Constructing focus alternatives from context and the limits of semantic priming

Christian Muxica & Jesse A. Harris*

Abstract. Interpreting focus requires a comprehender to identify the set of alternatives intended by the speaker. Previous psycholinguistic research has characterized this process in terms of a two-stage model that initially forms an alternative set via the context-insensitive mechanism of semantic priming (Gotzner et al. 2016, Husband & Ferreira 2016). We have instead advanced a one-stage immediate-access model, in which alternatives are immediately constructed from the discourse context (Muxica & Harris to appear). In two cross-modal probe recognition task experiments, we further test our prediction that the discourse context strongly influences response speed at early moments of focus interpretation. The results are interpreted as uniquely supporting the immediate-access model.

Keywords. Focus; Alternatives; Sentence processing; Probe recognition

1. Introduction. A long tradition of research in semantics addresses the interpretive effect of the focus of an utterance. In general, focus serves to highlight an element in some way against the background of the discourse (Krifka 1992). Focus typically refers to a constituent that is marked by prosodic, syntactic, or, in some languages, morphological means. As focus can be ambiguous in text, we mark it here with the F-feature, as in (2) below.

Since at least Jackendoff (1972), focus has been understood as evoking a contextually salient set of alternative expressions. Rooth's (1985, 1992) Alternative Semantics framework formalizes the alternative set as the focus semantic value, consisting of expressions of the same formal semantic type that can be substituted for the expression in focus. While focus itself does not alter the truth value of an utterance directly, focus can nonetheless affect the inferences associated with the utterance and, in the case of focus-sensitive semantic operators like *only* and *even*, determine the background against which those operators are interpreted (e.g. Beaver & Clark 2009).

Multiple kinds of focus have been identified in the literature, including informational and contrastive, among others (for overview, see Büring 2016). *Informational* focus marks non-given constituents of an utterance, typically, though not exclusively, as new information. Such uses can be illustrated with question-answer pairs, as in (1). In B's reply, the element in focus (*Willie*) provides new information while the remainder of the utterance is given in the discourse, in that it is information that has been previously mentioned.

(1) **A.** Who did Dolly sing to?
B. Dolly (only) sang to [Willie]$_F$

In (1), the set of relevant alternatives to the focused word *Willie* depends on the context. It might be limited to members of Dolly's band, other country stars, a particular concert, or even humanity

*Acknowledgments: We would like to thank audiences at the Third Meeting of Experiments in Linguistic Meaning and the UCLA Psycholinguistics/Computational Linguistics Seminar. We would also like to thank our wonderful team of undergraduate RAs at the UCLA Language Processing Lab. Last, but certainly not least, we are indebted to Thomas Bye for generously funding this and other research projects. Authors: Christian Muxica, University of California Los Angeles (cmuxica@g.ucla.edu) & Jesse A. Harris, University of California Los Angeles (jharris@humnet.ucla.edu).

large. In the absence of a focus operator like *only*, B's answer suggests that Dolly did not sing to anyone else relevant in the set of alternatives, though the strength and cancelability of such *exhaustive* implicatures is perhaps debatable.

On the other hand, *contrastive* focus marks constituents that stand in polar contrast with an item that either has been mentioned previously in discourse or is highly accessible. In (2), for instance, the focus on *Willie* evokes *Merle* as an alternative. Consequently, Speaker B's utterance not only expresses the entailment that Dolly sang to Willie, but again the exhaustive implicature that Dolly did not sing to Merle. This implicature arises despite the fact that Speaker B did not explicitly mention Merle. Instead, the preceding discourse context makes Merle a highly salient alternative. In the remainder of this paper, we concentrate on contrastive focus, with the expectation that many of our conclusions would extend to other types of focus, as well.

(2) **A.** I heard that Dolly sang to Merle
B. Dolly (only) sang to $[\text{Willie}]_F$

Successful interpretation of any utterance containing focus requires a comprehender to identify the set of alternatives intended by the speaker. As mentioned, the relevant set of alternatives is determined by the context, and, as yet, it is unclear exactly how the set of alternatives might be identified by comprehenders. In recent years, a novel body of psycholinguistic research has characterized this process as a two-stage model that crucially relies on lexical activation to generate and select alternatives (Gotzner et al. 2016, Husband & Ferreira 2016). In the first, context-insensitive, stage, lexical *associates* – i.e., words which bear a strong lexical association with the word in focus, become highly activated through semantic priming immediately upon encountering a focused constituent. In the second stage, a context-sensitive mechanism identifies the relevant alternatives from among these associates and maintains their activation. Eventually, the activation of non-alternatives will decay, leaving the relevant alternative set behind.[1]

As we understand it, the two-stage (sometimes known as the *alternative activation*) model advances two main hypotheses about the selection of alternatives, which we refer to as (i) *priming dependence* and (ii) *late generation*. Priming dependence is the hypothesis that constructing a representation of the alternative set depends on semantic priming from the element in focus. Late generation is the hypothesis that additional time, after the focus is encountered, is required to select only those alternatives that are contextually relevant. As a result, the two-stage model can be considered a *destructive* model of the alternative set, in which alternatives are first proposed by the lexicon and then disposed of by context.

