Iris Chuoying Ouyang and Elsi Kaiser*

Tracking who knows what: epistemic gaps and the prosodic realization of corrective focus

https://doi.org/10.1515/ling-2021-0019
Received March 9, 2018; accepted August 12, 2020; published online March 8, 2021

Abstract: Corrective information is produced with higher prosodic prominence than non-corrective information. However, it remains unclear how corrective prosody is realized in different communicative settings. We conducted two production experiments to investigate whether interlocutors' prosodic realization of corrective focus depends on each other's knowledge state. Participants carried out a statement-response task in pairs (e.g., Speaker B: *Tina had shrimp at a restaurant*. Speaker A: *No, she had beef at a restaurant*.). Our focus is on the prosody of Speaker A's utterance. We manipulated whether Speaker B's statement was implausible in the context (e.g., a context where it is known that Tina actually hates seafood). Furthermore, the two experiments differed in whether Speaker B knew that their statement (e.g., about Tina eating shrimp at a restaurant) was (im)plausible. In Experiment 1, both speakers had access to the crucial context concerning the probability of Speaker A's statement (Tina's preferences about food). In Experiment 2, only Speaker A had access to this background information. We found that Speaker A's prosody when responding to Speaker B was influenced by both (i) the contextual probability of Speaker B's statements and (ii) Speaker B's knowledge (or lack thereof) about the contextual probability. We present an analysis where the prosodic prominence associated with corrective information reflects the gap between expectation and reality – in this case, what Speaker A had expected Speaker B to say and what Speaker B actually says.

Keywords: addressee's knowledge state; contextual probability; corrective focus; epistemic surprisal; prosody

***Corresponding author: Elsi Kaiser**, Department of Linguistics, University of Southern California, Grace Ford Salvatori Hall 301, Los Angeles, CA 90089-1693, USA, E-mail: emkaiser@usc.edu
Iris Chuoying Ouyang, Department of Linguistics, University of Southern California, Los Angeles, CA, USA, E-mail: iris.oy@gmail.com

1 Introduction

It is widely accepted that linguistic elements that are more "informative" tend to be more prosodically prominent, as indicated by acoustic properties such as duration, f0, intensity, and spectral characteristics (see Wagner and Watson [2010] for a review). Prior prosodic work has defined the notion of "informativity" in different ways. In the widespread information-structural tradition, which we follow in our paper, the assumption is that acoustic prominence is associated with foregrounded linguistic material – broadly speaking, material that adds new information to the conversation. Depending on the preceding discourse, speakers may prosodically emphasize particular words in an utterance to direct their addressee's attention to the important message they are trying to convey. One type of information structure that has been extensively studied is contrastive focus, of which various subtypes have been identified (e.g., Dik and Hengeveld 1997; Vallduví and Vilkuna 1998; Zimmermann and Onea 2011). For example, consider the word *toys* in the following contexts:

(1) a. *Did David find books or toys on the stairs?*
b. *He found toys on the stairs*. [*toys* = narrow, contrastive/selective focus]

(2) a. *Did David find socks on the stairs?*
b. *No, he found toys on the stairs.* [*toys* = narrow, contrastive/corrective focus]

The word *toys* in both (1b) and (2b) is an explicit alternative to something else in the discourse. *Toys* in (1b) – responding to (1a) – picks *toys* from the set consisting of *books* and *toys* that has been established via (1a). *Toys* in (2b) – responding to (2a) – is intended to contradict the information *socks* conveyed in (2a). In this paper, we concentrate on the latter type of contrastive focus – which is often referred to as corrective focus (e.g., Dik and Hengeveld 1997; Gussenhoven 2008) – because its information-structural properties are well-understood and it is prevalent in communication. Corrective focus conveys messages contradicting information that is already present in the discourse. A large body of work has found that contrastive/corrective elements receive greater acoustic prominence than non-contrastive/non-corrective elements (e.g., Breen et al. 2010; Cooper et al. 1985; Couper-Kuhlen 1984; Katz and Selkirk 2011; Krahmer and Swerts 2001).

Crucially, since language communication involves not only a talker/speaker but also a listener/addressee, the informativity of a word could depend on the addressee's knowledge about the world. Because people often have different knowledge backgrounds, this brings up the question of whether speakers'

encoding of information is *addressee-oriented*, taking into account the audience's knowledge state, or whether it is *egocentric*, driven by speaker-internal considerations. The general question of perspective-taking has received considerable attention in the literature (e.g., Arnold et al. 2012; Bard et al. 2000; Galati and Brennan 2010; Horton and Keysar 1996; Kahn and Arnold 2012; Kaland et al. 2013; Lockridge and Brennan 2002; Rosa and Arnold 2011; Roßnagel 2000, 2004). See also van Bergen and Hogeweg, this issue, on the importance of acknowledging differences in the beliefs and opinions of conversational participants. Research shows that speakers may vary their prosody based on their addressee's knowledge state – in other words, top-down information about addressee's knowledge state can modulate speakers' prosody – although a consensus has not been reached on the cognitive processes and mechanisms behind this phenomenon.

When speakers are communicating with two addressees who are informed to different extents, speakers are able to track both addressees' knowledge states. This shown by the fact that speakers prosodically mark informativity from the perspective of the particular addressee they are talking to at a given point of time. For example, Galati and Brennan (2010) show that, when speakers tell a story twice to one addressee and once to another addressee using the same sentences, their utterances are perceptually less intelligible when they are talking to the addressee who has already heard the story once before. Similarly, Kaland et al. (2013) show that, in a task where Dutch speakers give different instructions to two addressees about where to place objects (e.g., instructing one addressee to move *the blue triangle*, and then either asking the same addressee or the different one to move *the red triangle*), contrastive information (*red*) is pronounced with higher perpetual prominence when it is also contrastive to the addressee (i.e., when the addressee is the one who has moved a triangle before) than when it is not.

Perspective-taking has been shown to have at least two dimensions: (i) what speakers assume about their interlocutors before the conversation or at the very start of the conversation (e.g., Baker and Bradlow 2009; Ferguson 2004; Ferguson and Kewley-Port 2007; Fussell and Krauss 1989; Hummert and Shaner 1994), and (ii) what speakers learn about their interlocutors during the conversation, as signaled by the interlocutors' utterances or behavior (e.g., Brennan 1991; Isaacs and Clark 1987). Existing work on the question of how well speakers update their assumptions about what their interlocutors know has led to divergent results. For example, Russell and Schober (1999) found that speakers stick to the assumptions they created based on information they received prior to the conversation, regardless of what their interlocutors' utterances suggest over the course of the conversation. In contrast, Brennan (1991) found that speakers abandon their initial assumptions and adjust to what their interlocutors' utterances reveal about their knowledge states.

Additional evidence in favor of the view that speakers are able to update their representations of addressees' knowledge states based on the actual behavior of addressees comes from Kuhlen and Brennan (2010). In a joke-retelling task, participants produced more vivid details when they (i) expected attentive addressees (based on information that had been provided before the start of the task) and (ii) were in fact talking to attentive addressees. In contrast, participants produced less vivid narrations when they either (i) expected distracted addressees or (ii) were in fact talking to distracted addressees. Moreover, participants put in more effort when their addressees' behavior matched their expectations than when it did not. That is, participants who expected attentive addressees spent more time telling the jokes to attentive addressees than distracted addressees, while those who expected distracted addressees spent more time telling the jokes to distracted addressees than attentive addressees.

These complex patterns reveal that (i) what speakers know about their interlocutors *prior* to the conversation and (ii) what speakers learn about their interlocutors *during* the conversation both play important roles in determining the way speakers say their utterances. Nevertheless, little attention has been paid to the interplay between these two factors (but see Brennan 1991; Kuhlen and Brennan 2010; Russell and Schober 1999). Moreover, it remains unclear how perspective-taking factors interact with information structure in shaping the prosody of utterances, since few studies have investigated this question in general (but see Galati and Brennan 2010; Kaland et al. 2013). These issues are related to the broader theme of how top-down and bottom-up factors interact during language processing and production, as they tap into questions of how (i) speakers' top-down knowledge/expectations (and potentially changes in these expectations) about what the addressee knows influence (ii) the prosodic realization of their utterances, an aspect of speech production influenced both by bottom-up factors (e.g., segment identity, lexical stress, phrase length and so on: e.g., Fowler 1992; Jun 2003) and more top-down factors (e.g., information structure and perspective).

1.1 Aims and expected outcome

Prior work shows that information-structural factors, such as corrective focus, as well as speaker's *expectations* (e.g., about what other people are likely to do or say) and the *reality* speakers encounter (what other people actually do or say) all play a role in the prosodic encoding of informativity. However, little work has been done on the potential interaction between these factors. To shed light on this issue, we conducted a psycholinguistic production study that investigated whether

top-down, perspective-taking factors – in particular, speakers' awareness of what addressees know or don't know – could affect the prosodic prominence associated with corrective focus.

