Semantic-Pragmatic Adaptation to Variable Use of Temporal Expressions

Yuxin Cao
Department of Linguistics, University College London, United Kingdom
yuxin.cao.23@alumni.ucl.ac.uk

Sebastian Schuster
Faculty of Computer Science, University of Vienna, Austria
sebastian.schuster@univie.ac.at

Abstract

Previous work has shown that listeners rapidly update their interpretations of vague expressions such as quantifiers and expressions of uncertainty when they observe a speaker's usage of such terms. However, previous studies focused on instances involving two reasoning steps: inferring a world state from a visual scene and communicating the world state. Based on these experiments, it has been argued that listeners infer speaker-specific mappings between world states and vague expressions rather than listeners making inferences about how the speaker infers the world state from a visual scene. Here, we extend the work on semantic-pragmatic adaptation to a new class of expressions, namely vague temporal expressions, such as *for a bit* and *for a while*, and employ an experimental paradigm in which the inference from the visual scene to the world state is deterministic. We replicate previous findings in this setting, suggesting that adaptation indeed targets semantic representations.

Keywords: semantic/pragmatic adaptation; experimental pragmatics; temporal expressions

Introduction

Successful communication has a collaborative nature (Grice, 1975). For instance, in establishing a shared reference, interlocutors converge and condense their descriptions of the same object as the iterative communication process takes place (Clark & Wilkes-Gibbs, 1986). This is also true for vague expressions, such as quantifiers to describe proportions or expressions of uncertainty to describe probabilities. For example, Yildirim, Degen, Tanenhaus, and Jaeger (2016) showed that listeners rapidly update their expectations of a speaker's use of the quantifiers *some* and *many* when they observe the speaker's use of these expressions: When participants were exposed to a speaker who always uttered "*Some* of the candies are green" to describe a scene where a bowl of candies was filled with approximately equal number of green and blue candies, listeners expected the speaker to use *some* to describe a greater proportion than when they were exposed to a speaker who always uttered "*Many* of the candies are green" to describe the same situation (see also Heim, Peiseler, and Bekemeier (2020)). In a similar vein, Schuster and Degen (2020) showed that listeners also update their expectations of a speaker's use of uncertainty expressions such as *might* and *probably* when they were used to describe different probabilities of getting a gumball of a certain color. They furthermore showed that the updated speaker expectations also affect the interpretation of the speaker's utterances, and based on simulations with a Bayesian model of this adaptation process, they argued that this kind of semantic-pragmatic adaptation is a result of listeners updating their beliefs about the speaker's utterance preferences and the speaker's semantic representations, i.e., the mapping between world states and utterances.

Recent work has also shown that this kind of semantic-pragmatic adaptation takes place with gradable adjectives. Xiang, Kramer, and Kennedy (2020) showed that listeners update their interpretations of gradable adjectives such as *tall*, *bent*, and *plain* based on observed usage, and Pezzelle and Fernández (2023) showed that listeners also update their interpretations in an interactive setting where they have to infer what a speaker means by gradable adjectives such as *big*.

One property that all of the existing studies have in common is that in the experimental setting, there are two processes involved: First, the fictional speaker has to infer the world state w from a visual scene. For example, in the case of Yildirim et al. (2016), the speaker has to infer the proportion of blue and green candies from a bowl with blue and green dots. Second, the speaker has to choose an utterance u to communicate the world state w. Similarly, in the study by Schuster and Degen (2020), the speaker has to infer a probability of getting a blue gumball from a visual representation of a gumball machine filled with a certain proportion of blue and orange gumballs, and in the studies by Pezzelle and Fernández (2023) and Xiang et al. (2020), the speaker has to infer the size, degree of bentness, or degree of plainness of different objects from a visual representation before choosing an utterance. Considering that inferring the world state in all of these scenarios is a potentially noisy process, it is possible that participants in the adaptation experiments do not actually draw inferences about the speaker's language use but rather assume a fixed semantics and make inferences about the speaker's ability to infer the world state. For example, it could be that in the experiments with quantifiers, participants in the *some*-biased condition, in which the speaker uses *some* to describe greater proportions than in the *many*-biased condition, assume that the speaker underestimates the true proportion and that the speaker's use of *some* is a result of this assumed proportion. Similarly, in the case of uncertainty expressions, participants may actually make inferences about the speaker's ability to

