses. In the filled-gap condition, an increase in embedded-verb response rate was observed in both child and adult groups. Children provided an embedded-verb response in 48% of the trials, whereas adults provided an embedded-verb response in 21% of the trials. In other words, for adults there was a 21% increase of embedded-verb responses in the filled-gap condition compared to the ambiguous condition, while for children the increase of embedded-verb responses was much greater, at 43%.

Table 3. *Distribution of the offline responses in Experiment 2*

Age group	Ambiguous		Filled gap	
	Main verb	Embedded verb	Main verb	Embedded verb
Children	95 (2)	5 (2)	52 (5)	48 (5)
Adults	100 (0)	0	79 (4)	21 (5)

Note: Offline responses in percentages. Standard errors in parentheses.

Table 4. *Model of empirical logit transformed embedded-verb responses in Experiment 2*

Effect	β	SE	df	t	p
Intercept	−0.975	0.085	77.500	−12.789	***
Question type	−0.512	0.050	527.700	−15.168	***
Age group	0.259	0.085	77.500	3.397	**
Question Type × Age Group	−0.137	0.050	527.700	−4.047	***

$**p < .01. ***p < .001.$

The mixed-effects model analysis of empirical logit transformed embedded-verb responses (see Table 4) revealed an effect of question type ($p < .001$), confirming the higher proportion of embedded-verb responses in the filled-gap condition than in the ambiguous condition, an effect of age group ($p = .001$), and a significant interaction between question type and age group ($p < .001$), revealing that the increase of embedded-verb responses from the ambiguous to the filled-gap condition was greater in children than in adults. Planned pairwise comparisons by age group showed a higher proportion of embedded-verb responses in the filled-gap condition than in the ambiguous condition both in children (β = −0.648, $SE = 0.078$, $df = 263$, $t = −12.314$, $p < .001$) and in adults (β = −0.375, $SE = 0.063$, $df = 266$, $t = −8.976$, $p < .001$).

Table 5 presents the response time data, and a summary of the statistical analyses of the log-transformed response times is provided in Table 6. The mixed-effects model analysis showed a significant effect of question type

($p = .011$), with slower response times in the filled-gap condition ($M = 4533$ ms) than in the ambiguous condition ($M = 3476$ ms). There was a marginal effect of age group ($p = .067$), with slower response times in children ($M = 4255$ ms) than in adults ($M = 3847$ ms), but no interaction between question type and age group ($t < 1$). The offline response type also had an effect on response latency ($p = .058$), with embedded-verb responses ($M = 4600$ ms) being slower than main-verb responses ($M = 3690$ ms). Furthermore, the model revealed a marginal interaction between offline response and question type ($p = .084$) as well as between offline response and age group ($p = .046$). A planned pairwise comparison revealed that the question type had a significant effect on main-verb responses only ($M = 2952$ ms in ambiguous vs. $M = 4342$ ms in filled gap; β = −0.150, $SE = 0.015$, $df = 319$, $t = −10.002$, $p < .001$), but did not modulate embedded-verb response latencies ($M = 4581$ ms vs. $M = 4472$ ms, respectively; $t < 1$).

Similarly to Experiment 1, the offline response preference in the filled-gap condition varied substantially not only between age groups, but also between subjects. In the child group, we found 14 participants preferring embedded-verb to main-verb responses, 13 preferring main-verb over embedded-verb responses, and three selecting embedded-verb responses as often as main-verb responses. By contrast, in the adult group, five participants preferred embedded-verb over main-verb responses, 23 preferred main-verb over embedded-verb responses, and two showed no

Table 5. *Response times in Experiment 2*

Age group	Responses	Ambiguous	Filled gap	Both conditions
Children	Main verb	3044 (329)	4599 (551)	3822 (440)
	Embedded verb	4488 (400)	4888 (392)	4688 (396)
	Both	3766 (365)	4744 (472)	4255 (418)
Adults	Main verb	2895 (186)	4223 (368)	3559 (277)
	Embedded verb	—	4423 (695)	4423 (695)
	Both	2895 (186)	4323 (532)	3847 (416)
Total		3476 (305)	4533 (502)	

Note: Response times in ms. Standard errors in parentheses.

Table 6. *Models of log-transformed response times in Experiment 2*

Effect	β	SE	df	t	p
Intercept	8.186	0.042	192.700	194.910	***
Question type	−0.086	0.034	401.300	−2.547	*
Age group	−0.054	0.029	114.300	−1.849	†
Offline response	0.071	0.037	426.600	1.900	†
Question Type × Age Group	−0.009	0.016	393.700	−0.552	>.10
Question Type × Offline Response	0.060	0.035	405.500	1.732	†
Age Group × Offline Response	−0.050	0.025	427.700	−2.003	*

$^{†}p < .10.$ $^{*}p < .05.$ $^{***}p < .001.$

preference. Thus, the majority of participants (90% in children and 93% in adults) showed a systematic interpretation pattern whereas only very few selected the embedded-verb response at chance in the filled-gap condition.

Combined analysis of Experiments 1 and 2

Both experiments taken together, the offline response data suggested that children's proportion of revision—that is, the proportion of embedded-verb responses to filled-gap questions—was higher in Experiment 2 with argument wh-questions than in Experiment 1 using adjunct wh-questions. In adults, however, revision performance appeared to be stable across experiments, so the adjunct versus argument manipulation had a different impact on children than on adults. In line with that observation, the comparative analysis of empirical logit transformed embedded-verb responses of both experiments showed a significant three-way interaction ($β = −0.313$, $SE = 0.071$, $df = 1015$, $t = −6.556$, $p < .001$), indicating that the interaction between question type and experiment was different across the two age groups. Subsequent models conducted separately for each age group revealed a significant interaction between question type and experiment in children ($β = −0.702$, $SE = 0.102$, $df = 486$, $t = −10.024$, $p < .001$), showing that children revised significantly more in Experiment 2 than in Experiment 1. In contrast, no interaction was found in adults ($β = −0.072$, $SE = 0.098$, $df = 535$, $t = −1.107$, $p = .269$), suggesting that they revised similarly across the two experiments.