Corrective focus essentially entails an incorrect message (e.g., an incorrect claim made by someone) and a corrective response. In daily conversation, these two components of corrective focus structure are often spoken by different conversational participants. For example, imagine that Sibling A says that the curtains in the living room of their childhood home were green. But then Sibling B corrects her stating that the curtains were actually red. How does Sibling B prosodically mark the corrective focused word *red* in her utterance? Does this depend on Sibling B's perception of the other person's knowledge state, i.e., how unexpected the other person's mistake is? For example, consider a context where A and B had just the other day looked at color photographs of their childhood home and discussed the color of the draperies. In this context, it seems very unexpected for Sibling A to make this kind of mistake. Conversely, if there had been no such prior discussion, and if Sibling B also happens to know that Sibling A suffers from some amount of red/green color blindness, then perhaps it is not so unexpected that Sibling A would report that the curtains were green, even though Sibling B knows for a fact that they were actually red.

Thus, the question is *whether a speaker's prosodic encoding of corrective focus is modulated by how likely/expected the addressee's mistake was, from the speaker's perspective*. If the speaker considers the mistake to be unexpected, would she attenuate or perhaps strengthen the prosodic prominence she produces in her corrective response? Alternatively, what if the other person's mistake is expected, from the speaker's perspective? Would this have consequences for how the speaker realizes the prosodically focused element? (See also van Bergen and Hogeweg, this issue, whose investigation of Dutch discourse particles relates to questions about when speakers take their interlocutor's belief state into consideration).

In terms of acoustic correlates of prosodic prominence, we concentrate on the size of excursions in an f0 contour (which will be called "f0 range" henceforth) in this paper. We chose f0 because it is an acoustic dimension that has been extensively studied in the information-structural tradition, but potential effects of perspective-related factors are not yet well-understood. By conducting this study, we hoped to provide further evidence for the effects of perspective-related factors on f0.

We chose to analyze f0 ranges, a simpler measure than the combination of f0 maxima, minima, and/or means, as various f0 measures have been shown to correlate with information-structural focus (e.g., Breen et al. 2010; Brown 1983; Couper-Kuhlen 1984; Eady and Cooper 1986; Katz and Selkirk 2011). Based on existing studies, we predict that corrective responses would have larger f0 ranges than non-corrective responses.

If our prediction for corrective focus is borne out, we can then look into how perspective-taking factors modulate the prosodic realization of corrective focus. We expect to see effects of both the speaker's prior expectations (about what their interlocuter is likely to say, given the interlocutor's knowledge state) and the reality the speaker encounters (what the interlocutor actually says during the conversation, and how this relates to the interlocutor's knowledge state), based on existing research discussed in the preceding section. Furthermore, if corrective prosody is indeed sensitive to both perspective-taking dimensions, it would be interesting to see whether the two dimensions interact.

2 Methods

We conducted two production experiments (Experiment 1 and Experiment 2), both of which had an interactive set-up. Each trial consisted of a production task and a subsequent comprehension task. Naïve participants worked in pairs on the production task and independently on the comprehension task; further details are provided in Section 2.2 about the set-up of the experiments. The production task provided the critical recordings, namely the target sentence in each dialogue. The comprehension task was included to engage participants in the production task, as paying attention to the dialogue was necessary to successfully perform the comprehension task.[1]

2.1 Participants

Forty-four self-reported native speakers of American English participated, eleven pairs in each experiment. One participant was excluded from the data analysis because of a clearly non-native accent. All participants were students or staff at the University of Southern California.

2.2 Design and procedures

Two naïve participants (Speakers A and B) worked with each other in reading aloud dialogues. They were given the following instructions: *In this study, you will*

1 We do not discuss the comprehension task in detail here because it is not relevant for the results, but participants essentially had to answer a wh-question about the dialogue. For further details, please see Ouyang (2015).

be working with a partner. You will tell your partner what is written on your sheets of paper, and likewise, your partner will tell you what is written on their sheets of paper. In both experiments, the primary speaker of interest is Speaker A. We will first present the set-up of Experiment 1, and then turn to Experiment 2. A full list of the target items can be found in Appendix. In both experiments, our analyses focus on corrective focus as used in statements. Corrective focus in other kinds of speech acts is an important question for future work.

In Experiment 1, each dialogue consisted of five sentences, as shown in (3) and (4). The critical sentence whose phonetic properties we are focusing on is Speaker A's response, labelled as "critical sentence" in (3) and (4).

As we explain in more depth below, we manipulated the correctiveness of Speaker A's response in Sentence 5 (Corrective vs. Non-Corrective) and the contextual probability of Speaker B's statement in Sentence 4 (Probable vs. Improbable). Participants saw the text of the sentences on paper. Sentences 1 and 2 were spoken by Speaker A, introducing a character (e.g., Jacky) and a preference or need associated with that character (e.g., *Jacky likes fruit in her salads but not vegetables*, or Zac likes vegetables in his salads but not fruit). They were shown on one page (seen by Speaker A only). Sentences 3 and 4 were spoken by Speaker B, and presented on a separate page (seen by Speaker B only). Sentence 3 provided information about a recent event (e.g., Jacky/Zac going grocery shopping). Sentence 4 commented on this event, starting with *I heard that…* and describing something the character had done. Crucially, this described event was either plausible or implausible in the context of Sentences 1–2 (e.g., plausible: Jacky got apples; implausible: Zac got apples). Once Sentence 4 was produced, the other speaker, Speaker A, said Sentence 5, starting with either *Yes* or *No* to confirm or correct the previous sentence. Sentence 5 was presented on a new, separate page, seen only by Speaker A. Participants thus interacted with each other to produce the dialogues, and each participant only had access to the text of sentences that he or she was responsible for.

(3) CORRECTIVE FOCUS

Speaker A: *Jacky prefers her salad a certain way.* [Sentence 1]

Speaker A: *She loves fruit but hates vegetables.* [Sentence 2]

Speaker B: *She went grocery shopping yesterday evening.* [Sentence 3]

Speaker B: *I heard that she got some apples*$_{PROBABLE}$ *at the farmer's market.* [Sentence 4]

Speaker A: *No, she got some [lettuce]*$_{CORRECTIVE}$ *at the farmer's market.* [Sentence 5: **critical sentence**]

(4) NON-CORRECTIVE INFORMATION

Speaker A:	*Zac tends to put certain things in his salad.*	[Sentence 1]
Speaker A:	*He loves vegetables but hates fruit.*	[Sentence 2]
Speaker B:	*He went grocery shopping this morning.*	[Sentence 3]
Speaker B:	*I heard that he got some apples*$_{IMPROBABLE}$ *at the supermarket.*	[Sentence 4]
Speaker A:	*Yes, he got some [apples]*$_{NON\text{-}CORRECTIVE}$ *at the supermarket.*	[Sentence 5: **critical sentence**]

The experiment had 192 targets (i.e., the theoretically meaningful experimental items that we analyzed to test our hypotheses) and 96 fillers (distractor items that were not analyzed and are not relevant for our hypotheses). Each pair of participants encountered 48–96 items and did not see any item more than once.[2] Each participant served as Speaker A in half of the dialogues and Speaker B in the other half, i.e., the roles of Speaker A and B were intermixed throughout the experiment. These two halves had different sets of characters and scenarios. For example, one participant began and finished all the dialogues that involved Jacky, Zac, the kind of salad they like (e.g., fruit vs. vegetable), and the type of furniture they need to get rid of due to moving (e.g., bedroom vs. kitchen), whereas the other participant began and finished all the dialogues about Gary, Lauren, the kind of entrées they like (e.g., meat vs. seafood), as well as the part of their house that they need to buy new things for (e.g., bathroom vs. patio). More details are provided in Appendix.

Each scenario contained two sub-scenarios. In the scenario of salads, for example, one sub-scenario (let's call it X) described that Elisa, Jacky, and Tina prefer fruit salads, while the other sub-scenario (let's call it Y) described that Bob, Daniel, and Zac prefer vegetable salads. Each scenario was associated with a set of four object nouns, e.g., *apples*, *cherries*, *lettuce*, and *spinach* for the salad scenario. Two of the four object nouns are probable in the context of sub-scenario X but improbable in sub-scenario Y. The other two object nouns, in contrast, are probable in the context of sub-scenario Y but improbable in sub-scenario X.

In all dialogues, Sentence 4 (Speaker B's statement) and Sentence 5 (Speaker A's response) play a crucial role in our design. They contained transitive clauses with the following structure: a third-person singular pronoun subject, a simple past tense verb, an object noun phrase, and a prepositional phrase indicating a

2 The number of items varied across different pairs of participants because participants' speaking rate varied widely, meaning that some pairs had time to complete more items than others. Nevertheless, all participants encountered items in every condition of Correctiveness and Statement Type. We ensured that all pairs completed at least the minimum number of items necessary per condition for meaningful statistical analysis.

location or beneficiary. The critical word of interest is the head noun of the object noun phrase (e.g., *apples* or *lettuce*). The prepositional phrase was included so that the critical word was not at the very end of the sentence, where prosodic variance might be hard to detect due to phonological marking of finality and the potentially low energy level at the end of a breath.