However, we have previously argued against both of these claims. In Muxica & Harris (to appear), we proposed the *immediate-access* model, in which membership in the alternative set is guided by contextual constraints immediately after focus is encountered, without first being mediated by semantic priming. Therefore, priming and the generation of focus alternatives can be separated conceptually, as alternatives are *constructed* immediately in concert with the context.

For this study, we adapted the materials and design from Muxica & Harris (to appear) in order to further probe the priming independent aspect of the immediate-access model. We tested

[1] It should be noted that Husband & Ferreira (2016) remain agnostic as to whether the second stage involves a passive process of decay or an active process of suppression. For ease of presentation, we will exclusively describe the two-stage model in terms of decaying activation.

the prediction that response times in a probe recognition task are immediately influenced by the alternative status of a probe word, even when that probe word is not a semantic associate of the word in focus. We explicitly manipulated the alternative status of such non-associate probe words through the discourse context. In doing so, we were able to more directly address the relative contributions of context and semantic priming in the selection of focus alternatives.

2. The two-stage model. By and large, the studies which have investigated the selection of alternatives have supported the properties of priming dependence and late generation. Multiple cross-modal forced-choice task experiments have yielded results compatible with the idea that contextually relevant alternatives are activated alongside associate non-alternatives in the earliest moments of processing focus (Husband & Ferreira 2016, Gotzner et al. 2016, Gotzner & Spalek 2019, Lacina et al. 2023, Jördens et al. 2020, Braun & Tagliapietra 2010). Until recently, only with additional time has an advantage for contextually relevant alternatives alone been observed (Husband & Ferreira 2016, Gotzner et al. 2016). Three foundational studies from this literature are reviewed below.

Husband & Ferreira (2016) conducted a cross-modal lexical decision experiment with a between-subjects stimulus onset asynchrony (SOA) manipulation. On each trial, subjects listened to an utterance and then responded to one of three lexical decision targets (3-b). For half of the subjects, targets appeared immediately after (i.e., at 0ms SOA) the focused word (*sculptor*). For the other half, targets appeared after a brief 750ms delay. Target words differed in their relationship to the focused word. Targets were either plausible associate alternatives (*painter*), implausible associate alternatives (*statue*), or implausible non-associate alternatives (*register*), included as a control.

(3) a. *Sample item from Husband & Ferreira (2016)*
The museum thrilled the [sculptor]$_F$ ...
b. *Lexical decision targets*
Alternative: PAINTER
Associate Non-alternative: STATUE
Control: REGISTER

At probe points presented immediately after the focused word, Husband & Ferreira (2016) found a simple effect of semantic priming: subjects responded faster to associates of the focus (*painter* and *statue*) than non-associates (*register*). However, after a 750ms delay, there was an effect of focus: subjects responded faster to the associate alternative (*painter*) than either the non-alternative (*statue*) or the control (*register*). These results are clearly compatible with the two-stage model. The earliest moments after encountering focus appear to be context-insensitive purely reflecting the semantic priming of the first stage. After a delay, the activation of irrelevant items decays, revealing the context-sensitive selection of the second stage.

One limitation of Husband & Ferreira (2016) is that their materials do not include an explicit discourse context. As discussed, the alternative status of any given element is primarily determined with respect to contextual relevance. However, the alternative and non-alternative status of targets in Husband & Ferreira (2016) was determined purely with respect to plausibility, which is only one component of contextual relevance.

The experiments in Gotzner et al. (2016) and Gotzner & Spalek (2019) addressed this concern in two cross-modal probe recognition experiments in German with a similar between-subjects SOA manipulation. Their materials consisted of two speaker dialogues. The first speaker introduced a set of alternatives (*peaches*, *cherries*, and *bananas*) relevant for the focus used by the second speaker (*peaches*). On each trial, subjects listened to these dialogues, and then were presented with one of three probe words, indicating whether or not they heard that probe word in the preceding dialogue. In Gotzner & Spalek (2019), the probe recognition task was administered immediately after (i.e., 0ms SOA) the focused word (*peaches*). In Gotzner et al. (2016), the task was administered following a 2050ms delay. The probe words again differed in their relationship to the focused word. Probe words were either associate mentioned alternatives (*cherries*), associate unmentioned alternatives (*melons*), or non-associate unmentioned non-alternatives (*clubs*), included as a control.