Discussion

Results from the ambiguous condition in Experiment 2 corroborate the strong main-clause interpretation bias that we found in Experiment 1. This finding provides further evidence for incremental processing in both adults and children. In contrast, the interpretation patterns observed in the filled-gap condition are different from those of Experiment 1. Surprisingly, children showed more revision than adults (43% increase of embedded-verb responses in children vs. 21% in adults), indicating that children may have a better ability to overcome their main-clause interpretation bias than adults. Moreover, whereas children showed higher sensitivity to argument filled gaps than to adjunct filled gaps, no improvement was found for adults who displayed a low rate of embedded-clause responses in both experiments (21% increase of embedded-verb response in both experiments). In summary, we found converging evidence for incrementality and revision difficulties in wh-argument and wh-adjunct questions. In both cases, adults only rarely succeed in adopting the ultimately correct embedded-clause interpretation, and these findings resemble previous reports of miscomprehension of garden-path sentences (e.g., Christianson et al., 2001). Children's dramatic improvement in Experiment 2 suggests that the high rate of main-verb attachment in Experiment 1 may be due to the possibility to attach the wh- to the filled gap within the main clause. The finding that adults' embedded-clause responses were lower than children's in Experiment 2 suggests that adults' sentence revision mechanisms

may be guided by a different set of constraints (see Experiment 3). In the General Discussion, we address various interpretations of this low revision rate in adults.

The robust main-clause interpretation bias in ambiguous and filled-gap questions may raise concerns about the general accessibility of the embedded-clause questions; it is theoretically possible that extraneous factors may have made the embedded-clause interpretations generally inaccessible in our experiment context. This concern was addressed by Lassotta, Omaki, Panizza, Villata, and Franck (2015), who conducted a similar question-after-story comprehension study with French-speaking 5- to 8-year-old children and adults. Questions involved wh-argument filled-gap questions similar to those in Experiment 2, but the wh-phrase was located in its in situ position (e.g., *Marie a dit à sa maman qu'elle avait servi une tasse de chocolat à qui*? "Marie said to her Mom that she had served a cup of hot chocolate to whom?"). Both children and adults demonstrated the expected preference for the embedded-clause, grammatical interpretation (94% and 97%, respectively). These data show that both age groups are in principle able to comprehend questions that target an embedded verb, and the inaccessibility of that interpretation arises only in the context in which the wh-phrase is fronted, as in Experiments 1 and 2. This is consistent with our argument that difficulties of accessing the embedded-clause interpretation are due to the main-clause interpretation bias that gives rise to garden-path effects.

In both children and adults, longer response times were observed in trials with main-verb responses to filled-gap questions than in trials with main-verb responses to ambiguous questions. This response time pattern suggests that even though the incorrect interpretation was ultimately adopted, participants did not completely ignore the filled gap and actually struggled in adopting the incorrect, main-verb interpretation. Although the cause of the increase in response time is unclear at present, we propose the following two explanations. First, participants may build the grammatical embedded-verb attachment structure

in addition to the ungrammatical main-verb attachment structure, but nevertheless select the latter in their offline response. This possibility cannot be ruled out by empirical evidence, although it is not obvious why participants would select the ungrammatical parse if the grammatical analysis was actually entertained. The second possibility is that the slowdown for main-verb responses in the filled-gap condition is due to the detection of the ungrammaticality in the structure. In fact, exactly like in Experiment 1, during exit interviews many of the adult participants reported to the experimenters that "the answer was in the question" and that these sentences were fairly unnatural. These anecdotal reports lend support to the suggestion that participants did indeed detect the ungrammaticality of their interpretation, but nevertheless failed to find the correct, alternative interpretation.

Before we turn to the broader implications of the current findings, it is important to further examine adults' unexpected difficulties in adopting the embedded-clause interpretation. There are two possible explanations, both of which relate to the nature of adults' linguistic knowledge of French wh-question formation. First, the low rate of embedded-clause interpretation in adults raises the possibility that the filled-gap sentences that we used in Experiments 1 and 2 are simply illicit in the French grammar. If the wh-phrase association with the embedded-clause verb was ungrammatical in the stimuli used in Experiments 1 and 2, it is natural that French-speaking adults did not choose the corresponding embedded-clause interpretation. The syntax research on French wh-questions suggests that such long-distance, cross-clausal wh-association is actually grammatical (e.g., Cheng & Rooryck, 2000), but it is important to empirically verify that naïve participants share the same acceptability intuition.

Second, the severe revision difficulties in adults may reflect their preference for an alternative question formation strategy. As mentioned above, French is an optional wh-movement language, where the wh-phrase can appear in sentence initial position (wh-fronting), or in the original thematic position after the verb (wh-in situ). It

has been noted in the syntax literature that wh-in situ questions are grammatical even when the wh-phrase is in the embedded clause (Mathieu, 2004; Oiry, 2010) and that they may even be a preferred strategy for forming a cross-clausal filler–gap dependency (Myers, 2007). Thus, even if an embedded-clause association of a fronted wh-phrase is grammatical in filled-gap sentences, French speakers may have a preference for an in situ wh-phrase for embedded-clause questions. This potential preference may in turn block the association of a fronted-wh-phrase to the embedded clause. Experiment 3 uses an acceptability judgement task to test these two possible explanations.

EXPERIMENT 3: ACCEPTABILITY JUDGEMENT STUDY

In Experiments 1 and 2, we found that adults often failed to inhibit their main-clause interpretation bias despite the presence of a filled-gap error signal. Moreover, self-reports showed that many of them considered these sentences as fairly unnatural. This may indicate that some French speakers consider cross-clausal filler–gap dependencies to be unavailable as a grammatical option in French, or that they have a strong preference to use a wh-in situ question formation in those particular syntactic contexts. This experiment used an acceptability judgement task to explore whether wh-fronting questions with filled gaps are acceptable for French-speaking adults, and whether the corresponding in situ versions may actually represent a more acceptable formulation of these questions (Myers, 2007).

