Guessing is Believing: Semantic/Pragmatic Adaptation of Temporal Expressions

Yuxin Cao

Abstract

On hearing a friend says *I'll be there in a bit*, do you expect 15 minutes or anticipate a longer wait? Listeners have expectations of how a speaker would use certain expressions to mean something. They form these expectations from the speaker's prior usage. This work follows an online experiment paradigm to explore whether such update of expectations occurs for temporal expressions. Findings reveal that: 1) "a bit" is commonly used for short durations, while "some time" and "a while" are used interchangeably for longer periods; 2) people are more confident in showing their judgements when they have an adaptation target.

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

1 Introduction

From a statistical learning point of view, listeners keep track of a *talker-specific* language use (Yildirim et al., 2016). Based on this prior knowledge, listeners generate an idea of how a particular speaker associates expressions with meanings (Creel et al., 2008; Jaeger and Snider, 2013). When the next time similar situations occur, listeners have their expectations of which expressions the speaker is more likely to use. This updating process of expectations is described as adaptation (Kleinschmidt and Jaeger, 2015).

Pragmatic theories have been established on the basis of reasoning about possible alternatives (Degen, 2013). According to Grice's maxim of quantity and Horn's scale theory, if an expression A is more informative than B but B is the chosen one, then using B over A implies the negation of A (Grice, 1975; Horn, 1984). However, adaptation adds another process in pragmatic inferencing as to the speaker's choice of B: the speaker has a preference to use B in that context. For example, knowing someone habitually says "I think" before voicing their opinion lowers the possibility of expecting them to be making an implication that facts may suggest otherwise. Adaptation, as an update process of expectations, elucidates another mechanism in the pragmatic inference process.

Past studies have shown that adaptation occurs at all linguistic levels. Experimental work covers a wide range, from pairing of acoustic cues and phonemes (Kleinschmidt and Jaeger, 2015; Roettger and Franke, 2019), persistence in syntactic structure (Bock, 1986; Gustafson et al., 1997; Fraundorf and Jaeger, 2016; Blanchette et al., 2024), lexical and acoustic convergence (Ward and Litman, 2007), and among others. The field of semantic/pragmatic adaptation is still at its early stage. Existing research has addressed expressions including quantifiers like *few, many* (Yildirim et al., 2016), uncertainty expressions such as *might, probably* (Schuster and Degen, 2020), and gradable adjectives like *plain, bent, tall* (Xiang, Kramer, and Kennedy, 2020).

This work follows a web-based experiment approach to investigate 1) Does semantic/pragmatic adaptation extend to another type of expressions – temporal expressions? and 2) If it does, are these temporal expressions similar to or different with other types of expressions that have already been studied?

2 Background

2.1 Adaptation

Successful communication has a collaborative nature (Grice, 1975). Adaptation, as a mutual process in communication, can go both ways of comprehension and production (Blanchette et al., 2024). 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 and Wilkes-Gibbs, 1986).

From the production side of adaptation, Verschueren (1995) observes that, when an English speaker is in a foreign country with limited grasp of the local language, they may opt for a direct, literal English translation of the foreign word, willingly sacrificing idiomatic expressions so that the meaning is more likely to be understood. Similarly, experiments of dialogue system demonstrate that people's answers show a convergence with the prompt questions in both syntactic structures and lexical choices (Stoyanchev and Stent, 2009).

As for the comprehension side, existing work has provided evidence that this adaptation process happens at all linguistic levels (Chang et al., 2012). In speech recognition, even though some speaker's /b/ and /p/ are phonetically impossible to differentiate, listeners manage to use former [b] or [p] evidence to infer which phonological category is more or less likely (Kleinschmidt and Jaeger, 2015).

Cross-level work has also been carried out. Blanchette et al. (2024) explores syntax-semantics adaptation of an American English vernacular structure, where negative auxiliary inversion (e.g. *Didn't everybody like the movie*) can mean wide-scope negation "*not everybody* liked the movie" than narrow-scope negation "*nobody* liked the movie." Participants show rapid adaptation by rating wide-scope negation over narrow-scope after exposed to wide-scope biased sentences, e.g. "...even though *I have a couple responses*, *didn't everybody* call me to RSVP."

It is worth noting that preference for syntactic structures and lexical choices are often treated similarly under the framework of repetition priming (Gustafson et al., 1997; Pickering and Ferreira, 2008; Xie and Kurumada, 2024), but Fine et al. (2013) argue that adaptation is a combination of "*experience-based processing, syntactic priming and statistical learning*" and priming alone fails to see a cumulative side of language comprehension and processing.

2.2 Semantic/Pragmatic Adaptation

Semantic/pragmatic adaptation falls in the field of dynamic semantics, as both focus on the updating of contexts and its impact on the communication process. Dynamics semantics, originally motivated by studies of unbound anaphora and presupposition, explores how interpretation and communication common ground interact and influence each other (Lewis, 2014).

There has been a shift from static semantics to dynamic semantics, with the development of computational models enabling some level of calculation to predict pragmatic judgements in context. It takes a leap from only acknowledging that listeners take prior and new information in their expectations to actual calculation of how these expectations are borne (Frank and Goodman, 2012; Franke and Jäger, 2016; Schuster and Degen, 2020).

In terms of the contents, semantic/pragmatic adaptation[1] has focused on three types of expressions. Yildirim et al. (2016) tested quantifiers such as *few, many*, Schuster and Degen (2020) investigated probabilistic expressions like *might, probably*, and Xiang, Kramer, and Kennedy (2020) researched on gradable adjectives of three categories: relative adjectives *tall*, absolute adjectives with a maximum threshold *plain* and with a minimum threshold *bent*. They share a web-based experiment paradigm that simulates conversation by having participants, as observers, experience varying exposure and demonstrate shifts in their expectations accordingly.

