Speech Prosody 2026
26-29 May 2026, Philadelphia, Pennsylvania, USA

The prosodic realization of English *since-when* questions

Youngdong Cho[1], *Chloe Dokyung Kwon*[2], *Fengyue Lisa Zhao*[3]

[1,2,3]Cornell University
yc2544@cornell.edu, dk837@cornell.edu, fz227@cornell.edu

Abstract

Prosody signals context-dependent interpretations of string-identical sentences, but only a few acoustic studies tested for a systematic association between interpretation and prosodic realization. This study examines English *since-when* questions, which can be used to ask for the time reference of an event (information-seeking) or express speaker's incredulity toward the preceding utterance. These interpretations have been suggested to exhibit distinct prosodic patterns: incredulous *since-when* questions carry emphatic stress on *when* and a low utterance-final boundary tone. However, no experimental work has examined their acoustic realization yet. The present study addresses this gap by testing whether speakers reliably produce distinct prosodic patterns for these interpretations. Regression results showed that in incredulous *since-when* questions, prominence falls on *when*, realized with significantly lower F0 mean, wider F0 range, and longer duration compared to information-seeking *since-when* questions. In addition, we fitted generalized additive mixed models (GAMMs) to capture the dynamic realization of F0 over time, further revealing distinct F0 contour shapes for the two interpretations: Incredulous *since-when* questions showed steeper F0 valley on *when* and the strong syllable of the final noun, and lower overall F0 levels, especially on the utterance-final word. Our findings thus suggest that speakers clearly distinguish two interpretations of English since-when questions prosodically.

Index Terms: since-when questions, question intonation, prosody-semantics interface, questions, incredulity

1. Introduction

The current study explores the relationship between form and meaning through the prosody of English *since-when* questions. Like other question constructions such as echo questions [1,2,3,4] and rhetorical questions [5], *since-when* questions are intriguing in that they give rise to two different interpretations depending on discourse context. They can either ask for the onset time of an event or state (*information-seeking interpretation*) or convey the speaker's incredulity toward the preceding utterance (*incredulous interpretation*) [6,7].

(1) A: John and Mary are linguists.
B: Since when are they linguists?

For example, in (1), B's question can be construed as either asking for onset time of John and Mary being linguists or expressing B's incredulity toward this claim, for instance, because B has never seen them engage in linguistic research. In the latter case, B is signaling that A's utterance conflicts with their prior belief. It has been suggested that incredulous *since-when* questions carry emphatic stress on *when* and exhibit a low utterance-final boundary tone [6]. However, this claim is based solely on introspection, and no experimental evidence on English *since-when* questions yet exists. The present study aims to address this gap by conducting a production experiment. Specifically, we address two research questions: (i) Do speakers employ different prosodic realizations for *since-when* questions depending on context? (ii) Do speakers' prosodic choices align with the characteristics discussed in previous literature?

To establish our predictions, it is informative to consider other string-identical constructions that are prosodically distinguished across different contexts. One such construction is incredulous declarative sentences. For example, when the sentence 'You like John' is uttered in a context where the speaker expresses disbelief in the addressee liking John, it is typically produced with double prosodic prominence: an initial steep rise on the subject (e.g., *you*), followed by an even steeper final rise on the object (e.g., *John*) [8]. In addition, the prosody of English rhetorical *wh*-questions further illustrates a similar point. It has been reported that rhetorical *wh*-questions exhibit prosodic realizations distinct from their string-identical information-seeking counterparts: rhetorical *wh*-questions display prosodic prominence on the *wh*-word (L*+H in ToBI) and on the object (L+!H*), as well as longer durations on both constituents [5]. Lastly, crosslinguistic evidence (e.g., Hungarian and Cantonese) shows that speakers distinguish incredulous *since-when* questions from information-seeking ones [7].

Building upon these findings, we propose the following hypotheses:

H_G: Incredulous *since-when* questions exhibit prosodic realization distinct from information-seeking ones.

H_1: Incredulous *since-when* questions show an emphatic stress on *when*, realized with longer duration, wider F0 range, and a steeper F0 valley. Additional steeper valleys are also predicted on the subject (e.g., *they* in (1B)) and the strong (i.e., stressed) syllable of the final noun (e.g., *linguists* in (1B)).

H_2: Incredulous *since-when* questions exhibit a low utterance-final boundary tone, as suggested by previous work on *since-when* questions. An alternative hypothesis is that it is instead realized with a rising boundary tone, as observed in incredulous declaratives.