Method

Participants

Fifty-two French-speaking adults (39 female) aged 20–45 years (mean age 27 years) from French-speaking regions of Switzerland and France participated in this study. Thirty-two (62%) of them were bilingual, but French was always their mother tongue.

Materials

The materials consisted of sentence–question pairings: Each question was preceded by a declarative sentence describing a situation comprising a main-clause event and an embedded-clause event in which a character is involved (e.g., *Paul a expliqué dans le salon qu'il allait pêcher du poisson au port*; "Paul explained in the living room that he was going to catch fish **in the harbor**"). These sentences were introduced to clarify what the intended interpretation of the target question was, given the finding in Experiments 1 and 2 that participants often misinterpret filled-gap questions. The critical noun phrase addressed by the question was in bold characters in order to ensure that participants correctly interpreted the question as bearing on that noun phrase. The declarative sentence was then followed by a question (as illustrated in Table 7), which varied by the following factors: the wh-phrase type (adjunct vs. argument), the complexity of the question (filled gap vs. simple), and the position of the wh-phrase (in situ vs. fronted). All test questions addressed the event from the embedded clause in the declarative sentence. Question type was treated as a between-subjects factor in order to reduce interference between argument and adjunct wh-questions, but sentence complexity and wh-phrase position were treated as within-subject factors. Two experimental lists were created, one with adjunct questions, the other one with argument questions. Each list contained 24 items, six items in each of the four conditions per list. An additional set of 30 filler items were created, identical in both lists, which consisted of 12 simple questions bearing on the event from the main clause in the declarative sentences, nine acceptable *quel* "which" wh-fronted questions (e.g., *Quel signal est-ce que Juliette a expliqué que le conducteur devait attendre?*, literally: "Which signal Juliette explained that the conductor had to await?"), and nine clearly ungrammatical *comment* "how" questions in which the wh-element was fronted within the embedded clause (e.g., *Anna a expliqué que comment le voleur avait laissé des traces?*, literally: "Anna explained that how the robber left traces?").

Table 7. *Examples of questions in the eight experimental conditions of Experiment 3*

Question		Adjunct	Argument
Simple	In situ	Paul allait pêcher du poisson <u>où</u>?	Alicia avait prêté le vélo <u>à qui</u>?
		"Paul was going to catch fish where?"	"Alicia lent the bike to whom?"
	Fronted	<u>Où</u> est-ce que Paul allait pêcher du poisson?	<u>A qui</u> Alicia avait prêté le vélo?
		"Where was Paul going to catch fish?"	"To whom did Alicia lend the bike?"
Filled gap	In situ	Paul a expliqué dans le salon qu'il allait pêcher du poisson <u>où</u>?	Alicia a raconté à son frère qu'elle avait prêté le vélo <u>à qui</u>?
		"Paul explained in the living room that he was going to catch fish where?	"Alicia told her brother that she lent the bike to whom?"
	Fronted	<u>Où</u> est-ce que Paul a expliqué dans le salon qu'il allait pêcher du poisson?	<u>A qui</u> Alicia a raconté à son frère qu'elle avait prêté le vélo?
		"Where did Paul explain in the living room that he was going to catch fish?"	"To whom did Alicia tell her brother that she had lent the bike?"

Procedure

Participants were requested to judge the acceptability of each question on a 7-point rating scale. Acceptability ratings were collected with the Ibex Farm program available on the Internet (Drummond, 2013). Six examples of sentence–question pairs were presented prior to testing with the expected acceptability ratings. None of these examples used the critical adjunct or argument wh-phrases of the target questions. Completing the online questionnaire took approximately 10–15 min.

Data coding and analysis

The 7-point-scale rating data were converted to z-scores in order to correct for individual variations in how the scale is used (for discussion, see Schütze & Sprouse, 2013) once responses with latencies over 2 standard deviations over the subject's mean had been discarded (<1%). Linear mixed-effects regression models were conducted on z-score transformed acceptability judgements with complexity (simple vs. filled gap), wh-position (in situ vs. fronted), and wh-type (adjunct vs. argument) as fixed effects, and random intercepts for subjects and items (e.g., Baayen et al., 2008). The fully crossed model containing random slopes failed to converge. All variables were dichotomous, so sum contrasts were used with one level of the factor coded as 0.5 and the other as 0.5. Estimates (β), standard errors (SE), degrees of freedom (df), and t- and p-values were obtained via the lmer function of the lmerTest package (Kuznetsova et al., 2014).

Results

The distribution of acceptability judgements (z-transformed) is illustrated in Figure 2 while Table 8 provides detailed information about the inferential statistics. The mixed-models analysis showed an effect of sentence complexity ($p < .001$), suggesting that participants gave overall lower ratings to complex questions ($M = -0.071$) than to simple questions ($M = 0.526$), and an effect of wh-type ($p = .009$), indicating lower ratings for adjunct questions ($M = 0.184$) than for argument questions ($M = 0.284$). Furthermore, we found a significant triple interaction between complexity, wh-phrase position, and wh-type ($p = .004$), which led us to conduct pairwise comparisons of wh-phrase position and wh-type effects at each level of complexity.
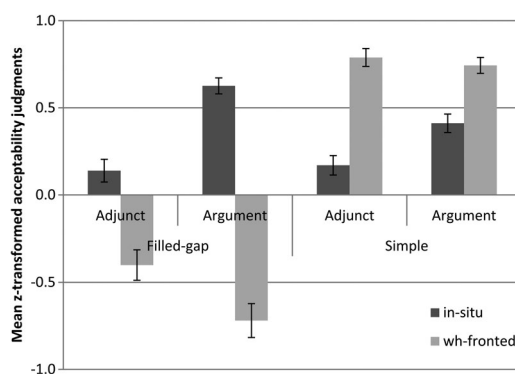