2.3 Vagueness and Temporal Expressions

Listeners' expectations towards speakers' denotational use of expressions are at the heart of semantic/pragmatic adaptation. Vagueness plays a key role in determining the denotation of expressions. Kennedy (2007) contended that a vague sentence

[1] This study adopts the term "semantic/pragmatic adaptation" as used in Schuster and Degen's work (2020), which reflects the view that adaptation involves both semantics and pragmatics. There are pragmatic assumptions made by the listener about the speaker and semantic beliefs regarding the lexicon.

exhibits contextual variability in truth conditions. According to his analysis on gradable adjectives, there were two assumptions contribute to the idea of contextual variability/sensitivity. First, there was a mapping onto the abstract notion of **degrees**. Second, ordered degrees realized themselves in a **dimension** that constitute a scale. Both the degree and the dimension are sensitive to the context. For instance, on hearing or producing "Someone is tall", people acknowledge that *tall* has a degree, and the exact centimetres in the dimension of height are ordered to form a scale. When expressed in a formal way: $[\![\text{pos}]\!] = \lambda g \lambda x. g(x) \succ d_c$. POS is short for the positive (unmarked) form of the adjective, and d_c is "the contextually appropriate" standard of comparison. If the context is a setting of basketball players and 185 cm is the d_c, then, to say someone is tall is to say their height equals to or surpasses 185 cm ($[\![\text{tall}]\!] = \lambda g \lambda x. g(x) \geq 185$ cm).

Just as gradable adjectives where the standard of comparison is highly context-based, quantifiers and uncertainty expressions also share this context sensitivity. For example, the cardinality of a set does not decide if the amount is *few* or *many*. And identical probability can result in different descriptions.

As the three types of expressions studied share the trait of vagueness, from the two assumptions of vagueness, I will use the two properties of context-sensitivity and measurement (having a degree) as guiding principles, and explore another category of expressions that fits into the paradigm: temporal expressions.

Peres (1993) observed three uses of time-related expressions: *location in time* (*in 1980*), *frequency* (*three times a week, annual*), and *duration* (*for three hours*).

The quantificational nature of *frequency* and *duration* revealed two underlying functions: *measurement* and *locating* (Kamp and Reyle, 1993). *Measurement* indicated a duration of time (Brée et al., 1993), while *locating* described the relations between the duration and the time axis (Móia, 2005). This relational aspect was further summarized into four types: *precedence*, *overlapping*, *abutment* (immediate precedence) and *inclusion* (subset) (Kamp and Reyle, 1993, as cited in Móia, 2002).

In general, measure adverbials shared a composition of quantifiers ranging over some measuring entity (Guenthner, 2014). Similarly, temporal measure expressions consisted of two parts: a quantificational phrase sitting in the determiner position, and a temporal noun occupying the noun position, such as *3* *years, some time.*

Móia (2005) had added a dimension of vagueness to the analysis of temporal expressions. There were non-vague temporal nouns such as *year, month, hour, second*, and vague temporal nouns such as *instant, while, eternity, blink of an eye.* This vagueness extended to quantifiers as well, with vague quantifiers like *a lot of, some, few, many.*

And certain kinds of temporal expressions possess the properties of context-sensitivity by being quantifiable in terms of time span. Here I will categorize temporal measure adverbials into four types based on the structure mentioned above.

1) Numerals + temporal nouns, such as *15 minutes, 3 years.* These adverbials provide a precise measurement of duration. If exact numerals are to be combined with vague temporal nouns, such as *five instants*, there is a clash. The use of numerals marks an intention of being precise.

2) Quantifiers + temporal expressions, such as *some minutes, a few instants, some while.* Temporal nouns in its plural forms, such as *minutes*, are included in this category, given that they are also cumulation of the unit of the temporal nouns.

3) Existential quantifier "a" + temporal nouns, such as *a minute, a while.* When the temporal noun is not vague, they are flexible as to rough or precise estimation.

4) Idiomatic phrases , such as *a blink of an eye, in a snap, an arrow from a bow.* These phrases often match the measure adverbial structure, where an NP sits in the complement of DP, with an "a" in the determiner position. The action (NP) after the determiner can be treated as a temporal noun. Most of these phrases describe an extreme case of measurement.

These categorizations provide a framework for understanding how temporal expressions function in language. In the next two sections, I will conduct experiments on temporal measure expressions to test the hypothesis that listeners update their expectations of the speaker's usage for temporal expressions based on the speaker's prior usage.

3 Experiment 1

Experiment 1 is to obtain a general use of the selected temporal nouns and provide some of the materials for the later adaptation experiment (Experiment 2).

3.1 Participants

I recruited 40 participants (20 per scenario), and they were paid £1 for an estimated 5 minutes of participation. They were recruited via Prolific, filtered for native English speakers born in the UK.

3.2 Procedures

Participants were first presented a context page for scenario introduction. In this experiment, I designed two scenarios for gathering general usage preference. Every participant only completed one of the two. In each trial, participants were presented with a picture showing both the setting (the scenario) and the situation (clear indication of time spans). Below the picture were sliders, each representing a choice. Choices were sentences consisted of temporal expressions and one other option (*something else*). Participants were asked to distribute a total of 100% points among the listed choices under the instruction that they should make the judgements based on how likely they thought speakers would produce the listed utterances in the given situation. Each scenario had 9 test time spans, repeating twice in a random order.

3.3 Materials

3.3.1 Target Utterances

The most ideal candidates in temporal measure expressions were the ones with a clear start and a fuzzy boundary line. After exclusion of potential hyperbolic expressions, I picked three vague phrases that were also commonly used. They were *some time, a while*, and *a bit.*

While having a VP such as *take some time* was possible, I chose to embed a PP to avoid semantic influence from the main verb[2]. Two prepositions *in* and *for* were possible as they were used to show a relation between the measure duration and the time axis (the locating function). I settled on *for* because it possessed an extra meaning of completion (Brée et al., 1993). The target temporal phrases were placed at the end of sentences.