To investigate how the addressee's knowledge state affects Speaker A's corrective prosody, we manipulated (i) the information-structural status of Speaker A's response with respect to Speaker B's statement, as exemplified in (3) and (4), and (ii) the contextual probability of Speaker B's statement, relative to the character's preference and need. The contextual probability of Speaker A's responses was counterbalanced, except on non-corrective trials, where the object noun was the same between B's statement and A's response. In other words, there were six types of object pairs between B's statement and A's response. Four of the six types occurred on the corrective trials:

(a) probable object in the statement, a different but still probable object in the response;
(b) probable object in the statement, improbable object in the response;
(c) improbable object in the statement, a different but still improbable object in the response;
(d) improbable object in the statement, probable object in the response.

The other two types occurred on the non-corrective trials:

(e) probable object in the statement, the same probable object in the response;
(f) improbable object in the statement, the same improbable object in the response.

The critical words rotated between different conditions of information structure and contextual probability. For example, *lettuce* was corrective on some trials (such as (3)) and non-corrective on the others (where Sentence 5 was *Yes, she got some lettuce at the farmer's market*). Similarly, *lettuce* appeared in responses to statements involving fruit on some trials (such as (3)) and to statements involving vegetables on some other trials (where Sentence 4 was *I heard that she got some spinach/lettuce at the farmer's market*). The full list of 192 items was compiled by pairing the four object nouns with one another in each sub-scenario (i.e., four object nouns in Speaker B's statements * four object nouns in Speaker A's responses * six scenarios and their "switched" versions).

Furthermore, in order to tease apart the two perspective-taking factors, i.e., what people initially expected about their interlocutor versus what they learn about their interlocutor during the conversation, we conducted a parallel experiment, Experiment 2, and manipulated Speaker B's knowledge state between the

two experiments. In essence, Experiments 1 and 2 differ with regard to whether or not Speaker B is aware of the (im)probability of their statement:

In Experiment 2, we created a knowledge gap between Speakers A and B by removing Speaker B's access to Sentence 2 (e.g., *She loves fruit but hates vegetables*) – this was the sentence that introduced the character's preference/need and thereby established a contextual bias. This was done simply by instructing Speaker A to not to read Sentence 2 aloud, which essentially made it Speaker A's private (privileged) knowledge. Thus, in Experiment 2, in contrast to Experiment 1, the contextual probability (probable/improbable) of the critical words (e.g., *apples* in *I heard that she got some apples at the farmer's market*) was privileged information known *only to Speaker A and unavailable to Speaker B*. In other words, in Experiment 2, because Speaker B did not know anything about whether Jacky or Zac likes or dislikes fruits or vegetables, Speaker B had no special reason to treat Jacky or Zac buying apples as probable or improbable. Apart from the fact that Sentence 2 was privileged information known only to Speaker A in Experiment 2, everything else was held the same between the experiments.

Thus, three independent variables were implemented:

(i) Correctiveness of the response (Corrective vs. Non-Corrective): whether the object head noun in Speaker A's response (Sentence 5) was in corrective focus. For instance, in (3), the object "lettuce" in Speaker A's response (*No, she got some lettuce at the farmer's market*) is corrective information because Speaker A was correcting Speaker B's statement (Sentence 4, *I heard that she got some apples at the farmer's market*). In contrast, in (4), the object *apple*s in Speaker A's response (*Yes, he got some apples at the farmer's market*) is non-corrective information because Speaker A was confirming Speaker B's statement (*I heard that he got some apples at the supermarket*).

(ii) Contextual probability of the statement; Statement Type (Probable vs. Improbable): whether the object head noun in Speaker B's statement (e.g., *apples* in (3)) matched or conflicted with the character's preference/need known to Speaker A. For instance, in (3), Speaker B says that s/he heard Jacky bought apples, and Speaker A knows Jacky loves fruit, so (3) shows a contextually probable statement. However, consider an alternative where Speaker B says that s/he heard Jacky bought lettuce. Because Speaker A knows from the preceding context that Jacky hates vegetables, *she bought lettuce* is a contextually improbable statement here.

(iii) Knowledge Type (Shared vs. Privileged): whether the contextual probability of the object head nouns was shared knowledge between the interlocutors (Experiment 1) or Speaker A's privileged knowledge (Experiment 2).

In sum, all three independent variables are defined in terms of the speaker: (i) correctiveness refers to whether the *speaker's* response corrects the addressee or not, (ii) statement type refers to about whether the addressee's statement is contextually (im)probable based on what the *speaker* knows, and (iii) knowledge type refers to whether the knowledge about Speaker B's statement's contextual probability is privileged only to the *speaker* or shared by both speaker and addressee.

In our primary analysis, the dependent variable we measured was the f0 range (calculated by subtracting the f0 minimum from the f0 maximum) in the object head noun of Speaker A's response in Sentence 5 (e.g., *lettuce* in (3)). In a secondary analysis, we also analyzed the noun in Speaker B's utterance in Sentence 4.

2.3 Predictions

We designed two experiments to examine how Speaker A's expectations and observations about Speaker B's knowledge state play a role in Speaker A's production of corrective focus. In other words, do speakers' expectations and observations about addressees' knowledge state influence the realization of corrective focus? As shown in (3) and (4), our predictions and analyses focus on Speaker A's response (labelled as the *critical sentence* in (3) and (4)).

As discussed in Section 1.1, we expected the critical sentence uttered by Speaker A to show effects of (i) Correctiveness (corrective vs. non-corrective), (ii) Statement Type (Did the conversational partner express probable vs. improbable information?), and (iii) Knowledge Type (Is the conversational partner aware of the (im)probability of their statement?). We predicted that all three independent variables would influence the dependent variable, namely the f0 range of the critical noun in Sentence 5, uttered by Speaker A.

As mentioned in Section 1.1, correctively-focused words tend to have larger f0 ranges than non-corrective words. This pattern could potentially hold across different conditions of Knowledge Type and Statement Type (i.e., a "main effect" in statistical terms). Furthermore, the effects of Knowledge Type and Statement Type could potentially be dependent on each other (i.e., an "interaction" in statistical terms), as suggested in Kuhlen and Brennan (2010). If interactions are found in our study, we could then look at their directions. A better understanding of such potential interactions would allow us to better understand the interlocutor- and speaker-based factors that guide how correction – a fundamental aspect of human interaction – is prosodically encoded.

2.4 Norming study

The contextual probabilities of the critical words were estimated through a web-based norming study. Participants performed a rating task where they saw sequences of four sentences and judged the probability of the fourth sentence using a seven-point scale, where 1 = very unlikely and 7 = very likely. The first three sentences in a sequence constructed a scenario involving someone's preference (e.g., *Jacky likes her salad a certain way. She loves fruit but hates vegetables. Yesterday evening she went grocery shopping.*) or need (e.g., *Gary just moved into his new house. He has enough bathroom stuff but needs more things for the patio. This afternoon he went out shopping.*) The last sentence described something this person did (e.g., *She got some apples at the farmer's market.*) Participants were told to rate "how likely it is that the following event [referring to the last sentence] took place" "if the above statements [referring to the first three sentences] are true".

We tested various choices with respect to the person's preference/need, the verbs and nouns used for each sentence, and the background setting of the scenario. Six scenarios were ultimately selected for the production study, each with four objects and two background settings. All the selected items had median ratings not smaller than 5 in the probable condition and not larger than 3 in the improbable condition (recall that ratings were done using a seven-point scale where 1 = very unlikely and 7 = very likely). The median rating of each selected item is reported in Appendix.

2.5 Data analysis

The central analysis of interest that is relevant for our predictions has to do *with Speaker A's responses* (Sentence 5, marked as the critical sentence in (3) and (4)). 1,632 utterances were collected from the 43 participants, each producing 24–48 target responses. Out of the full set of data, 171 utterances (10.5%) were not included in the data analysis due to speech errors, disfluencies, or technical issues with the audio recording.

We also conducted a further, secondary analysis of *Speaker B's second statements (Sentence 4)*. This secondary analysis is not related to our hypotheses or predictions. Instead, it was done to address two potential concerns: First, entrainment (also called accommodation, adaptation, or alignment) between interlocutors has been found in the prosody domain (e.g., Lee et al. 2016; Levitan and Hirschberg 2011; Levitan et al. 2011). Since Sentences 4 and 5 were similar in content and immediately followed each other, we wanted to check whether Speaker A's responses prosodically resembled Speaker B's statements. Second, a word's contextual probability given the

sentence has been shown to affect its prosodic realization (e.g., Clopper and Pierrehumbert 2008; Lieberman 1963; Scarborough 2010). If the prosody of the statements clearly indicated whether Speaker B was aware of the characters' preferences, it could have affected the way Speaker A responded. We wanted to check whether this potential confound interfered with our manipulation of Knowledge Type. In sum, we performed a supplemental analysis to examine whether Speaker A's prosody could be skewed by Speaker B's prosody in any way relevant to our claims (as we will see below, it turns out that no, it was not being skewed, so this is not a problem). Importantly, *Speaker A*'s responses are the main focus of our research; Speaker B's statements themselves are not pertinent to our hypotheses.