Figure 2. *Mean z-transformed acceptability judgements in Experiment 3.*

Table 8. *Models of z-transformed acceptability judgements in Experiment 3*

Question	Effect	β	SE	df	t	p
All questions	Intercept	0.223	0.060	100.300	3.744	***
	Complexity	−0.598	0.044	1086.200	−13.522	***
	Position	−0.100	0.064	1089.500	−1.567	>.10
	Wh-type	−0.223	0.084	101.800	−2.643	**
	Complexity × Position	0.579	0.062	1087.500	9.324	***
	Complexity × Wh-type	−0.141	0.063	1126.500	−2.218	*
	Position × Wh-type	0.633	0.090	1110.700	7.029	***
	Complexity × Position × Wh-type	0.253	0.089	1129.100	2.852	**
Filled-gap questions	Intercept	−0.375	0.104	89.100	−3.616	***
	Position	0.484	0.099	527.000	4.870	***
	Wh-type	−0.369	0.147	89.700	−2.507	*
	Position × Wh-type	0.891	0.141	521.300	6.326	***
Simple questions	Intercept	0.824	0.078	82.200	10.601	***
	Position	−0.687	0.065	552.100	−10.640	***
	Wh-type	−0.096	0.110	82.900	−0.868	>.10
	Position × Wh-type	0.383	0.093	547.900	4.136	***

*$p < .05$. **$p < .01$. ***$p < .001$.

The model on simple questions showed an effect of position ($p < .001$) with higher acceptability rates for wh-fronted questions ($M = 0.766$) than for in situ questions ($M = 0.289$), and no effect of wh-phrase type ($t < |1|$). The significant interaction between position and wh-type ($p < .001$) showed that while the acceptability ratings differ in the in situ conditions ($M = 0.170$ for adjunct wh-questions and $M = 0.411$ for argument wh-questions), they are comparable in the wh-fronting conditions ($M = 0.789$ for adjunct wh-questions and $M = 0.743$ for argument wh-questions).

The model on filled-gap questions showed an effect of wh-phrase position ($p < .001$), with higher rates for in situ questions ($M = 0.386$) than wh-fronting questions ($M = -0.559$), and an effect of wh-phrase type ($p = .014$), with higher rates for argument wh-questions ($M = -0.017$) than adjunct wh-questions ($M = -0.125$). The significant interaction between wh-phrase position and wh-phrase type ($p < .001$) further revealed that for wh-fronting questions, acceptability ratings were higher in the adjunct ($M = -0.401$) than in the argument condition ($M = -0.719$) while the opposite pattern was observed for in situ questions, receiving lower ratings in the adjunct ($M = 0.139$) than in the argument

condition ($M = 0.626$). A post hoc pairwise comparison confirmed that for filled-gap questions, the acceptability of wh-fronted structures was significantly lower for argument than for adjunct wh-questions (Mann–Whitney $U = 11{,}431$, $p = .022$, two-tailed).

Finally, the comparison between filled-gap wh-fronted questions and the ungrammatical *comment* ("how") filler questions showed that the former were rated as significantly more acceptable ($M = 0.559$) than the latter ($M = -1.145$; Mann–Whitney $U = 83{,}276$, $p < .001$, two-tailed).

Discussion

The main goal of Experiment 3 was to test how French-speaking adults would rate the acceptability of adjunct and argument filled-gap questions used in Experiments 1 and 2. First, as expected, simple questions were rated as more acceptable than complex filled-gap questions. Second, wh-fronting sentences are perfectly acceptable in simple questions, and even significantly more so than in situ versions. Importantly, however, the reverse pattern was observed for filled-gap questions, for which in situ questions are more acceptable than wh-fronted questions. This finding suggests that with biclausal

wh-questions with argument or adjunct filled-gaps, French-speaking adults typically prefer the in situ versions to the wh-fronting versions. Moreover, among the filled-gap questions, the acceptability improvement with in situ wh-phrases was greater for argument than for adjunct questions. Finally, despite their relatively low acceptability, filled-gap questions with wh-fronting are more acceptable than the ungrammatical filler questions, suggesting that the filled-gap wh-fronting questions are not treated on a par with the ungrammatical filler sentences that we used in this study.

The preference for the wh-in situ version of biclausal wh-questions may have important implications for how the wh-fronting questions were processed in Experiments 1 and 2. First, this preference could plausibly reflect distributional patterns of wh-in situ and wh-fronting options in French, namely that wh-in situ questions are more common grammatical realizations of embedded-clause constituent questions (Myers, 2007). This distributional knowledge could lead French-speaking adults to infer that if there is a fronted wh-phrase, it is likely to be associated with the verb in the same clause: In other words, in a biclausal sentence with wh-fronting, they should expect the fronted wh-phrase to have originated from the main clause. This explanation is consistent with psycholinguistic theories that attribute syntactic expectations during language comprehension to the use of distributional information about their native language (Hale, 2003; Levy, 2008). If adults prioritize this distributional constraint on filler-gap dependencies in deciding where the gap position is, the quality of the filled-gap error signal becomes less relevant. On the other hand, children may not have developed the knowledge of this distributional constraint, or else they may not have learned to use this information to generate expectations during sentence processing. This would explain their higher sensitivity to the syntactic error signal provided by the filled gap.

The question of why adults find in situ filled-gap questions more acceptable goes beyond the scope of the present paper. Nevertheless, we note that the in situ structure may incur fewer processing demands than the wh-fronted version, due to the absence of filler-gap dependencies with intervening discourse referents (e.g., Fiebach, Schlesewsky, & Friederici, 2002; Gibson, 1998; Gordon, Hendrick, & Levine, 2002; King & Just, 1991; Van Dyke & McElree, 2006). However, given that wh-in situ involves covert long-distance dependencies that are subject to memory interference effects (e.g., Xiang, Dillon, Wagers, Liu, & Guo, 2014), further empirical work is needed to systematically compare processing cost differences (if any) between wh-fronting and wh-in situ dependencies.