2. Method

2.1. Design and materials

10 experimental items were constructed using the frame *Since when are they* NOUN. The target questions were evenly divided between those containing disyllabic nouns (e.g., *lawyers*, n=5) and trisyllabic nouns (e.g., *beginners*, n=5), all with penultimate stress, allowing us to examine whether the SYLLABLE NUMBER affects the F0 contour. In addition to the experimental items, 30 fillers (declaratives, polar questions, and *wh*-questions in equal proportion) were included.

Each item was presented in three levels of incredulity toward the preceding utterance, manipulated as CONTEXT, as illustrated in Table 1. Each context consisted of a situational

10.21437/SpeechProsody.2026-114

description, a previous speaker's utterance, and contextual information that licensed distinct interpretations of the target *since-when question.* Two contexts served as the primary focus of the present analysis: (i) **100% incredulity**, in which the speaker is fully confident that the preceding utterance is mistaken and (ii) **information-seeking**, in which the speaker has no relevant prior knowledge and genuinely asks for the onset time of the stated event. We also included 50% incredulity context, where the speaker vaguely recalls evidence that might contradict the preceding utterance but remains uncertain. It allowed us to examine potential effects of intermediate degrees of disbelief, although the analyses reported here focus on the contrast between the 100% incredulity (INCREDULOUS) and the INFORMATION-SEEKING contexts.

Table 1. *Experimental materials example*

Situation: You and John are discussing who should organize an upcoming group trip, reviewing people who are good at planning and taking initiative. John says:
John: Why don't we let Alex and Brianna handle it? They're leaders now.
CONTEXT **1: 100%** INCREDULITY You were at last week's club meeting to vote for the new leadership team and know Alex and Brianna weren't even running. You're confident John is mistaken. So, you say:
CONTEXT **2: 50%** INCREDULITY You vaguely recall Alex and Brianna saying they prefer to stay behind the scenes and avoid leadership roles. But you're not entirely certain about this conversation. So, you say:
CONTEXT **3:** INFORMATION-SEEKING You haven't kept up with club matters lately and don't know about any leadership changes. So, you say:
You: Since when are they leaders?

An online norming study assessed the naturalness of the designed items (n=37), and all conditions received mean ratings above 5 on a 7-point Likert scale.

2.2. Participants and procedure

27 speakers of American English (17 female; mean age 20) completed an online production experiment using Labvanced [9] for course credit, using their own laptops with microphones. Participants were instructed to work in a quiet setting with a stable internet connection. After reading each scenario, they recorded the target utterance (e.g., *Since when are they beginners?*). They were asked to speak in a way that felt natural and appropriate for the situation without hesitation. Participants could record as many attempts as they wanted. Following three practice trials, experimental items were presented in randomized order, with conditions counterbalanced so each participant saw only one condition per item. The entire session took approximately 15 minutes.

2.3. Data processing and acoustic measurements

The data was forced-aligned with Montreal Forced Aligner [10] and subsequently hand-inspected and corrected to ensure accurate syllable boundaries. This process led to the removal of seven tokens (2.6%) due to speech errors. Syllable durations were then extracted from the corrected alignments. F0 was measured over the entire utterance using the autocorrelation method in Praat [11], with gender-specific floor and ceiling parameters. Outlier F0 tokens were removed if F0 values lay outside three standard deviations from the speaker's mean, showed abrupt frame-to-frame jumps, and were isolated with fewer than five consecutive F0 measurements. Four F0 descriptors, including F0 mean, F0 maximum, F0 minimum, and F0 range (F0 max - F0 min), were calculated for each syllable. Smoothed F0 trajectories were obtained using a smoothing spline method [12] with F0 recentered to individual means. Following visual inspection of the smoothed F0 trajectories, 24 tokens (9.0%) were excluded due to unreliable F0 values, primarily caused by creaky voice. A total of 237 tokens were used for analysis.

2.4. Statistical analysis

To examine prosodic differences across conditions, we conducted two complementary analyses: (i) linear-mixed effects (LME) models for F0 descriptors and duration, and (ii) generalized additive mixed models (GAMMs) to compare dynamic F0 trajectories across conditions. For both analyses, we focus on reporting differences between 100% incredulity (INCREDULOUS) and INFORMATION-SEEKING contexts.