F0 measurements were obtained using the YAAPT (Yet Another Algorithm for Pitch Tracking) algorithm (Zahorian and Hu 2008). The raw f0 values were then smoothed using the *smoothn* function in MATLAB (MATLAB release 2004a; *smoothn* in MATLAB: Garcia 2010) to remove f0 tracking errors and segmental effects. The smoothed values were then converted into a semitone scale, as semitones reflect pitch perception better than the Hertz scale (e.g., Nolan 2003). Finally, the data were normalized by subject using *z*-scores, to factor out individual differences in f0 registers (e.g., women usually have wider and higher registers of f0 than men). The *z*-scores represented each data point in terms of its number of standard deviations above or below the mean across all utterances produced by a given speaker.

Mixed-effects models were conducted on f0 ranges using statistical packages in the R software (R version 3.3.1, 2006-06-21; *anova* in R: Chambers and Hastie 1992; *lme4* in R: Bates et al. 2015; *lmerTest* in R: Kuznetsova et al. 2017). Correctiveness, Statement Type, and Knowledge Type were included as fixed effects; Subject and Scenario were included as random effects. When specifying the structure of random effects, we started with a full model (i.e., including intercepts and slopes for Subject and Scenario) and excluded a random slope when it did not significantly contribute to the model. The final analyses did not have random slopes but included random intercepts for Subject and Scenario.

3 Results: speaker A's responses

Overall, our predictions regarding Speaker A's responses were borne out. Figures 1 and 2 show the average f0 ranges (in terms of *z*-scored semitones) of the critical nouns in Speaker A's responses (Sentence 5). Critical nouns that were in Corrective focus (e.g., *lettuce* in "*No, he got some lettuce at the supermarket*) and those that were conveying Non-Corrective information (e.g., *lettuce* in *Yes, he got some lettuce at the supermarket*) are presented in separate figures. The left half of each figure is data from Experiment 2, where the contextual probability of the critical nouns was Privileged knowledge, whereas the right half of each figure is data from

Experiment 1, where everything was Shared knowledge. Speaker A's responses to Speaker B's Probable statements are represented by the bars with red, horizontal stripes (e.g., Speaker B: *I heard that he got some spinach at the supermarket*. Speaker A: *No, he got some lettuce at the supermarket*" – in scenarios where *he*/Zac loves vegetables). In contrast, responses to improbable statements are represented by the bars with blue, vertical stripes (e.g., Speaker B: *I heard that he got some cherries at the supermarket*. Speaker A: *No, he got some lettuce at the supermarket*. – given that *he*/Zac hates fruit).

As can be seen in the figures, Speaker A's corrective responses (Figure 1) had larger f0 ranges than their non-corrective responses (Figure 2) across the board

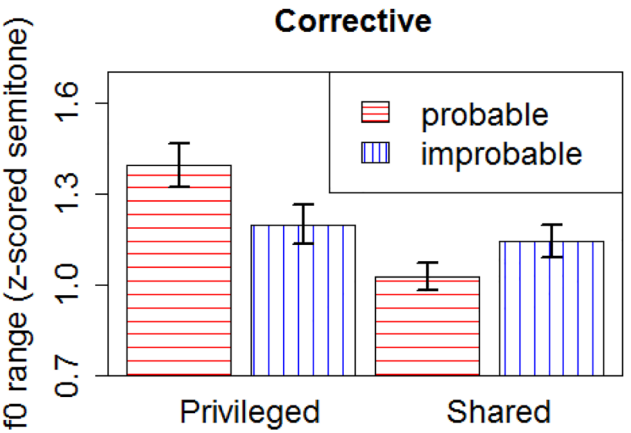

Figure 1: Mean f0 range of the critical noun in Speaker A's response (Sentence 5) in Corrective conditions.

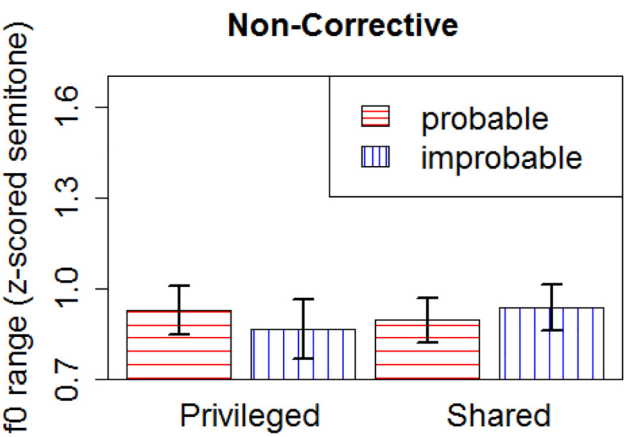

Figure 2: Mean f0 range of the critical noun in Speaker A's response (Sentence 5) in Non-Corrective conditions.

(corrective mean = 1.168; non-corrective mean = 0.908). These patterns are also confirmed statistically: Pitch ranges on the critical noun in Sentence 5 are significantly greater when the noun is corrective than when it is not (a main effect of correctiveness, $t = 2.919$, $p < 0.01$). This pattern holds regardless of whether the noun being corrected was probable or improbable (Statement Type), and whether or not Speaker B was aware of this (Knowledge Type): Correctiveness does not interact with either Statement Type ($t = 1.082$, $p = 0.280$) or Knowledge Type ($t = 0.744$, $p = 0.457$).

However, we do find an interaction of Statement Type (Probable/Improbable) and Knowledge Type (Privileged/Shared; $t = 2.827$, $p < 0.01$). Although the three-way interaction between Correctiveness, Statement Type and Knowledge Type was not significant ($t = 1.170$, $p = 0.242$), planned comparisons revealed that the interaction between Statement Type and Knowledge Type affected f0 ranges in the corrective responses ($t = 2.799$, $p = 0.005$), but not in the non-corrective responses ($t = 0.417$, $p = 0.677$).[3] As shown in Figure 1, f0 ranges of Speaker A's corrective responses varied depending on whether the statements being responded to were contextually probable, and whether Speaker B knew the contextual probability of his or her statements. In contrast, f0 ranges of the non-corrective responses are invariant across conditions, as can be observed from Figure 2. The statistics reported earlier confirm these visual observations.

More specifically, when the contextual probability of the object nouns was shared knowledge, Speaker B's improbable misstatements elicited significantly larger f0 ranges (in Speaker A's corrective responses) than their probable misstatements ($t = 1.966$, $p < 0.05$; improbable mean = 1.140; probable mean = 1.024). In contrast, when the contextual probability of the object nouns was Speaker A's privileged knowledge, Speaker B's probable misstatements elicited significantly larger f0 ranges (in Speaker A's corrective responses) than their improbable misstatements ($t = 2.043$, $p < 0.05$; probable mean = 1.392; improbable mean = 1.196). This interaction was not found in non-corrective responses (Statement Type in Experiment 1: $t = 0.552$, $p = 0.581$; Statement Type in Experiment 2: $t = 0.095$, $p = 0.924$).

3.1 Secondary analysis: Speaker B's statements

Now let us take a look at the results regarding Speaker B's statements (Sentence 4). As mentioned in Section 2.5, we analyzed the statements to check whether the

3 Here we primarily focused on the differences between Statement Types (probable vs. improbable) and did not directly compare between Knowledge Types (shared vs. privileged), because Knowledge Type was a between-subjects variable manipulated between experiments. Nevertheless, we provide speculative discussion on this topic in Section 4.

prosody of Speaker A's responses was correlated with the prosody of Speaker B's statements. Speaker B's sentences were *not* the target sentences of this study, but we checked them to rule out potential concerns.

Recall that there is a main effect of Correctiveness and an interaction between Statement Type (probable vs. improbable) and Knowledge Type (does Speaker B know whether their statement is probable or improbable?) in *Speaker A's* responses. Speaker B's statements do not exhibit this pattern. We found neither a main effect of Correctiveness (t = 0.452, p = 0.651) nor an interaction between Statement Type and Knowledge Type (t = 0.267, p = 0.790) in Speaker B's statements. In fact, no factor examined in this study had any significant effect (t's ≤ 0.452, p's ≥ 0.651). F0 ranges of Speaker B's statements were consistent throughout the experiments, regardless of whether the statements were probable or improbable, whether they were misstatements (later corrected by Speaker A), and whether Speaker B knew the contextual probability of the statements or not. In other words, we did not observe any evidence for prosodic influence between the interlocutors that would be pertinent to the aims of this study.

