languages

MDPI

Article

Adaptation of Gap Predictions in Filler-Gap Dependency Processing during Reading

Emily Atkinson [1,*] and Akira Omaki [2,†]

[1] Department of Language Studies, University of Toronto Mississauga, Mississauga, ON L5L 1C6, Canada
[2] Department of Linguistics, University of Washington, Seattle, WA 98195, USA
* Correspondence: e.atkinson@utoronto.ca
† Deceased author.

Abstract: Syntactic adaptation effects have been demonstrated for an expanding list of structure types, but the mechanism underlying this effect is still being explored. In the current work on filler-gap dependency processing, we examined whether exposing participants to a less common gap location—prepositional object (PO) gaps—altered their gap predictions, and whether these effects would transfer across tasks when this input was presented in a quasi-naturalistic way (i.e., by reading stories). In Experiment 1, we demonstrated that comprehenders dampened their direct object (DO) gap predictions following exposure to PO gaps. However, Experiments 2A and 2B suggest that these adaptation effects did not transfer when the quasi-naturalistic exposure phase was presented as a separate task (Experiment 2A) and when they also needed to generalize from a syntactic to a semantic measure of direct object gap predictions (i.e., filled gap vs. plausibility mismatch sentences; Experiment 2B). Overall, these experiments add filler-gap dependency processing, as well as the gap predictions associated with it, to the growing list of structures demonstrating adaptation effects, while also suggesting that this effect may be specific to a singular experimental task environment.

Keywords: sentence processing; syntactic adaptation; filler-gap dependencies; prediction

check for updates

Citation: Atkinson, Emily, and Akira Omaki. 2023. Adaptation of Gap Predictions in Filler-Gap Dependency Processing during Reading. *Languages* 8: 285. https://doi.org/10.3390/languages8040285

Academic Editors: Naomi Havron and Chiara Gambi

Received: 28 March 2023
Revised: 20 November 2023
Accepted: 1 December 2023
Published: 12 December 2023

1. Introduction

Language input is inherently variable, as individual speakers may differ in their pronunciations, lexical choices, or selection of syntactic structures. Given this, comprehenders need to adjust their expectations when presented with non-stereotypical language input. The faster that this adjustment can occur, the less cognitive resources will be necessary for effective comprehension. Thus, rapid adaptability to variability is required for efficient communication.

In the speech perception literature, it has been demonstrated that listeners can temporarily adapt their phonetic categorizations to variable acoustic input (Kraljic and Samuel 2007). Listeners adapt after limited exposure to non-native speech (Bradlow and Bent 2008; Samuel and Larraza 2015), dialect variants (Dahan et al. 2008; Kraljic et al. 2008; Sumner and Samuel 2009), and acoustically distorted speech (Davis et al. 2005). This phenomenon has also been modeled using Bayesian belief-updating (Kleinschmidt and Jaeger 2011, 2015). Recently, adaptation methodologies from speech perception research have been extended to other linguistic domains, namely, expectation of syntactic structure (e.g., Fine et al. 2013; Myslín and Levy 2016), quantifier meaning (Yildirim et al. 2016), and informativity of referring expressions (Pogue et al. 2016). Additional work has shown that biases in real-time ambiguity resolution can be altered by manipulating the probability of the competing structures in the input (see Kaan and Chun 2018, for a review).

Another major source of evidence for comprehenders' ability to generate and update expectations based on their previous linguistic input and the current linguistic environment comes from the extensive syntactic priming literature. Many studies have demonstrated that speakers tend to repeat syntactic structures across utterances (Bock 1986; Bock and

Loebell 1990; Branigan et al. 2000a, 2000b; Pickering and Branigan 1998, among many others). While not originally intended to investigate adaptation to structural preferences, the literature on cumulative priming effects provides a large body of evidence for a tendency toward structural repetition that extends beyond the following utterance (Bock and Griffin 2000; Kaschak 2007; Kaschak et al. 2006, 2014). Kaschak et al. (2006) demonstrated that recent experience with double object (DO) and prepositional object (PO) datives affected the strength of structural priming in written sentence productions. Participants completed sentences skewed toward DO dative completions (*Meghan gave her mother*...), PO dative completions (*Meghan gave the doll*...), or either DO or PO dative completions (*The soldier gave*...). During a "Recent Experience" phase, participants completed either an equal number of DO and PO constructions or only one of the constructions. Those with equal experience with the two dative completions were strongly primed. However, when participants' recent experience was skewed toward a single construction, priming was greatly reduced for the alternate structure. This finding suggests that there are cumulative effects of language experience on structural priming; specifically, the relative frequency of the structures in recent experience affected the strength of priming.

A similar line of work examines whether comprehenders adapt their expectations about upcoming syntactic structure after exposure to a priori unexpected structures (Chun 2018; Farmer et al. 2014; Fine et al. 2010, 2013; Fine and Jaeger 2013; Kamide 2012; Kaschak and Glenberg 2004; Linzen and Jaeger 2016; Noppeney and Price 2004; Tooley and Traxler 2018; Wells et al. 2009). Fine et al. (2013) explored whether the comprehension of temporary syntactic ambiguities changed based on repeated exposure to these structures. In particular, they examined the temporary ambiguity that is generated by the past participle form of certain verbs, which can be used both as a main verb (1a) and as the verb in a reduced relative clause (1b).

(1) The experienced soldiers...
  a. *Main verb:* ...warned about the dangers before the midnight raid.
  b. *Reduced relative:* ...warned about the dangers conducted the midnight raid.

In Fine et al.'s Experiment 2, a group of participants was exposed to 16 sentences with reduced relative clauses like those in (1b).[1] Normally, the main clause/reduced relative ambiguity leads to a significant reading time slowdown on the disambiguating region, *conducted the midnight* in (1b), in sentences with reduced relative clauses. Participants exposed to reduced relative clauses demonstrated reduced processing difficulty on the disambiguating region in ambiguous sentences including this structure compared to a control group that was exposed to filler sentences.

Fine et al. suggest that these results indicated adaptation to the distribution of syntactic structures in the input. When reduced relatives were more frequent in the input, participants updated their expectations about the probability of a reduced relative continuation upon encountering a verb in its past participle form. Increased exposure to reduced relatives led to an increased expectation of that structure in the future and decreased processing difficulty when it was encountered. This finding has led to a burst of interest in examining syntactic adaptation effects. However, recent attempts by Harrington Stack et al. (2018) to investigate whether these adaptation effects are context sensitive failed to find transfer of learning to a new context.

The Current Studies

Additional recent work has replicated the original finding with reduced relative clauses (Dempsey et al. 2020, 2023; Yan and Jaeger 2020) as well as expanded the range of structures involved in successful adaptation to include relative clause attachment ambiguities (Chun 2018; Kamide 2012), subject versus object relative clauses (Wells et al. 2009), *and* coordination (Kaan et al. 2019) and dialectal differences (Boland et al. 2023; Fraundorf and Jaeger 2016; Kaschak and Glenberg 2004).

This paper seeks to add another structure to this list: gap predictions in filler-gap dependency processing. Filler-gap dependencies are long-distance dependencies found in constructions like *wh*-questions (e.g., *What did Julie paint __?*) or relative clauses (e.g., *The book that the author wrote __ was on sale*). The *filler* is a constituent that has been fronted to a non-canonical structural position (e.g., *what* or *the book*) and must be associated with its canonical structural position (i.e., the *gap*). Extensive research has established that after processing a filler, comprehenders predict a gap position before bottom-up information is available, a processing strategy known as active gap filling (e.g., Crain and Fodor 1985; Frazier 1987; Garnsey et al. 1989; Omaki et al. 2015; Stowe 1986; Traxler and Pickering 1996). For example, Stowe (1986) compared reading times at the direct object in sentences with a filler-gap dependency (2a) to those without such a dependency (2b).

(2) a. My brother wanted to know who Ruth will bring **us** home to __ at Christmas.
b. My brother wanted to know if Ruth will bring **us** home to __ at Christmas.

The direct object, *us*, was read more slowly in the filler-gap dependency conditions (2a) than in those without such a dependency (2b). Increased reading time on this region indicates surprise that the direct object position was filled with a pronominal NP. This *filled gap effect* indicates that a gap had been posited in the direct object position before confirming, via the information provided by a gap in the argument structure of the verb, that this position was unoccupied.

Crucially, filler-gap dependencies are unbounded as the gap can potentially occur in any number of structural positions (e.g., subject, direct object, indirect object, prepositional object). The preference to predict a direct object gap could arise from one of several sources. Maintaining the filler in memory until it can be integrated is costly, and the parser may be driven to reduce memory loads (Frazier and Clifton 1989; Gibson 1998, 2000). Alternatively, it could reflect statistical regularities about gap positions in the input. If the latter, it seems reasonable that distributional information about syntactic structures (namely, gap positions) might also have an effect on the predictive processing associated with filler-gap dependencies.

While one previous study examined adaptation of gap position predictions in both native English speakers and L2 learners of English (Kaan et al. 2019), no definitive conclusions about syntactic adaptation of this structure can be drawn. Kaan and colleagues compared reading times on the direct object of verbs within embedded questions that either contained a filler-gap dependency or did not (*The builder wondered <u>what</u> the worker repaired <u>the leak</u> with __* versus *The builder wondered <u>whether</u> the worker repaired <u>the leak</u> with some tape*). They found no evidence of syntactic adaptation in either group. However, the native English group also did not demonstrate the expected filled gap effect; their reading times were not reliably slower on the direct object region (*the leak*) in the filler-gap dependency conditions. Because this group did not demonstrate the underlying filled gap effect, it is difficult to draw any conclusions about the lack of adaptation effects. Thus, our first experiment examined whether adaptation of gap positions occurs when participants were presented filler-gap dependencies in relative clause structures.

Additionally, we attempted to address the open question about generalization of the learning associated with syntactic adaptation. In their Experiment 1, Harrington Stack et al. (2018) did not observe transfer effects across contexts when context was defined by experimental testing rooms, but alternative operational definitions of context are possible. In Experiment 2, we explored whether a more naturalistic environment for language exposure—the reading of passages of text—leads to transfer of learning.

2. Experiment 1: Blocked Adaptation

Experiment 1 was an eye tracking during reading experiment utilizing a between-subject blocked design adopted from Fine et al.'s (2013) second experiment to examine whether real-time filler-gap dependency processing procedures adapt to the statistics of the input. As described above, one of the hallmarks of syntactic adaptation is decreased difficulty processing an a priori less preferred structure, so the distribution of relevant syntactic structures is crucial. Thus, filler-gap dependency processing is only a candidate for syntactic adaptation effects if the relative frequency of gap positions matches comprehenders' initial preference to associate the filler with a direct object gap. A probabilistic account of active gap filling would require direct object gaps to be more frequent than (at least) other post-verbal gaps like prepositional object gaps. In order to confirm this intuition, a brief corpus analysis was conducted.

2.1. Gap Position Corpus Analysis

Two naturalistic corpora of adult spoken language were examined: the CallHome corpus (Kingsbury et al. 1997) and a selection from the Switchboard corpus (Marcus et al. 1999). These corpora were chosen because they consist of naturalistic, conversational speech between two participants. Filler-gap dependencies are more likely to occur in a conversational environment than in non-conversational spoken or written corpora. The search was limited to syntactically parsed files, which resulted in the examination of 120 CallHome files (28,967 lines of speech) and 199 Switchboard files (44,696 lines of speech). The Tregex utility (Levy and Andrew 2006) was used to extract the parsed trees for sentences that utilize *which* or *that* as the relative clause head. Table 1 presents the results of this corpus analysis (see also Experiment 1 in Atkinson et al. 2018).

Table 1. Distribution of gap positions in relative clauses.