We fitted LME models using *lmer* in R [13] to analyze F0 descriptors (mean, maximum, minimum, and range) and duration, with CONTEXT (100% incredulity vs. 50% incredulity vs. Information-seeking) as the primary fixed effect. Separate models were fitted for each syllable: *since*, *when*, *are*, *they*, the strong and weak syllables of disyllabic nouns (syll1(S), syll2(W)), and the three syllables of trisyllabic nouns (syll1(W), syll2(S), syll3(W)). Starting with a maximal model including CONTEXT, SYLLABLE NUMBER, GENDER, their interactions, attempt number, and trial number, we used stepwise AIC-based model comparison to select the best-fitting model. All maximal models included random intercepts for participants and items.

We fitted GAMMs to the female speakers' data using *mgcv* [14] and visualized the predicted trajectories using *itsadug* [15]. GAMM analyses capture the non-linear F0 patterns across the entire utterance. Following stepwise reduction from a maximal model, the final model included fixed effects of CONTEXT and SYLLABLE NUMBER, as well as several smooth terms: a general smooth of time captured the baseline contour (aligned at the offset of *when*), and additional smooths allowed the contour to vary by CONTEXT, SYLLABLE NUMBER, and their interaction. This structure enabled the model to identify context-dependent and syllable-specific differences in the shape of the F0 contour over time. Random intercepts were included for subjects and items, and subject-specific smooths of time were added to account for individual differences in F0 contour shape.

3. Results

3.1. Duration

Model-predicted durations for each syllable are shown in Figure 1. A robust effect of CONTEXT (INCREDULOUS vs. INFORMATION-SEEKING) emerged only on *when* for both disyllabic and trisyllabic noun conditions. The best-fitting models for the duration of *when* included fixed effects of CONTEXT, SYLLABLE NUMBER, and their interaction, as well as trial number, with a random intercept for participants. Across SYLLABLE NUMBER, *when* was significantly longer in INCREDULOUS than in INFORMATION-SEEKING contexts (2-syllable: estimate = –0.032, SE = 0.01, $t(244) = -3.97$, $p < .001$; 3-syllable: estimate = –0.022, SE = 0.01, $t(243) = -2.88$, $p <$

.005), indicating that speakers reliably lengthened *when* to express incredulity. No significant durational differences across CONTEXT were found for the subject *they*, the strong syllable of the final noun, or the utterance-final syllable, suggesting that durational marking of incredulity is localized to *when*.

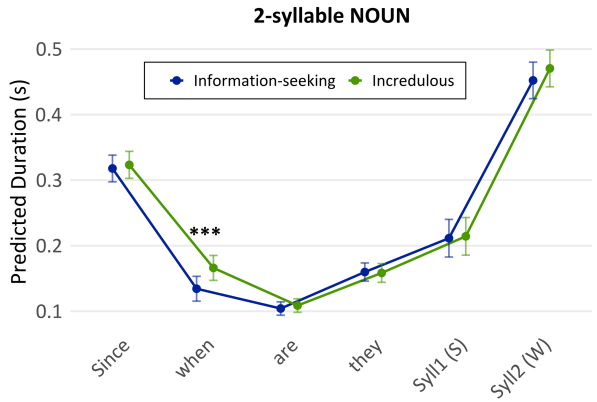


Figure 1. *Model predicted durations with 95% CIs of each syllable for 2-syllable NOUN condition.*

3.2. F0 descriptors

Robust prosodic differences were found on ***when*** between INCREDULOUS and INFORMATION-SEEKING contexts across measures. In INCREDULOUS contexts, *when* had significantly lower F0 mean (estimate = 10.68, SE = 3.20, t(238) = 3.34, $p < .001$), lower F0 minimum (estimate = 13.4, SE = 3.66, t(239) = 3.66, $p < .001$), and wider F0 range (estimate = -8.71, SE = 3.96, t(240) = -2.20, $p < .05$). These effects were held across both SYLLABLE NUMBER and across genders. Together with the duration effects, these findings indicate that *when* consistently receives enhanced prominence in INCREDULOUS contexts, realized through a lower pitch target followed by a steeper rise to the following syllable. Section 3.3 examines this dynamic pitch movement in greater detail.

Turning to regions after *when*, no F0 descriptors showed significant differences on the subject *they* across CONTEXT. For the strong syllable of the final noun, CONTEXT effects were limited in both measure and gender. A marginal reduction in F0 range was observed for disyllabic nouns across all speakers in INCREDULOUS contexts. Effects on F0 mean were restricted to female speakers: females produced a lower F0 mean on the strong initial syllable of disyllabic nouns (estimate = 18.85, SE = 6.61, t(114) = 2.85, $p < .01$) and on the strong medial syllable of trisyllabic nouns (estimate = 22.86, SE = 6.71, t(106) = 3.41, $p < .001$) in INCREDULOUS contexts. Lastly, no individual F0 descriptor showed significant CONTEXT effects on the utterance-final syllable.