(4) a. *Sample dialogue from Gotzner et al. (2016)*
A. In the fruit bowl, there are peaches, cherries, and bananas
I bet Carsten has eaten cherries and bananas
B. No, he only ate [peaches]$_F$

b. *Probe words*
Mentioned: CHERRIES
Unmentioned: MELONS
Control: CLUBS

When tested immediately after the sentence, there was a simple semantic priming effect; subjects responded faster to the associates (*cherries* and *melons*) than the non-associates (*clubs*). After a delay, there was an effect of context such that responses to mentioned alternatives (*cherries*) were faster than either the unmentioned (*melons*) or the control (*clubs*) probe word. These results receive a natural explanation under the two-stage model: effects at the early SOA reflect the first stage of context-insensitive semantic priming, while effects at the late SOA reflect the context-sensitive selection of focus alternatives in the second stage.

Additional cross-modal forced choice task studies have investigated the selection of alternatives (e.g., Lacina et al. 2023, Jördens et al. 2020). For the most part, these studies have argued in favor of the two-stage model. In fact, until recently, the two-stage model was perhaps the only existing model for the selection of alternatives. We now briefly present our recent study which presented evidence against the two-stage model and advances the immediate-access model as an alternative.

3. The immediate-access model. In Muxica & Harris (to appear), we identified a number of conceptual challenges for the two-stage model. First, focus is an extremely flexible phenomenon; almost any element of the same type-theoretic category as the focus can serve as an alternative given proper contextual support. In cases of broad focus, complex constituents that contain multiple words can be marked for focus, as can entire utterances. At this point, it is not clear how lexical-level priming is meant to generate alternatives in cases of broad focus.

Second, recall that the two-stage model is priming dependent. This means that the context-sensitive selection of alternatives in the second stage depends upon the lexical activation generated by semantic priming in the first stage. For example, imagine a context in which a group of artist has painted a mural that depicts a tank driving through a meadow. In such a context, *tank* is clearly a relevant alternative to *flowers* in (5). Presumably, *flowers* and *tank* are semantically unrelated, and thus these words cannot be associates in the two-stage model.

(5) **A.** What did Simon paint on the mural?
B. Simon only painted [the flowers]$_F$ on the mural

It's unclear how non-associate alternatives such as *tank* could enter the alternative set without first entering as a semantic prime under the two-stage model. On these grounds, we have argued that context immediately guides the selection of alternatives in focus processing and that the priming effects observed in past studies are at least partially independent effects of low-level lexical activation, unrelated to the interpretation of focus.

A number of implementational questions arise once the *priming dependent* aspect of the two-stage model is relaxed. In Muxica & Harris (to appear), we addressed the issue of when non-associate, i.e., semantically unrelated, words are given as alternatives in the context. Thirty items consisting of two speaker dialogues were presented to listeners in a cross-modal probe recognition experiment. The first speaker in the dialogue introduced an associate alternative (*muffin*) and a non-associate alternative (*pistol*) relevant for the second speaker's use of focus (*violin*). Additionally, the first speaker mentioned a non-associate (*house*) which served as a control. Unlike previous studies, all the probe words were given in the discourse, arguably allowing us to better disentangle the effects of focus from the effects of a discourse new word (see also Hoeks et al. 2023). These words served as probes in the recognition task and were controlled for various lexical factors (e.g., frequency, number of morpheme, orthographic neighborhood size, etc.). On each trial, subjects performed a probe recognition task immediately (i.e., 0ms SOA) after the presentation of focus.

(6) *Sample dialogue from Muxica & Harris (to appear)*
A. Jonah brought the guitar and the pizza to band practice at the new house
B. No, he only brought the [violin]$_F$

(7) *Probe words*
Associate: MUFFIN
Non-Associate: PISTOL
Control: MOVIE

As expected, responses to probe words were faster for alternatives (*muffin* and *pistol*) than for non-alternatives (*house*). Crucially, there was no evidence of a difference between the associate and non-associate alternatives. In fact, Bayes factors provided evidence against the hypothesis that response times to associate and non-associate conditions were different. We took these results to be incompatible with the two-stage model and in support of the immediate-access model. We suggested that the previous literature may have had confounded semantic association with alternative status, obscuring the early effect of contextual relevance on focus computation.

The immediate-access model crucially predicts that the discourse context is recruited to select alternatives immediately upon encountering focus. However, we did not explicitly manipulate the discourse context to explore another crucial prediction of the model, namely, that focus alternatives are directly determined by contextual information. The current study aims to address this key prediction. We adapted the design and materials from Muxica & Harris (to appear), explicitly manipulating the alternative status of the non-associate probe word in the discourse context. To preview, we found that response times to non-associate probe words was modulated by their contextual relevance as alternatives. We take this finding as further evidence that a contextually-relevant alternative set is generated immediately after focus is encountered, partially replicating our previous results and further supporting the main predictions of the immediate-access model.

4. Experiment.

4.1. MATERIALS AND METHOD. The 3 condition design from Muxica & Harris (to appear) was adapted into a 2x2 factorial design (Context x Probe Word). Our materials consisted of 28 pairs of audio dialogues and 28 pairs of written probe words.