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

	Queuing Scenario	Timer Scenario
Utterances	I'm in the queue for check-in, but it looks like I'll be here (for a bit / for a while / for some time).	I don't feel well, so I'm gonna lie down (for a bit / for a while / for some time). I'll get back to you later.
Time Spans	{5, 10, 15, 20, 30, 45, 60, 90, 120}	{5, 10, 15, 30, 45, 60, 75, 90, 120}

Table 1: Details of the two scenarios used in Experiment 1.

estimate probabilities rather than making inferences about their semantic representations of uncertainty expressions.

In this work, we aim to tease apart these two potential explanations and investigate whether the observed adaptation behavior in semantic-pragmatic adaptation is a result of listeners making inferences about the speaker's semantic representations or whether the results from previous studies are better explained by participants making inferences about the speaker's ability to infer the true world state.

We do this by adapting the exposure-test paradigm that was used in Yildirim et al. (2016) and Schuster and Degen (2020) to scenarios involving the communication of a time span using vague temporal expressions such as *for a bit* and *for a while*. Crucially, in our setting, it is clear to participants that the fictional speaker sees a precise duration, so the speaker does not have to estimate the world state from a visual scene.

In the following sections, we first provide a brief overview of the semantics of temporal expressions. In Experiment 1, we estimate what listeners expect of a generic speaker to determine naturalistic scenarios for the adaptation experiment, and in Experiment 2, we then use an exposure-test paradigm in which listeners observe the speaker's usage before providing their expectations of the exposure speaker's use of temporal expressions in different situations.

Temporal Expressions

In this work we are considering temporal adverbial clauses such as *for a bit* or *for a while*. The key functions of these expressions are *measurement* and *locating* (Kamp & Reyle, 1993). The *measurement* function refers to the duration of time (Brée, Feddag, & Pratt, 1993), while the *locating* function describes the relationship between this duration and the time axis (Móia, 2005).

While some temporal expressions (e.g., *for 3 minutes*) are very precise, many are vague expressions (Kennedy, 2007), which makes them an interesting domain for studying semantic-pragmatic adaptation. Since there is no agreed meaning of how long *a bit* or *a while* is, we expect that different language users have different expectations about their use.

For the purpose of this work, we will assume that quantifiers have a threshold semantics (Swanson, 2006). This means that a phrase such as *for a bit* is true if the time duration that the adverbial clause is modifying is above a contextually specified threshold θ_{bit}.

Following the model by Schuster and Degen (2020), we hypothesize that one of the contextual factors affecting thresholds such as θ_{bit} is the speaker's identity and that listeners have malleable beliefs of a specific speaker's thresholds.

Experiment 1

The goal of Experiment 1 is to estimate listeners' expectations of the use of temporal expressions to describe different durations. Here we used two different scenarios and the secondary purpose of this experiment is to determine which of the two scenarios seems better suited for the subsequent adaptation experiment.

Participants

We recruited 40 participants (20 per scenario) via Prolific (https://www.prolific.com/). We filtered for native English speakers born in the UK. Participants were paid £1 for an estimated 5 minutes of participation (∼£12/hr).

Procedure

Participants were first presented with a description of the context in order to introduce them to the scenario. In one scenario, the *queuing* scenario, participants were presented with a scene at an airport where a man was queuing for a check-in counter and crucially, there was a sign showing the estimated wait time (see left panel of Figure 1). In the other scenario, the *timer* scenario, participants were presented with a scene with a woman and her phone showing the countdown of a timer (see right panel of Figure 1). Each participant only completed one of the two scenarios. On each trial, participants were presented with a picture showing the overall scenario. In the *queuing* scenario, there was a sign indicating the expected wait time in minutes; in the *timer* scenario, there was a timer on the phone showing the remaining time. Participants were asked to distribute 100 points across four potential utterances to indicate what they thought the speaker would utter in the given situation. The choices consisted of three temporal expressions and a blanket *something else* option to indicate that none of the utterance choices was appropriate in this situation. Table 1 shows the utterances used in the two scenarios along with the 9 time durations used across trials. Each participant provided two ratings for each time duration, resulting in 18 trials per participant in total. Participants completed trials on their own devices in full-screen mode to minimize distractions.