Corpus	Subject Gap	DO Gap	PO Gap	Total
CallHome	126	301	76	503
Switchboard	642	412	111	1165
Overall	768 (46.0%)	713 (42.8%)	187 (11.2%)	1668

Subject and direct object gaps accounted for approximately 90% of gaps within relative clauses. Subject gaps were more frequent than direct object gaps (46.0% vs. 42.8%), and this slight preference for subject gaps is consistent with the finding that subject relative clauses are easier to process than direct object relative clauses (e.g., Gibson 1998, 2000; Grodner et al. 2002; Levy et al. 2013; Staub 2010). Focusing on post-verbal gap positions (i.e., DO and PO gaps), which are the critical ones for active gap filling, DO gaps were more frequent in relative clauses than prepositional object gaps (79.2% vs. 20.8%). Thus, the current experiments are justified in assuming that the input favors DO gaps in relative clauses (compared to prepositional object gaps). Given this distribution, direct object gaps are a priori highly probable in the syntactic structures used in this experiment. This experiment attempted to alter this parsing bias by exposing participants to input skewed toward prepositional object gaps.

2.2. Eye Tracking Materials and Methods

2.2.1. Participants

Sixty native English speakers from the Johns Hopkins University community were randomly assigned to one of three exposure groups and were paid USD 10 or received course credit. The local Institutional Review Board (IRB) reviewed and approved the study.

2.2.2. Design and Materials

The design of this study is a two block version of Fine et al.'s (2013) Experiment 2. Table 2 describes the blocked design of Experiment 1, and Table 3 provides examples of the sentence types used in this experiment. The critical items for all three experiments presented here are available in Supplement S1.

The first block consisted of 24 sentences and was used to manipulate the distribution of gaps. There were three groups of participants defined by the input they received in this exposure block (see the Exposure Block section of Table 3): (1) the DO gap exposure group read 24 sentences with direct object gaps, which matched participants' a priori distribution of gap positions; (2) the PO gap exposure group read 24 sentences with PO gaps, which increased the probability of these gaps; and (3) the filler exposure group received neutral input (i.e., 24 sentences that did not contain a filler-gap dependency). Sentences in the

DO gap exposure block contained a direct object gap followed by a prepositional phrase. Sentences in the PO gap exposure block contained an overt direct object and a prepositional object gap. The fillers read by the filler exposure group were of a comparable length and complexity as the sentences read by the exposure groups.

Table 2. Design of Experiment 1.

Exposure Type	Exposure Block	Experimental Block
DO gap exposure	24 direct object gap sentences	24 target pairs + 48 fillers
PO gap exposure	24 prepositional object gap sentences	
Filler exposure	24 fillers	

Table 3. Examples of sentence types used in the exposure and experimental phases of Experiment 1.

	Exposure Block
3a. *DO gap exposure*	The book that the famous non-fiction author wrote __ about the adventure was named for an explorer.
3b. *PO gap exposure*	The book that the famous non-fiction author wrote the interesting article about __ was named for an explorer.
3c. *Filler exposure*	The scientist investigated how car engines would need to be modified to use the new fuel.
	Experimental Block
4a. *NP-fronting*	The suitcase **that** the stealthy, wanted thief stole the precious jewels from __ was full of sentimental items.
4b. *PP-fronting*	The suitcase **from which** the stealthy, wanted thief stole the precious jewels __ was full of sentimental items.

The second block was identical for all groups and consisted of 24 target sentences randomly interspersed with 48 fillers. The target sentences in the experimental block consisted of filled gap sentences (4a), labeled as NP-fronting in the Experimental Block section of Table 3. These sentences had prepositional object gaps, which were preceded by a direct object NP, and they were identical in structure to the sentences that the PO gap exposure group read in the exposure block (e.g., 3b). In the experimental block, the NP-fronting target sentences were compared to PP-fronting sentences in which the preposition was pied-piped with the *wh*-phrase (4b), which are incompatible with a direct object gap (Lee 2004; Wagers and Phillips 2014). If the parser actively completes the dependency, reading time should increase on the direct object of the NP-fronting condition compared to the PP-fronting condition because the presence of a direct object is not compatible with the direct object gap prediction that is generated in the NP-fronting condition. Such processing difficulty should not be found in the PP-fronting condition because it unambiguously indicates a prepositional object gap.

In both blocks, each sentence was followed by a *yes/no* comprehension question. The number of *yes* and *no* answers was balanced across the exposure block, the targets, and the experiment as a whole. Two experimental lists were created by crossing fronted phrase type (NP-fronting vs. PP-fronting). These lists, in turn, were crossed with exposure type, for a total of six lists.

2.2.3. Procedure

Eye movements were recorded using an EyeLink 1000 eye tracker (SR Research: Mississauga, ON, Canada). Participants' heads were stabilized on a chin rest and a forehead rest. Only participants' right eye was tracked at a sampling rate of 1000 Hz. The display allowed a maximum of 120 characters per line in 10 point Monaco font. Stimuli were displayed on a 26 inch monitor, and participants sat 70 cm away from the display.

Before the experiment began, participants received instructions. A nine-point calibration routine was performed at the beginning of the experiment and was monitored throughout with automatic drift checks preceding each trial. Participants were recalibrated as necessary. The experiment began with written instructions on the display and five practice items. At the beginning of each trial, a black circle was displayed on the left side of the monitor and was aligned with the beginning of the sentence. The sentence text was displayed after the participant fixated on the circle. After reading the sentence, participants pressed a button on a game controller to remove the sentence from the display and to trigger a *yes/no* comprehension question. These questions were answered by pressing the left (*yes*) or right (*no*) trigger buttons. Comprehension questions never concerned the critical filler-gap dependency portion of the sentence. In total, the experiment lasted approximately 40 min.

2.2.4. Analysis

Vertical drift in the positions of fixations was hand corrected. Fixations shorter than 80 ms were either merged with contiguous fixations within one character or deleted because readers are unable to extract much information from these very short fixations (Rayner et al. 2012). Additionally, fixations longer than 800 ms were removed because they usually result from tracker losses or other atypical events (Rayner et al. 2012).

For the purposes of analysis, the target sentences were divided into regions, as shown in Table 4. Four common eye tracking measures were analyzed (Rayner 1998)—first fixation duration, first pass reading time, regression path time, and percent regressions—in two main regions of interest: the verb region (region 4) and the critical filled gap region (region 5), i.e., the direct object. While it is common for the region following the critical region to be analyzed, i.e., the spillover region, this region was not identical across the two conditions (region 6 in the NP-fronting condition versus region 7 in the PP-fronting condition). Given this difference, it is difficult to draw conclusions about any findings in this region. Thus, the data from the spillover region were not considered in the main analysis, but are instead presented in Appendix A.

Table 4. Sample materials and analysis regions for the target sentences in Experiment 1.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
NP-fronting	The suitcase	that	the stealthy, wanted thief	stole	the precious jewels	from	was	full of sentimental items
PP-fronting	The suitcase	from which	the stealthy, wanted thief	stole	the precious jewels		was	full of sentimental items

First fixation duration is the length of the first fixation on a particular region, no matter the number of words in that region. First pass reading time is the sum of all fixations on the region before exiting the region to the right or to the left. Regression path time is the length of all fixations on the region and any regions earlier in the sentence, and it includes the time spent rereading previous regions in the sentence. Percent regressions is the percentage of trials in which a regression to a previous region occurred.

Statistical analyses differed between the reading time measures (i.e., first fixation, first pass, and regression path duration) and the percent regressions measure. For the reading time data, the logged data for each region and measure pair were fit to a linear mixed effect model with the exposure group and fronted phrase type as fixed effects and participants and items as random effects (Baayen et al. 2008). Trial number was centered and also included as a fixed effect. The exposure group factor was treatment coded with the filler exposure group serving as the baseline. Fronted phrase type was sum coded with NP-fronting coded positively and PP-fronting coded negatively. A logistic mixed effects model was fit for the percent regressions analyses (Jaeger 2008). Maximum random effects including random slopes were utilized when the models converged (Barr et al. 2013). When the maximal models did not converge, a set procedure was used to minimize the models until the models converged: (1) interactions in the random slopes were removed, (2) random intercepts and slopes were decorrelated, (3) slope effects were removed, and (4) random intercepts for items were removed. In all cases, these simplifications were applied first to the item random effect and then to the participant one. These models were run in the R environment (R Core Development Team 2022) using the lme4 package (Bates et al. 2015). *p*-value estimates for the fixed and random effects in the linear models were calculated using the Sattherwaite approximation in the lmerTest package (Kuznetsova et al. 2015). Additional planned pairwise comparisons within the exposure group were performed by individually fitting linear or logistic mixed effects models to the data from each group, as appropriate, with fronted phrase type and trial number as fixed effects with random participants and items.

In addition to data from the experimental block, we also analyzed reading time data from the PO gap exposure group in the exposure block. This additional analysis was used to confirm that the PO gap exposure sentences indeed caused reading disruption as we intended. Because there were no PP-fronting sentences in the exposure block, the NP-fronting sentences from the exposure block were compared to the PP-fronting sentences from the experimental block. The same eye tracking measures were analyzed for each of the critical regions shown in Table 2. Models included the fronted phrase type as the fixed effect and participants and items as random effects. Raw data and the analysis scripts for this experiment and the following experiments are available through the first author's OSF.

2.3. Results

2.3.1. Exposure Block

In order to draw any conclusions about the data from the experimental block, we must ensure that the group exposed to PO gaps actively associated the filler with the verb before exposure, i.e., that the PO gap exposure group was initially subject to the filled gap effect. While there were no PP-fronting sentences in the exposure block, the filled gap effect in this block can be approximated by comparing the NP-fronting sentences in the exposure block to the PP-fronting sentences in the experimental block. For each reading time measure (i.e., first fixation, first pass, and regression path duration) and each region, individual reading times longer than three standard deviations greater than the mean were excluded. This resulted in 5.2% of the data being excluded, which were approximately evenly distributed across measure, region, exposure group, and fronting-type. Table 5 presents the data relevant for this analysis from the PO gap exposure group, and Table 6 summarizes the analysis.

Table 5. PO gap exposure group participant mean reading times in milliseconds (standard error) and percent regressions.

	Verb Region	**Filled Gap Region**
First fixation duration		
NP-fronting	238 (8)	229 (6)
PP-fronting	242 (5)	224 (4)
First pass time		
NP-fronting	286 (15)	497 (21)
PP-fronting	270 (8)	444 (13)
Regression path time		
NP-fronting	323 (23)	721 (50)
PP-fronting	306 (10)	665 (35)
Percent regressions		
NP-fronting	8.6 (2.1)	29.2 (4.5)
PP-fronting	10.0 (1.9)	25.9 (2.4)

Table 6. Experiment 1 PO gap exposure group fixed effects summary for the four eye tracking measures and the two regions of analysis: the verb region and the filled gap region.

	Verb Region		**Filled Gap Region**	
	β	***SE***	***β***	***SE***
First fixation duration				
Intercept	5.43 ***	0.03	5.37 ***	0.03
Fronting type	−0.04	0.04	0.02	0.03
First pass time				
Intercept	5.54 ***	0.05	6.02 ***	0.05
Fronting type	0.02	0.05	0.09 *	0.04
Regression path time				
Intercept	5.63 ***	0.06	6.36 ***	0.08
Fronting type	0.03	0.05	0.11 †	0.06
Percent regressions				
Intercept	−2.66 ***	0.31	−1.16 ***	0.24
Fronting type	−0.04	0.43	0.21	0.20

Note: † $p < 0.1$, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

In the verb region, there were no significant effects for any of the reading time measures or for percent regressions (all *p*s > 0.1). The expected filled gap effect was observed in the critical filled gap region. In this region, first pass time was significantly longer in the NP-fronting condition compared to the PP-fronting condition ($\beta = 0.09$, $SE = 0.04$, $p < 0.05$). Regression path time was also marginally longer in the NP-fronting condition ($\beta = 0.11$, $SE = 0.06$, $p = 0.09$). There were no significant effects on first fixation duration ($\beta = 0.02$, $SE = 0.03$, $p > 0.1$) or percent regressions ($\beta = 0.12$, $SE = 0.24$, $p > 0.1$).

The results from the exposure block indicate that the PO gap exposure group demonstrated the filled gap effect prior to exposure. Thus, we can consider whether exposure to PO gaps affected the appearance of this effect by analyzing the experimental block.

2.3.2. Experimental Block

Comprehension accuracy for all three groups on the target trials was 95.1%, and trials during which participants answered the comprehension question incorrectly were removed from the analysis, as these trials likely reflect distracted reading.