3.3. F0 contours

Figure 2 shows averaged F0 contours of female speakers by CONTEXT, with the upper panel showing utterances ending in disyllabic nouns and the lower panel showing those ending in trisyllabic nouns. Time was normalized to the median number of frames per syllable, and the contours were interpolated accordingly. Both female and male speakers generally exhibited lower F0 in INCREDULOUS than INFORMATION-SEEKING contexts. While both INCREDULOUS and INFORMATION-SEEKING *since-when* questions exhibited rising pitch accents on *when*, they differed in the magnitude. That is, in INCREDULOUS contexts, *when* was realized with deeper F0 valleys, followed by a steeper rise onto the subsequent syllable. With respect to peak location, female speakers' contours showed a clear F0 peak only on *are*, whereas male speakers' contours (not shown) exhibited distinct peaks on both *are* and *they*.

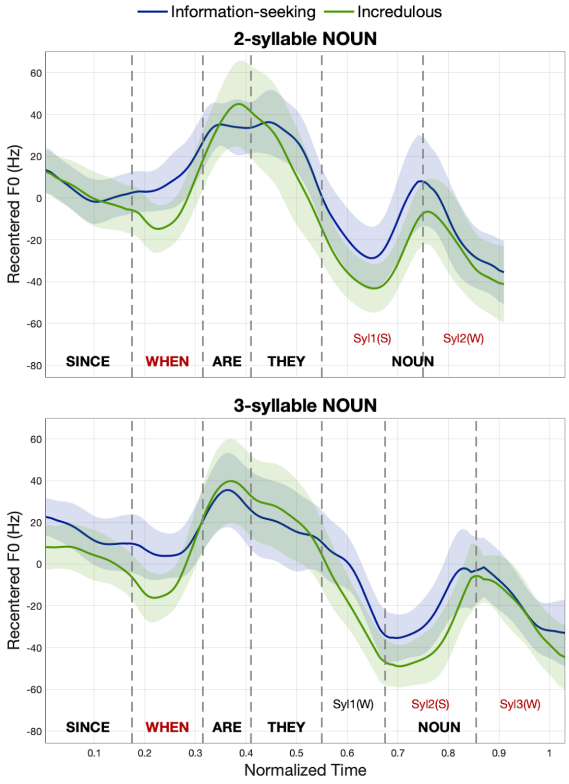


Figure 2. *Averaged F0 contours of female speakers with 95% CIs. Time was normalized to median frames per syllable. F0 values were recentered by speaker mean.*

To statistically assess the dynamic F0 differences observed in Figure 2, we fitted GAMMs to the female speakers' data. Figure 3 shows the model predictions for the utterances ending in disyllabic nouns. The top panel displays the fitted recentered F0 contours for INFORMATION-SEEKING and INCREDULOUS contexts, aligned at the offset of *when*. The bottom panel plots the predicted context difference (INCREDULOUS – INFORMATION-SEEKING), with statistically significant regions highlighted in red. Shaded bands indicate 95% confidence intervals, and median syllable durations are indicated as reference boxes along the x-axis. Fitted outliers were removed using a 90% CI quantile threshold.

GAMM analyses showed that prosodic marking of incredulity is conveyed not only through static F0 descriptors but also through dynamic adjustments to the contour. In particular, beyond the lower F0 values and steeper rises observed around *when*, INCREDULOUS contexts also exhibited lower F0 and a steeper rise on the strong syllable of the final noun, a pattern that was not captured by the F0 descriptor analyses in Section 3.2. The GAMM results further indicated significantly lower F0 in the utterance-final region in INCREDULOUS contexts. However, this effect should be interpreted with caution. Speakers frequently produced creaky phonation utterance-finally, so datapoints in the low-F0 range

were excluded based on the criteria described earlier in the method. This reduction in available data resulted in substantially wider confidence intervals in the final portion of the contour or getting excluded in plotting.