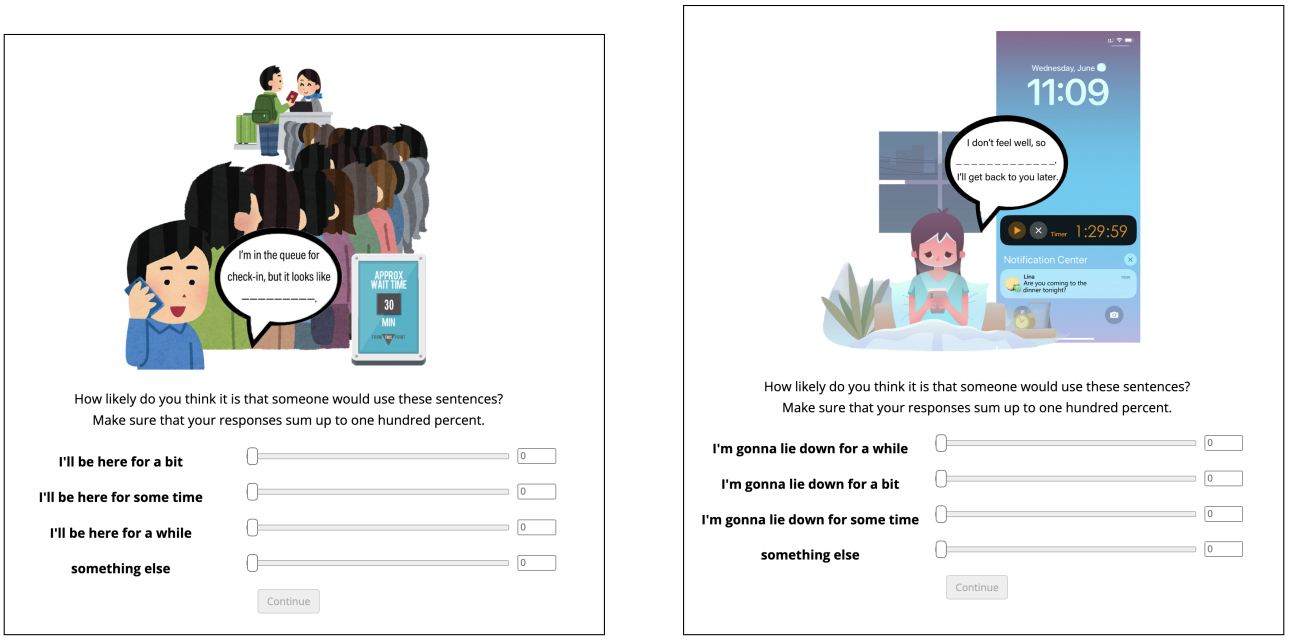

Figure 1: Trials in Experiment 1 for *queuing* scenario (left) and *timer* scenario (right).

Materials

We chose three common vague temporal phrases, namely *some time*, *a while*, and *a bit*, where we expected that they would be used to describe time durations of different length. The target temporal phrases were placed at the end of a sentence. We used an informal speech style to reflect everyday language, using colloquial contracted form of *I'll* / *I'm*. The utterances were highlighted in a speech bubble next to the sign or the timer that displayed the length of the time interval.

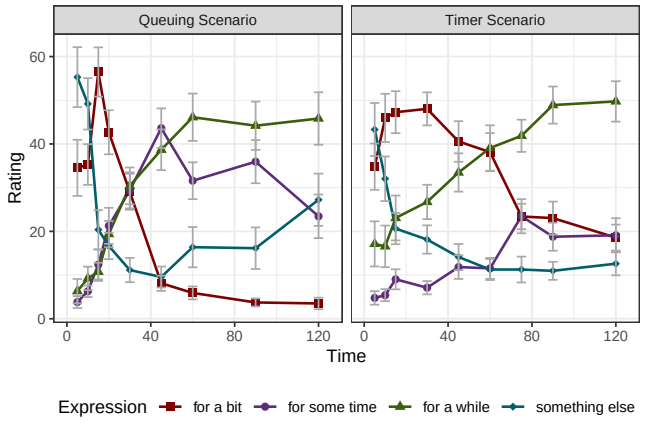

Figure 2: Mean ratings of the three tested utterances for different time durations in the two scenarios in Experiment 1. Error bars indicate standard errors of the mean.