For each reading time measure (i.e., first fixation, first pass, and regression path duration) and each region, individual reading times longer than three standard deviations greater than the mean were excluded. This resulted in 4.9% of the data being excluded. Table 7 presents the participant means on each measure for each region of analysis as well as the standard errors of the participant means.

Table 8 presents a summary of the statistical analyses. In the verb region, the only significant effect was of trial number on the percent regressions ($\beta = -0.01$, $SE = 0.01$, $p < 0.05$). As the experiment continued, participants became less likely to make a regression from this region.

In the critical filled gap region, there were significant effects in both regression path time and percent regressions. There was a marginal effect of trial number ($\beta = -0.002$, $SE = 0.001$, $p = 0.07$) and a significant effect of fronting type ($\beta = 0.20$, $SE = 0.05$, $p < 0.001$) on regression path times. In the filler exposure group (the reference level), participants' regression path times decreased later in the experiment. The effect of fronting type is indicative of the fact that the filled gap effect holds for the filler exposure group; regression path times were longer in the NP-fronting condition when a direct object gap was possible and predicted. Additionally, there was a significant interaction of exposure group and fronting type for the PO gap exposure group only ($\beta = -0.14$, $SE = 0.06$, $p < 0.05$). For clarity, this interaction was further explored with planned pairwise comparisons. As demonstrated in the main analysis, regression path times in the filled gap region were longer when the NP was fronted for the filler exposure group ($\beta = 0.20$, $SE = 0.05$, $p < 0.001$). This was a marginal effect for the direct object exposed group (β = 0.06, SE = 0.05, p = 0.06). Crucially, the filled gap effect was not demonstrated by the PO gap exposure group, as there was no difference between the two fronting types for this group (β = 0.06, SE = 0.05, $p > 0.1$).

For percent regressions in the filled gap region, the filled gap effect in the filler exposure group was demonstrated by the significant effect of fronting type (β = 0.81, SE = 0.25, $p < 0.01$). There was also a significant interaction of the exposure group and fronting type for the PO gap exposure group (β = −0.80, SE = 0.34, $p < 0.05$). Again, this interaction was explored with planned pairwise comparisons. Yet again, the main model demonstrated that there were significantly more trials with regressions for the NP-fronting condition than the PP-fronting condition for the filler exposure group, and the same was true for the object gap exposure group (β = 0.65, SE = 0.30, $p < 0.05$). In contrast, there was no significant difference in percent regressions based on fronting type for the PO gap exposure group (β = −0.12, SE = 0.27, $p > 0.1$). These differences between the control groups and the PO gap exposure group suggest that active gap filling was diminished after concentrated exposure to PO gaps.

Table 7. Experiment 1 participant mean reading times in milliseconds (standard error) and percent regressions.

	Verb Region	**Filled Gap Region**
First fixation duration		
Filler exposure, NP-fronting	234 (5)	223 (5)
Filler exposure, PP-fronting	230 (5)	224 (6)
DO gap exposure, NP-fronting	235 (4)	237 (5)
DO gap exposure, PP-fronting	243 (5)	233 (3)
PO gap exposure, NP-fronting	241 (5)	229 (5)
PO gap exposure, PP-fronting	242 (5)	224 (4)
First pass time		
Filler exposure, NP-fronting	274 (7)	539 (21)
Filler exposure, PP-fronting	267 (7)	491 (18)
DO gap exposure, NP-fronting	262 (6)	486 (16)
DO gap exposure, PP-fronting	273 (7)	486 (14)
PO gap exposure, NP-fronting	267 (8)	470 (15)
PO gap exposure, PP-fronting	270 (8)	444 (13)
Regression path time		
Filler exposure, NP-fronting	305 (8)	747 (30)
Filler exposure, PP-fronting	327 (15)	625 (25)
DO gap exposure, NP-fronting	291 (8)	686 (27)
DO gap exposure, PP-fronting	299 (8)	610 (29)
PO gap exposure, NP-fronting	303 (11)	680 (30)
PO gap exposure, PP-fronting	306 (10)	665 (35)
Percent regressions		
Filler exposure, NP-fronting	8.2 (1.3)	26.0 (2.3)
Filler exposure, PP-fronting	10.7 (2.0)	14.7 (1.6)
DO gap exposure, NP-fronting	9.8 (1.3)	28.2 (3.1)
DO gap exposure, PP-fronting	9.7 (1.7)	15.6 (1.6)
PO gap exposure, NP-fronting	9.0 (1.3)	25.6 (2.9)
PO gap exposure, PP-fronting	10.0 (1.9)	25.9 (2.4)

Because these findings rely on null results, we calculated Bayes factors (BFs) for the pairwise comparisons for the regression path data for all of the exposure groups to attempt to provide some additional evidence in favor of the null result for the PO gap exposure group. Bayes factors are the ratio of the likelihood of the alternate hypothesis to the likelihood of the null; a BF of 1 indicates no evidence for either hypothesis because the likelihoods are the same, a BF less than 1 indicates evidence for the null, and a BF greater than 1 indicates evidence for the alternative hypothesis. All BF analyses of linear mixed effect models (e.g., regression path duration) were conducted using the BayesFactor package (Morey and Rouder 2018) set to 100,000 iterations. Unfortunately, there is not, as of yet, an R package that can easily calculate Bayes factors for mixed effect logistic regressions, so BFs were not obtained for the percent regression data.

Table 8. Experiment 1 experimental block fixed effects summary for the five eye tracking measures and three regions of analysis: the verb region, the filled gap region, and the spillover region.

	Verb Region		**Filled Gap Region**	
	β	***SE***	***β***	***SE***
First fixation duration				
DO gap exposure	0.02	0.04	0.06	0.04
PO gap exposure	0.04	0.04	0.02	0.04
Fronting type	−0.01	0.03	−0.004	0.03
Trial	−0.01	0.001	−0.0002	0.0008
DO gap × fronting	−0.04	0.05	0.0005	0.04
PO gap × fronting	−0.001	0.05	0.03	0.04
DO gap × trial	0.003 †	0.001	0.001	0.001
PO gap × trial	0.001	0.001	−0.001	0.001
First pass time				
DO gap exposure	−0.02	0.05	−0.04	0.08
PO gap exposure	0.003	0.05	−0.10	0.08
Fronting type	0.004	0.04	0.08	0.05
Trial	0.0002	0.001	−0.001	0.001
DO gap × fronting	−0.05	0.05	−0.11	0.07
PO gap × fronting	−0.02	0.05	−0.03	0.07
DO gap ×trial	0.001	0.002	0.001	0.002
PO gap ×trial	−0.001	0.002	−0.002	0.002
Regression path time				
DO gap exposure	−0.05	0.06	−0.06	0.10
PO gap exposure	−0.003	0.06	−0.06	0.10
Fronting type	−0.02	0.04	0.20 ***	0.05
Trial	−0.002	0.001	−0.002 †	0.001
DO gap × fronting	0.004	0.06	−0.09	0.06
PO gap × fronting	0.02	0.06	−0.14 *	0.06
DO gap × trial	0.002	0.002	0.002	0.002
PO gap × trial	0.002	0.002	0.001	0.002
Percent regressions				
DO gap exposure	0.25	0.27	0.02	0.33
PO gap exposure	0.21	0.27	0.34	0.32
Fronting type	−0.06	0.38	0.81 **	0.25
Trial	−0.03 *	0.01	−0.01	0.01
DO gap × fronting	0.10	0.52	0.04	0.35
PO gap × fronting	0.16	0.53	−0.80 *	0.34
DO gap × trial	0.01	0.02	−0.006	0.01
PO gap × trial	0.02	0.02	0.01	0.01

Note: † $p < 0.1$, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

For the regression path data, the full pairwise model for each exposure group (containing fronting type and trial number as factors) was compared to a null model without the fronting type factor. This resulted in BFs providing various levels of evidence for the alternative hypothesis for the filler and DO exposure groups (filler exposure: very strong evidence, 32.44 ± 59.03%; DO gap exposure: anecdotal evidence, 2.39 ± 0.63%). For the PO exposure group, the BF of 0.09 ± 35.09% indicates strong evidence for the null model. These Bayes factors provide evidence in favor of the null result for the PO exposure group; participants in this condition did not demonstrate the filled gap effect in the experimental phase.

2.4. Discussion

Experiment 1 utilized the blocked adaptation paradigm of Fine et al. (2013) to examine whether filler-gap dependency processing can be affected by the distribution of gaps in the input. For the control groups, exposure to fillers and DO gaps did not affect active gap filling; regression path times and percent regressions on the critical filled gap region were significantly greater for NP-fronting sentences than for PP-fronting sentences. For the PO gap exposure group, however, there was no difference between the two types of sentences, indicating that the PO gap exposure group was not actively associating the filler with the verb during the second, experimental block. This difference cannot be attributed to an inherent difference in active gap filling between the groups; participants in this group had longer first pass and regression path times in the NP-fronting, PO gap sentences from the exposure block compared to the PP-fronting condition from the experimental block. Initially, they were actively associating the filler with the verb and were surprised by the presence of a direct object NP. Taken together, these results suggest that syntactic adaptation effects extend to filler-gap dependency processing (cf. Kaan et al. 2019). Additionally, they provide evidence that probabilistic information from the input can override memory biases that favor shorter dependencies.

3. Experiment 2: Adaptation with Masked Input

Experiment 1 demonstrated that probabilistic information can affect gap predictions, but it is still unclear whether the adaptation of syntactic predictions generalizes outside the laboratory and into naturalistic language processing. The skewed distribution of PO gaps in the blocked design of Experiment 1 may have led participants to develop task-specific expectations about upcoming input. In other words, the input was presented in a context (i.e., as part of a sentence processing experiment) where participants may expect manipulations of this sort. These experimental factors suggest that the results from Experiment 1 and those from other studies demonstrating successful syntactic adaptation may not reflect an effect of language experience on online processing strategies. Rather, they may indicate an effect of distributional information on the likelihood of task-specific strategies. Exposure to less probable input may lead participants to generate more task-specific parsing strategies because the experimental environment does not reflect the distribution of structures outside the laboratory. Experiment 2 examined whether gap distributions were generalized across tasks by masking the exposure block as a separate task.

3.1. Experiment 2A: Filled Gap Effect

Experiment 2A tested the generalizability of the input statistics by replicating the basic design of Experiment 1, but replacing the exposure block with a sentence recognition task. In this task, the critical input sentences were presented within a story. As story reading is a more naturalistic learning environment than the blocked exposure task, evidence for adaptation of gap predictions in this study would suggest that the syntactic adaptation effects reported above and in other work were not the result of task-specific strategies.

3.1.1. Materials and Method

Participants

Forty-eight native English speakers from the University of Washington community were paid USD 15 for their participation. One additional participant completed the experiment but was excluded for low accuracy (4%) on the comprehension questions in the experimental block. The local Institutional Review Board (IRB) reviewed and approved the study.

Design and Materials

Experiment 2A was presented to participants as two unrelated tasks. The first task was a sentence recognition study that served as the exposure phase comparable to the first block in Experiment 1 and masked the presentation of the skewed input. The second task was an eye tracking study utilizing the filled gap effect to examine participants' gap predictions after exposure.

Sentence recognition. The materials for the sentence recognition study consisted of twelve short stories. The stories contained the sentences that manipulated the input distribution of gap positions. Participants were divided into two groups: a prepositional object (PO) gap exposure group and a direct object (DO) gap exposure group. Each story contained four critical sentences, for a total of 48 input filler-gap dependencies. An example PO gap story is presented in (5); the critical filled gap sentences are bolded.

(5) Jill and Justin planned to spend a day exploring New York City. Over the past few weeks, they had been reading all the information they could find about things to do there. **The newspaper article that their friend wrote the blog post about __ gave great tips about the most popular attractions in the city.** They decided that they definitely wanted to go shopping in Times Square and that, in the evening, they would see a Broadway play. They left on the train the next morning. After they arrived in New York, they made their way to Times Square. **The shops that they encountered the crowds in __ were enormous.** They looked around for a while, but decided not to buy anything so that they would not have to carry bags with them the rest of the day. After all of their time in the crowds, Jill and Justin were exhausted and they decided to find a place where they could eat lunch. **The deli's menu that Jill discovered the delicious sandwich on __ was much more expensive than she expected.** The couple decided to splurge, though, since it was their first time in the city. Then, they walked through Central Park until it was time for them to take their seats for the show. **The musical that the couple watched the famous actress in __ made them want to come back and see a Broadway performance again.** Jill and Justin were sad to leave after such an exciting day.