An informal speech style was adopted to reflect everyday language, using colloquial contracted form of *I'll / I'm*. To conclude, the temporal expressions under test were *"for a while", "for a bit" and "for some time."*

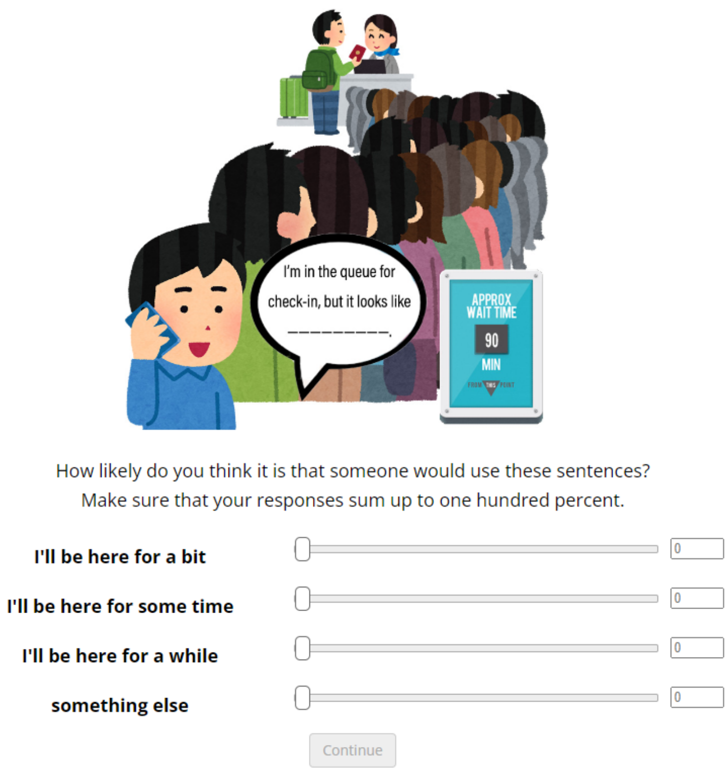

Figure 1: Queuing Scenario for Experiment 1

[2]The following three sentences, "It's gonna take a while", "It's gonna take a bit", and "It's gonna take some time", appear to achieve a similar interpretation that weakens the differences in the temporal phrases.

3.3.2 Stimuli

Figure 1, the queuing scenario, was a man responding to a friend's call while queuing near the check-in counter. Next to the man was a sign with a clear number for the estimated waiting time. Between the man and the counter were people in line. The queuing lines were kept identical to ensure that participants focused on the sign for the key information of the event.

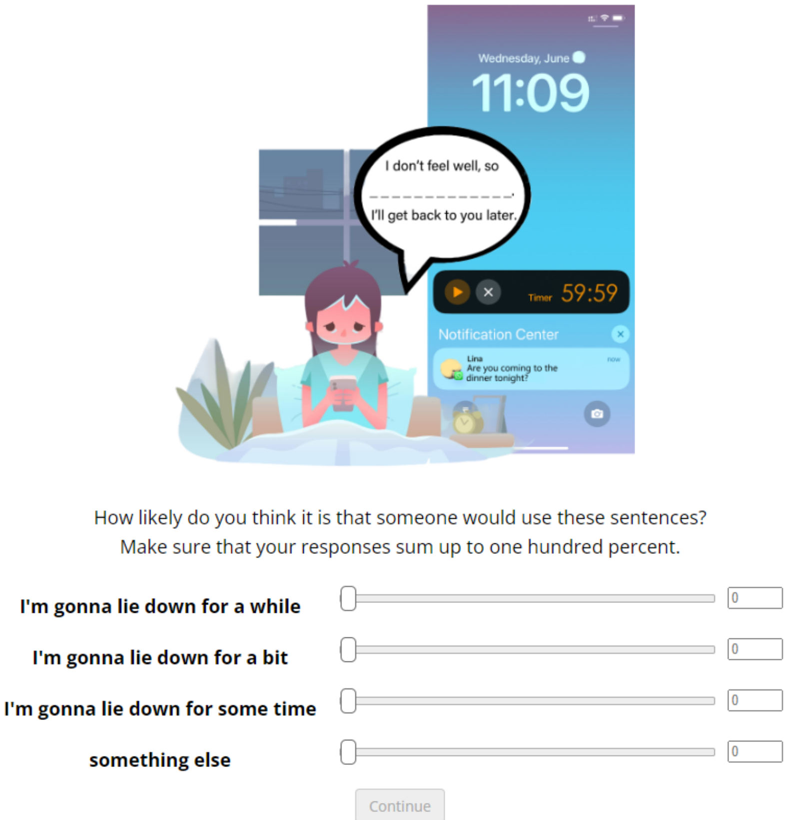

Figure 2: Timer Scenario for Experiment 1

Figure 2, the timer scenario, was a woman replying to a friend's message after just setting up a timer for a nap. The utterances were highlighted in a speech bubble next to a sign or a timer that displayed the length of the time interval, hinting at the relationship between the duration and the utterances. The test time span varied between the two scenarios, ranging from 5 minutes to 2 hours.

3.4 Results

3.4.1 Exclusions

Participants completed the trials on their own devices with the experiment in full-screen mode throughout to minimize distractions. Those who gave similar likelihood ratings for the same expression across all nine contexts were excluded from the analysis, as this likely indicated insufficient attention. This manual filtering served as substitution for attention check trials.