4 Discussion

The two experiments reported in this paper investigated the modulation of corrective prosody based on the speaker's expectation and realization about the addressee's knowledge state. As discussed in Section 1, previous research has not paid close attention to the interplay between these two dimensions of perspective-taking. A better understanding of these issues is important because they tap into fundamental questions regarding the communicative functions of prosody, as well as the relation between top-down and bottom-up aspects of language production. This work relates to questions of how and whether speakers' top-down expectations – in particular, assumptions about their addressee's knowledge state (about the contextual probability of different statements) – influence the prosodic realization of their utterances, an aspect of speech production that seems susceptible to both bottom-up and top-down influences.

Our results show that corrective prosody is influenced both by (i) speakers' prior expectations about their addressees' knowledge status and by (ii) what speakers learn about their addressees' knowledge status during the conversation. Our participants consistently emphasized corrective information over non-corrective information in terms of f0 ranges. However, the degree of emphasis they placed on corrective information varied depending on their conversational partner's knowledge state (i.e., whether the conversational partner is aware of the (im)probability of their statement). In Experiment 1, where both participants were

expected to be fully aware of the context, they corrected each other with greater prosodic prominence when the misstatements were contextually improbable. In contrast, in Experiment 2, where the speaker of the misstatements [Speaker B] was *not* expected to have any knowledge about the context, the speaker of the corrections [Speaker A] produced greater prosodic prominence when the misstatements were contextually probable. This supports our general prediction that use of corrective prosody by the speaker is modulated by the addressee's knowledge state, and involves an interplay between the speaker's prior expectations of the addressee's knowledge state and what the speaker learns about the addressee's knowledge state during the conversation.

4.1 Prosodic consequences of an epistemic gap

Why should the speaker's prior expectations and the addressee's actual behavior interact and influence corrective prosody in the way we observed? We suggest that our results can be explained in terms of the epistemic gap between expectation and reality – i.e., in terms of the gap between what the speaker had expected the addressee to know (and therefore what they had expected to addressee to say) and what the addressee actually appears to know (as indicated by what they actually say). The level of prosodic prominence, such as f0 ranges (what would be argued to be bottom-up aspects of the speech signal, as they describe purely acoustic-phonetic properties of the speech signal), might reflect the extent to which speakers are 'surprised' at what they encounter during the conversation (arguably a more top-down aspect of communication). Our use of the word *surprise* harks back to the notion of surprisal/surprise in information-theoretic work, where (simplifying somewhat) *surprisal* has to do with the amount of information provided by a particular element, for example how (un)predictable/(im)probable a word is in a given context (e.g., Hale 2001; Levy 2008; Mahowald et al. 2013; see also Koornneef, this issue, for related discussion on lexical prediction): Low-probability words have higher surprisal than high-probability words. Previous research has investigated various types of predictability/probability such as word frequency, phonological neighborhood density, and conditional probability, among others, and found that these characteristics of a word affect how people say it prosodically (Gregory et al. 1999; Munson and Solomon 2004; van Son and Pols 1999; Wright 2004). In our study, participants had assumptions about what the other person knew and thus were able to form expectations about the kinds of things their interlocutor might say in a conversation. If the other person's utterances contradicted those prior assumptions (in other words, if the content of a

person's utterances was low-probability, given the context of the other person's prior knowledge), speakers marked this surprisal prosodically in their responses.

Let us consider the communicative setting in Experiment 1, where the characters' preferences and needs were mentioned in the dialogues. Expecting their partner to have this knowledge, participants might not have found it unexpected when their partner had misbeliefs consistent with the characters' preferences and needs. For example, since participants had told their partner that Jacky loves fruit, it would be natural that their partner thought that Jacky bought apples. In contrast, it might have struck participants as unexpected when their supposedly-informed partner had misbeliefs conflicting with the context. When both people knew that Jacky hates vegetables, it would be rather unexpected that the partner made a mistake and stated that Jacky bought lettuce. This epistemic gap might be what elicited extra prosodic prominence in the corrective responses to contextually improbable misstatements.

Probably for similar reasons, the opposite patterns were yielded by the different communicative setting in Experiment 2, where the characters' preferences and needs were privileged information known only to Speaker A. Here, participants might have found their partner's probable misstatements *unexpected*, because it might have seemed that their partner was somehow able to form probable beliefs *without* the critical knowledge (in other words, if someone appears to know something quite specific when I did not expect them to have that specific knowledge, this may strike me as unexpected). In contrast, improbable misstatements might not have struck participants as odd, since improbable beliefs might have fit their assumption about their partner's lack of critical knowledge.[4] In other words, the extra prosodic prominence still reflects an epistemic gap, but compared to Experiment 1, (im)probability of the misstatements had the opposite effect in Experiment 2, because the participants had the opposite expectation about their partner's knowledge state.

Note that the results of this study cannot be explained if speakers are either completely egocentric or completely addressee-oriented. Egocentric speakers would behave the same way regardless of whether the critical information is shared knowledge or not. In contrast, addressee-oriented speakers would not be "surprised" by their addressee's assumptions when the critical information is privileged knowledge, since all assumptions would be equally probable to the addressee.

4 The nature of the relationship between surprisal and speakers' reactions to (uninformed) interlocutors' statements in the Privileged knowledge situation is an area that would benefit from future work. Here we observe that unexpected knowledge of something is more surprising than lack of knowledge, but it is important to investigate this further.

4.2 Further evidence for the prosodic realization of an epistemic gap

Results of the current study confirm the ideas we presented in Ouyang and Kaiser (2014, 2015) – one of our previous studies on the prosodic encoding of informativity. In that earlier work, we examined the interaction between information-structural and information-theoretic factors. We found complex patterns showing that a word's information-theoretic properties, such as word frequency and contextual probability, modulate the prosodic realization of its information-structural status, such as being new or corrective information in the discourse. We explained our findings in terms of "epistemic surprisal": "The higher the degree of epistemic surprisal experienced by the speaker, the higher the degree of prosodic prominence on the relevant words" (Ouyang 2015: 29).

As mentioned in Ouyang and Kaiser (2014), our theory about the epistemic gap between expectation and reality is related to the notion of "surprise" developed by researchers in intonation and social interaction (e.g., Dombrowski 2003; Local 1996; Niebuhr and Zellers 2013; Selting 1996; Wilkinson and Kitzinger 2006). Crucially, existing work distinguishes "the social expression of surprise (the public display of finding something counter to expectation)" from "the psychology of surprise (the emotional experience of encountering the unexpected)" (Wilkinson and Kitzinger 2006: 152). Our thinking is connected to the former.

These ideas regarding the epistemic gap are also supported by our other work (Ouyang et al. 2017), where we used an interactive task to look at another information-structural distinction in addition to corrective focus, namely *given* (already-mentioned) versus *new* information. Our aim was to see if speakers' prosodic encoding of the given-new distinction is also modulated by their assumptions and observations about the addressee's knowledge state. In Ouyang et al. (2017), we also tested whether speakers dynamically update their expectations and adjust their prosody based on addressees' behavior. Our results show that speakers' prosodic realization of givenness is indeed sensitive to both (i) speakers' prior assumptions about the addressee and (ii) speakers' observations of the addressee over the course of the conversation. When the observations do not match the assumptions, the gap between expectation and reality leads to prosodic emphasis, but speakers are able to (at least partially) recalibrate their initial assumptions to reflect the addressee's true knowledge-state characteristics that are revealed over the course of the conversation. These findings are consistent with the notion of 'epistemic surprisal' that we propose in the present paper.

4.3 Conclusions

It is well known that corrective information is produced with higher prosodic prominence than non-corrective information. However, it remains unclear how corrective prosody is realized in different communicative settings, in particular whether interlocutors' prosodic realization of corrective focus depends on each other's knowledge state. Existing studies have shown that, broadly speaking, speakers' responses are influenced by both what they assume their addressee to know and what their addressee appears to know. We conducted two production experiments to examine how speakers' expectations and observations about their interlocutor's knowledge state play a role in speakers' production of corrective focus. We used an experimental approach, but hope that future work on naturalistic data can provide converging evidence relevant to our findings.

The results from our production experiments point to a nuanced interplay between speakers' prior expectations and interlocutors' actual communicative behavior: We suggest that the prosody of speakers' responses can reveal what we call an "epistemic gap" between speakers' expectations (e.g., what a speaker thinks other people are likely to say) and the reality they encounter (e.g., what other people actually say). These findings add to our understanding of an under-researched but communicatively fundamental area, namely how speakers update their assumptions about their addressee by using incoming cues from their addressee's behavior (e.g., Brennan 1991; Kuhlen and Brennan 2010; Russell and Schober 1999).

Our idea that speakers' expectations modulate the encoding of corrective information is in line with work in other areas of experimental linguistics. Indeed, a substantial amount of research has investigated the effects of expectations in language comprehension (e.g., Altmann and Kamide 1999; Boland 2005; Brown et al. 2011; DeLong et al. 2005; Ito and Speer 2008; Watson et al. 2008; van Berkum et al. 2005; Wicha et al. 2003, see also Koehne et al., this issue, and Koornneef, this issue) and production (e.g., Bell et al. 2003; Clopper and Pierrehumbert 2008; Pan and Hirschberg 2000; Scarborough 2010; van Son et al. 1999). Predictive processing has been shown to occur at various linguistic levels, and our work contributes to this literature by exploring the prosodic domain.