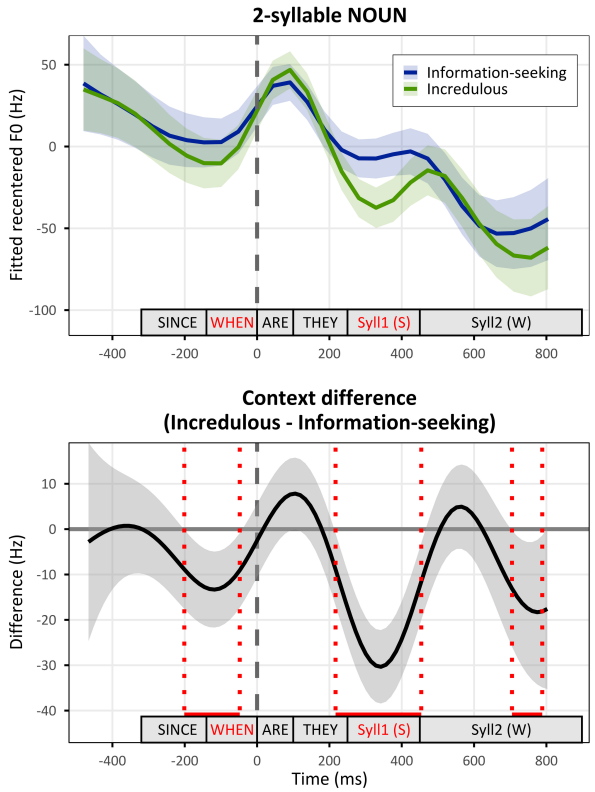


Figure 3. *Top: GAMM-fitted recentered F0 contours of female speakers with 95% CIs, aligned at the offset of when. Bottom: predicted* CONTEXT *differences (*INCREDULOUS − INFORMATION-SEEKING*) with 95% CIs, with significant regions highlighted in red.*

4. Discussion

Our findings support the general hypothesis H_G: incredulous *since-when* questions exhibit distinct prosodic realization from that of information-seeking ones. Differences arise in duration, several F0 measures, and F0 contour shapes. The largest effects occur on the *wh*-word *when,* strong syllables of the final noun, and the utterance-final region.

For *when*, speakers produced longer durations, lower F0 values (minimum/mean), and wider F0 range in the INCREDULOUS context. These phonetic realizations confirm hypothesis H_1, showing that *when* carries an emphatic stress realized as a lower pitch target and steeper rise. By contrast, the subject (e.g., *they*) did not show reliable prosodic differences across contexts, and no steeper rise was observed.

Turning to the strong syllable of the final noun, both male and female speakers produced significantly lower F0 values in INCREDULOUS contexts for both disyllabic nouns (e.g., *leaders*) and trisyllabic nouns (e.g., *beginners*), followed by a steeper rise to the subsequent syllable. These findings partially support the hypothesis H_1 in that prosodic differences between INCREDULOUS and INFORMATION-SEEKING *since-when* questions emerge on *when* and on the strong syllable of the final noun, but not on the subject.

Importantly, both INCREDULOUS and INFORMATION-SEEKING *since-when* questions exhibited rising pitch accents on *when* and on the strong syllable of the final noun, consistent with an L*+H ToBI analysis in both contexts. These contexts did not differ systematically in the category or the alignment of the targets; rather, they differed in the magnitude. In INCREDULOUS contexts, *when* and the strong syllable of the final noun were realized with deeper F0 valleys (lower L targets), followed by steeper rising movements toward the H peak, often realized on the subsequent syllables. These gradient differences in pitch scaling underscore the importance of quantitative modeling approaches such as GAMMs, which are well-suited to capturing subtle, time-varying differences in F0 contour shape that may not be apparent from categorical analyses alone.

For the utterance-final boundary tone, speakers consistently produced low or falling boundary tones in both INCREDULOUS and INFORMATION-SEEKING contexts, with no statistical differences in individual F0 descriptors. However, GAMM analyses revealed significantly lower F0 trajectories in INCREDULOUS contexts beginning at the final syllable. Taken together, the observed tendencies toward a narrower F0 range, lower F0 mean on the final syllable, and overall lower F0 trajectories provide weak but converging support for hypothesis H_2, namely that incredulous *since-when* questions exhibit a low utterance-final boundary tone rather than the alternative hypothesis of a rising boundary tone.