3.4.2 Results and Discussion

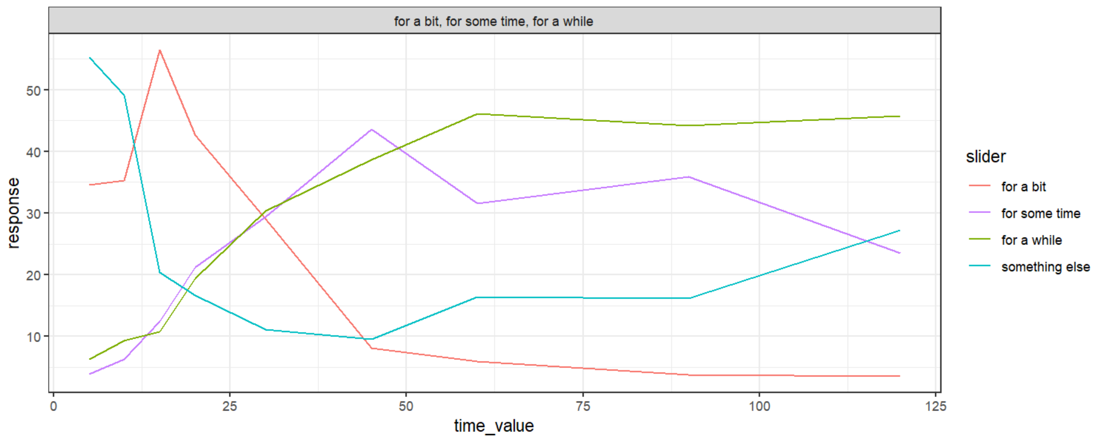

Figure 3: Experiment 1 Results of the Queuing Scenario

Figure 3 and 4 illustrate the findings[3]. In the queuing scenario (Figure 3), *for a bit* (red line) shows the highest expected likelihood of 56.47% at 15 minutes, which then drops sharply to 8.12% by 45 minutes. In contrast, *for a while* (green) and *for some time* (purple) have overlapping rising trends from 5 minutes to 45 minutes. *For a while* increases from 6.32% to 38.68%, peaking at 46.12% at 60 minutes. As for *for some time*, the likelihood gradually rises from 3.82% to 43.62%, fluctuating between 23.44% and 43.62% after 45 minutes.

[3]I randomized the order of the choices presented to the participants. To make the figures more consistent: the red line represents *for a bit*, the purple line represents *for some time*, the green line represents *for a while*, and the blue line represents something else.

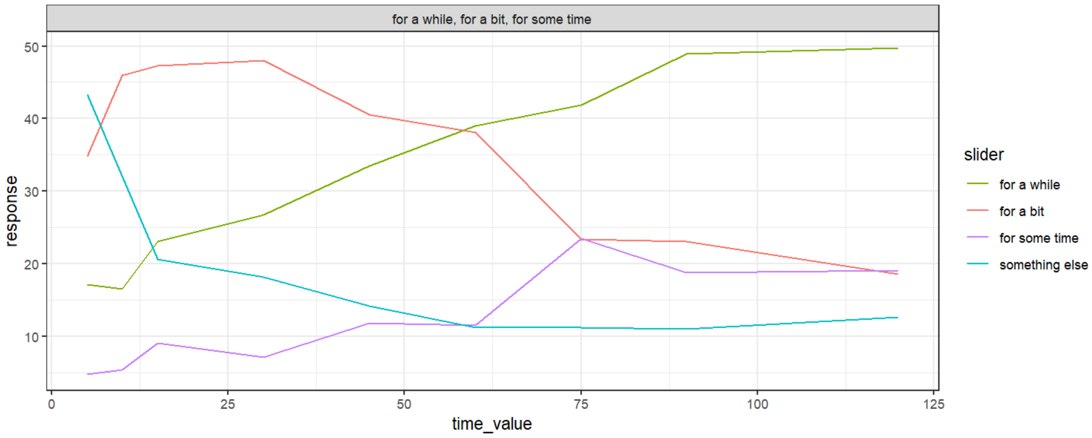

Figure 4: Experiment 1 Results of the Timer Scenario

Similarly in the timer scenario (Figure 4), *for a bit* is rated relatively higher, ranging from 34.8% to 48.03% between 5 minutes and 60 minutes, then decreases for longer periods. Both *for a while* and *for some time* show increasing likelihoods, with *for a while* consistently rated higher than *for some time*. *For a while* rises from 17.13% to 49.73%, while *for some time* increases from 4.77% to 19.10%.

In both scenarios, *for a while* and *for some time* show similar trends: as the time span increases, the likelihood rises. Conversely, *for a bit* is rated as more likely for shorter time periods. This also matches an etymological view where *bit* means biting a piece, indicating that the current time span is actually a small piece of the whole span[4], and *while* is regarded as a portion of time, more neutral.

While in the timer scenario, *for some time* is rated less likely throughout, considering the two expressions' likelihood curves cross frequently, both being rated highly around 40%, I speculate that the two expressions share an interchangeable usage preference.

Context sensitivity is also reflected in the varied decreasing points for *for a bit*. In the queuing scenario, *for a bit* starts to decrease after 30 minutes while in the timer scenario, it is only after 15 minutes.

I pick the queuing scenario for the adaptation experiment for the following two reasons. First, the queuing scenario had a less pragmatic influence. According to Grice's Maxim of Quantity (1975), the speaker should be as informative as possible.

[4] *Online Etymology Dictionary*, https://www.etymonline.com/word/bit

In the timer scenario, since the woman has set a timer out of her intention, she knows the exact duration (or at least the upper limit) of her absence time period. But instead of providing the information, such as "I'm gonna lie down *for half an hour* or so", she chooses to not narrow it down. This lack of specificity leads participants to infer an implied meaning for the given sentences. This effect is particularly evident in the results for "I'm gonna lie down *for some time*" (the purple line). In the timer scenario, the mean response rate is only 11.89% (ranging from 4.77% to 28.47%), while in queuing scenario, the mean is higher at 21.70%, (ranging from 3.82% to 43.62%). For the queuing scenario, because there is no other information source besides the sign, the information is assumed to be more objective. A second reason is that a more clear-cut exposure condition is preferred, as the crossing point represents borderline cases where the two expressions share a similar possibility in the production. In the queuing situation, the key exposure span is easier to identify, as there is only one crossing point (around 30 minutes) for the three expressions. In contrast, the timer scenario presents two, making it less straightforward to pinpoint the key exposure span.

I select two temporal expressions for the adaptation experiment, *"for a bit"* and *"for a while"*, as they have the most stable ratings across two scenarios, and their statistic point (time value of 30 minutes) in the queuing scenario as the exposure material to conduct the second experiment.