GENERAL DISCUSSION

This study used two story-based comprehension experiments with adults and 5- to 8-year-old children to investigate their interpretation biases in globally ambiguous biclausal wh-questions like *Où est-ce qu'Aline a expliqué qu'elle allait attraper des papillons?* ("Where did Aline explain that she was going to catch butterflies?"). Following the previous Japanese study (Omaki et al., 2014), Experiment 1 used an adjunct wh-phrase (*où* "where") and produced two important findings. French-speaking adults and children both have a strong main-clause interpretation bias. Moreover, children failed to inhibit their main-clause interpretation bias when an overt filled-gap PP modifier of the main-clause verb encouraged the embedded-clause interpretation. In this experiment, adults succeeded in inhibiting their main-clause bias more often than children did, but not as successfully as expected based on previous findings in Japanese. Experiment 2 used argument wh-questions (*a qui* "to whom"), which provided a stronger error signal that the main-clause attachment should be syntactically blocked. Here, we found that the argument filled-gap manipulation indeed improved children's inhibition of the main-clause interpretation bias. However, adults again showed little sensitivity to the filled-gap information and did not inhibit their main-clause interpretation any more than they did in Experiment 1. Finally, Experiment 3 used an acceptability judgement task to ensure that the surprisingly high rate of misinterpretation in adults is not due to low acceptability of our target sentences. It was found that adults

consider both adjunct and argument wh-questions with filled gaps to be grammatical sentences of French, although they systematically preferred the in situ counterpart of these questions over the wh-fronted versions used in Experiments 1 and 2.

French–Japanese contrast and implications for models of sentence revision mechanisms

One of the important findings in this study was that French-speaking adults failed to inhibit their main-clause interpretation bias in the majority of the sentences. For the adjunct wh-questions in Experiment 1, adults provided incorrect, main-clause verb attachment interpretations in 49% of the filled-gap sentences, in contrast to the grammatically licit interpretations (18% of sublocation answers, and 33% of embedded-clause responses). In the argument wh-questions in Experiment 2, adults provided the incorrect main-clause verb attachment interpretation 79% of the time. This was a striking finding, especially because adults' comprehension accuracy was lower than that of children in Experiment 2. These findings indicate that adults' interpretation mechanisms can be severely misled by the initial, preferred analysis of garden-path sentences (Ferreira & Patson, 2007). In fact, the data reported here with the question-after-story procedure provide even stronger evidence for misinterpretation effects than previous comprehension studies. As noted in the introduction, Christianson et al. (2001) used garden-path sentences like *While Anna dressed the baby spit up on the bed*, and found that adults often answer "yes" to the question *Did Anna dress the baby?* even though this interpretation is not grammatically available. However, it has been pointed out that the yes–no questions used in those studies may have artificially induced participants to consider the ungrammatical interpretation (Tabor et al., 2004). In contrast, our study used wh-questions that do not promote either of the potential interpretations. Thus, our experimental findings are free of potential methodological biases and provide unambiguous evidence for severe misinterpretations in adults.

On the other hand, Japanese adults in previous work showed little difficulty in comprehending such sentences. Given that the story design in the present study was modelled after the Japanese study by Omaki et al. (2014), the source of this contrast is likely to arise from differences in the structure of the target sentences. One major difference between the two languages concerns the word order: In French the filled gap was postverbal as it followed the first verb in the sentence, whereas in Japanese the filled gap was preverbal. We suggest that this word order effect may imply that inhibiting the initial interpretation is much more costly than revising the initial syntactic analysis. Given the robust filled-gap effects seen in Japanese- and French-speaking children, it is very plausible that the adult parser also triggers a syntactic revision process at the point of encountering a filled gap (for related reading time evidence for Japanese adults, see Aoshima et al., 2004). Under this assumption, the cost for revising the syntactic structure should not differ between French and Japanese. However, the two languages differ in terms of the timing of the revision signal. In Japanese where the filled gap precedes the verb, no interpretation has been assigned to the filler-gap dependency at the point of encountering the filled gap, whereas in French the filler-gap dependency can be assigned an interpretation at the verb, before encountering the filled gap. In this sense, revision in French not only involves revising the syntactic structure, it also involves inhibiting the interpretation assigned to this structure, whereas only syntactic revision is required in Japanese. We thus propose that the cost of inhibiting this initial interpretation played an important role in the severe misinterpretations observed in French.

There are relevant findings from previous work that the timing of error signal presentation modulates sentence revision difficulties. For example, Ferreira and Henderson (1991) manipulated the length of the ambiguous region of garden-path sentences [e.g., *When the men hunt the birds (that cheetahs eat) typically scatter*] and found that revision errors increased when the ambiguous region was made longer, such that the disambiguation region

arrived much later in the sentence (see also Christianson et al., 2001; Ferreira & Henderson, 1998). It is important to note that these studies held the initial misinterpretation (e.g., interpreting *the birds* as the patient of *hunt*) constant across conditions, whereas the French–Japanese comparison of garden-path wh-questions effectively allowed us to vary the presence or absence of temporary misinterpretation. Taken together, these findings highlight the cost of interpretation revision; simply revising the syntactic structure incurs little cost, but revision of interpretative commitment causes a severe cost that increases as the revision process is delayed.