The direct object gap versions of these critical sentences are given in (6). They were created by simply fronting the direct object from the critical PO gap sentences rather than the prepositional object.

(6) a. The blog post that their friend wrote __ about the newspaper article gave great tips about the most popular attractions in the city. b. The crowds that they encountered __ in the shops were enormous. c. The delicious sandwich that Jill discovered __ on the deli's menu was much more expensive than she expected. d. The famous actress that the couple watched __ in the musical made them want to come back and see a Broadway performance again.

The stories were paired with sentences either exactly duplicating ones from the story (7a) or altered by a single word (7b). These small changes were all of content words, so the task should have been challenging but not impossible if the participants were fully comprehending the sentences in the stories. None of these sentences involved the critical direct object or prepositional object gap sentences. The participants' task was to identify whether these sentences appeared in the story that they just read. Each story was paired with two sentences for a total of 24 sentence recognition trials.

(7) a. Jill and Justin were sad to leave after such a tiring day in the city.
b. Jill and Justin were happy to leave after such a tiring day in the city.

Eye tracking. The second study was an eye tracking experiment, identical to the second block in Experiment 1. Participants in both exposure groups read 24 target filled gap sentences (4) and 48 fillers. Each sentence was followed by a *yes/no* comprehension question. If the parser is actively predicting a direct object gap, there should be a reading time slowdown on the filled gap region in the NP-fronting condition because the predicted direct object gap is impossible given the overt object noun phrase.

For the eye tracking portion of the experiment, two experimental lists were generated by crossing fronting type. These lists were also crossed with exposure group (PO gaps vs. DO gaps) for a total of four lists. If the input distribution of gap positions transfers from the sentence recognition experiment to the eye tracking experiment, the results should be similar to those of Experiment 1: the group exposed to PO gaps should not demonstrate the filled gap effect because their active association of the filler with the verb is weakened. Thus, they should not be surprised when they encounter an overt object NP. Alternatively, participants may treat input distributions as local and thus only apply them within an experiment. Were this the case, the PO gap exposure group should not differ from the DO gap exposure group. Both groups should actively predict a direct object gap and should, therefore, have slower reading times on the filled gap region in the NP-fronting condition.

Procedure

Sentence recognition. This portion of the experiment was presented using the Ibex online experiment platform (Drummond 2010), which allows the stories and sentences to be displayed on a browser. It was presented on a computer in the same room as the eye tracker. Participants were instructed to read the stories aloud to verify that they were reading all of the critical sentences as this was not otherwise monitored. After each story, a sentence was displayed on the screen, and participants were asked to identify two sentences as novel or duplicates from the previous story.

Eye tracking. After completing the sentence recognition portion of the experiment, participants shifted to the eye tracker. Thus, there was a brief break between the exposure task and the experimental task that was not present in Experiment 1, but we attempted to keep this transition time to a minimum. The eye tracking procedure was identical to that from Experiment 1.

Analysis

Sentence recognition. Because the underlying purpose of the sentence recognition task was to present skewed distributional information on gap positions, it is important to analyze accuracy on this task to make sure that participants were paying attention. Accuracy in recognizing sentences was analyzed using a logit mixed effect model (Jaeger 2008) with the story exposure group as the fixed effect and random intercepts for participants and items (i.e., stories).

Eye tracking. The analysis procedure for the eye tracking portion was identical to that from Experiment 1. The one exception is that there were only two exposure groups in this experiment, so they were sum coded such that PO gap exposure was positive while DO gap exposure was negative.

3.1.2. Results

Sentence Recognition

The sentence recognition experiment was difficult because the sentences that were not from the story differed minimally from the actual sentences on which they were based; thus, accuracies around 70% were to be expected. Table 9 presents the accuracy by story exposure group.

Table 9. Percent accurate recognition and standard error by story exposure group.

Story Exposure Group	Accuracy	*SE*
DO gap sentences	70.3%	1.1%
PO gap sentences	67.9%	1.5%

There was no significant accuracy difference between the two groups ($\beta = -0.11$, $SE = 0.15$, $p > 0.1$). The fact that participants successfully identified the sentences almost three-quarters of the time and did not differ based on exposure group suggests that both groups were reading the stories carefully enough to identify fairly minimal changes.

Eye Tracking

Comprehension accuracy for the target trials was 94.5%. Trials during which participants answered the comprehension question incorrectly were removed from the analysis. For each reading time measure (i.e., first fixation, first pass, and regression path duration) and each region, individual reading times longer than three standard deviations greater than the mean were excluded. This resulted in 4.3% of the data being excluded, which were approximately evenly distributed across measure, region, exposure group, and fronting type. Table 10 presents the participant means on each measure for each region of analysis as well as the standard errors of the participant means.

Table 10. Experiment 2A participant mean reading times in milliseconds (standard error) and percent regressions.

	Verb Region	Filled Gap Region
First fixation duration		
DO gap exposure, NP-fronting	255 (8)	238 (5)
DO gap exposure, PP-fronting	261 (8)	235 (6)
PO gap exposure, NP-fronting	250 (5)	236 (7)
PO gap exposure, PP-fronting	244 (7)	240 (6)
First pass time		
DO gap exposure, NP-fronting	309 (10)	533 (22)
DO gap exposure, PP-fronting	296 (22)	514 (23)
PO gap exposure, NP-fronting	296 (8)	536 (21)
PO gap exposure, PP-fronting	292 (9)	532 (17)
Regression path time		
DO gap exposure, NP-fronting	388 (19)	781 (43)
DO gap exposure, PP-fronting	356 (19)	665 (35)
PO gap exposure, NP-fronting	352 (15)	732 (26)
PO gap exposure, PP-fronting	353 (13)	645 (18)
Percent regressions		
DO gap exposure, NP-fronting	15.1 (2.9)	29.5 (3.2)
DO gap exposure, PP-fronting	8.2 (1.7)	21.9 (2.6)
PO gap exposure, NP-fronting	9.3 (1.4)	26.5 (3.0)
PO gap exposure, PP-fronting	11.3 (1.4)	14.9 (2.0)

Table 11 presents a summary of the statistical analyses. In the verb region, there were no statistically significant effects on first fixation duration or first pass time (all *p*s > 0.1). On regression path times, there was only a significant effect of trial number ($\beta = -0.002$, $SE = 0.001$, $p < 0.05$). Regression path times decreased as the experiment continued. Regressions from the verb region also significantly decreased as trial number increased ($\beta = -0.01$, $SE = 0.01$, $p < 0.05$). For percent regressions, while the effects of the exposure group and fronting type were not significant (exposure group: $\beta = -0.10$, $SE = 0.21$, $p > 0.1$; fronting type: $\beta = 0.15$, $SE = 0.21$, $p > 0.1$), the interaction of these factors was marginally significant ($\beta = -0.78$, $SE = 0.43$, $p = 0.07$). However, pairwise comparisons revealed that there was no significant effect of fronting type ($\beta = 0.67$, $SE = 0.48$, $p > 0.1$) or trial number ($\beta = -0.01$, $SE = 0.01$, $p > 0.1$) on percent regressions for those exposed to object gaps during the story phase. Similarly, there was no significant effect of fronting type ($\beta = -0.28$, $SE = 0.35$, $p > 0.1$) for the PO gap exposure group, though regressions did significantly decrease as trial number increased ($\beta = -0.02$, $SE = 0.01$, $p < 0.5$). Thus, neither exposure group demonstrated a filled gap effect in this verb region.

In the critical filled gap region, there was a significant effect of fronting type ($\beta = 0.13$, $SE = 0.02$, $p < 0.001$) and a marginal effect of trial number ($\beta = -0.001$, $SE = 0.001$, $p = 0.07$) on regression path time. Regression path time was longer in the temporarily ambiguous NP-fronting sentences than in the non-ambiguous PP-fronting sentences. Also, as in other regions and with other measures, regression path times marginally decreased as trial number increased. Finally, there were significantly more regressions from this critical region in the NP-fronting condition ($\beta = 0.58$, $SE = 0.21$, $p < 0.01$). Both exposure groups demonstrated the filled gap effect on the critical filled gap region.

Table 11. Experiment 2A fixed effects summary for the five eye tracking measures and two regions of analysis: the verb region and the filled gap region.

	Verb Region		Filled Gap Region	
	β	SE	β	SE
First fixation duration				
Exposure group	−0.04	0.05	0.02	0.04
Fronting type	−0.01	0.03	−0.003	0.02
Trial	−0.0005	0.001	0.0004	0.001
Exposure × fronting	0.06	0.05	−0.04	0.05
First pass time				
Exposure group	−0.03	0.05	0.06	0.06
Fronting type	0.01	0.03	0.01	0.04
Trial	−0.001	0.001	−0.00001	0.001
Exposure × fronting	−0.03	0.06	−0.07	0.07
Regression path time				
Exposure group	−0.05	0.07	−0.01	0.07
Fronting type	0.02	0.04	0.13 ***	0.02
Trial	−0.002 *	0.001	−0.001 †	0.001
Exposure × fronting	−0.08	0.08	−0.02	0.05
Percent regressions[2]				
Exposure group	−0.10	0.21	−0.31	0.26
Fronting type	0.15	0.21	0.62 **	0.20
Trial	−0.01 *	0.01	−0.01	0.004
Exposure × fronting	−0.78 †	0.43	0.36	0.38

Note: † $p < 0.1$, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

This study failed to find an effect of exposure group, so Bayes factors were calculated to potentially provide evidence for this null effect (and for consistency across experiments). The full overall model was compared to a null model without the exposure group factor (and its interaction with the fronting type factor). For the regression path data, the resulting BF of $0.05 \pm 0.32\%$ indicated strong evidence for the null hypothesis. In other words, exposure during the sentence recognition portion of the experiment did not alter participants' gap predictions.

3.1.3. Discussion

Experiment 2A examined whether the syntactic adaptation effects observed in Experiment 1 could be induced by a more naturalistic exposure environment. In Experiment 1, the PO gap sentences were presented in a single block and may have led participants to generate an experiment-specific expectation about the types of sentences that would appear later in the experiment. Experiment 2A suggests that this is a plausible explanation for the diminished active gap filling in Experiment 1. When the exposure block was disguised as a separate task (with separate goals), the group exposed to PO gap sentences no longer demonstrated decreased active gap filling. Both exposure groups had more and longer regressions from the filled gap region in the NP-fronting conditions. This suggests that both groups were actively associating the filler with the verb and were surprised when the overt NP blocked the direct object gap interpretation.

Unlike Experiment 1, there was no measure of the strength of the filled gap effect prior to exposure. This was not a problem for interpreting these results for several reasons. First, in contrast with Experiment 1, the PO gap group demonstrated the filled gap effect in the eye tracking task. It is unlikely that exposure to PO gaps would lead to the generation of filled gap effects if they did not previously exist in this population. Thus, this would be more of a concern if they were not demonstrating the effect. Second, there was anecdotal evidence that the PO gap exposure group were in fact demonstrating the filled gap effect during the sentence recognition task. Specifically, many of the participants produced disfluencies (e.g., pausing, repeating, non-fluent speech) when reading these critical sentences, specifically when reading the filled direct object region. Unfortunately, participants were not recorded in this task, so these features cannot be analyzed in any more quantitative depth.

3.2. Experiment 2B: Plausibility Mismatch Effect

If syntactic adaptation effects are general, they should not only transfer across tasks but also across the types of test sentences. Thus, Experiment 2B attempted to test for generalizability in two ways. First, as in Experiment 2A, the exposure phase was masked as a separate task from the eye tracking task used to measure adaptation. Second, the structures used to measure active gap filling in the exposure and experimental blocks of both Experiments 1 and 2A were identical; participants were exposed to filled gap sentences, and their adaptation was tested with filled gap sentences. In Experiment 2B, participants were exposed to filled gap sentences (as in the previous experiments) but were tested on a different measure of active gap filling: the *plausibility mismatch effect* (Traxler and Pickering 1996). If participants were truly adapting to the abstract structural representation of the gap position, utilizing different measures of active gap filling in the exposure and experimental phases should not have affected the likelihood of adaptation.