Results

Exclusions Participants who provided likelihood ratings that were almost identical across all nine situations were excluded, as this likely indicated insufficient attention. This manual filtering served as a substitution for attention check trials. This resulted in 3 exclusions in the *queuing* scenario, and 5 exclusions in the *timer* scenario.

Figure 2 shows the mean ratings across participants in Experiment 1. In the queuing scenario (left panel), *for a bit* is rated the highest by participants when the sign shows 15 minutes (mean rating of 56.47), which then drops sharply to 8.12 when the sign shows 45 minutes. In contrast, *for a while* and *for some time* have overlapping rising trends for spans from 5 to 45 minutes. The mean ratings for *for a while* increase from 6.32 to 38.68, peaking at 46.12 at 60 minutes. As for *for some time*, the mean rating rises from 3.82 to 43.62, fluctuating between 23.44 and 43.62 for spans greater than 45 minutes.

We observe similar trends in the *timer* scenario (right panel): *For a bit* is rated relatively higher between 5 minutes and 60 minutes, ranging from 34.8 to 48.03, then the ratings decrease when it comes to longer periods. Meanwhile, the ratings for both *for a while* and *for some time* increase for longer periods, with *for a while* consistently rated higher than *for some time*.

In both scenarios, *for a while* and *for some time* show similar trends: as the time span increases, the mean ratings increase. Conversely, *for a bit* is rated as more likely for shorter time periods.

While in the timer scenario, *for some time* is rated less likely throughout, considering that the ratings for the two expressions show a similar usage tendency, we speculate that

	Utterances
For a bit	*... I'll be here for a bit.*
For a while	*... I'll be here for a while.*
Filler	*I can't talk right now.* I'm at the check-in for my flight, but I'll call you back later.

Table 2: Utterances used in Experiment 2. The recording during the exposure trials only consisted of the part that is in *italics*.

	for a bit		***for a while***		**Filler**	
	n	t	n	t	n	t
bit-biased	**5**	**30**	5	45	3	0
while-biased	5	15	**5**	**30**	3	0

Table 3: Number of exposure trials (n) per utterance (*for a bit*, *for a while*, *Filler*) and associated time duration (t) in the bit-biased vs. while-biased exposure blocks. Critical trials bolded.

the two expressions are seen to be interchangeable to a large extent.

Discussion

The main purpose of this experiment was a) to establish whether and where there is an inflection point at which ratings for one expression drop and ratings for another increase, and b) to determine which of the two scenarios might be better suited for the subsequent adaptation experiment.

Based on the results, we chose the *queuing* scenario for the adaptation experiment. In this scenario, there was a clear inflection at a duration of 30 minutes where both *for a bit* and *for a while* received similarly high ratings on average. Furthermore, both of these utterances received high ratings for very short or long durations, respectively. Additionally, the expected waiting time sign provides a more direct link to the duration the speaker is communicating since it is the only information source. The timer in the *timing* scenario may be seen as a lower bound for when the woman gets back to the other interlocutor.

Since the ratings for *some time* and *a while* showed similar patterns but *a while* generally received higher ratings, we limit the expressions in the adaptation experiment to *a bit* and *a while*.

Finally, this experiment also provides that there is evidence for variability in the use of temporal expressions. A manual analysis of the responses across participants suggests that different language users use different thresholds for these expressions.

Experiment 2

The goal of Experiment 2 is to test whether participants update their expectations of the speaker's use of temporal expressions after being exposed to a specific speaker. If this is the case, then this would provide evidence for adaptation behavior in the domain of temporal expressions, and furthermore, provide evidence that listeners actually make inferences about the speaker's semantics.

Participants

We recruited 60 participants (30 per condition) and paid each participant £1.70 for approximately 10 minutes of participation time (∼£10/hr). None of the participants from the previous experiment participated in this one.