4 Experiment 2

Experiment 2 aims to test whether participants experience any expectation shift after exposure. The different exposure conditions can be treated as speakers with different usage preferences. This setup simulates how listeners adapt to different speakers by examining how participants adjust their expectations.

4.1 Participants

I recruited 60 participants (30 per condition) for the second experiment on adaptation. Participants were paid £1.70 for approximately 10 minutes of participation time and those from previous experiment were ruled out.

4.2 Procedures

Participants were presented with an exposure stage before the trials. They were shown pictures, accompanied by recordings, of the speaker producing one of the two utterances (*for a while* / *for a bit*) at the most critical time beforehand. The continue button to enter the next page was disabled until the recording finished. The rating trials were identical to those in Experiment 1. After obtaining the results from the two adaptation conditions, I calculated their areas under the curve (AUC) for both expressions and conducted statistical analysis to determine whether the differences were significant.

4.3 Exposure Materials

4.3.1 The Critical Span

From Experiment 1, I took "30 minutes" as the unit for the critical exposure span and set up a 15-minute gap for the other expression. The choice of adding ("45 minutes" for *a while*) or subtracting ("15 minutes" for *a bit*) corresponds to the contrasting usage preference found in Experiment 1. One condition was *"bit"* taking over 30 minutes, and the other condition used *"while"* for that period. They are referred as bit-condition and while-condition in later analysis.

The two critical periods for producing *for a bit* and *for a while* were played five times each. To demonstrate that the speaker (male voice) was a rational individual, a filler condition — where the man did not need to queue at all — was included in the exposure. This filler condition, with *I can't talk right now*, was played three times (see Figure 5). A sum of thirteen times were played in a random order for each participant (see Figure 6 for details of exposure trials).

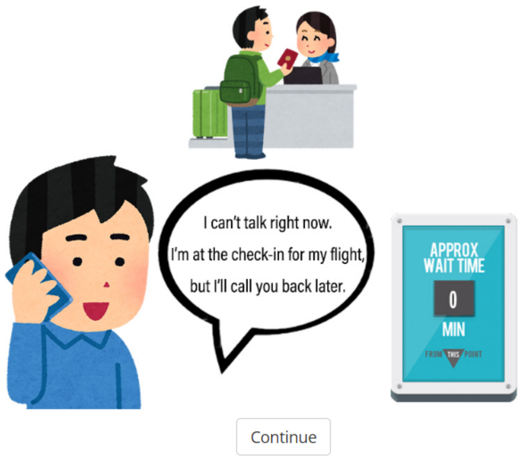

Figure 5: The *Cannot* Exposure Example

Exposure Conditions	***for a bit***		***for a while***		**Filler (*cannot*)**	
	n	t	n	t	n	t
while-condition	5	15	**5**	**30**	3	0
bit-condition	**5**	**30**	5	45	3	0

n = number of appearances; t = time span (minutes)

Figure 6: Numbers of Exposure Trials

4.3.2 Stimuli

To highlight the exposure expressions, recordings[5] of a male English speaker were played simultaneously with presenting pictures of the corresponding utterances. And to avoid repetition, the recordings only contained key parts maintaining forms of a natural speech sentence (e.g. "I'll be here for a while"). To separate exposure from trials, instructions were rearranged as a transition page.

In the trial phase, to emphasize the distinction between generic speakers and a specific speaker, the subject of the embedded prompt question was replaced with "the man" instead of "a speaker" as used in the first experiment.[6]

[5]The recordings, while created by a synthesized speech generator Text-to-Speech AI, closely mimicked human speech to the extent that distinguishing them from human-generated recordings was challenging.

[6]Yildirim et al. (2016) has shown in their experiment that the use of definite vs. indefinite

4.4 Results and Discussion

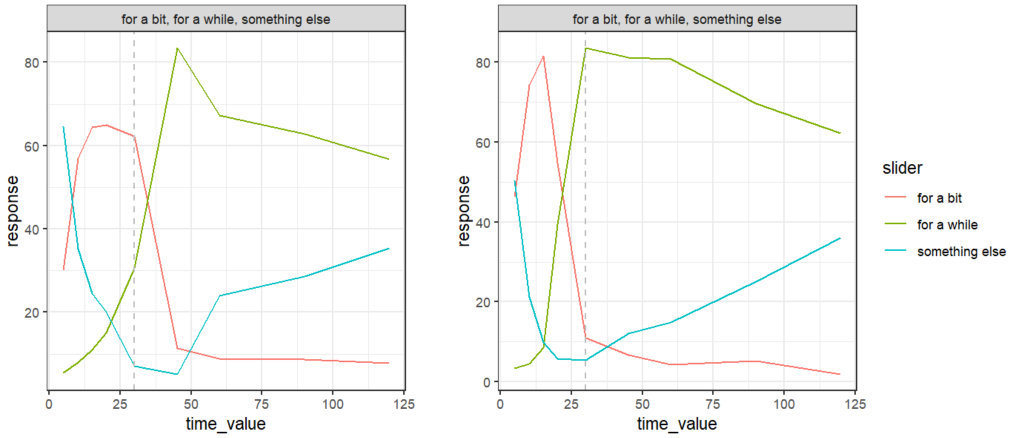

Figure 7: Comparison Between the Two Conditions in Experiment 2.
On the left side is the bit-condition, and on the right side is the while-condition. The grey line at 30 minutes highlights the exposure time span.

The initial rating for *for a bit* is slightly higher in the while-condition (46.28%) than in the bit-condition (30.00%) (Figure 7). This difference may be influenced by the later highest response rating (81.47% > 64.91%). Both conditions show a similar gap of approximately 35% between the initial and highest ratings. This could indicate that adaptation occurs in both scenarios but starts from different baselines.