3.2.1. Materials and Methods

Participants

Forty-seven native English-speaking Johns Hopkins University undergraduates participated for course credit. The local Institutional Review Board (IRB) reviewed and approved the study.

Design and Materials

As in Experiment 2A, Experiment 2B was presented to participants as two tasks: a sentence recognition task and an eye tracking task. However, the eye tracking portion of this experiment utilized the plausibility mismatch effect to examine participants' gap predictions after exposure rather than the filled gap effect (as in Experiments 1 and 2A).

Sentence recognition. The sentence recognition portion of Experiment 2B was identical to the sentence recognition study in Experiment 2A.

Eye tracking. The eye tracking portion of Experiment 2B was similar to the eye tracking portion of Experiment 2A and the second block of Experiment 1. Unlike the previous adaptation experiments, active gap filling was tested using the plausibility mismatch effect (Traxler and Pickering 1996). The target sentences included fillers that are plausible direct objects of the verb, e.g., *wrote* and *the book* (8a), or that are implausible objects, e.g., *wrote* and *the city* (8b).

(8) a. *Plausibility match:* The book that the author wrote thoughtfully about __ was named for an explorer.
b. *Plausibility mismatch:* The city that the author wrote thoughtfully about __ was named for an explorer.

Participants in both exposure groups read 24 target sentences and 48 fillers; each was followed by a *yes/no* comprehension question. If the parser actively predicted a direct object gap, there should be a reading time slowdown on the verb in the plausibility mismatch condition because a direct object gap interpretation is impossible.

For the eye tracking portion, two experimental lists were generated by crossing plausibility. These lists were also crossed with the exposure group (DO gaps vs. PO gaps) for a total of four lists. As in Experiment 2A, if the input distribution of gap positions transferred from the sentence recognition experiment to the eye tracking experiment, the group exposed to PO gaps should not demonstrate the plausibility mismatch effect (similar to the findings from Experiment 1). Their active association of the filler with the verb should be weakened, and therefore, they should not be surprised when the filler is not a plausible direct object of the verb. Alternatively, participants may treat input distributions as local, as suggested by the results of Experiment 2A, and only apply them within an experiment. Were this the case, the PO gap exposure group should actively predict a direct object gap and, therefore, demonstrate the plausibility mismatch effect by having slower reading times in the plausibility mismatch conditions (as should the DO gap exposure group).

Procedure

The sentence recognition and eye tracking procedures were identical to those from Experiment 2A.

Analysis

Sentence recognition. The sentence recognition analysis was identical to that from Experiment 2A.

Eye tracking. The analysis procedure for the eye tracking portion of the study was identical to that for Experiments 1 and 2A with different regions of analysis (see Table 12).

Table 12. Sample materials and analysis regions for the target sentences in Experiment 2B.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Plausibility match	The book	that	the author	wrote	thoughtfully	about	was	named for an explorer
Plausibility mismatch	The suitcase	that	the author	wrote	thoughtfully	about	was	named for an explorer

Unlike filled gap sentences, the verb is the critical region for the plausibility mismatch effect. Whether or not the filler is a plausible direct object can be evaluated when the argument structure is available, i.e., when the verb is processed. Immediate effects of a semantic mismatch between the filler and the verb were expected. The analysis concentrated on three regions of interest: the pre-verb region (region 3, i.e., the subject of the relative clause), the verb region (region 4), and the spillover region (region 5, i.e., the adverb).

3.2.2. Results

Sentence Recognition

As in Experiment 2A, this task was difficult, so accuracies around 70% were expected. Table 13 presents the accuracy by story exposure group condition.

Table 13. Percent accurate recognition and standard error by story exposure group.

Story Exposure Group	**Accuracy**	***SE***
DO gap sentences	71.2%	6.5%
PO gap sentences	73.8%	6.4%

There was no significant accuracy difference between the two groups ($\beta = 0.14$, $Z = 0.96$, $p > 0.1$). Again, participants successfully identified the sentences almost three-quarters of the time and did not differ based on exposure group, which suggests that both groups were reading the stories carefully enough to identify fairly minimal changes.

Eye Tracking

Comprehension accuracy for the target trials was 94.4%. Trials during which participants answered the comprehension question incorrectly were removed from the analysis, as these trials likely reflected distracted reading.

For each reading time measure (i.e., first fixation, first pass, and regression path duration) and each region, reading times longer than three standard deviations greater than the mean were excluded. This resulted in 4.5% of the data being excluded, which were approximately evenly distributed across measure, region, exposure group, and fronting-type. Table 14 presents the participant means on each measure for each region of analysis as well as the standard errors of the participant means.

Table 14. Experiment 2B participant mean reading times in milliseconds (standard error) and percent regressions.

	Pre-Verb Region	Verb Region	Spillover Region
First fixation duration			
DO gap exposure, plausible	211 (4)	259 (6)	237 (6)
DO gap exposure, implausible	215 (4)	256 (8)	242 (9)
PO gap exposure, plausible	199 (5)	237 (5)	225 (6)
PO gap exposure, implausible	206 (5)	246 (5)	241 (6)
First pass time			
DO gap exposure, plausible	342 (14)	299 (10)	278 (9)
DO gap exposure, implausible	324 (10)	303 (9)	296 (15)
PO gap exposure, plausible	314 (13)	283 (7)	279 (11)
PO gap exposure, implausible	335 (11)	289 (8)	288 (8)
Regression path time			
DO gap exposure, plausible	469 (22)	346 (11)	380 (16)
DO gap exposure, implausible	466 (19)	372 (13)	396 (16)
PO gap exposure, plausible	469 (18)	320 (10)	336 (15)
PO gap exposure, implausible	449 (18)	345 (14)	401 (15)
Percent regressions			
DO gap exposure, plausible	23.5 (2.3)	10.7 (1.4)	17.5 (2.1)
DO gap exposure, implausible	24.8 (2.3)	14.5 (1.9)	22.4 (2.1)
PO gap exposure, plausible	27.3 (2.5)	7.4 (1.1)	11.5 (1.4)
PO gap exposure, implausible	17.7 (2.5)	9.7 (1.6)	21.8 (2.2)

Table 15 presents a summary of the statistical analyses. In the pre-verb region (i.e., the subject of the relative clause), there was a significant interaction of the exposure group and plausibility in both first pass times ($\beta = -0.12$, $SE = 0.06$, $p < 0.05$) and percent regressions ($\beta = 0.76$, $SE = 0.33$, $p < 0.05$). Planned pairwise comparisons, however, indicated that there was no significant difference in first pass time based on plausibility for the DO gap exposure group ($\beta = 0.03$, $SE = 0.05$, $p > 0.1$) and marginally longer first past times for the PO gap exposure group ($\beta = -0.09$, $SE = 0.05$, $p = 0.06$). On the other hand, planned pairwise comparisons for percent regressions indicated no significant difference based on plausibility for the DO gap exposure group ($\beta = -0.16$, $SE = -0.23$, $p > 0.1$), but significantly more regressions from this region in the PO gap exposure condition when the filler was a plausible direct object of the verb ($\beta = 0.82$, $SE = 0.36$, $p < 0.05$). This result does not demonstrate the plausibility mismatch effect before the critical region, however, because it was in the opposite direction from the prediction: more regressions occurred when the filler was a *plausible* direct object rather than an implausible one.

In the critical verb region, there was only a marginal effect of the exposure group on percent regressions ($\beta = -0.41$, $SE = 0.24$, $p = 0.08$); participants exposed to DO gaps were more likely to regress out of this region. While we did not find effects of plausibility in the critical verb region, results from the spillover region suggest that the plausibility manipulation was successful. In this region, sentences with implausible fillers had significantly longer regression path times ($\beta = -0.10$, $SE = 0.03$, $p < 0.01$) and significantly more regressions ($\beta = -0.55$, $SE = 0.18$, $p < 0.01$). This main effect was not tempered by any interactions for any of the analyzed eye tracking measures. Thus, the plausibility mismatch effect was visible in the reading times on the spillover region but was not affected by exposure to PO gap sentences.

Table 15. Experiment 2B fixed effects summary for the five eye tracking measures and three regions of analysis: the pre-verb region, the verb region, and the spillover region.

	Pre-Verb Region		Verb Region		Spillover Region	
	β	*SE*	*β*	*SE*	*β*	*SE*
First fixation duration						
Exposure group	−0.05	0.04	−0.06	0.04	−0.03	0.05
Plausibility	−0.02	0.02	−0.01	0.03	−0.03	0.02
Trial	0.001	0.001	0.00001	0.001	−0.001	0.001
Exposure × plausibility	−0.01	0.04	−0.06	0.05	−0.03	0.04
First pass time						
Exposure group	−0.02	0.06	−0.05	0.05	−0.01	0.07
Plausibility	−0.03	0.03	−0.01	0.03	−0.04	0.03
Trial	0.001	0.001	−0.0001	0.001	−0.001	0.001
Exposure × plausibility	−0.12 *	0.06	−0.04	0.06	0.02	0.05
Regression path time						
Exposure group	−0.03	0.08	−0.09	0.05	−0.04	0.08
Plausibility	0.02	0.03	−0.05	0.03	−0.10 **	0.03
Trial	−0.002 *	0.001	−0.0003	0.001	−0.002 *	0.001
Exposure × plausibility	0.002	0.07	−0.004	0.06	−0.09	0.07
Percent regressions						
Exposure group	−0.11	0.24	−0.42 †	0.24	−0.32	0.24
Plausibility	0.42	0.25	−0.32	0.25	−0.55 **	0.18
Trial	−0.01 *	0.005	−0.01	0.01	−0.01 *	0.01
Exposure × plausibility	0.81 *	0.35	0.26	0.50	−0.49	0.35

Note: † $p < 0.1$, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

As with the previous experiences, the null results (i.e., the lack of an effect of exposure group) were confirmed with a Bayes factor analysis. For the regression path data, the full overall model was compared to a null model without the exposure group factor as well as its interaction with the plausibility manipulation; the BF of $0.13 \pm 0.38\%$ indicated moderate evidence for the null hypothesis. The exposure group did not affect participants' regression path durations in the spillover region and therefore did not affect gap predictions.

3.2.3. Discussion

Experiment 2B examined the generalizability of syntactic adaptation in two ways: the exposure phase was presented as a separate task, and active gap filling was measured using the plausibility mismatch effect rather than the filled gap effect. As in Experiment 2A, when exposure was more naturalistic, the DO gap expectation persisted even for the group exposed to PO gaps. Both exposure groups had more and longer regressions in the region following the verb when the filler was an implausible direct object of the verb. This suggests that both groups were actively associating the filler with the verb and were surprised when the semantic fit between the verb and the filler did not allow the direct object gap interpretation. As in Experiment 2A, in contrast to Experiment 1, there was no measure of the strength of the plausibility mismatch effect prior to exposure. As in Experiment 2A, the PO gap group demonstrated the plausibility mismatch effect in the eye tracking task, and it was unlikely that exposure to PO gaps would lead to the generation of this effect if it did not previously exist in this population. Again, this would be more of a concern if they were not demonstrating the effect. However, unlike Experiment 2A, this cannot be directly corroborated by anecdotal evidence, as the sentence types in the sentence recognition task were not the same as those in the eye tracking task. Nonetheless, the disfluencies occurred for participants in the PO gap exposure group in this experiment as well, which does indicate that these participants were actively filling the gap position prior to adaptation.

Taken together, Experiments 2A and 2B suggest that syntactic adaptation of filler-gap dependency processing does not generalize across tasks. Given that we do not see adaptation in either of these studies, it is difficult to draw any compelling conclusions about the generalizability of effects from a structural test of active gap filling—filled gap sentences, which syntactically block a DO gap interpretation—to a semantic test of the same: plausibility mismatch sentences, which semantically block a DO gap interpretation.