4.4 Directions for future work

Many intriguing questions still remain open for future work. For example, we did not directly compare between Knowledge Types (shared vs. privileged) in Section 3, because this study was not designed with the statistical power for a

between-experiments and between-subjects comparison. A reviewer pointed out that the difference between the correction of probable and improbable assumptions seemed to be smaller when knowledge of the probability was shared (t = 1.966) than when it was not (t = 2.043), although the difference was statistically significant in both cases. Here we address this observation, but the reader should keep in mind that a statistical comparison between Knowledge Types would not be meaningful. As the reviewer also pointed out, there seemed to be less difference between shared and privileged knowledge conditions when correcting improbable assumptions than when correcting probable assumptions. If both observations are true, corrections are prosodically most prominent when the speaker is correcting probable statements of which the probability is unknown to the interlocutor, and less prominent in all the other Knowledge and Statement Types. This would lead to alternative theories of how speakers take their addressee's viewpoint into account. For example, participants in the shared knowledge condition might not have felt the need to prosodically mark corrections based on the probability of their interlocutor's statement, since all information was shared. Furthermore, when contextual probability was privileged knowledge, participants might have tried to share the information through prosody by emphasizing their response when their interlocutor made probable assumptions. We leave the question open for future work.

When studying correction, questions regarding politeness, emotional involvement and face-saving strategies may also come to mind. Our study did not test for or to manipulate factors related to politeness, and our dialogs were such that correcting someone was presented as a natural part of the experiment. However, a better understanding of how and whether considerations related to politeness modulate prosodic realization of corrective focus is a rich avenue for future work.

Another important area for future work has to do with speech acts other than declaratives. Much of the psycholinguistic, experimentally-oriented work on the production and perception of focus has investigated declarative statements. Speech acts such as requests and commands are, to the best of our knowledge, very under-researched in this domain, although they can potentially yield important insights concerning the communicative dynamics of correction as it relates to people's expectations about their interlocutors' knowledge states.

Acknowledgments: We gratefully acknowledge helpful feedback and comments from the audience at DETEC 2015, held at the University of Alberta, Canada, as well as anonymous reviewers. This material is based upon work supported by the National Science Foundation under Grant No. BCS-1451596. An earlier version of some of this work was presented at the 2015 conference on Architectures and Mechanisms for Language Processing (AMLaP) and at the 8th International Conference on Speech Prosody, and is included in the Speech Prosody 2016 proceedings.

Appendix Target items

The 192 critical dialogues in Experiments 1 and 2 are recoverable as follows. Each dialogue has five sentences (Sentences 1–5); the last two sentences of each dialogue (Sentence 4 and Sentence 5) constitute a statement-response pair. In Experiment 1, Sentences 1, 2, and 5 are seen and said aloud by one participant (Speaker A), whereas Sentences 3 and 4 are seen and said aloud by the other participant (Speaker B). Experiment 2 follows the same design as Experiment 1, except Sentence 2 is never said aloud.

There are six conditions, formed by combining two kinds of statement-response relationship (Corrective vs. Non-Corrective), two types of object nouns in the statements (Probable vs. Improbable), and two types of object nouns in the responses (Probable vs. Improbable). A corrective response begins with the word 'no' and has an object noun that is different from the object noun in the statement. In contrast, a non-corrective response begins with the word 'yes' and has the same object noun as the statement.

Every condition has items from six scenarios (Scenarios 1–6). Each scenario contains two sub-scenarios (X vs. Y, e.g., Christopher prefers meat vs. Abbey prefers seafood) and is associated with a set of four object nouns (e.g., *beef*, *lamb*, *fish*, and *shrimp*). Two of the four object nouns are probable in the context of sub-scenario X but improbable in sub-scenario Y. The other two object nouns, in contrast, are probable in the context of sub-scenario Y but improbable in sub-scenario X.

These target items were selected based on the results of a norming study, where the contextual probability of each f object noun given the sub-scenario was judged on a seven-point scale (see Section 2.1.2). We provide the median likelihood ratings in the tables below.

Scenario 1: Preference: meat versus seafood

Sub-scenario 1X: Probable Objects: *beef, lamb*; Improbable Objects: *fish, shrimp*.

Likelihood	Object	Median likelihood rating
Probable	Beef	6
Probable	Lamb	6
Improbable	Fish	1
Improbable	Shrimp	1

Sentence 1 [Speaker A]: {Christopher; Gary; Joseph} is particular about food.
Sentence 2 [Speaker A]: (He loves meat but hates seafood.)
Sentence 3 [Speaker B]: He went out for lunch yesterday.
Sentence 4 [Speaker B]: I heard that he had {beef; lamb; fish; shrimp} at a restaurant.
Sentence 5 [Speaker A]: {Yes; No}, he had {beef; lamb; fish; shrimp} at a restaurant.

Sub-scenario 1Y: Probable Objects: *fish, shrimp*; Improbable Objects: *beef, lamb*.

Likelihood	Object	Median likelihood rating
Probable	Fish	6.5
Probable	Shrimp	6
Improbable	Beef	1
Improbable	Lamb	1.5

Sentence 1 [Speaker A]: {Abbey; Lauren; Sabrina} is a picky eater.
Sentence 2 [Speaker A]: (She loves seafood but hates meat.)
Sentence 3 [Speaker B]: She went out for dinner last night.
Sentence 4 [Speaker B]: I heard that she had {beef; lamb; fish; shrimp} at a restaurant.
Sentence 5 [Speaker A]: {Yes; No}, she had {beef; lamb; fish; shrimp} at a restaurant.

Scenario 2: Preference: fruit versus vegetables

Sub-scenario 2X: Probable Objects: *apples, cherries*; Improbable Objects: *lettuce, spinach*.

Likelihood	Object	Median likelihood rating
Probable	Apples	6
Probable	Cherries	6
Improbable	Lettuce	3
Improbable	Spinach	2.5

Sentence 1 [Speaker A]: {Elisa; Jacky; Tina} prefers her salad a certain way.
Sentence 2 [Speaker A]: She loves fruit but hates vegetables.
Sentence 3 [Speaker B]: She went grocery shopping yesterday evening.
Sentence 4 [Speaker B]: I heard that she got some {apples; cherries; lettuce; spinach} at the farmer's market.
Sentence 5 [Speaker A]: {Yes; No}, she got some {apples; cherries; lettuce; spinach} at the farmer's market.

Sub-scenario 2Y: Probable Objects: *lettuce, spinach*; Improbable Objects: *apples, cherries*.

Likelihood	Object	Median likelihood rating
Probable	Lettuce	6
Probable	Spinach	6
Improbable	Apples	2
Improbable	Cherries	2.5

Sentence 1 [Speaker A]: {Bob; Daniel; Zac} tends to put certain things in his salad.
Sentence 2 [Speaker A]: (He loves vegetables but hates fruit.)
Sentence 3 [Speaker B]: He went grocery shopping this morning.
Sentence 4 [Speaker B]: I heard that he got some {apples; cherries; lettuce; spinach} at the supermarket.
Sentence 5 [Speaker A]: {Yes; No}, he got some {apples; cherries; lettuce; spinach} at the supermarket.

Scenario 3: Need: getting bathroom versus patio stuff

Sub-scenario 3X: Probable Objects: *bath mats, face wash*; Improbable Objects: *lawn chairs, yard lights*.

Likelihood	Object	Median likelihood rating
Probable	Bath mats	7
Probable	Face wash	6
Improbable	Lawn chairs	2
Improbable	Yard lights	2

Sentence 1 [Speaker A]: {Abbey; Lauren; Sabrina} just moved into her new house.
Sentence 2 [Speaker A]: (She has enough patio stuff but needs more things for the bathroom.)
Sentence 3 [Speaker B]: She went out shopping today.
Sentence 4 [Speaker B]: I heard that she bought some {bath mats; face wash; lawn chairs; yard lights} at a store.
Sentence 5 [Speaker A]: {Yes; No}, she bought some {bath mats; face wash; lawn chairs; yard lights} at a store.

Sub-scenario 3Y: Probable Objects: *lawn chairs, yard lights*; Improbable Objects: *bath mats, face wash*.

Likelihood	Object	Median likelihood rating
Probable	Lawn chairs	6
Probable	Yard lights	6
Improbable	Bath mats	2
Improbable	Face wash	2

Sentence 1 [Speaker A]: {Christopher; Gary; Joseph} just moved into his new house.
Sentence 2 [Speaker A]: (He has enough bathroom stuff but needs more things for the patio.)
Sentence 3 [Speaker B]: He went out shopping this afternoon.
Sentence 4 [Speaker B]: I heard that he bought some {bath mats; face wash; lawn chairs; yard lights} at the mall.
Sentence 5 [Speaker A]: {Yes; No}, he bought some {bath mats; face wash; lawn chairs; yard lights} at the mall.