In the bit-condition, the peak of *for a bit* is extended, compared to general usage preference from Experiment 1, forming a continuum from 15 to 30 minutes. Additionally, while the likelihood rating for crossing points in both conditions is around 45%, the crossing time span is further away from the exposure mark of 30-minutes for the while-condition (20 minutes) than the bit-condition (around 35 minutes). Also given that *for a bit* received the highest rating in Experiment 1 and maintains this only in the while-condition, the above reasons suggest that participants might be more familiar with the while-condition usage in this context, making it easier for them to adapt to and assign higher ratings.

When the event duration significantly exceeds the exposure time (e.g., after 60

articles *the* / *a* marks the differentiation of generic vs. specific to some extent.

minutes), a duration possibly deemed impossible for the event, participants exhibit nearly identical rating trends for the three choices. In Experiment 1, ratings for the two expressions tended to slightly increase towards the very end (maintaining a higher likelihood, especially for *for a while*). However, in Experiment 2, this upward trend is taken by a rising preference for *something else*. This shift could be attributed to the smaller number of available choices, leading participants to favour the alternative *other* option.

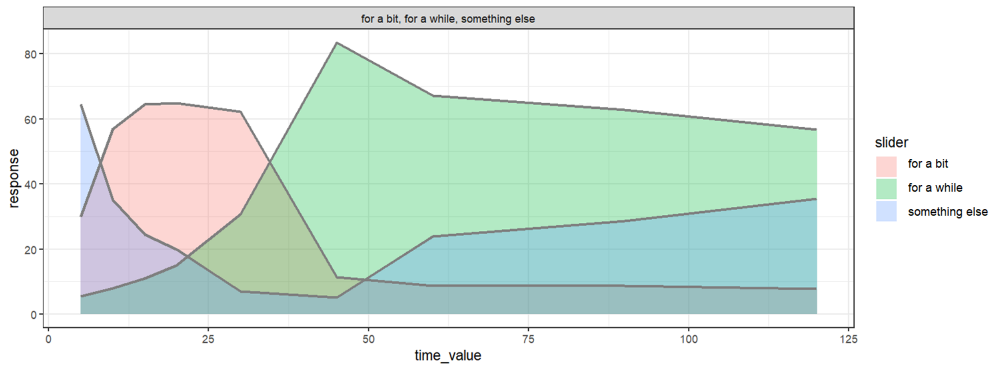

Figure 8: AUC for Bit-Condition

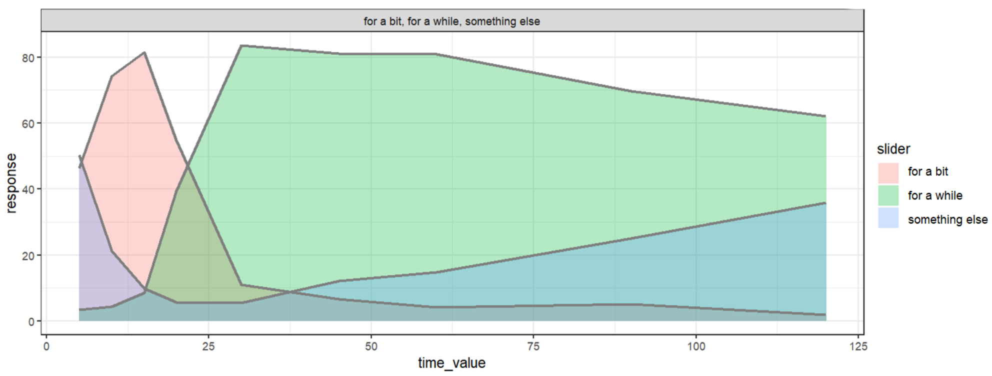

Figure 9: AUC for While-Condition

After conducting a t-test on the differences between the two expressions under the same condition, AUC_{bit} and AUC_{while} (i.e., the space of the red area minus the space of the green area as shown in Figure 8 & 9), the results demonstrate a significant difference, t(55) = 3.48, $p < 0.001$. This finding supports the effect of exposure and indicates semantic/pragmatic adaptation in temporal expressions.

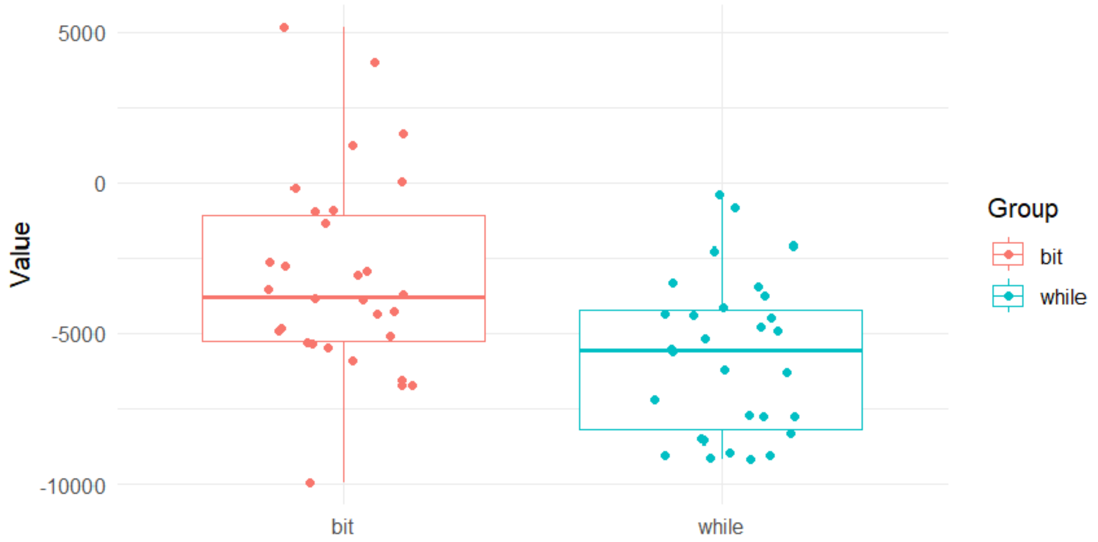

Figure 10: Boxplot for AUC Differences in Bit-Condition and While-Condition

The median line inside each box in Figure 10 indicates the central value of the AUC differences for the two conditions. The long line across the box in the bit-condition highlights greater variability in participants' responses, aligning with the idea that this condition is more familiar to general usage preferences. These findings are consistent with the conclusion that exposure leads to semantic/pragmatic adaptation in temporal expressions.