(8) *One-Alternative (One-Alt) context*
A1. After eating leftover pizza, Jonah brought the guitar to band practice at the new house
Two-Alternative (Two-Alt) context
A2. Jonah brought the guitar and the pizza to band practice at the new house

(9) *Target sentence for both contexts*
B. No, he only brought the [violin]$_F$

(10) *Probe words*
Associate: GUITAR
Non-Associate: PIZZA

Written probe words were presented in one of two conditions. In the Associate condition, the probe word (*guitar*) was closely related to the word in focus (*violin*). While in the Non-Associate condition, the probe word (*pizza*) was not closely related to the word in focus. In both conditions, the probe word was mentioned in the preceding audio dialogue. And thus, the correct response to the probe recognition task was always "Yes" on critical trials, precluding the possibility of a response bias confound between conditions. Filler items balanced the overall distribution of "Yes" and "No" responses.

Audio dialogues consisted of a context sentence in one of two conditions (8), followed by a target sentence (9). In both contexts, Speaker A's utterance described a situation using the Associate and Non-Associate probe words. Speaker B responded using corrective associated focus (i.e., *No + only*). In the response, the focused word provides the only new information as the rest of the content words were previously given in the context.

In the Two-Alt context, the Associate and Non-Associate probe words were conjoined arguments of a main verb (e.g., *Jonah brought the guitar and the pizza*), making both probe words contextually relevant focus alternatives to the word in contrastive/corrective focus (*violin*). However, in the One-Alt context, only the Associate probe word (*guitar*) appeared as an argument of a main verb, making it the only contextually relevant focus alternative in the target sentence. Crucially, although the Non-Associate probe word was not a contextually relevant alternative, it was still mentioned by Speaker A in the One-Alt condition.

The same recordings from the previous study were used for all of Speaker B's utterances and for all of Speaker A's utterances in the Two-Alt context. As the One-Alt context is novel to this study, twenty-eight new contexts were recorded from Speaker A. As in our previous study, Speaker B was a male speaker trained in the production and transcription of English intonation using the Tones and Breaks Indices system (ToBI; Pierrehumbert & Hirschberg 1990). He was instructed to use an L+H* pitch accent when producing the final focused word.[2] Speaker A was the same female speaker from the previous study. She had not received any formal prosodic training previously and was instructed to produce the items naturally, rather than with any specific contour.

The 28 pairs of probe words (Associate and Non-Associate) were identical to the previous study and were controlled for length, frequency, number of morphemes, and orthographic neighborhood size (Balota et al. 2007, Brysbaert & New 2009).[3] We also controlled for the semantic association between each probe word and their corresponding focus using Latent Semantic Analysis (LSA; Landauer & Dumais 1997) and an internet norming study in which UCLA undergraduates provided semantic similarity judgements on a 7-point Likert scale.[4]

The same fifty-six two-speaker filler dialogues from the previous study were used, resulting in a final list of 84 items. The probe word was not mentioned in 42 of these filler items in order to balance the distribution of responses across the study. Both of our speakers were instructed to produce the filler items naturally rather than with any specific contour.

4.2. ANALYSIS. Before discussing the results of the pilot and the main in-person studies, the general procedure for data cleaning and analysis is described. All subjects included in the final data set answered at least 75% of questions on the probe task and comprehension questions correctly. Only correct responses to the probe recognition task were retained for the response time analysis. Response times below 200ms and above 2,500ms were excluded from our analysis. Responses faster than 200ms were assumed to reflect insufficient processing of the stimulus. Responses slower than 2,500ms were not taken to exclusively involve the early moments of processing focus relevant to our research question. This exclusion criteria resulted in less than 10% data loss across conditions.

Bayesian mixed-effects models were used to analyze accuracy and log transformed response time data via the brms package (Bürkner 2017) in the R software environment (R Core Team 2023).[5] No divergent chains were observed and all models converged with $\hat{R} \approx 1$ and sufficient

[2]In English, the L+H* pitch accent is associated with both the presence of focus and sentence final Nuclear Pitch Accent (Büring 2016). Given that focus always occurred sentence final, our stimuli are technically ambiguous with respect prosody. However, the presence of the focus particle *only* and the givenness of the surrounding non-focused material eliminated any possible interpretive ambiguity. See Muxica & Harris (to appear) for further discussion.

[3]Pairwise differences between each of the probe word conditions were evaluated with a Bayesian t-test and no reliable differences in any measure was observed (89% CrI, BF<1) for each comparison.

[4]For the LSA norming, pairwise differences between between each of the probe word conditions were evaluated with a Bayesian t-test. As intended, Associate probe words were found to have a higher cosine similarity to the focused word than Non-Associate probe words (Med=0.5, $CrI_{89\%}$=[0.45, 0.54], BF>1,000). The results of the norming study were integrated into our analysis. We fit Bayesian mixed-effects models for both response time and accuracy using semantic similarity ratings as a random effect, specifically the log transformed difference between between the associate and non-associate probe. In all cases, this addition neither improved model fit nor changed the qualitative pattern of results. We have included these models on the OSF repository for this paper (https://osf.io/kmdu3/?view_only=92f204af070c433d9db6f61ab7c30123).