Procedure

The experiment consisted of two blocks: an exposure block and a test block. In the exposure block, participants were shown pictures accompanied by recordings[1] of the speaker producing one of three utterances, as shown in Table 2. Two of these utterances contained one of the two temporal expressions, *for a while* or *for a bit*. Participants in the *bit-biased* condition always heard the speaker produce the utterance with *for a bit* when the wait time was 30 min; participants in the *while-biased* condition always heard the speaker produce the utterance with *for a while* when the wait time was 30 min. The filler utterances and the utterances with the other expression were presented as shown in Table 3. The "I can't talk right now" filler trials where the man did not need to wait at all were included to make it clear that the speaker is a rational speaker. The choice of time durations that was used in combination with the non-critical expression was chosen based on the results of Experiment 1: All combinations of time durations and expressions were rated very highly in that experiment.

To highlight the critical parts of the utterance and to avoid too repetitive recordings, the audio only contained the phrase with the temporal expression (e.g., *I'll be here for a while*). See Table 2 for the phrases that appeared in the recording. Participants had to listen to the recording before being able to proceed to the next exposure trial by clicking a button.

The test trials were identical to those in Experiment 1, except that they included only three response options: the utterance with *for a bit*, the utterance with *for a while*, and the blanket *something else* option.

Analysis

If listeners adapt and update their expectations, we expect that participants in the *bit-biased* condition provide high ratings for the utterance with *for a bit* for a wider range of time durations, including the duration of 30mins, compared to participants in the *while-biased* condition. Following Yildirim et al. (2016), we quantify this difference by fitting a spline with four knots for the rating curve for each expression and each participant. We then compute the area under the curve (AUC) for each of these splines. This is a measure

[1]The recordings, while created by a synthesized speech generator, closely mimicked human speech to the extent that the authors found that distinguishing them from human-generated recordings was challenging.

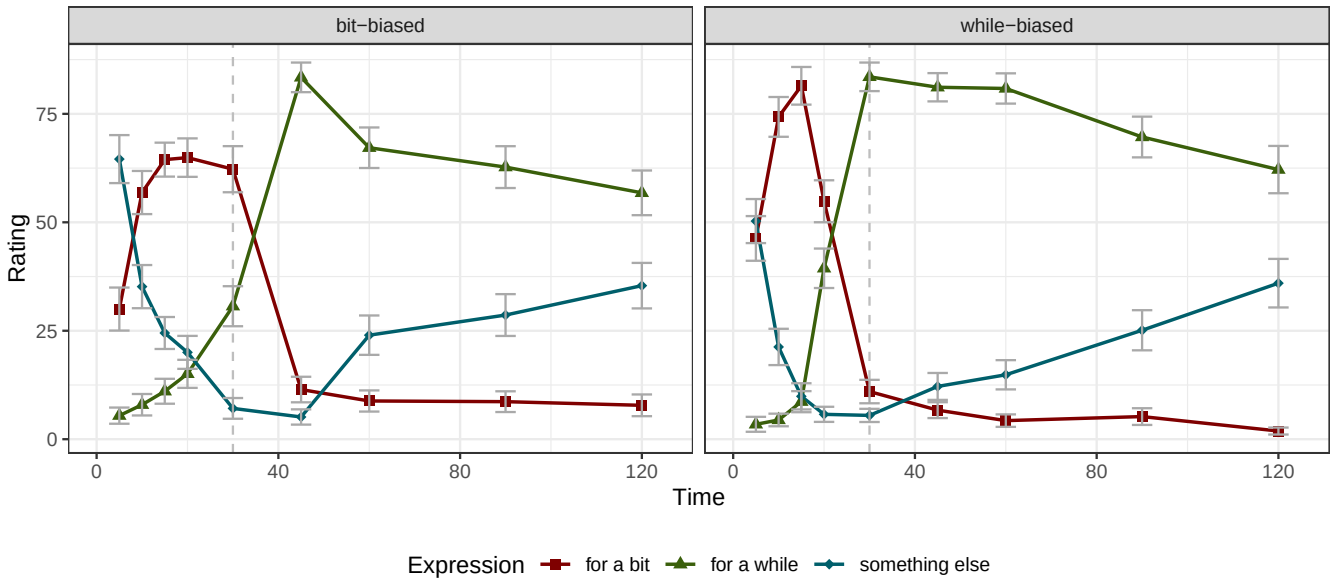

Figure 3: Ratings in both conditions of Experiment 2. The left panel shows the rating for the *bit-biased* condition and the right panel shows the rating for *while*-biased condition. The dotted vertical line indicates 30 min, the time span where the speakers across the two conditions used different expressions. Error bars indicate standard errors.