One more finding is the difference in median completion times between the two experiments: 8:40 for Experiment 1, and 6:57 for Experiment 2. Given that Experiment 2 includes an additional exposure phase estimated to last around 2 minutes, it is surprising that participants completed Experiment 2 more faster. The discrepancy may be attributed to a reduced number of choices in Experiment 2. However, overall difference in the mean trial time (the completion time minus an estimated exposure time of 1 minute) of more than two minutes remains substantial.[7]

[7] One potential concern is the change in the context page format from a single page text to

5 General Discussion

5.1 Summary and Discussion

The individual data from Experiment 1 reveal talker variability in the usage preferences for all three temporal expressions. This corresponds to previous findings on quantifiers, uncertainty expressions and gradable adjectives (Yildirim et al., 2016; Schuster and Degen, 2020; Xiang, Kramer, and Kennedy, 2020. Moreover, Experiment 1 supports the account that vague temporal measure expressions are context-sensitive, as usage preferences for the same time span vary across different scenarios.

In Experiment 2, the significant shift in listeners' expectations provided evidence for semantic/pragmatic adaptation in temporal expressions.[8] Additionally, the quicker response times in Experiment 2 suggest that participants are more confident in making judgements when they have an anchor for the context. This finding aligns with Schuster and Degen's (2020) study, where participants gave more extreme ratings like 100% in the adaptation experiment, a pattern also observed in this study. Both the reduced response times and the extreme ratings reflect listeners' confidence in their responses when they have a clear target to adapt to. On the other hand, they also mentioned that a slower reaction time could also be evidence for an experiment limitation, where participants might be engaged in too much meta-linguistic reasoning in production expectation experiment.

One of the biggest oppositions to the experiment design is that the repetitive exposure process cannot represent everyday situation at all. It is hardly ever the case that someone would repeat the same utterance over and over in such a short period of time. However, Xiang et al.'s experiment (2020) shows that participants are quick adapters who exhibit adaptation after the earliest exposure documented (the second time) and repetitive exposure hardly enhances the result.

three pages with continue buttons for smoother transition. This change should not significantly affect completion time, as the overall word count remains the same, suggesting that reading time should not differ markedly.

[8] As the two exposure conditions are utterly the same in utterances and their frequency, results of this paradigm can be considered separately from priming.

If so, repetition in this paradigm is only to suffice possible lack of attention for an on-line experiment.

This study also supports that adaptation results from both prior experience and exposure likelihood as the Bayesian model. Here I will try to give account in natural language. The processing of context-sensitivity primarily takes a probabilistic approach with regard to expectations. Moltmann (1991) points out that a contextually determined expectation value plays a role in the lexical meaning difference of *many* and *few*. On deciding a situation, prior experiences weigh in. Similar to having a prevailing standard of comparison in the semantic representation (Kennedy, 2007), there is also a most anticipated outcome for temporal measure expressions. When the actual duration length takes shorter or longer than the expected outcome, speakers use context-sensitive expressions to mark the extent. In rare cases where relevant situations are difficult to find, people might resort to mechanism other than prior knowledge for decision-making, such as mathematical concepts like half, but most adaptation phenomena involve both factors.

5.2 Limitations and Future Directions

One limitation is the size of tested temporal expressions. With a limited expense, I only picked three of the many temporal expressions with context-sensitivity and measurement constraints. The study would produce more generalized outcome if more temporal expressions were included in Experiment 1.

Second, the analysis for vagueness in temporal expressions can be further explored. For the analysis of temporal expressions, I mostly followed semantic functions (Móia, 2005) for classification and measure adverbials (Moltmann, 1991) for the structure. However, what I consider as *non-vague* temporal nouns, such as *minutes, days, years*, is still based on a vague standard. While some might argue that an external basis such as a clock-counter system, with its objective emissions for time tracking (Grondin et al., 1999) is not vague, it fundamentally is not a process people consciously consider when using these expressions. Further accounts can be made as to how people measure and to what extent do they stop and use the result as a measure unit. According to Meck and Church (1983), rats may use the same brain areas for both counting and timing. If humans function similarly, vagueness on temporal adverbials could be close to quantifiers.

For future research, how prior knowledge affects adaptation is an open question. To lessen influence from prior knowledge, Xiang et al. (2022) expanded the range of adjective expressions studied and categorized visual stimuli into two categories: *artifact* and *shape*. The former was identical as in Xiang et al. (2020), but the latter was abstract, geometric shapes such as cylinder. They managed to do so as the use of adjective is often combined with a noun phrase where abstraction of shape from the physical is allowed. For other types that are difficult to visualize, such as temporal expressions, how to abstract remains challenging.

One thing about the experiment procedure that can be improved is as Pezzelle and Fernández (2023) points out, the exposure is very passive and does not have active interaction. While they address this by letting participants to ask questions, there should be room for improving within the likelihood distributor trial paradigm.

6 Conclusion

Following an online experiment paradigm, I explored semantic/pragmatic adaptation in temporal expressions. Three categories studied under this paradigm shared the trait of vagueness and talker-variability. Using context-sensitivity and measurement from vagueness as guiding principles, I explored another category of expressions that fitted into the paradigm: temporal measure expressions.

Findings show that 1) there exists talker variability in the use of temporal expressions, 2) the effect of semantic/pragmatic adaptation also applies to the category of temporal measure expressions, 3) *a bit* is used to describe a short period of time, while *some time* and *a while* are preferred for a longer time; 4) *some time* and *a while* are used interchangeably; 4) people show more confidence and higher efficiency when they have an adaptation target.