[5]Comparable frequentist mixed-effects models were also fit using the lme4 package (Bates 2010). In all cases, the results were qualitatively the same and thus we do not report these models in the main text. We have made these models available on the OSF repository for this paper.

Effective Sample Sizes (ESS) for each parameter. Posterior predictive checks graphically confirmed that the model was an appropriate fit of the response variable.

4.3. INTERNET PILOT. A pilot experiment was conducted over the internet using a subset of the materials. This not only allowed us to test the design and validate the central effects in a different setting, but also to generate an informative prior for use in the Bayesian analysis of the main in-person experiment.

4.3.1. PARTICIPANTS. Forty-five self-reported native English speaking undergraduates were recruited from the University of California Los Angeles Psychology Department Subject Pool and given course credit in exchange for participation.

4.3.2. PROCEDURE. Four lists of 12 items (selected from the set of 28 critical items) were created in a counterbalanced design. Eighteen items from the set of 56 filler items were added to these lists yielding 30 trials per list. PCIbex was used to host the experiment (Zehr & Schwarz 2018). On each trial, subjects were presented with a central fixation cross while the audio dialogue played. Immediately after the audio completed, subjects were presented with a written probe word in the center of the screen. Subjects were instructed to provide their response to the probe word as quickly as possible without sacrificing accuracy. There was no explicit timeout for long responses. After each trial, subjects were given the opportunity to take a self-paced break. The pilot took approximately 12 minutes to complete on average.

4.3.3. RESULTS. In the One-Alt context, probe task accuracy task was higher on average for the Associate (M=94%, SE=2) than the Non-Associate (M=78%, SE=4) probe word. The same pattern held for the Two-Alt context; probe task accuracy was higher on average for the Associate (M=95%, SE=2) than the Non-Associate (M=86%, SE=3) probe word. However, according to a Bayesian logistic regression model, there was no evidence for an effect of probe word (Med=-0.898, $CrI_{89\%}$=[-2.306, 0.483]), context (Med=-0.134, $CrI_{89\%}$=[-0.77, 0.511]), or an interaction between the two (Med=-0.188, $CrI_{89\%}$=[-0.844, 0.425]).

Log transformed response times were subjected to a Bayesian linear mixed effects model. Main effects are presented first, followed by the interaction. As shown in Table 1 in the next section, the model indicated that, across contexts, the Non-Associate probe word (M=1189ms, SE=45) yielded slower response times than the Associate probe word (M=1058ms, SE=38). There was no evidence that One-Alt and Two-Alt contexts elicited different reaction times.

The magnitude of the response time difference was much larger in the One-Alt context (131ms) than the Two-Alt context (54ms), which was in line with an interactive effect between probe words in the two contexts. Although there was no evidence of an interaction in the main model, investigation of the estimated marginal means of the model was consistent with the predicted interactive effect. Response times to the Associate probe were faster than the Non-Associate probe in the One-Alt context (Med=-0.020, HPD=[-0.035, -0.004]), but there was no difference between probe words in the Two-Alt context (Med=-0.009, HPD=[-0.023, 0.005]).

4.3.4. DISCUSSION. Despite its limited the power, the pilot provided preliminary support of a key prediction of the immediate-access model. Response times to a probe word depended on whether the word was presented as a contextually-relevant alternative, rather than on its lexical association with the word in focus. Crucially, the effect was observed immediately after presentation of focus.

Under the two-stage model, response times to a given probe word should solely depend upon semantic priming at early stages of focus interpretation, rather than contextually-determined alternatives status. At this point, it is not entirely clear how such a model could explain the immediate advantage for contextually-determined alternatives. However, evidence for the crucial interaction was only observed in the marginal means, and we now turn to the main, in-person experiment with increased power and a more controlled experimental setting.

4.4. IN-PERSON EXPERIMENT.

4.4.1. PARTICIPANTS. Fifty-one self-reported native English speaking undergraduates were recruited from the same population population as the pilot and given course credit in exchange for participation.

4.4.2. PROCEDURE. The experiment was presented using Linger (Rhode 2001) and was hosted on a Linux desktop computer in a sound-attenuated booth. Subjects listened to the audio through Seinheiser HD280 Pro wired headphones and provided all responses using a PS/2 keyboard. On each trial, subjects were presented with a central fixation cross while the audio dialogue played. Immediately after the audio completed, a written probe word was presented in the center of the screen. Subjects were instructed to provide this response as quickly as possible without sacrificing accuracy, but there was no explicit timeout for long responses. In addition, multiple choice comprehension questions were presented after a third of the trials. Subjects were instructed to prioritize accuracy over speed in answering these questions. As in the pilot, subjects were given the opportunity to take a self-paced break after each trial. The experiment took approximately 30 minutes to complete on average.