The large cost associated with revising the interpretive commitment also has implications for explanations of misinterpretations in garden-path sentences. The current finding is consistent with proposals that misinterpretations in garden-path sentences arise due to the interpretation that was assigned based on the initial syntactic and semantic analyses, which lingers in long-term memory and interferes with subsequent interpretation of the sentence (Slattery et al., 2013; Sturt, 2007). Thus, the misinterpretation data in French garden-path questions may also be explained by a lingering memory representation of the initial main-clause interpretation, which is completed before the parser encounters the postverbal filled-gap error signal. In the Japanese case, however, no interpretation has been assigned to the initial (mis)analysis of the filler-gap dependency, and there is thus no misinterpretation that lingers in memory.

Our findings are also compatible with the view that "good-enough" syntactic representations are the main cause of misinterpretations (Christianson et al., 2001; Ferreira, 2003; Ferreira & Patson, 2007). Under that account, both Japanese- and French-speaking participants initially attached the wh-element to the main verb. However, whereas the former would significantly succeed in revising that initial parse, the latter would fail and therefore persist with a "good-enough" representation in which both the wh-element and the filled-gap PP compete for attaching to the main verb. Here also, the difference in the timing of occurrence of the error signal could have played a role: A signal arriving

before the verb would allow a more shallow attachment of the wh-element to the main verb, which is relatively easy to revise, whereas a signal arriving after the verb would allow a more robust attachment, harder to revise. Another possibility is that the extreme difficulties observed in French-speaking participants are due to their distributional knowledge of wh-extractions uncovered in Experiment 3. This possibility is discussed in the next section.

Developmental changes in mechanisms underlying the comprehension of filled-gap sentences

One of the striking findings in our experiments concerns the developmental changes in the inhibition of the main-clause interpretation bias. In the wh-adjunct questions of Experiment 1, adults showed more embedded-verb attachment than children. However, the difference between the two groups is potentially explained by the possibility of attaching the wh-phrase to the filled-gap PP within the main clause, a possibility that could only be observed in adults since children were required to point at the whole picture. In the wh-argument questions, where this alternative sublocation interpretation was ruled out by the grammar, children showed a higher sensitivity to the error signal than adults, whose comprehension pattern was roughly the same in wh-argument and wh-adjunct questions. This complex pattern of developmental changes in interpretation behaviours suggests that the underlying mechanisms for sentence revision in children and adults may rely on different types of constraints.

With respect to children, the interpretation patterns reported in Experiments 1 and 2 are consistent with our expectations based on previous findings on sentence revision mechanisms. First, the fact that children do not always succeed in inhibiting the main-clause interpretation bias is consistent with the general suggestion in the literature that children do not have sufficient cognitive resources to inhibit incorrect representational alternatives (e.g., Mazuka et al., 2009; Novick et al., 2005, 2010). Second, the increased sensitivity to the filled-gap error signal in argument wh-questions is also

consistent with observations in the adult sentence-processing literature that the quality of error signals influences sentence revision performance (Fodor & Inoue, 1994, 1998, 2000; Frazier & Clifton, 1998; Lewis, 1998). In our filled-gap error signal manipulation, the adjunct filled gap only served as a weak error signal: The wh-phrase could potentially be attached directly to the filled-gap PP (under the interpretation "where in the [location]?") and thus did not syntactically force wh-attachment to the embedded-clause verb. Here, the primary motivation for adopting the embedded-clause interpretation comes from the absence of specific sublocations in the given context; the adjunct filled-gap does not syntactically indicate where the contextually felicitous gap position is. In contrast, argument filled gaps provide clear error signals, as there is no grammatical way to attach the wh-phrase anywhere in the main clause. The complete elimination of main-clause attachment options effectively forces the wh-phrase to be attached to the embedded-clause verb. The fact that children's embedded-clause response rate was significantly higher in argument than in adjunct questions suggests that linguistic factors, and more specifically the presence of a strong syntactic error signal, plays an important role in their sentence revision mechanisms: The child's incremental parser privileges an early attachment if syntax allows it, even though it is pragmatically infelicitous (wh-adjunct condition), but it revises this attachment, to some extent, if it is syntactically illicit (with wh-arguments).

Let us turn now to the mechanisms underlying the surprisingly frequent misinterpretation effects reported in French-speaking adults. The different comprehension patterns observed in adults and children suggest that the interpretation processes of adults and children may be guided by different types of constraints. Based on the acceptability judgement data gathered in Experiment 3, we suggest that French-speaking adults are primarily guided by their knowledge of distributional information on filler-gap dependencies themselves. As shown in the acceptability judgement data, in biclausal wh-questions with filled gaps, adults generally prefer the in situ version to the wh-fronting version, even though they prefer wh- extraction in the case of simple questions (Myers, 2007). We suggest that this judgement contrast is what guided their general syntactic biases on filler-gap dependencies: Adults expect fronted wh-questions in biclausal sentences to involve a filler-gap dependency that does not cross a clausal boundary, and thus that the dependency must be completed within the main clause. Furthermore, the data suggest that this syntactic expectation on filler-gap dependencies is prioritized over other syntactic cues for the gap position, such as the presence of a filled-gap error signal. These two proposals together can explain why French-speaking adults are generally poor at inhibiting the main-clause interpretation: Based on the syntactic distribution of filler-gap dependencies in French, they expect it to be completed in the main clause. This expectation overrides information coming from syntactic error signals, explaining the same low rate of embedded-clause interpretation with adjuncts and arguments.

One implication of the current proposal is that at some point in development, children's sentence revision mechanisms evolve to become sensitive to distributional constraints on the syntax of filler-gap dependencies that appear to guide adults' comprehension mechanisms. The possibility that children are not sensitive to distributional information may at first glance appear to contradict the vast developmental literature on children's sophisticated distributional analysis mechanisms (e.g., Gerken, 2006; Gomez, 2002; Maye, Werker, & Gerken, 2002; Saffran, Aslin, & Newport, 1996; Thompson & Newport, 2007). However, these studies are concerned with the acquisition of linguistic rules and representations (e.g., phonemic categories, word units, phrase structure rules, nonadjacent dependencies), but not with the real-time procedures for assigning those representations during sentence comprehension. To our knowledge, there is very little empirical work on the role of distributional information in child sentence processing. Recently, adult sentence-processing studies have suggested that the same type of distributional analysis mechanisms may explain how adults' sentence-processing biases change during the course of an experimental

session (Fine, Jaeger, Farmer, & Qian, 2013; Kaschak & Glenberg, 2004). It is plausible that French-speaking children gradually analyse the distribution of in situ and wh-fronting questions and eventually acquire adult-like biases to prioritize question type distribution information over the information coming from error signals in the sentence.