Scenario 4: Need: selling bedroom versus kitchen stuff

Sub-scenario 4X: Probable Objects: *dresser, mattress*; Improbable Objects: *blender, mixer*.

Likelihood	Object	Median likelihood rating
Probable	Dresser	6
Probable	Mattress	6.5
Improbable	Blender	2.5
Improbable	Mixer	2

Sentence 1 [Speaker A]: {Elisa; Jacky; Tina} is moving in with her boyfriend.
Sentence 2 [Speaker A]: (They want to keep her kitchen stuff but get rid of her bedroom furniture.)
Sentence 3 [Speaker B]: They are not going to donate anything.
Sentence 4 [Speaker B]: I heard that she sold her {dresser; mattress; blender; mixer} at a garage sale.
Sentence 5 [Speaker A]: {Yes; No}, she sold her {dresser; mattress; blender; mixer} at a garage sale.

Sub-scenario 4Y: Probable Objects: *blender, mixer*; Improbable Objects: *dresser, mattress*.

Likelihood	Object	Median likelihood rating
Probable	Blender	6
Probable	Mixer	6.5
Improbable	Dresser	2
Improbable	Mattress	2

Sentence 1 [Speaker A]: {Bob; Daniel; Zac} just moved into his new house.
Sentence 2 [Speaker A]: (They want to keep his bedroom furniture but get rid of his kitchen stuff.)
Sentence 3 [Speaker B]: They are not going to give anything away.
Sentence 4 [Speaker B]: I heard that he sold his {dresser; mattress; blender; mixer} in the classified ads.
Sentence 5 [Speaker A]: {Yes; No}, he sold his {dresser; mattress; blender; mixer} in the classified ads.

Scenario 5: Preference: playing carpenter versus chef

Sub-scenario 5X: Probable Objects: *hammers, wrenches*; Improbable Objects: *burgers, pizzas*.

Likelihood	Object	Median likelihood rating
Probable	Hammers	6.5
Probable	Wrenches	6
Improbable	Burgers	2
Improbable	Pizzas	2

Sentence 1 [Speaker A]: {Christopher; Gary; Joseph}'s nephew is six, and he likes to imagine what he wants to do when he grows up.
Sentence 2 [Speaker A]: (He loves to pretend to be a carpenter, but never plays chef.)
Sentence 3 [Speaker B]: {Christopher; Gary; Joseph} went to a flea market over the weekend.
Sentence 4 [Speaker B]: I heard that he bought toy {burgers; pizzas; hammers; wrenches} for his nephew.
Sentence 5 [Speaker A]: {Yes; No}, he bought toy {burgers; pizzas; hammers; wrenches} for his nephew.

Sub-scenario 5Y: Probable Objects: *burgers, pizzas*; Improbable Objects: *hammers, wrenches*.

Likelihood	Object	Median likelihood rating
Probable	Burgers	6
Probable	Pizzas	6
Improbable	Hammers	2
Improbable	Wrenches	2

Sentence 1 [Speaker A]: {Abbey; Lauren; Sabrina}'s niece is four, and she likes to play make-believe.
Sentence 2 [Speaker A]: (She loves to pretend to be a chef, but never plays carpenter.)
Sentence 3 [Speaker B]: {Abbey; Lauren; Sabrina} found some make-believe props on eBay recently.
Sentence 4 [Speaker B]: I heard that she bought toy {burgers; pizzas; hammers; wrenches} for her niece.
Sentence 5 [Speaker A]: {Yes; No}, she bought toy {burgers; pizzas; hammers; wrenches} for her niece.

Scenario 6: Preference: farm versus jungle animals

Sub-scenario 6X: Probable Objects: *cow, sheep*; Improbable Objects: *bear, lion*.

Likelihood	Object	Median likelihood rating
Probable	Cow	6.5
Probable	Sheep	6
Improbable	Bear	2.5
Improbable	Lion	2

Sentence 1 [Speaker A]: {Elisa; Jacky; Tina}'s son is not a fan of all animals.
Sentence 2 [Speaker A]: (He is obsessed with farm animals but completely uninterested in jungle animals.)
Sentence 3 [Speaker B]: She took her son to a toy store downtown.
Sentence 4 [Speaker B]: I heard that he got a stuffed {cow; sheep; bear; lion} at the shop.
Sentence 5 [Speaker A]: {Yes; No}, he got a stuffed {cow; sheep; bear; lion} at the shop.

Sub-scenario 6Y: Probable Objects: *bear, lion*; Improbable Objects: *cow, sheep*.

Likelihood	Object	Median likelihood rating
Probable	Bear	5.5
Probable	Lion	6
Improbable	Cow	1.5
Improbable	Sheep	1.5

Sentence 1 [Speaker A]: {Bob; Daniel; Zac}'s daughter only likes certain animals.
Sentence 2 [Speaker A]: (She is obsessed with jungle animals but completely uninterested in farm animals.)
Sentence 3 [Speaker B]: He took his daughter to a kid's store the other day.
Sentence 4 [Speaker B]: I heard that she got a stuffed {cow; sheep; bear; lion} at the shop.
Sentence 5 [Speaker A]: {Yes; No}, she got a stuffed {cow; sheep; bear; lion} at the shop.

References

Altmann, Gerry T. M. & Yuki Kamide. 1999. Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition* 73(3). 247–264.

Arnold, Jennifer E., Jason M. Kahn & Giulia C. Pancani. 2012. Audience design affects acoustic reduction via production facilitation. *Psychonomic Bulletin & Review* 19(3). 505–512.

Baker, Rachel E. & Ann R. Bradlow. 2009. Variability in word duration as a function of probability, speech style, and prosody. *Language and Speech* 52(4). 391–413.

Bard, Ellen Gurman, Anne H. Anderson, Catherine Sotillo, Matthew Aylett, Gwyneth Doherty-Sneddon & Alison Newlands. 2000. Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language* 42(1). 1–22.

Bates, Douglas, Martin Mächler, Bolker Ben & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48.

Bell, Alan, Daniel Jurafsky, Eric Fosler-Lussier, Cynthia Girand, Michelle Gregory & Daniel Gildea. 2003. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America* 113(2). 1001–1024.

van Berkum, Jos J. A., Colin M. Brown, Pienie Zwitserlood, Valesca Kooijman & Peter Hagoort. 2005. Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31(3). 443.

Boland, Julie E. 2005. Visual arguments. *Cognition* 95(3). 237–274.

Breen, Mara, Evelina Fedorenko, Michael Wagner & Edward Gibson. E. 2010. Acoustic correlates of information structure. *Language and Cognitive Processes* 25(7–9). 1044–1098.

Brennan, Susan E. 1991. Conversation with and through computers. *User Modeling and User-Adapted Interaction* 1(1). 67–86.

Brown, Gillian. 1983. Prosodic structure and the given/new distinction. In Anne Cutler & D. Robert Ladd (eds.), *Prosody: Models and measurements (Springer series in language and communication 14)*, 67–77. Berlin & Heidelberg: Springer.

Brown, Meredith, Anne Pier Salverda, Laura C. Dilley & Michael K. Tanenhaus. 2011. Expectations from preceding prosody influence segmentation in online sentence processing. *Psychonomic Bulletin & Review* 18(6). 1189–1196.

Chambers, John M. & Trevor J. Hastie. 1992. *Statistical models in S*. Pacific Grove, CA: Wadsworth & Brooks/Cole.

Clopper, Cynthia G. & Janet B. Pierrehumbert. 2008. Effects of semantic predictability and regional dialect on vowel space reduction. *The Journal of the Acoustical Society of America* 124(3). 1682–1688.

Cooper, William E., Stephen J. Eady & Pamela R. Mueller. 1985. Acoustical aspects of contrastive stress in question–answer contexts. *The Journal of the Acoustical Society of America* 77(6). 2142–2156.

Couper-Kuhlen, Elizabeth. 1984. A new look at contrastive intonation. In Ernst Leisi (ed.), *Modes of interpretation: Essays presented to Ernst Leisi on the occasion of his 65th birthday (Tubinger Beitrage zur Linguistik 260)*, 137–158. Tübingen: Gunter Narr.

DeLong, Katherine A., Thomas P. Urbach & Marta Kutas. 2005. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience* 8(8). 1117–1121.

Dombrowski, Ernst. 2003. Semantic features of accent contours: Effects of f0 peak position and f0 time shape. In *Proceedings of the 15th international congress of phonetic sciences*, 1217–1220. ICPhS 2003.

Dik, Simon C. & Kees Hengeveld. 1997. *The theory of functional grammar: The structure of the clause*. Berlin & New York: Mouton De Gruyter.

Eady, Stephen J. & William E. Cooper. 1986. Speech intonation and focus location in matched statements and questions. *The Journal of the Acoustical Society of America* 80(2). 402–415.

Ferguson, S. Hargus. 2004. Talker differences in clear and conversational speech: Vowel intelligibility for normal hearing listeners. *The Journal of the Acoustical Society of America* 116(4). 2365–2373.

Ferguson, S. Hargus & Diane Kewley-Port. 2007. Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research* 50. 1241–1255.