4. General Discussion

The present studies examined two aspects of syntactic adaptation effects. The first experiment modified the blocked, between-subjects design from Fine et al.'s (2013) Experiment 2 to investigate whether the gap predictions associated with filler-gap dependency processing are subject to adaptation effects. Participants that were exposed to prepositional object gaps ceased generating direct object gap predictions. Experiments 2A and 2B aimed to address questions of the generalizability of syntactic adaptation effects by making the exposure phase more like a naturalistic learning environment (i.e., within a narrative) and by presenting this phase as a separate task. Despite the fact that participants adapted their gap position expectations in Experiment 1, there was no evidence of adaptation when the exposure phase was masked as a separate task.

4.1. Comparison to Previous Studies of Filler-Gap Dependency Adapation

The results of Experiment 1 are contradictory to those of Kaan et al. (2019); in our study, participants' predictions about the position of the gap were altered by exposure, while Kaan et al. found no such adaptation effects in either native speakers or L2 learners of English. There are several key differences between these studies, however, that may explain the discrepancy in their findings. First, the critical dependencies in Kaan et al.'s stimuli were embedded questions and, therefore, within a subordinate clause (e.g., *The builder wondered <u>what</u> the worker fixed the leak with __*). Alternatively, our target items utilized relative clauses in the subject position of the main clause of the sentences. Incongruities between active gap filling behavior in main clause versus embedded clause contexts have been observed developmentally; 5-year-olds have been shown not to actively complete these dependencies when they were within singular clauses (Atkinson et al. 2018), but do show some evidence of active completion in the main clause when an embedded clause completion was the alternative (Omaki et al. 2014). It is plausible, then, that the difference in adaptation effects reflects these clause level differences in the target stimuli. Also, having the dependency in the subject position may have been particularly salient for participants in the current study, which may have led to observable adaptation effects.

Perhaps more importantly, Kaan et al.'s (2019) study and Experiment 1 utilized different methodologies to test for syntactic adaptation. As in Fine et al.'s (2013) Experiment 1, Kaan et al. defined adaptation as decreased difficulty processing filled gap sentences as demonstrated by a decreasing difference in reading time between the sentences containing a filler-gap dependency and those without. In this design, exposure was cumulative over the length of the experiment. The critical filled gap sentences were interspersed with fillers and experimental items for a different adaptation study (i.e., adaptation to sentential level *and* coordination). In other words, the ability to locate the relevant sentences for adaptation to gap positions was made more difficult by the surrounding items. Conversely, the current Experiment 1 utilized the blocked adaptation design of Fine et al.'s (2013) Experiment 2, such that the relevant items for adaptation were presented sequentially at the beginning of the experiment. Again, this presentation of unexpected structures was likely quite salient for participants; several of them commented during the debrief that they noticed a number of similar sentences at the beginning of the experiment, despite the fact that this blocked structure was not made explicit. Thus, the successful adaptation to gap positions in the current paper may be the result of the combined saliency of relative clauses in the subject position and exposure to the critical targets in a blocked design.

4.2. Successful Adaptation of Gap Predictions and Probabilistic Parsing

Experiment 1 was the first study to demonstrate that gap predictions can be altered by the distribution of gaps in the input. These results could be taken as evidence for the probabilistic parsing account of processing: when the distribution of gaps in the input changed from one favoring DO gaps (see Section 2.1) to one favoring PO gaps, DO gap predictions decreased significantly. However, the current study had no way of measuring if there was an associated increase in PO gap predictions associated with this decrease in DO gap predictions. The probabilistic parsing account predicts that processing should directly reflect the structural distribution; any increase in expectation of one structure should be accompanied by a decrease in expectation of a competing structure (and vice versa). For example, the main clause and reduced relative continuations studied by Fine et al. (2013) were in direct competition. Thus, decreasing the probability of one structure automatically increased the probability of the other. This was reflected in the fact that processing difficulty on the disambiguating region increased for main clause continuations and decreased for reduced relative continuations.

Given that there is no way to know if the participants in Experiment 1 increased their prepositional object gap predictions, the results of this study could instead be attributed to a general diminishing of the generation of gap predictions rather than a shift from one specific prediction to another (i.e., from direct object gaps to prepositional object gaps). In fact, a pilot study examining whether participants exposed to PO gaps increased their PO gap predictions suggests that gap predictions are not simply being transferred from one position to another (see Chapter 3, Atkinson 2016). This pilot study used a similar design to Experiment 1, except the experimental block evaluated PO gap predictions instead of DO gap predictions. PO gap predictions were evaluated using a novel sentence type: filled prepositional object gaps (*The suitcase that the thief stole __ from <u>the hotel room</u> contained precious jewels*). If participants were shifting their gap predictions from direct object gaps to prepositional object gaps after exposure, they should not be expecting a complement NP in the prepositional phrase and should slow down when reading this region. This was not the case: participants exposed to PO gaps did not generate a filled PO gap effect, which suggests that they were not predicting a PO gap. Importantly, this also suggests that participants are not simply trying to match their gap predictions to the input and that the adaptation effects obtained in Experiment 1 were more likely the result of a general dampening of gap predictive behavior.

An even stronger hypothesis is that this reflects a diminishing of predictions in general due to the number of prediction-based errors that the parser is making during the exposure block. This interpretation is consistent with other theories of prediction and syntactic adaptation including predictive utility (see Kuperberg and Jaeger 2016 for discussion) and syntactic satiation (Boland et al. 2023; Do and Kaiser 2017; Fraundorf and Jaeger 2016; Kaschak and Glenberg 2004). According to predictive utility accounts (Kuperberg and Jaeger 2016), predictions are only generated when they are useful for the goal that the comprehender is attempting to accomplish. In the case of Experiment 1, DO gap prediction (or perhaps gap prediction in general) is no longer a reliable strategy for efficient comprehension, so they are not maintained. Relatedly, syntactic satiation is a phenomenon in which ungrammatical, difficult, or novel sentences become more acceptable (e.g., Do and Kaiser 2017) or easier to process after increased exposure (e.g., Boland et al. 2023; Fraundorf and Jaeger 2016; Kaschak and Glenberg 2004). Such an account would attribute the reduction in DO gap prediction not to an expectation of gaps, but to an expectation of unexpected stimuli. The account is also compatible with the lack of transfer effects in Experiments 2A and 2B. Perhaps, participants were satiated to unexpected stimuli while reading the stories, but this satiation did not carry over to the eye tracking task. The current studies are not able to adjudicate between these alternatives, so future work is needed to do so. Additionally, future work should examine whether the effects presented here are structurally specific or whether they generalizes to other predictions as well (e.g., predicting an appropriate direct object based on the semantics of the verb).

The fact that adaptation of gap predictions may be attributable to a different source than the adaptation of ambiguity resolution is somewhat unsurprising. While there are two clear alternatives in the ambiguity tested by Fine et al. (2013), filler-gap dependencies are unbounded. Thus, there is no obvious, finite set of alternatives if the gap is ultimately not in the initially predicted position (i.e., active gap filling fails at the direct object position). While one alternative is presented by the input distribution—PO gaps—there is nothing in the grammar requiring this to be the gap location once DO gaps are ruled out. Further work is needed to disentangle whether syntactic adaptation effects that generate new, specific predictions versus those that generate more general changes to processing routines should be attributed to the same or different sources.

4.3. Task-Specific or Talker-Specific?

Thus far, we have discussed the lack of adaptation effects in Experiments 2A and 2B as being the result of task-specific learning. In other words, participants were exposed to an input distribution of gap positions in one task (i.e., sentence recognition), but that distribution did not seem to be relevant for their behavior in another task (i.e., eye tracking during reading). An alternative account is possible, however; rather than being a *task*-specific effect, it may instead be a *talker*-specific effect. We utilize "talker" instead of "speaker" following Kamide (2012). She suggests that the term "speaker" indicates a focus on the speech production system of the participants rather than the comprehension system. Thus, we utilize the "talker" terminology, despite the fact that participants were reading written sentences (not listening to spoken speech) because our participants were comprehending sentences.[3]

In the domain of speech perception, there are robust effects of speaker-specific adaptation (Bradlow and Bent 2008; Kraljic et al. 2008; Kraljic and Samuel 2007; Sumner and Samuel 2009, among others). In Kraljic and Samuel (2007), participants were exposed to two speakers who produced a sound that was ambiguous between /s/ and /ʃ/. One speaker used this ambiguous sound in /s/-requiring contexts, while the other used it in /ʃ/-requiring contexts. Speakers were also distinguished by gender. During an identification task, participants demonstrated speaker-specific categorization knowledge; they gave significantly more /ʃ/ categorizations for the voice that used the ambiguous sound in /ʃ/-contexts during exposure and vice versa (i.e., more /s/ categorizations for the voice that used the ambiguous sound in this environment). In other words, participants adapted to speaker-specific phonetic realizations of /s/ or /ʃ/ and used this information in their later sound categorizations.

In addition to speaker effects in phonetic adaptation, there are also examples of talker-specific adaptation at the lexical (Creel et al. 2008; Ryskin et al. 2020) or syntactic level (Kamide 2012) and to spatial perspective (Ryskin et al. 2016). For example, Kamide (2012) conducted a visual world eye tracking experiment in which participants were exposed to two talkers: one preferred high attachment (*The <u>uncle</u> of the girl who will ride the <u>motorbike</u> is from France*) and the other preferred low attachment (*The uncle of the <u>girl</u> who will ride the <u>carousel</u> is from France*). During the test phase, participants made anticipatory looks to the target consistent with the talker's attachment preference during the verb region (e.g., the motorbike for high attachment and the carousel for low attachment) more than to the competitor (e.g., vice versa: the carousel for high attachment and the motorbike for low attachment). A new talker was also introduced during the test phase and served as a control condition; participants were equally likely to look at either image when this talker produced the target sentence. Taken together, these results suggest that participants were adapting to the talker's syntactic preferences during the exposure phase and generating predictions for the upcoming input based on these preferences during the test phase. Recent work, however, has failed to replicate this result (Liu et al. 2017; Ryskin et al. 2017).

In Experiments 2A and 2B, there was no indication that the talker in the sentence recognition task was the same as the talker in the eye tracking task. Thus, the lack of transfer of adaptation effects could have been due to different tasks (i.e., sentence recognition and eye tracking during reading) or to different talkers (i.e., the writer of the stories and the writer of the standalone sentences). The concern about which talker produced the sentences is not a problem for Experiment 1 if we presume that participants interpret all sentences presented in a single task as being produced by a single talker, perhaps the experimenter. In Experiment 1, the exposure and experiment phases were presented within a single task, so the talker was likely interpreted as being the same across all items. It is impossible to know if successful adaptation in this experiment is due to a consistent task or consistent talker because task and talker are confounded. In Experiments 2A and 2B, when the task changed from sentence recognition to eye tracking (and also changed machines on which the task was performed), it was possible that participants also interpreted the talker as having changed. Thus, the current experiments cannot distinguish between an explanation of the gap prediction adaptation effects presented herein as task specific (i.e., no transfer across tasks) or talker specific (i.e., no transfer across talkers). One way to distinguish task and talker specificity would be to replicate Experiment 2 except to explicitly tell the participants that the talker in the sentence recognition task is the same as that in the eye tracking task. For instance, you could tell them that the same person wrote the stories and the sentences. If talker specificity really was not a concern in Experiment 1, then an alternative would be to complete the sentence recognition task as a block within the larger eye tracking task (i.e., as a less obvious separate task).

5. Conclusions

The current work examined whether exposing participants to PO gaps, a less frequent gap location, altered their gap predictions, and whether these effects would transfer across tasks when this input was presented in a quasi-naturalistic way (i.e., by reading stories). In Experiment 1, comprehenders decreased their DO gap predictions following exposure to PO gaps, as evidenced by the lack of a filled gap effect only for this exposure group. However, Experiments 2A and 2B suggest that these adaptation effects may be either task or talker specific; they do not generalize when the quasi-naturalistic exposure phase is presented as a separate task (Experiments 2A and 2B) nor when the measure of gap filling changes from a structural anomaly in the exposure phase (i.e., filled gap sentences) to a semantic anomaly in the test phase (i.e., plausibility mismatch sentences; Experiment 2B). Overall, these experiments add filler-gap dependency processing, as well as the gap predictions associated with it, to the growing list of structures demonstrating adaptation effects, while also suggesting that this effect may be specific to a singular experimental environment.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/languages8040285/s1. Supplement S1: Experimental items.