4.4.3. RESULTS. Accuracy results of the main study was comparable to that of the pilot study. In the One-Alt context, probe task accuracy task was higher on average for the Associate (M=92%, SE=1) than the Non-Associate (M=73%, SE=3) probe word. In the Two-Alt context, probe task accuracy was also higher on average for the Associate (M=90%, SE=2) than the Non-Associate (M=94%, SE=1) probe word.

According to a Bayesian logistic regression model, there was no reliable evidence indicating an effect of probe word (Med=-0.58, $CrI_{89\%}$=[-1.15, 0.01], BF=0.89) or an effect of context (Med=-0.34, $CrI_{89\%}$=[-0.68, 0.01], BF=1.77). However, there was strong evidence for an interaction between probe word and context (Med=-0.66, $CrI_{89\%}$=[-1.00, -0.33], BF=51.29). The interaction was further supported by the estimated marginal means of the model, which indicated that the accuracy for the Non-Associate probe was lower than the Associate probe in the One-Alt context (Med=2.48, HPD=[0.81, 4.24]), but did not differ in the Two-Alt context (Med=-0.163, HPD=[-1.79, 1.39]).

Log transformed response times were subjected to a Bayesian linear mixed effects model. The model, summarized in Table 1, indicated that, across contexts, the Non-Associate probe word yielded slower response times than the Associate probe word (Non-Associate: M=1163ms, SE=39; Associate: M=1014ms, SE=24; BF>100). There was no evidence that response times differed between the One-Alt and the Two-Alt context (BF=0.662). However, there was strong evidence in favor of the crucial interaction effect between probe word and context predicted by the immediate-access model (BF>100). The interaction was further supported by the estimated marginal means of the model. Response times to the Associate probe were faster (149ms) than the Non-Associate probe in the One-Alt context (Med=-0.024, HPD=[-0.032, -0.015]), but the difference between probe words in the Two-Alt context (45ms) was not reliable (Med=-0.006, HPD=[-0.014, 0.001]).

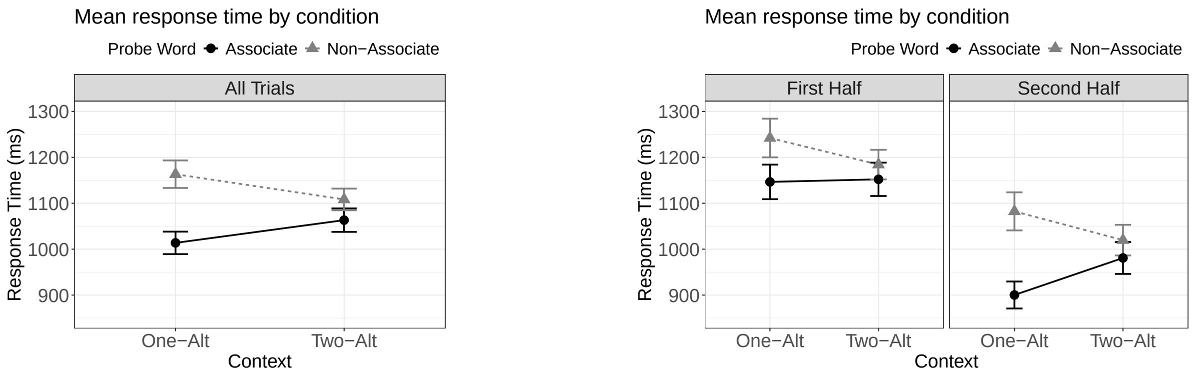
	Pilot		In-person		
Parameter	**Median**	**89% CrI**	**Median**	**89% CrI**	**BF**
Intercept	1.941	[1.930, 1.952]	1.938	[1.931, 1.945]	NA
Non-Associate vs. Associate	0.007	[0.002, 0.012]	0.007	[0.005, 0.010]	>100
One-Alt vs. Two-Alt	0.000	[-0.004, 0.004]	0.001	[-0.001, 0.002]	0.662
Probe Word x Context	0.003	[-0.001, 0.006]	0.004	[0.002, 0.006]	>100

Table 1: Results for pilot and in-person studies from Bayesian linear mixed effects regression model on log response times with maximal random effect structures and sum-coded predictors. In the pilot study, uninformative (flat) priors were used and so no Bayes Factor could be computed. The model was run with 5,000 iterations and a 1,000 iteration warm up, and converged with $\hat{R} = 1$ and at least an 2,500 ESS per parameter. In the in-person study, informative priors from the pilot study were used. The model was run with 12,000 iterations and a 2,000 iteration warm up, and converged with $\hat{R} = 1$ and an ESS $\geq$4,000 per parameter. Bayes Factor (BF) was computed over a null point estimate using the Savage-Dickey density ratio.