Fowler, Carol A. 1992. Vowel duration and closure duration in voiced and unvoiced stops: There are no contrast effects here. *Journal of Phonetics* 20(1). 143–165.

Fussell, Susan R. & Robert M. Krauss. 1989. The effects of intended audience on message production and comprehension: Reference in a common ground framework. *Journal of Experimental Social Psychology* 25(3). 203–219.

Galati, Alexia & Susan E. Brennan. 2010. Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language* 62(1). 35–51.

Garcia, Damien. 2010. Robust smoothing of gridded data in one and higher dimensions with missing values. *Computational Statistics & Data Analysis* 54(4). 1167–1178.

Gregory, Michelle L., William D. Raymond, Alan Bell, Eric Fosler-Lussier & Daniel Jurafsky. 1999. The effects of collocational strength and contextual predictability in lexical production. *CLS* 35(2). 151–166.

Gussenhoven, Carlos. 2008. Types of focus in English. In Chungmin Lee, Matthew Gordon & Daniel Büring (eds.), *Topic and focus: Cross-linguistic perspectives on meaning and intonation*, 83–100. Dordrecht: Springer.

Hale, John. 2001. A probabilistic earley parser as a psycholinguistic model. In *NAACL '01, Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics*, 1–8.

Horton, William S. & Boaz Keysar. 1996. When do speakers take into account common ground? *Cognition* 59(1). 91–117.

Hummert, Mary Lee & Jaye L. Shaner. 1994. Patronizing speech to the elderly as a function of stereotyping. *Communication Studies* 45(2). 145–158.

Isaacs, Ellen A. & Herbert H. Clark. 1987. References in conversation between experts and novices. *Journal of Experimental Psychology: General* 116(1). 26–37.

Ito, Kiwako & Shari R. Speer. 2008. Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language* 58(2). 541–573.

Jun, Sun-Ah. 2003. The effect of phrase length and speech rate on prosodic phrasing. In *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS 2003)*, 483–486.

Kahn, Jason M. & Jennifer E. Arnold. 2012. A processing-centered look at the contribution of givenness to durational reduction. *Journal of Memory and Language* 67(3). 311–325.

Kaland, Constantijn, Marc Swerts & Emiel Krahmer. 2013. Accounting for the listener: Comparing the production of contrastive intonation in typically-developing speakers and speakers with autism. *The Journal of the Acoustical Society of America* 134(3). 2182–2196.

Katz, Jonah & Elisabeth Selkirk. 2011. Contrastive focus vs. discourse-new: Evidence from phonetic prominence in English. *Language* 87(4). 771–816.

Krahmer, Emiel & Marc Swerts. 2001. On the alleged existence of contrastive accents. *Speech Communication* 34(4). 391–405.

Kuhlen, Anna K. & Susan E. Brennan. 2010. Anticipating distracted addressees: How speakers' expectations and addressees' feedback influence storytelling. *Discourse Processes* 47(7). 567–587.

Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13). 1–26.

Lee, Yoonjeong, Samantha Gordon Danner, Benjamin Parrell, Sungbok Lee, Louis Goldstein & Dani Byrd. 2016. Acoustic and articulatory measures of prosodic convergence in a cooperative maze navigation task. *The Journal of the Acoustical Society of America* 140(4). 3344.

Levitan, Rivka, Agustín Gravano & Julia Bell Hirschberg. 2011. Entrainment in speech preceding backchannels. In *Proceedings of the 49th annual meeting of the Association for Computational Linguistics*, 113–117.

Levitan, Rivka & Julia Hirschberg. 2011. Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In *Proceedings of the 12th annual conference of the International Speech Communication Association (Interspeech)*, 3081–3084.
Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106(3). 1126–1177.
Lieberman, Philip. 1963. Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech* 6(3). 172–187.
Local, John. 1996. Conversational phonetics: Some aspects of news receipts in everyday talk. *Studies in Interactional Sociolinguistics* 12. 177–230.
Lockridge, Calion B. & Susan E. Brennan. 2002. Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin & Review* 9(3). 550–557.
Mahowald, Kyle, Evelina Fedorenko, Steven T. Piantadosi & Edward Gibson. 2013. Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition* 126(2). 313–318.
Munson, Benjamin & Nancy Pearl Solomon. 2004. The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research* 47(5). 1048–1058.
MATLAB. 2004. *Release 2004a*. Natick, MA: The MathWorks Inc.
Niebuhr, Oliver & Margaret Zellers. 2013. Late pitch accents in hat and dip intonation patterns. In Oliver Niebuhr (ed.), *Understanding prosody: The role of context, function and communication (Language, Context and Cognition 13)*, 159–186. Berlin & Boston: De Gruyter.
Nolan, Francis. 2003. Intonational equivalence: An experimental evaluation of pitch scales. In *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS 2003)*, 771–774.
Ouyang, Iris Chuoying & Elsi Kaiser. 2014. Prosodic encoding of informativity: Word frequency and contextual probability interact with information structure. In *Proceedings of the 36th annual meeting of the Cognitive Science Society (CogSci 2014)*, 1120–1125. https://cogsci.mindmodeling.org/2014/papers/199/.
Ouyang, Iris Chuoying. 2015. *Prosody and informativity: A cross-linguistic investigation*. Los Angeles, CA: University of Southern California dissertation.
Ouyang, Iris Chuoying & Elsi Kaiser. 2015. Individual differences in the prosodic encoding of informativity. In Susanne Fuchs, Daniel Pape, Caterina Petrone & Pascal Perrier (eds.), *Individual differences in speech production and perception*, 147–188. Peter Lang: Bern.
Ouyang, Iris Chuoying, Sasha Spala & Elsi Kaiser. 2017. Speakers' rapidly-updated expectations influence prosodic realization of information structure. *Proceedings of the Linguistic Society of Americ*a 2(51). 1–10.
Pan, Shimei & Julia Hirschberg. 2000. Modeling local context for pitch accent prediction. In *Proceedings of the 38th annual meeting on Association for Computational Linguistics*, 233–240.
Rosa, Elise C. & Jennifer E. Arnold. 2011. The role of attention in choice of referring expressions. In *Proceedings of PRE-Cogsci: Bridging the gap between computational, empirical and theoretical approaches to reference*.
Roßnagel, Christian Stamov. 2000. Cognitive load and perspective-taking: applying the automatic-controlled distinction to verbal communication. *European Journal of Social Psychology* 30(3). 429–445.
Roßnagel, Christian Stamov. 2004. Lost in thought: Cognitive load and the processing of addressees' feedback in verbal communication. *Experimental Psychology* 51(3). 191–200.
Russell, Alex W. & Michael F. Schober. 1999. How beliefs about a partner's goals affect referring in goal-discrepant conversations. *Discourse Processes* 27(1). 1–33.

R Core Team. 2006. *R: A language and environment for statistical computing*. Version 3.3.1. Vienna: R Foundation for Statistical Computing. http://www.R-project.org/.

Scarborough, Rebecca. 2010. *Lexical and contextual predictability: Confluent effects on the production of vowels*, Vol. 10, 557–586. Laboratory Phonology.

Selting, Margret. 1996. Prosody as an activity-type distinctive cue in conversation: The case of so-called 'astonished'. In Elizabeth Couper-Kuhlen & Margret Selting (eds.), *Prosody in conversation: Interactional studies*, 231–270. Cambridge: Cambridge University Press.

van Son, Rob J. J. H. & Louis C. W. Pols. 1999. An acoustic description of consonant reduction. *Speech Communication* 28(2). 125–140.

Vallduví, Enrique & Maria Vilkuna. 1998. On rheme and kontrast. In Peter Culicover & Louise McNally (eds.), *The limits of syntax*, 79–108. Leiden: Brill.

Wagner, Michael & Duane G. Watson. 2010. Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes* 25(7–9). 905–945.

Watson, Duane G., Michael K. Tanenhaus & Christine A. Gunlogson. 2008. Interpreting pitch accents in online comprehension: H* vs. L + H*. *Cognitive Science* 32(7). 1232–1244.

Wicha, Nicole Y. Y., Elizabeth A. Bates, Eva M. Moreno & Marta Kutas. 2003. Potato not Pope: human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters* 346(3). 165–168.

Wilkinson, Sue & Celia Kitzinger. 2006. Surprise as an interactional achievement: Reaction tokens in conversation. *Social Psychology Quarterly* 69. 150–182.

Wright, Richard. 2004. Factors of lexical competition in vowel articulation. In John Local, Richard Ogden & Rosalind Temple (eds.), *Phonetic interpretation: Papers in laboratory phonology VI*, 75–87. Cambridge: Cambridge University Press.

Zahorian, Stephen A. & Hongbing Hu. 2008. A spectral/temporal method for robust fundamental frequency tracking. *The Journal of the Acoustical Society of America* 123(6). 4559–4571.

Zimmermann, Malte & Edgar Onea. 2011. Focus marking and focus interpretation. *Lingua* 121(11). 1651–1670.