Author Contributions: Conceptualization, E.A. and A.O.; methodology, E.A. and A.O.; software, E.A.; formal analysis, E.A.; investigation, E.A.; resources, A.O.; writing—original draft preparation, E.A.; writing—review and editing, E.A. and A.O.; visualization, E.A.; supervision, A.O.; acquisition, A.O. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by an NSF grant awarded to Akira Omaki, NSF BCS-1423117 and BCS-1737736.

Institutional Review Board Statement: The studies were conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Boards of Johns Hopkins University (HIRB00000496, 17 June 2014) and the University of Washington (23 March 2017).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data and code presented in this study are openly available in OSF at DOI 10.17605/OSF.IO/4PYDA.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Experiments 1 and 2A Spillover Region Analyses

As explained above in the main analyses for Experiments 1 and 2A, the spillover regions were not identical across the conditions when filled gap sentences were used. In the NP-fronting conditions, the first word after the critical direct object NP region was the preposition that headed the PP with the eventual gap. In the PP-fronting conditions, however, this preposition was pied-piped with the filler and the spillover region consisted of the main verb of the sentence. This main verb was often the past tense of the copula *be* (i.e., *was*), which does not carry much semantic meaning. Thus, this region may have been skipped more frequently. These two spillover regions serve different purposes; the preposition provides the phrase that will ultimately host the gap and resolve the dependency, while the main verb provides the initial content of the predicate (which is only provided by the region following the spillover region in NP-fronting conditions). Given these differences, it is not reasonable to directly compare these regions across conditions. For transparency, however, the data and analysis of these regions from Experiments 1 and 2A are presented below (in Appendices A.1 and A.2, respectively).

Appendix A.1. Experiment 1: Spillover Region Analysis

Appendix A.1.1. Exposure Block

Table A1 presents the data relevant for the analysis from the spillover region of the Experiment 1 PO gap exposure group, and Table A2 summarizes the analysis.

Table A1. Experiment 1 PO gap exposure group participant mean reading times in milliseconds (standard error) and percent regressions for the spillover region.

	First Fixation Duration	First Pass Time	Regression Path Time	Percent Regressions
NP-fronting	215 (7)	247 (11)	287 (18)	9.5 (1.8)
PP-fronting	215 (5)	246 (9)	381 (31)	22.9 (3.4)

Table A2. Experiment 1 PO gap exposure group fixed effects summary for the four eye tracking measures in the spillover region.

	First Fixation Duration		First Pass Time		Regression Path Time		Percent Regressions	
	β	*SE*	*β*	*SE*	*B*	*SE*	*β*	*SE*
Intercept	5.34 ***	0.03	5.44 ***	0.05	5.59 ***	0.07	−1.93 ***	0.24
Fronting type	−0.01	0.04	0.005	0.06	−0.10	0.10	−0.86 *	0.36

Note: † $p < 0.1$, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

In the spillover region, participants made more regressions during PP-fronting sentences ($\beta = -0.86$, $SE = 0.36$, $p = 0.013$). However, it is not possible to draw any conclusions from these results because of the differences in the spillover regions for the two fronting types (e.g., "from" in NP-fronting conditions versus "was" in PP-fronting conditions; see (4)).

Appendix A.1.2. Experimental Block

Table A3 presents the participant means on each measure (first fixation duration, first pass time, regression path time, and percent regressions) for the spillover region as well as the standard errors of the participant means.

Table A4 presents a summary of the statistical analyses for the spillover region. For percent regressions, the model including an interaction of exposure group and trial did not converge. Thus, a model without this interaction is reported. In this region, was a significant interaction of exposure group and fronting type for the DO gap exposure group on regression path times ($\beta = -0.24$, $SE = 0.12$, $p < 0.05$) and on percent regressions ($\beta = -1.23$, $SE = 0.57$, $p < 0.05$). As in the analysis of the exposure block, no reasonable conclusions can be drawn from these effects because the content of the spillover region was different for the two conditions.

Table A3. Experiment 1 participant mean reading times in milliseconds (standard error) and percent regressions for the spillover region.

	First Fixation Duration	First Pass Time	Regression Path Time	Percent Regressions
Filler exposure, NP-fronting	231 (7)	247 (9)	315 (17)	13.3 (2.9)
Filler exposure, PP-fronting	219 (6)	239 (7)	324 (20)	14.5 (2.0)
DO gap exposure, NP-fronting	221 (6)	239 (6)	346 (28)	17.5 (3.2)
DO gap exposure, PP-fronting	231 (5)	249 (6)	562 (45)	35.7 (3.3)
PO gap exposure, NP-fronting	224 (7)	243 (9)	320 (22)	16.8 (2.4)
PO gap exposure, PP-fronting	215 (5)	246 (9)	381 (31)	22.9 (3.4)

Table A4. Experiment 1 experimental block fixed effects summary for the four eye tracking measures in the spillover region.

	First Fixation Duration		First Pass Time		Regression Path Time		Percent Regressions	
	β	*SE*	β	*SE*	β	*SE*	β	*SE*
DO gap exposure	0.01	0.05	0.01	0.07	0.14	0.09	0.71 *	0.28
PO gap exposure	−0.01	0.05	0.02	0.07	0.08	0.08	0.47	0.28
Fronting type	0.05	0.04	0.04	0.05	0.06	0.08	0.07	0.43
Trial	−0.001	0.001	−0.001	0.001	−0.001	0.002	−0.002	0.01
DO gap × fronting	−0.06	0.06	−0.04	0.07	−0.24 *	0.12	−1.23 *	0.57
PO gap × fronting	−0.03	0.06	−0.03	0.07	−0.10	0.11	−0.29	0.57
DO gap × trial	0.0001	0.002	−0.0004	0.002	−0.003	0.003		
PO gap × trial	0.0002	0.002	0.001	0.002	−0.003	0.003		

Note: † $p < 0.1$, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

Appendix A.2. Experiment 2A Spillover Region Analysis

Table A5 presents the participant means and standard errors on each measure for the spillover region in Experiment 2A.

Table A5. Experiment 2A participant mean reading times in milliseconds (standard error) and percent regressions for the spillover region.

	First Fixation Duration	First Pass Time	Regression Path Time	Percent Regressions
DO gap exposure, NP-fronting	245 (9)	255 (10)	285 (12)	9.5 (2.0)
DO gap exposure, PP-fronting	239 (10)	263 (12)	331 (26)	15.5 (3.1)
PO gap exposure, NP-fronting	237 (9)	262 (10)	304 (21)	13.6 (3.8)
PO gap exposure, PP-fronting	233 (9)	250 (10)	290 (19)	12.0 (2.9)

Table A6 presents a summary of the statistical analyses. There were no significant effects on this region of analysis for any of the eye tracking measures.

Table A6. Experiment 2A fixed effects summary for the four eye tracking measures in the spillover region.

	First Fixation Duration		First Pass Time		Regression Path Time		Percent Regressions	
	β	*SE*	β	*SE*	β	*SE*	β	*SE*
Exposure group	0.01	0.06	0.03	0.06	0.01	0.07	−0.09	0.29
Fronting type	−0.01	0.04	−0.02	0.06	0.01	0.06	−0.26	0.29
Trial	−0.0001	0.001	0.001	0.001	0.001	0.001	0.01	0.01
Exposure × fronting	−0.11	0.08	−0.01	0.08	0.005	0.10	−0.11	0.58

Note: † $p < 0.1$, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

Notes

[1] Only half of these sentences (8 out of 16) were ambiguous. The other half included an overt relative clause marker, i.e., *that*, which rendered the sentences unambiguous (*The experienced soldiers that warned about the dangers conducted the midnight raid.*). These count as exposure to reduced relative clauses, though they are easier to process than the ambiguous sentences because they prevent misanalysis of the past participle verb as a main clause verb and subsequent reanalysis.

[2] The model for the verb region did not converge with random effects, so the simple logistic regression is reported here.

[3] In this case, "writer" is perhaps a more accurate term, but we continued to use "talker" for continuity with previous work.

References

Atkinson, Emily. 2016. Active Dependency Completion in Adults and Children: Representations and Adaptation. Doctoral dissertation, Johns Hopkins University, Baltimore, MD, USA, October 28.

Atkinson, Emily, Matthew W. Wagers, Jeffrey Lidz, Colin Phillips, and Akira Omaki. 2018. Developing incrementality in filler-gap dependency processing. *Cognition* 179: 132–49. [CrossRef]

Baayen, R. Harald, Douglas J. Davidson, and Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59: 390–412. [CrossRef]

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68: 255–78. [CrossRef]

Bates, Douglas M., Martin Maechler, Benjamin M. Bolker, and Steven Walker. 2015. Lme4: Linear Mixed-Effects Models Using Eigen and S4. R Package Version 1.1-9. Available online: https://CRAN.R-project.org/package=lme4/ (accessed on 31 October 2022).

Bock, J. Kathryn. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18: 355–87. [CrossRef]

Bock, Kathryn, and Helga Loebell. 1990. Framing sentences. *Cognition* 35: 1–39. [CrossRef] [PubMed]

Bock, Kathryn, and Zenzi M. Griffin. 2000. The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General* 129: 177–92. [CrossRef] [PubMed]

Boland, Julie E., Emily Atkinson, Guadalupe De Los Santos, and Robin Queen. 2023. What do we learn when we adapt to reading regional constructions? *PLoS ONE* 18: e0282850. [CrossRef] [PubMed]

Bradlow, Ann R., and Tessa Bent. 2008. Perceptual adaptation to non-native speech. *Cognition* 106: 707–29. [CrossRef]

Branigan, Holly P., Martin J. Pickering, and Alexandra A. Cleland. 2000a. Syntactic co-ordination in dialogue. *Cognition* 75: B13–B25. [CrossRef] [PubMed]

Branigan, Holly P., Martin J. Pickering, Andrew J. Stewart, and Janet F. McLean. 2000b. Syntactic priming in spoken production: Linguistic and temporal interference. *Memory & Cognition* 28: 1297–302.

Chun, Eunjin. 2018. The Role of Prediction in Adaptation: An Evaluation of Error-Based Learning Accounts. Doctoral dissertation, University of Florida, Gainesville, FL, USA.

Crain, Stephen, and Janet Dean Fodor. 1985. How can grammars help parsers? In *Natural Language Parsing: Psychological, Computational and Theoretical Perspectives*. Edited by David R. Dowty, Lauri Kartuunen and Arnold M. Zwicky. Cambridge: Cambridge University Press, pp. 94–128.

Creel, Sarah C., Richard N. Aslin, and Michael K. Tanenhaus. 2008. Heeding the voice of experience: The role of talker variation in lexical access. *Cognition* 106: 633–64. [CrossRef] [PubMed]

Dahan, Delphine, Sarah J. Drucker, and Rebecca A. Scarborough. 2008. Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition* 108: 710–18. [CrossRef] [PubMed]

Davis, Matthew H., Ingrid S. Johnsrude, Alexis Hervais-Adelman, Karen Taylor, and Carolyn McGettigan. 2005. Lexical Information Drives Perceptual Learning of Distorted Speech: Evidence from the Comprehension of Noise-Vocoded Sentences. *Journal of Experimental Psychology: General* 134: 222–41. [CrossRef]

Dempsey, Jack, Qiawen Liu, and Kiel Christianson. 2020. Convergent probabilistic cues do not trigger syntactic adaptation: Evidence from self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 46: 1906–21. [CrossRef] [PubMed]

Dempsey, Jack, Qiawen Liu, and Kiel Christianson. 2023. Syntactic adaptation leads to updated knowledge for local structural frequencies. *Quarterly Journal of Experimental Psychology*, 174702182311729. [CrossRef] [PubMed]

Do, Monica L., and Elsi Kaiser. 2017. The Relationship between Syntactic Satiation and Syntactic Priming: A First Look. *Frontiers in Psychology* 8: 1851. [CrossRef]

Drummond, Alex. 2010. *Ibexfarm (0.3)*. Available online: https://adrummond.net/ibexfarm (accessed on 3 November 2018).

