**Individual strategies for resolving lexical and prosodic cues to certainty**

Marie-Catherine de Marneffe, Miranda Dickerman, Micha Elsner and Shari R. Speer
(The Ohio State University), demarneffe.@osu.edu

Verbs like *think* have two different interpretations (Rooryck 2001, Simons 2007): they can indicate genuine uncertainty (*I think it's this way?*), or they can be used to hedge or soften an assertion (*I think we should drop this*). The two interpretations are prototypically associated with different prosodic tunes, which elicit reliably different certainty judgments on a continuous rating scale. Listeners also use the tunes to decide which of two speakers' opinions to trust in short dialogues (de Marneffe et al. 2017). Here, we use these two tunes in a different experimental paradigm to probe explicit judgments of whether a speaker is certain or not. In this metalinguistic task, we find large individual differences between participants in how they resolve conflicting lexical and prosodic cues. A reanalysis of data from de Marneffe et al. (2017) shows that their effects were not driven by individual differences, suggesting that while participants' behavior in social settings is gradient, their conscious judgments about their interlocutors may not be.

We used a "blicket" task (Gopnik & Sobel 2000): our "blicket" detector lights up when a blicket is put on it. Participants are told that our puppet, Kathryn, has been learning to distinguish "blickets" from ordinary pompoms. In each trial, the experimenter shows a pompom to Kathryn, who utters one of the four stimuli in Table 1 (telling us whether she believes the pompom is or isn't a blicket). Two tunes are used for the "I think it is" utterance: in both instances, *think* receives a H* accent; the "polite" hedging tune ends in H* L-L%, whereas the "uncertain" tune ends in L* H-H% typically associated with interrogatives. To provide a felicitous context for hedging utterances, Kathryn's utterances are preceded by a comment from a 'dumb puppet' who urges sorting the pompom based on its color; we demonstrate in a pre-trial phase that color is irrelevant. Participants test Kathryn's sorting skill. Their task is to place each pompom in one of three jars: *yes* or *no* when Kathryn states a clear opinion on which she should be scored, and *not sure* if she needs more practice on the item later. After they have sorted her responses, they verify her choices with the detector.

20 adult participants heard 20 trials. We first show the results averaged over participants (Figure 1), which are similar to those reported in de Marneffe et al. 2017 (Figure 2, where people trusted the speaker uttering the "polite" tune much more than when the "uncertain" tune was used). We found that bare assertions were overwhelmingly judged as certain (*yes/no* jars). Explicit statements of uncertainty (*I don't know*) were judged as uncertain. The critical stimuli fall between these, with the "polite" tune judged as more certain than the "uncertain" tune (Binomial GLM with random effects by participant [LME4]: $p < 0.001$). We next look for individual differences. The left panel in Figure 3 shows that, over the 6 trials where Kathryn produced the "polite" tune, most participants assigned all 6 to the *yes* jar. However, the distribution is bimodal: participants who deviate from this pattern are most likely never to choose *yes*. The right panel shows the same pattern for the "uncertain" tune; the modal response is *not sure* but some participants assigned all these trials to *yes*. Based on these findings, we assigned participants to four groups: half of the participants are prosodically tuned, assigning the "polite" tune to *yes* and the "uncertain" one to *not sure*; 20% of the participants are lexically tuned, assigning both tunes to *not sure* due to the lexical item *think*; 15% of the participants are lexically and prosodically insensitive, assigning both tunes to *yes* (presumably because of Kathryn expressing an opinion); and the remaining 15% switch answers midway for the *polite* and *uncertain* tunes. Preliminary results show a trend for less prosodic sensitivity in male participants.

Reanalysis of de Marneffe et al. (2017) data shows that such individual differences do not emerge (Figure 4); the distribution is not bimodal but gradient. When presented with the same tunes and asked which speaker's opinion was right, participants were more influenced by prosody than in the present experiment, where some participants tuned more to the lexical item *think*.

Overall our results show that the experimental paradigm has an effect in how participants weigh lexical and prosodic cues. When participants are explicitly asked to make a judgment about speaker certainty, individual differences emerge in the ways that participants weigh the cues.

**Table 1.** Stimuli in experiment

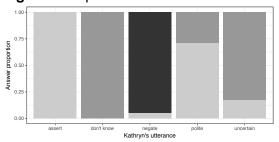|          | Condition  | # trials | Utterance                          | Expected behavior |
|----------|------------|----------|------------------------------------|-------------------|
| Control  | Assert     | 2        | Yes it is.                         | *Yes* jar         |
|          | Negate     | 2        | No, it is not.                     | *No* jar          |
|          | Don't know | 4        | I don't know.                      | *Not sure* jar    |
| Critical | Polite     | 6        | I think it is.<br>　　H*　　H* L-L% | *Yes* jar         |
|          | Uncertain  | 6        | I think it is?<br>　　H*　　L* H-H% | *Not sure* jar    |

**Figure 1.** Experiment results



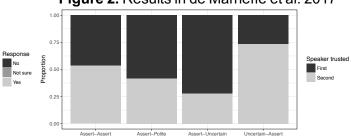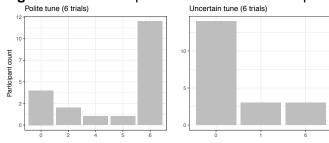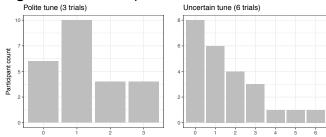**Figure 2.** Results in de Marneffe et al. 2017



**Figure 3.** Individual patterns in the "blicket" experiment



Each panel shows a different condition; the x-axis indicates the number of trials on which participants selected the *yes* response and the *y*-axis gives the count of participants. Participants who behave categorically are counted in the bars 0 (no *yes*) or 6 (all *yes*).

**Figure 4.** Individual patterns in de Marneffe et al. (2017)



Each panel shows a short dialogue in which one speaker uses *think* and the other disagrees, using a bare assertion. *x*-axis indicates number of trials on which participants trust the speaker using *think*. Participants who behave categorically are counted in the leftmost and rightmost bars.

**References**

de Marneffe, M.-C., Elsner, M. and Speer, S. (2017). "I think it's a tiger": Does prosody help distinguish different uses of belief verbs? *30th CUNY Conference on Human Sentence Processing*.

Gopnik, A. and Sobel, D.M. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. *Child development*, *71*(5):1205-1222.

Rooryck, J. (2001). Evidentiality, Part I. *Glot International* 5(4):3-11.

Simons, M. (2007). Observations on embedding verbs, evidentiality, and presupposition. *Lingua* 117(6):1034–1056.