for how large of a range of time spans receives high ratings since the wider the range of highly rated time spans, the more area will be under the fitted curve. To measure the relation between the ratings for utterances with *a bit* and *a while*, we compute the difference of the AUCs for these two expressions. Based on the prediction about adaptation, we expect that this difference in AUCs will be greater for the *bit-biased* condition than for the *while-biased* condition.

We did not exclude any participants from the analyses in this experiment.

Results

Figure 3 shows the mean ratings for both conditions. As this figure shows, participants provided higher ratings for the utterance with *a bit* than for the utterances with *a while* in the *bit-biased* condition when the duration was 30 minutes, and the reverse was true in the *while-biased* condition. This suggests that participants indeed updated their expectations of the speaker's productions depending on the behavior they observed during the exposure trials.

This observation is also confirmed quantitatively with the differences in AUCs between the spline for *a bit* and the spline for *a while*. This difference is greater in the *bit-biased* condition than in the *while-biased* condition ($t(55) = 3.48, p < 0.001$, see also Figure 4).

In both conditions, we also observed a slow decrease in ratings for *for a while* for long time durations, which we did not observe in Experiment 1. This may suggest that participants assumed that the speaker would use other expressions to communicate very long durations, since they did not observe any utterances by the speaker to describe long durations in the exposure phase.

General Discussion

In this work, we set out to resolve a potential confound in previous semantic-pragmatic adaptation studies and to determine whether listeners actually learn speaker-specific mappings between world states and vague expressions or it is more likely that they make inferences about the speaker's ability to infer the world state from a visual scene.

In Experiment 2, we found evidence for participants updating their expectations of the speaker's productions. This suggests that listeners indeed learn speaker-specific[2] mappings between world states and utterances as predicted by theories and models of semantic-pragmatic adaptation (Schuster & Degen, 2020). In Experiment 2, participants provided more certain responses after being exposed to a speaker and the results are compatible with assuming that listeners shift their beliefs about a speaker's thresholds as they observe their utterances. Both of these behaviors are expected under Bayesian models of adaptation: Through the exposure to the speaker, listeners update their beliefs about the speaker's thresholds which results in more peaked threshold distributions and hence leads to more certain predictions.

Beyond answering our main question, this work also

[2]While we did not run any experiments with multiple speakers, Yildirim et al. (2016), Schuster and Degen (2019), and Schuster, Mayn, and Demberg (2023) all found that semantic-pragmatic adaptation to variable use of vague expressions is indeed speaker-specific, and it seems likely that this would also be true for adaptation to variable use of temporal expressions.

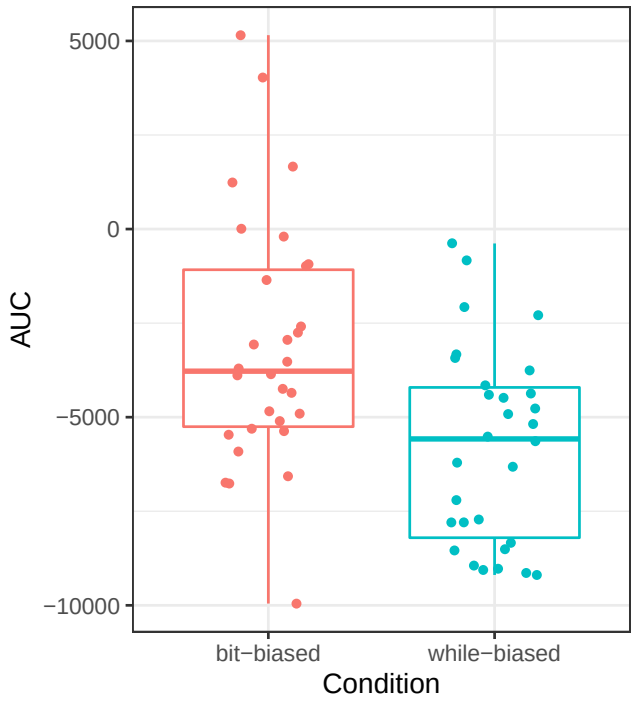

Figure 4: AUC difference between the spline for *for a bit* and *for a while* in both conditions.

shows that adaptation happens for a class of expressions that has not been investigated so far. While temporal expressions are assumed to have similar meaning representations as quantifiers, our experiments nevertheless provide an important data point to highlight that semantic-pragmatic adaptation can be observed for a wide range of expressions, furthermore highlighting the flexibility in semantic representations that interlocutors use in communication.