Before concluding, our findings suggest some similarity between the prosodic patterns of INCREDULOUS *since-when* questions and rhetorical *wh*-questions in English [5]. This may be attributed to the availability of a rhetorical interpretation for INCREDULOUS *since-when* questions. Rhetorical questions are characterized as having answers that are already known to interlocutors [16, 17] and as serving to make a rhetorical point [18]. Assuming that not all facts in the common ground are equally salient to interlocutors, INCREDULOUS *since-when* questions can be interpreted as rhetorical, signaling that the addressee has failed to infer an obvious answer already available in the common ground [19]. Importantly, however, having a rhetorical interpretation is not a sufficient condition for a construction to qualify as a rhetorical question. Accordingly, the present findings should be understood as showing that INCREDULOUS and INFORMATION-SEEKING *since-when* questions exhibit systematically distinct prosodic realizations, in a way that parallels the contrast between rhetorical and information-seeking *wh*-questions.

In summary, our study presents the first systematic experimental evidence on the phonetic realization of English *since-when* questions in INCREDULOUS and INFORMATION-SEEKING contexts, using a production experiment with both F0 descriptors and GAMM analyses. Overall, our findings suggest that speakers clearly prosodically distinguish INCREDULOUS from INFORMATION-SEEKING *since-when* questions. Specifically, our results support the prediction that *when* in INCREDULOUS *since-when* questions bears emphatic stress, realized as longer duration and a more extreme rising pitch movement from a lower pitch target. In addition, INCREDULOUS *since-when* questions exhibit lower utterance-final F0 values. Together, these findings provide new empirical evidence that speakers produce string-identical sentences with distinct prosodic realizations depending on context [4,5,7], contributing to our understanding of the prosody-semantics interface.

5. References

[1] D. Bolinger, *Interrogative Structures of American English: The Direct Questions. American Dialect Society. 28*. Birmingham: University of Alabama Press, 1957.

[2] C. Bartels, *The intonation of English statements and questions: A compositional interpretation*. New York: Routledge, 1999.

[3] M. Biezma, B. Braun, and A. James, "Prosody is adding what?: Echo questions are not a thing," *Proceedings of SALT 31*, pp. 241–261, 2021.

[4] S. Repp and L. Rosin, "The intonation of echo wh-questions," *Proceedings of Interspeech 2015*, pp. 938–942, 2015.

[5] N. Dehé and B. Braun, "The prosody of rhetorical questions in English," *English Language and Linguistics*, vol. 24, no. 4, pp. 607–635, 2020.

[6] A. Kiss, "Meta-conversational since when-questions and the common ground," *Proceedings of the 21st Workshop on the Semantics and Pragmatics of Dialogue*, Saarbrücken: Universität des Saarlandes, 2017, pp. 38–47.

[7] A. Kiss, *Form–meaning relations in non-canonical questions*. Ph.D. dissertation, Univ. of Toronto, 2024.

[8] D. Goodhue, L. Harrison, Y. T. C. Su, and M. Wagner, "Toward a bestiary of English intonational tunes," *Proceedings of NELS 46*, Concordia Univ., 2015, pp. 311–320.

[9] H. Finger, C. Goeke, D. Diekamp, K. Standvoß, and P. König, "LabVanced: a unified JavaScript framework for online studies," *Proceedings of International Conference on Computational Social Science*, Cologne, 2017. https://www.labvanced.com/publication.html

[10] M. McAuliffe, M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger, "Montreal Forced Aligner: Trainable text–speech alignment using Kaldi," *Proceedings of Interspeech 2017*, pp. 498–502, 2017.

[11] P. Boersma and D. Weenink, *Praat: Doing phonetics by computer*, ver. 6.4.47, Computer program.

[12] The MathWorks, Inc., "Smoothing splines." https://www.mathworks.com/help/curvefit/smoothing-splines.html

[13] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[14] S. N. Wood, "Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models," *Journal of Royal Statistical Society. B*, vol. 73, no. 1, pp. 3–36, 2011, doi:10.1111/j.1467-9868.2010.00749.x.

[15] J. van Rij, M. Wieling, R. Baayen, and H. van Rijn, *itsadug: Interpreting time series and autocorrelated data using GAMMs*, R package v2.4.1, 2022.

[16] H. Rohde, "Rhetorical questions as redundant interrogatives," *San Diego Linguistic Papers*, vol. 2, pp. 134–168, 2006.

[17] I. Caponigro and J. Sprouse, "Rhetorical questions as questions," *Proceedings of Sinn und Bedeutung 11*, pp. 121–133, 2007.

[18] D. Farkas, "Rhetorical questions revisited," *Proceedings of CLS 60*, 2024.

[19] Y. Cho, "Encoding speaker bias: the semantics of *wh*-polar questions," *Proceedings of Sinn und Bedeutung 30*, to appear.