CONCLUSION

The present study used garden-path wh-questions in French to probe the nature of sentence revision mechanisms in adults and children. The results from two comprehension experiments and an acceptability judgement experiment revealed two novel findings. First, French-speaking adults show robust misinterpretation errors, unlike Japanese adults, which suggests that the late arrival of an error signal increases the chance of misinterpretation errors. Second, children showed a higher sensitivity to syntactic error signals than adults did. This developmental change suggests that adults and children are guided by different types of constraints: Adults may prioritize distributional constraints on the syntax of filler-gap dependencies in the language, while children would prioritize syntactic information that comes directly from error signals in the sentence. While future studies are needed to explore how parsing mechanisms change over the time-course of language development, the present study demonstrates that detailed investigations of linguistic and age factors provide insights into the nature of sentence revision mechanisms.

REFERENCES

Aoshima, S., Phillips, C., & Weinberg, A. (2004). Processing filler-gap dependencies in a head-final language. *Journal of Memory and Language*, *51*, 23–54.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 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*, 255–278.

Cheng, L. L-S., & Rooryck, J. (2000). Licensing Wh-in-situ. *Syntax*, *3*(1), 1–19.

Choi, Y., & Trueswell, J. C. (2010). Children's (in)ability to recover from garden paths in a verb-final language: Evidence for developing control in sentence processing. *Journal of Experimental Child Psychology*, *106*, 41–61.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*, 368–407.

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 Villiers, J., Roeper, T., Bland-Stewart, L., & Pearson, B. (2008). Answering hard questions: Wh-movement across dialects and disorder. *Applied Psycholinguistics*, *29*, 67–103.

De Villiers, J., Roeper, T., & Vainikka, A. (1990). The acquisition of long-distance rules. In L. Frazier & J. De Villiers (Eds.), *Language processing and language acquisition* (pp. 257–297). Dordrecht: Kluwer.

Drummond, A. (2013). *Ibex farm*. Retrieved from http://spellout.net/ibexfarm/

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*(2), 164–203.

Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, *30*, 725–745.

Ferreira, F., & Henderson, J. M. (1998). Syntactic reanalysis, thematic processing, and sentence comprehension. In J. D. Fodor & F. Ferreira (Eds.), *Studies in Theoretical Psycholinguistics, Volume 21* (pp. 73–100). Dordrecht: Kluwer Academic Publishers.

Ferreira, F., & Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass*, *1*, 71–83.

Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German WH-questions. *Journal of Memory and Language*, *47*, 250–272.

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS ONE*, *8*(10), 1–18.

Fodor, J. D., & Inoue, A. (1994). The diagnosis and cure of garden paths. *Journal of Psycholinguistic Research*, *23* (5), 407–434.

Fodor, J. D., & Inoue, A. (1998). Attach anyway. In J. D. Fodor & Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 101–141). Dordrecht: Kluwer Academic Publishers.

Fodor, J. D., & Inoue, A. (2000). Syntactic features in reanalysis: Positive and negative symptoms. *Journal of Psycholinguistic Research*, *29*(1), 25–36.

Frazier, L., & Clifton, C. Jr. (1998). Sentence reanalysis, and visibility. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 143–176). Dordrecht: Springer.

Frazier, L. & Flores d'Arcais, G. B. F. (1989). Filler driven parsing: A study of gap filling in dutch. *Journal of Memory and Language*, *28*, 331–344.

Garnsey, S. M., Tanenhaus, M. K., & Chapman, R. M. (1989). Evoked potentials and the study of sentence comprehension. *Journal of Psycholinguistic Research*, *18*, 51–60.

Gerken, L. (2006). Decisions, decisions: Infant language learning when multiple generalizations are possible. *Cognition*, *98*(3), B67–B74.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1–76.

Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*, 431–436.

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-load interference in syntactic processing. *Psychological Science*, *13*(5), 425–430.

Hale, J. (2003). The information conveyed by words in sentences. *Journal of Psycholinguistic Research*, *32*(2), 101–123.

Huang, Y. T., Zheng, X., Meng, X., & Snedeker, J. (2013). Children's assignment of grammatical roles in the online processing of Mandarin passive sentences. *Journal of Memory and Language*, *69*(4), 589–606.

Hurewitz, F., Brown-Schmidt, S., Thorpe, K., Gleitman, L. R., & Trueswell, J. C. (2000). One frog, two frog, red frog, blue frog: Factors affecting children's syntactic choices in production and comprehension. *Journal of Psycholinguistic Research*, *29*, 597–626.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446.

Kaschak, M. P., & Glenberg, A. M. (2004). This construction needs learned. *Journal of Experimental Psychology*, *133*(3), 450–467.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory* 1. *Journal of Memory and Language*, *30*(5), 580–602.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2014). lmerTest: Tests for random and fixed effects for linearmixed effect models. R Package Version 2.0–11. Retrieved from http://CRAN.R-project.org/package=lmerTest.

Lassotta, R., Omak, A., Panizza, D., Villata, S., & Franck, J. (2015). *Linking cognitive control and syntactic revision of garden-path wh-questions in adults and children*. Manuscript in preparation.

Levin, B., & Rappaport-Hovav, M. R. (1995). *Unaccusativity: At the syntax-lexical semantics interface*. Cambridge, MA: The MIT Press.