Farmer, Thomas A., Alex B. Fine, Shaorong Yan, Spyridoula Cheimariou, and T. Florian Jaeger. 2014. Error-Driven Adaptation of Higher-Level Expectations During Reading. Paper presented at the 36th Annual Conference of the Cognitive Science Society, Quebec City, Canada, July 23–26; Edited by Paul Bello, Marcello Guarini, Marjorie McShane and Brian Scassellati. Seattle: Cognitive Science Society, pp. 2181–86.

Fine, Alex B., and T. Florian Jaeger. 2013. Evidence for implicit learning in syntactic comprehension. *Cognitive Science* 37: 578–91. [CrossRef] [PubMed]

Fine, Alex B., T. Florian Jaeger, Thomas A. Farmer, and Ting Qian. 2013. Rapid expectation adaptation during syntactic comprehension. *PLoS ONE* 8: e77661. [CrossRef] [PubMed]

Fine, Alex B., Ting Qian, T. Florian Jaeger, and Robert A. Jacobs. 2010. Is there syntactic adaptation in language comprehension? Paper presented at the 2010 Workshop on Cognitive Modeling and Computational Linguistics, Uppsala, Sweden, July 15; Stroudsburg: Association for Computational Linguistics, pp. 18–26.

Fraundorf, Scott H., and T. Florian Jaeger. 2016. Readers generalize adaptation to newly-encountered dialectal structures to other unfamiliar structures. *Journal of Memory and Language* 91: 28–58. [CrossRef]

Frazier, Lyn. 1987. Syntactic processing: Evidence from Dutch. *Natural Language & Linguistic Theory* 5: 519–59.

Frazier, Lyn, and Charles Clifton. 1989. Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes* 4: 93–126. [CrossRef]

Garnsey, Susan M., Michael K. Tanenhaus, and Robert M. Chapman. 1989. Evoked potentials and the study of sentence comprehension. *Journal of Psycholinguistic Research* 18: 51–60. [CrossRef] [PubMed]

Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68: 1–76. [CrossRef]

Gibson, Edward. 2000. The Dependency Locality Theory: A distance-based theory of linguistic complexity. In *Image, Language, Brain: Papers from the First Mind Articulation Project Symposium*. Edited by Alec Marantz, Yasushi Miyashita and Wayne O'Neil. Cambridge, MA: MIT Press, pp. 95–126.

Grodner, Daniel, Edward Gibson, and Susanne Tunstall. 2002. Syntactic complexity in ambiguity resolution. *Journal of Memory and Language* 46: 267–95. [CrossRef]

Harrington Stack, Caoimhe M., Ariel N. James, and Duane G. Watson. 2018. A failure to replicate rapid syntactic adaptation in comprehension. *Memory & Cognition* 46: 864–77. [CrossRef]

Jaeger, T. Florian. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59: 434–46. [CrossRef]

Kaan, Edith, and Eunjin Chun. 2018. Syntactic Adaptation. In *Psychology of Learning and Motivation: Current Topics in Language*. Edited by Kara D. Federmeier and Duane G. Watson. Amsterdam: Elsevier, vol. 68, pp. 85–116. [CrossRef]

Kaan, Edith, Corinne Futch, Raquel Fernández Fuertes, Sonja Mujcinovic, and Esther Álvarez De La Fuente. 2019. Adaptation to syntactic structures in native and nonnative sentence comprehension. *Applied Psycholinguistics* 40: 3–27. [CrossRef]

Kamide, Yuki. 2012. Learning individual talkers' structural preferences. *Cognition* 124: 66–71. [CrossRef]

Kaschak, Michael P. 2007. Long-term structural priming affects subsequent patterns of language production. *Memory & Cognition* 35: 925–37.

Kaschak, Michael P., and Arthur M. Glenberg. 2004. This construction needs learned. *Journal of Experimental Psychology: General* 133: 450–67. [CrossRef] [PubMed]

Kaschak, Michael P., Renrick A. Loney, and Kristin L. Borreggine. 2006. Recent experience affects the strength of structural priming. *Cognition* 99: B73–B82. [CrossRef] [PubMed]

Kaschak, Michael P., Timothy J. Kutta, and Jacqueline M. Coyle. 2014. Long and short term cumulative structural priming effects. *Language, Cognition and Neuroscience* 29: 728–43. [CrossRef] [PubMed]

Kingsbury, Paul, Stephanie Strassel, Cynthia McLemore, and Robert McIntyre. 1997. *CALLHOME American English Transcripts LDC97T14*. Web Download. Philadelphia: Linguistic Data Consortium.

Kleinschmidt, Dave, and T. Florian Jaeger. 2011. A Bayesian belief updating model of phonetic recalibration and selective adaptation. Paper presented at the 2nd Workshop on Cognitive Modeling and Computational Linguistics, Portland, OR, USA, June 23; Stroudsburg: Association for Computational Linguistics, pp. 10–19.

Kleinschmidt, Dave F., and T. Florian Jaeger. 2015. Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review* 120: 148–203. [CrossRef]

Kraljic, Tanya, and Arthur G. Samuel. 2007. Perceptual adjustments to multiple speakers. *Journal of Memory and Language* 56: 1–15. [CrossRef]

Kraljic, Tanya, Susan E. Brennan, and Arthur G. Samuel. 2008. Accommodating variation: Dialects, idiolects, and speech processing. *Cognition* 107: 54–81. [CrossRef]

Kuperberg, Gina R., and T. Florian Jaeger. 2016. What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience* 31: 32–59. [CrossRef]

Kuznetsova, Alexandra, Per Bruun Brockhoff, and Rune Haubo Bojesen Christensen. 2015. lmerTest: Tests for Random and Fixed Effects for Linear Mixed Effect Models. R Package Version 2.0-29. Available online: http://CRAN.R-project.org/package=lmerTest/ (accessed on 16 July 2015).

Lee, Ming-Wei. 2004. Another look at the role of empty categories in sentence processing (and grammar). *Journal of Psycholinguistic Research* 33: 51–73. [CrossRef]

Levy, Roger, and Galen Andrew. 2006. Tregex and Tsurgeon: Tools for querying and manipulating tree data structures. Paper presented at the 5th International Conference on Language Resources and Evaluation, Genoa, Italy, May 22–28.

Levy, Roger, Evelina Fedorenko, and Edward Gibson. 2013. The syntactic complexity of Russian relative clauses. *Journal of Memory and Language* 69: 461–95. [CrossRef]

Linzen, Tal, and T. Florian Jaeger. 2016. Uncertainty and expectation in sentence processing: Evidence from subcategorization distributions. *Cognitive Science* 40: 1382–411. [CrossRef]

Liu, Linda, Zachary Burchill, Michael K Tanenhaus, and T Florian Jaeger. 2017. Failure to replicate talker-Specific syntactic adaptation. Paper presented at the 39th Annual Conference of the Cognitive Science Society, London, UK, July 26–29; Edited by Glenn Gunzelmann, Andrew Howes, Thora Tenbrink and Eddy Davelaar. Seattle: Cognitive Science Society, pp. 2616–21.

Marcus, Mitchell, Beatrice Santorini, Mary Ann Marcinkiewicz, and Ann Taylor. 1999. *Treebank-3 LDC99T42*. Web Download. Philadelphia: Linguistic Data Consortium.

Morey, Richard D., and Jeffrey N. Rouder. 2018. BayesFactor: Computation of Bayes Factors for Common Designs. R Package Version 0.9.12-4.6. Available online: https://cran.r-project.org/web/packages/BayesFactor/ (accessed on 5 November 2023).

Myslín, Mark, and Roger Levy. 2016. Comprehension priming as rational expectation for repetition: Evidence from syntactic processing. *Cognition* 147: 29–56. [CrossRef]

Noppeney, Uta, and Catherine J. Price. 2004. An fMRI Study of Syntactic Adaptation. *Journal of Cognitive Neuroscience* 16: 702–13. [CrossRef]

Omaki, Akira, Ellen Lau, Imogen Davidson White, Myles L. Dakan, Aaron Apple, and Colin Phillips. 2015. Hyper-active gap filling. *Frontiers in Psychology: Language Sciences* 6: 384. [CrossRef] [PubMed]

Omaki, Akira, Imogen Davidson White, Takuya Goro, Jeffrey Lidz, and Colin Phillips. 2014. No fear of commitment: Children's incremental interpretation in English and Japanese wh-questions. *Language Learning and Development* 10: 206–33. [CrossRef]

Pickering, Martin J., and Holly P. Branigan. 1998. The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language* 39: 633–51. [CrossRef]

Pogue, Amanda, Chigusa Kurumada, and Michael K. Tanenhaus. 2016. Talker-Specific Generalization of Pragmatic Inferences based on Under- and Over-Informative Prenominal Adjective Use. *Frontiers in Psychology* 6: 2035. [CrossRef] [PubMed]

R Core Development Team. 2022. *R: A Language and Environment for Statistical Computing*. Computer Software. Vienna: R Foundation for Statistical Computing. Available online: http://www.R-project.org/ (accessed on 1 January 2022).

Rayner, Keith. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* 124: 372–422. [CrossRef] [PubMed]

Rayner, Keith, Alexander Pollatsek, Jane Ashby, and Charles Clifton, Jr. 2012. *Psychology of Reading*, 2nd ed. New York: Psychology Press.

Ryskin, Rachel A., Ranxiao Frances Wang, and Sarah Brown-Schmidt. 2016. Listeners use speaker identity to access representations of spatial perspective during online language comprehension. *Cognition* 147: 75–84. [CrossRef] [PubMed]

Ryskin, Rachel, Alex Fine, and Sarah Brown-Schmidt. 2017. Do Listeners Learn Speaker/Accent-Specific Syntactic Biases? Paper presented at the 30th Annual CUNY Converence on Human Sentence Comprehension, Cambridge, MA, USA, March 29–April 1.

Ryskin, Rachel, Shukhan Ng, Katherine Mimnaugh, Sarah Brown-Schmidt, and Kara D. Federmeier. 2020. Talker-specific predictions during language processing. *Language, Cognition and Neuroscience* 35: 797–812. [CrossRef]

Samuel, Arthur G., and Saioa Larraza. 2015. Does listening to non-native speech impair speech perception? *Journal of Memory and Language* 81: 51–71. [CrossRef]

Staub, Adrian. 2010. Eye movements and processing difficulty in object relative clauses. *Cognition* 116: 71–86. [CrossRef]

Stowe, Laurie E. 1986. Parsing wh-constructions: Evidence for on-line gap location. *Language and Cognitive Processes* 1: 227–45. [CrossRef]

Sumner, Meghan, and Arthur G. Samuel. 2009. The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language* 60: 487–501. [CrossRef]

Tooley, Kristen M., and Matthew J. Traxler. 2018. Implicit learning of structure occurs in parallel with lexically-mediated syntactic priming effects in sentence comprehension. *Journal of Memory and Language* 98: 59–76. [CrossRef] [PubMed]

Traxler, Matthew J., and Martin J. Pickering. 1996. Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language* 35: 454–75. [CrossRef]

Wagers, Matthew W., and Colin Phillips. 2014. Going the distance: Memory and control processes in active dependency construction. *The Quarterly Journal of Experimental Psychology* 67: 1274–304. [CrossRef] [PubMed]

Wells, Justine B., Morten H. Christiansen, David S. Race, Daniel J. Acheson, and Maryellen C. MacDonald. 2009. Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology* 58: 250–71. [CrossRef] [PubMed]

Yan, Shaorong, and T. Florian Jaeger. 2020. Expectation adaptation during natural reading. *Language, Cognition and Neuroscience* 35: 1394–422. [CrossRef]

Yildirim, Ilker, Judith Degen, Michael K. Tanenhaus, and T. Florian Jaeger. 2016. Talker-specificity and adaptation in quantifier interpretation. *Journal of Memory and Language* 87: 128–43. [CrossRef]