This work extends the study of semantic/pragmatic adaptation include temporal expressions and highlights future research fields such as the investigation between prior knowledge and adaptation.

References

Blanchette, F., Flannery, E., Jackson, C., & Reed, P. (2024). Adaptation at the syntax–semantics interface: Evidence from a vernacular structure. *Language and Speech*, *67*(1), 140–165. https://doi.org/10.1177/00238309231164972

Bock, K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *31*, 163–186.

Brée, D. S., Feddag, A., & Pratt, I. (1993). Towards a formalization of the semantics of some temporal prepositions. *Time & Society*, *2*(2), 219–240. https://doi.org/10.1177/0961463X93002002005

Chang, F., Janciauskas, M., & Fitz, H. (2012). Language adaptation and learning: Getting explicit about implicit learning. *Language and Linguistics Compass*, *6*. https://doi.org/10.1002/lnc3.337

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1–39. https://doi.org/https://doi.org/10.1016/0010-0277(86)90010-7

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition*, *106*(2), 633–664. https://doi.org/https://doi.org/10.1016/j.cognition.2007.03.013

Degen, J. (2013). Alternatives in pragmatic reasoning. https://api.semanticscholar.org/CorpusID:146216651

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0077661

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998–998. https://doi.org/10.1126/science.1218633

Franke, M., & Jäger, G. (2016). Probabilistic pragmatics, or why bayes' rule is probably important for pragmatics. *Zeitschrift für Sprachwissenschaft*, *35*(1), 3–44. https://doi.org/10.1515/zfs-2016-0002

Fraundorf, S. H., & Jaeger, T. F. (2016). Readers generalize adaptation to newly-encountered dialectal structures to other unfamiliar structures.

Journal of Memory and Language, *91*, 28–58. https://doi.org/10.1016/j.jml.2016.05.006

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

Grondin, S., Meilleur-Wells, G., & Lachance, R. (1999). When to start explicit counting in a time-intervals discrimination task: A critical point in the timing process of humans. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 993–1004. https://doi.org/10.1037/0096-1523.25.4.993

Guenthner, F. (2014). Dowty david r.. word meaning and montague grammar. the semantics of verbs and times in generative semantics and in montague's ptq. synthese language library, vol. 7. d. reidel publishing company, dordrecht, boston, and london, 1979, xvii + 415 pp. *The Journal of Symbolic Logic*, *48*, 501–502. https://doi.org/10.2307/2273576

Gustafson, J., Larsson, A., Carlson, R., & Hellman, K. (1997). How do system questions influence lexical choices in user answers? *Proceedings of Eurospeech '97, 5th European Conference on Speech Communication and Technology*, 2275–2278. https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-13323

Horn, L. (1984). Toward a new taxonomy for pragmatic inference: Q-based and r-based implicature.

Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition*, *127*(1), 57–83. https://doi.org/https://doi.org/10.1016/j.cognition.2012.10.013

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

Kennedy, C. (2007). Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics & Philosophy*, *30*, 1–45. https://doi.org/10.1007/s10988-006-9008-0

Kleinschmidt, D., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*, 148–203. https://doi.org/10.1037/a0038695

Lewis, K. S. (2014). Do we need dynamic semantics? In A. Burgess & B. Sherman (Eds.), *Metasemantics: New essays on the foundations of meaning* (pp. 1–19). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199669592.003.0010

Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes*, *9*(3), 320–334. https://doi.org/10.1037/0097-7403.9.3.320

Móia, T. (2005). Portuguese expressions of duration and its english counterparts. *Journal of Portuguese Linguistics*, *5*, 37. https://doi.org/10.5334/jpl.154

Moltmann, F. (1991). Measure adverbials. *Linguistics and Philosophy*, *14*(6), 629–660. https://doi.org/10.1007/bf00631962

Pezzelle, S., & Fernández, R. (2023). Semantic adaptation to the interpretation of gradable adjectives via active linguistic interaction. *Cognitive Science*, *47*. https://doi.org/10.1111/cogs.13248

Pickering, M., & Ferreira, V. (2008). Structural priming: A critical review. *Psychological bulletin*, *134*, 427–59. https://doi.org/10.1037/0033-2909.134.3.427

Roettger, T. B., & Franke, M. (2019). Evidential strength of intonational cues and rational adaptation to (un-)reliable intonation. *Cognitive Science*, *43*. https://doi.org/10.1111/cogs.12745

Schuster, S., & Degen, J. (2020). I know what you're probably going to say: Listener adaptation to variable use of uncertainty expressions. *Cognition*, *203*, 104285. https://doi.org/10.1016/j.cognition.2020.104285

Stoyanchev, S., & Stent, A. (2009, June). Lexical and syntactic adaptation and their impact in deployed spoken dialog systems. In M. Ostendorf, M. Collins, S. Narayanan, D. W. Oard, & L. Vanderwende (Eds.), *Proceedings of human language technologies: The 2009 annual conference of the north American chapter of the association for computational linguistics, companion volume: Short papers* (pp. 189–192). Association for Computational Linguistics. https://aclanthology.org/N09-2048

Verschueren, J. (1995). The pragmatic perspective. In J. Verschueren, J. O. H. Östman, J. Blommaert, & C. Bulcaen (Eds.), *Handbook of pragmatics* (pp. 1–19). John Benjamins. https://doi.org/10.1075/hop.m.prag

Ward, A., & Litman, D. (2007). Automatically measuring lexical and acoustic/prosodic convergence in tutorial dialog corpora. https://doi.org/10.21437/SLaTE.2007-13

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

Xiang, M., Kennedy, C., Xu, W., & Leffel, T. (2022). Pragmatic reasoning and semantic convention: A case study on gradable adjectives. *Semantics and Pragmatics*, *15*. https://doi.org/10.3765/sp.15.9

Xie, X., & Kurumada, C. (2024). From first encounters to longitudinal exposure: A repeated exposure-test paradigm for monitoring speech adaptation. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2024.1383904

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. https://doi.org/10.1016/j.jml.2015.08.003