Levy, R. (2008). *A noisy-channel model of rational human sentence comprehension under uncertain input*. Proceedings of the 13th Conference on Empirical Methods in Natural Language Processing (Association for Computational Linguistics, Stroudsburg, PA), 234–243.

Lewis, R. L. (1998). Reanalysis and limited repair parsing: Leaping off the garden path. In F. Ferreira & J. D. Fodor (Eds.), *Reanalysis in sentence processing* (pp. 247–286). Dordrecht: Springer.

Love, T. E. (2007). The processing of non-canonically ordered constituents in long distance dependencies by pre-school children: A real-time investigation. *Journal of Psycholinguistic Research*, *36*, 191–206.

Marslen-Wilson, W. (1975). Sentence perception as an interactive parallel process. *Science*, *189*, 226–228.

Mathieu, E. (2004). The mapping of form and interpretation: The case of optional wh-movement in french. *Lingua*, *114*(9–10), 1090–1132.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*, B101–B111.

Mazuka, R., Jincho, N., & Oishi, H. (2009). Development of executive control and language processing. *Language and Linguistics Compass*, *3*, 59–89.

Meroni, L., & Crain, S. (2003). On not being led down the kindergarten path. In B. Beachley, A. Brown, & F. Conlin (Eds.), *Proceedings of the 27th boston university conference on language development* (pp. 531–544). Somerville, MA: Cascadilla Press.

Myers, L. L. (2007). *WH-interrogatives in spoken French: A corpus-based analysis of their form and function*. Unpublished doctoral dissertation, University of Texas at Austin.

Novick, J. M., Hussey, E., Teubner-Rhodes, S., Harbison, J. I., & Bunting, M. F. (2013). Clearing

the garden-path: Improving sentence processing through cognitive control training. *Language, Cognition and Neuroscience*, 29(2), 186–217.

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of broca's area in sentence comprehension. *Cognitive, Affective & Behavioral Neuroscience*, 5, 263–281.

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2010). Broca's area and language processing: Evidence for the cognitive control connection. *Language and Linguistics Compass*, 4, 906–924.

Oiry, M. (2010). A case of true optionality: Wh-in-situ patterns like long movement in French. In A. Roussou & C. Vlachos (Eds.), The optionality of wh-Movement. *Linguistic Analysis*, 37(1–2), 112–136.

Omaki, A., Davidson White, I., Goro, T., Lidz, J., & Phillips, C. (2014). No fear of commitment: Children's incremental interpretation in English and Japanese wh questions. *Language Learning and Development*, 0, 1–28.

Omaki, A., & Lidz, J. (2015). Linking parser development to acquisition of syntactic knowledge. *Language Acquisition*, 22, 158–192.

Omaki, A., & Schulz, B. (2011). Filler-gap dependencies and island constraints in second-language sentence processing. *Studies in Second Language Acquisition*, 33, 563–588.

Perlmutter, D. (1978). *Impersonal passives and the unaccusative hypothesis*. Proceedings of the fourth annual meeting of the Berkeley Linguistics Society, Berkeley, pp. 157–189.

Phillips, C., & Wagers, M. (2007). Relating structure and time in linguistics and psycholinguistics. In G. Gaskell (Ed.), *Oxford handbook of psycholinguistics* (pp. 739–756). Oxford: Oxford University Press.

R Development Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Roberts, L., Marinis, T., Felser, C., & Clahsen, H. (2007). Antecedent priming at trace positions in children's sentence processing. *Journal of Psycholinguistic Research*, 36, 175–188.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, New Series*, 274, 1926–1928.

Schütze, C. T., & Sprouse, J. (2013). Judgment data. In R. J. Podesva & D. Sharma (Eds.), *Research methods in linguistics* (pp. 27–50). Cambridge: Cambride University Press.

Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language*, 69(2), 104–120.

Snedeker, J. (2013). Children's sentence processing. In R. P. G. van Gompel (Ed.), *Sentence processing* (pp. 189–220). New York, NY: Psychology Press.

Staub, A. (2007). The return of the repressed: Abandoned parses facilitate syntactic reanalysis. *Journal of Memory and Language*, 57(2), 299–323.

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

Sturt, P. (2007). Semantic re-interpretation and garden path recovery. *Cognition*, 105, 477–488.

Sussman, R. S., & Sedivy, J. (2003). The time-course of processing syntactic dependencies: Evidence from eye movements. *Language and Cognitive Processes*, 18(2), 143–163.

Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, 50(4), 355–370.

Tanenhaus, K., Spivey-Knowlton, J., Eberhard, M., & Sedivy, C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634.

Thompson, S. P., & Newport, E. L. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development*, 3(1), 1–42.

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

Trueswell, J. C., & Gleitman, L. R. (2007). Learning to parse and its implications for language acquisition. In Gaskell (Ed.), *Oxford handbook of psycholinguistics* (pp. 635–656). Oxford: Oxford University Press.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, 73, 89–134.

Unsworth, S. (2013). Assessing the role of current and *cumulative* exposure in simultaneous bilingual acquisition: The case of Dutch gender. *Bilingualism: Language and Cognition*, 16, 86–110.

Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, 55(2), 157–166.

van Gompel, R. P. G., Pickering, M. J., Pearson, J., & Jacob, G. (2006). The activation of inappropriate analyses in garden-path sentences: Evidence from structural priming. *Journal of Memory and Language*, *55*(3), 335–362.

Wagers, M. W., & Phillips, C. (2014). Going the distance: Memory and control processes in active dependency construction. *The Quarterly Journal of Experimental Psychology*, *67*(7), 1274–1304.

Weighall, A. R. (2008). The kindergarten path effect revisited: Children's use of context in processing structural ambiguities. *Journal of Experimental Child Psychology*, *99*, 75–95.

Xiang, M., Dillon, B., Wagers, M., Liu, F. Q., & Guo, T. M. (2014). Processing covert dependencies: An SAT study on Mandarin wh-in-situ questions. *Journal of East Asian Linguistics*, *23*, 207–232.