Since we set out to resolve a confound of previous work, we would like to highlight that, as pointed out by the reviewers, the data is still compatible with other explanations. An additional factor that we varied across conditions is the distribution of waiting times: in the *bit-biased* condition waiting times are on average longer than in the *while-biased* condition and it could be that participants learned something about this distribution and the change in production expectations is a result of listeners incorporating these statistics in computing the meaning of temporal expressions, as for example in the model proposed by Lassiter and Goodman (2017). We consider this explanation less likely given that it requires an additional level of reasoning but our current data cannot fully rule out this hypothesis.

Similarly, it could also be that listeners adapt to the discourse context or what is relevant to the situation. While we did not investigate this question in this work, Schuster and Degen (2019) suggests that in the context of uncertainty expressions, this is not the case and listeners are indeed learning speaker-specific meanings of uncertainty expressions. It seems likely that this would also be true for temporal expressions but once again, our data cannot fully rule out this hypothesis either.

Conclusion

In two web-based experiments, we explored semantic-pragmatic adaptation in temporal expressions. Our results suggest that there exists talker variability in the use of temporal expressions and that the effect of semantic-pragmatic adaptation also applies to the category of temporal measure expressions. Finally, we removed a confound of previous studies by removing any vagueness in the visual representation and provided additional evidence that the effects observed in semantic-pragmatic adaptation studies are a result of listeners updating beliefs about semantic representations.

References

Brée, D. S., Feddag, A., & Pratt, I. (1993). Towards a formalization of the semantics of some temporal prepositions. *Time & Society*, *2*(2), 219-240.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1-39.

Grice, H. P. (1975). *Logic and conversation* (D. Davidson, Ed.). Dickenson Pub. Co.

Heim, S., Peiseler, N., & Bekemeier, N. (2020). "few" or "many"? An adaptation level theory account for flexibility in quantifier processing. *Frontiers in Psychology*, *11*, 382.

Kamp, H., & Reyle, U. (1993). *From discourse to logic: Introduction to modeltheoretic semantics of natural language, formal logic and discourse representation theory*. Dordrecht: Kluwer Academic Publishers.

Kennedy, C. (2007). Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics & Philosophy*, *30*, 1-45.

Lassiter, D., & Goodman, N. D. (2017). Adjectival vagueness in a Bayesian model of interpretation. *Synthese*, *194*(10), 3801–3836.

Móia, T. (2005, 06). Portuguese expressions of duration and its english counterparts. *Journal of Portuguese Linguistics*, *5*, 37.

Pezzelle, S., & Fernández, R. (2023). Semantic adaptation to the interpretation of gradable adjectives via active linguistic interaction. *Cognitive Science*, *47*.

Schuster, S., & Degen, J. (2019). Speaker-specific adaptationto variable use of uncertainty expressions. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 41).

Schuster, S., & Degen, J. (2020, 10). I know what you're probably going to say: Listener adaptation to variable use of uncertainty expressions. *Cognition*, *203*, 104285.

Schuster, S., Mayn, A., & Demberg, V. (2023). Working memory updating modulates adaptation to

speaker-specific use of uncertainty expressions. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 45).

Swanson, E. (2006). *Interactions with context* Unpublished doctoral dissertation. MIT.

Xiang, M., Kramer, A., & Kennedy, C. (2020). *Semantic adaptation in gradable adjective interpretation.* (Unpublished, U Chicago)

Yildirim, I., Degen, J., Tanenhaus, M. K., & Jaeger, T. F. (2016). Talker-specificity and adaptation in quantifier interpretation. *Journal of Memory and Language*, *87*, 128-143.