A central prediction of the immediate-access model is that context can modulate whether a semantically unrelated word is immediately accessed as an alternative of a word in focus. The interaction we observed in response times is highly compatible with this prediction. The cost observed for Non-Associate over Associate probes in the One-Alt context was absent in the Two-Alt context. However, the strongest manifestation of this prediction is in a "3 against 1" pattern, in which the One-Alt Non-Associate condition elicits longer reaction times than the other three conditions. While the crucial penalty for Non-Associate over Associate probes was evident in the results, the precise pattern is less clear. To explore the interaction in more detail, we fit an exploratory model with trial half order as an interactive predictor and random effect, which yielded a better fit than the original.

In the trial half model, the overall qualitative pattern of results did not change: Non-Associate probes yielded slower response times than Associate probes (BF>100), response times did not differ between contexts (BF=0.686), and there was strong evidence in favor of an interaction between probe word and context (BF>100). In addition, response times in the first half of trials differed from those in the second half (BF>1,000). Despite an impressionistic difference between trial halves, a three-way interaction between context, probe type, and trial order was not supported by the model.

Figure 1b depicts the response times by condition along trial half. Unsurprisingly, response times in the second half of trials (M=987ms, SE=17) were much faster than in the first half of trials (M=1178ms, SE=18) overall, suggesting that subjects became more adept at the task over time. More importantly, the predicted interaction was observed in both halves. However, the pattern of interaction in the first half of trials more closely aligns with the precise three against one pattern predicted by the immediate-access model, in which the One-Alt Non-Associate condition elicited slower responses than the other three conditions. In the second half, there was an additional response time advantage for the One-Alt Associate condition. We further speculate on the interpretation of the differential effects across conditions in the discussion.



(a) Collapsing across trial order

(b) Splitting along trial order

Figure 1: Main study. Mean response time results by condition. Error bars indicate standard errors.

4.4.4. DISCUSSION. The results of the main experiment provide strong evidence in favor of the immediate-access model, largely replicating the pilot study. The most important finding is that the contextual status of a given probe word as a focus alternative determined response speed across conditions. Specifically, the Associate the Non-Associate probes elicited similar response times in the Two-Alt contexts, when both probe types were presented as contextually-relevant focus alternatives. However, in the One-Alt context, where only the Associate probe was a relevant focus alternative, we found a response time penalty for the Non-Associate probe word. Crucially, this advantage for alternatives over non-alternatives manifested immediately after the word in focus was encountered.

Further, our results cannot be explained in terms of semantic priming from the element in focus. In particular, the response time advantage for the Non-Associate probe word tracked our manipulation of the discourse context rather than semantic association with the element in focus. Only the immediate-access model directly predicts that context-sensitivity should manifest immediately. In contrast, the two-stage model predicts that response speed should solely depend upon semantic association at early stages.

As mentioned, an interaction was observed in both halves of the study. However, the character of the interaction differed slightly in later trials. Though there are many possible explanations for effects of exposure, we speculate that subjects developed a strategy to better predict what the upcoming focus will contrast with from the context and the target sentence. In each condition, the target sentence presented just one new content word (e.g., *violin*), while the remainder of the sentence frame presented discourse-given information (*No, he only brought the* ...). In the Two-Alt context, subjects may have been able to predict that the focused word would relate to one of the alternatives (e.g., *guitar* or *pizza*), but would not have enough information to prioritize one over the other. In the One-Alt context, only the Associate (*guitar*) appeared previously with the sentence frame, thus allowing the subject to make a more fine-grained prediction that the focused word would contrast with the Associate, potentially re-activating it in memory. In other words, subjects learned to make more specific predictions for the Associate in the One-Alt condition, which gave them an advantage when the prediction was confirmed. In support of this interpretation, subjects became more accurate just in the One-Alt Associate condition in the second half of the experiment (M=95%, SE=2) compared to the first half (M=89%, SE=2). In all other conditions, accuracy decreased in the second half.

This line of reasoning is supported by the literature. There is considerable evidence that readers and listeners continuously form predictions during comprehension (e.g., Staub 2015, Kuperberg & Jaeger 2016; for review) and that they exhibit processing difficulties when their predictions fail to be validated (e.g. Rich & Harris 2021, 2023; among many others). It is likely that many sources of information contribute to prediction formation and that focus, and perhaps information structure more generally, is another strong factor. In addition, subjects adapt their processing strategies when presented with mismatching contrastive accent (Roettger & Franke 2019, Nakamura et al. 2019) over the course of an experiment. Far from being an experimental nuisance, such strategies can be construed as rational adaptations to the task and may even reflect fundamental aspects of language processing. Whatever the case may be, the effect of trial order observed here raises interesting general questions about the potential relationship between focus and predictability, as well as how subjects may strategically adapt to the structure of the experiment.

5. Conclusion. Across two cross-modal probe recognition experiments, alternative status of a given probe word was found to immediately modulate response speed. The results strongly support the view that an unrelated word can be immediately construed as a focus alternative depending on the context.

We take these results to strongly support the immediate *construction* of focus alternatives by context, as predicted by the immediate-access model. The results provide further evidence against versions of a two stage model in which focus alternatives are initially determined by context-insensitive lexical association with the word in focus.

In all, we believe that the results showing the immediate availability of a contextually determined alternative set strongly coheres with the anaphoric component inherent in calculating the effect of focus on a sentence in the context of utterance (Rooth 1992). Indeed, constructing contextually-relevant focus alternatives might well constitute a grammatically mandatory operation that cannot be delayed during interpretation (Frazier 1999). We leave this, and other questions of focus interpretation and the architecture of the language processing system, to future research.

References

Balota, David A, Melvin J Yap, Keith A Hutchison, Michael J Cortese, Brett Kessler, Bjorn Loftis, James H Neely, Douglas L Nelson, Greg B Simpson & Rebecca Treiman. 2007. The English Lexicon Project. *Behavior Research Methods* 39. 445–459.

Bates, Douglas M. 2010. lme4: Mixed-effects modeling with R.

Beaver, David I & Brady Z Clark. 2009. *Sense and sensitivity: How focus determines meaning*. John Wiley & Sons.

Braun, Bettina & Lara Tagliapietra. 2010. The role of contrastive intonation contours in the retrieval of contextual alternatives. *Language and Cognitive Processes* 25(7-9). 1024–1043.

Brysbaert, Marc & Boris New. 2009. Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* 41(4). 977–990.

Büring, Daniel. 2016. *Intonation and meaning*. Oxford University Press.

Bürkner, Paul-Christian. 2017. brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software* 80(1). 1–28.

Frazier, Lyn. 1999. *On sentence interpretation*, vol. 22. Springer Science & Business Media.

Gotzner, Nicole & Katharina Spalek. 2019. The life and times of focus alternatives: Tracing the activation of alternatives to a focused constituent in language comprehension. *Language and Linguistics Compass* 13(2). e12310.

Gotzner, Nicole, Isabell Wartenburger & Katharina Spalek. 2016. The impact of focus particles on the recognition and rejection of contrastive alternatives. *Language and Cognition* 8. 59 – 95.

Hoeks, Morwenna, Maziar Toosarvandani & Amanda Rysling. 2023. Processing of linguistic focus depends on contrastive alternatives. *Journal of Memory and Language* 132. 104444.

Husband, Matthew & Fernanda Ferreira. 2016. The role of selection in the comprehension of focus alternatives. *Language, Cognition and Neuroscience* 31(2). 217–235.

Jördens, Kim A., Nicole Gotzner & Kathrina Spalek. 2020. The role of non-categorical relations in establishing focus alternative sets. *Language and Cognition* 12(4). 729–754.

Krifka, Manfred. 1992. A framework for focus-sensitive quantification. *Semantics and Linguistic Theory* 2. 215–236.

Kuperberg, Gina R & T Florian Jaeger. 2016. What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience* 31(1). 32–59.

Lacina, Radim, Patrick Sturt & Nicole Gotzner. 2023. The comprehension of broad focus: Probing alternatives to verb phrases.

Landauer, Thomas K & Susan T Dumais. 1997. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review* 104(2). 211.

Muxica, Christian J. & Jesse A. Harris. to appear. Constructing alternatives: Evidence for the early availability of contextually relevant focus alternatives. In Nicole Gotzner, Jesse Harris, Richard Breheny & Yael Sharvit (eds.), *Alternatives in grammar and cognition*, Palgrave.

Nakamura, Chie, Jesse A. Harris & Sun-Ah Jun. 2019. Listeners' beliefs influence prosodic adaptation: Anticipatory use of contrastive accent during visual search. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren. (eds.), *Proceeding of the 19th International Congress of Phonetic Sciences*, 447–451.

Pierrehumbert, Janet & Julia Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In *Intentions in communication*, 271–311. MIT Press.

R Core Team. 2023. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing Vienna, Austria. https://www.R-project.org/.

Rhode, Doug. 2001. Linger. *A flexible platform for language processing experiments* .

Rich, Stephanie & Jesse A. Harris. 2021. Unexpected guests: When disconfirmed predictions linger. In *The proceedings of the 43rd annual meeting of the cognitive science society*, 2246–2252. Vienna, Austria.

Rich, Stephanie & Jesse A Harris. 2023. Global expectations mediate local constraint: evidence

from concessive structures. *Language, Cognition and Neuroscience* 38(3). 302–327.
Roettger, Timo B & Michael Franke. 2019. Evidential strength of intonational cues and rational adaptation to (un-) reliable intonation. *Cognitive science* 43(7). e12745.
Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1(1). 75–116.
Staub, Adrian. 2015. The effect of lexical predictability on eye movements in reading: Critical review and theoretical interpretation. *Language and Linguistics Compass* 9(8). 311–327.
Zehr, Jeremy & Florian Schwarz. 2018. Penncontroller for Internet Based Experiments (IBEX).