Navigating ambiguity: The usefulness of context and prosody for naturalistic scope interpretations

Natural languages are full of potentially ambiguous expressions—at least, when these expressions are considered as text out of context—but we seem to be very good at navigating ambiguity and understanding each other. Two key sources of potentially disambiguating information are context and prosody. Our question is, to what extent can listeners use context and prosody to interpret a potentially ambiguous utterance in everyday conversation? We focus on *every*-negation scope ambiguity (e.g., *Every vote doesn't count*) as a case study of ambiguity. In prior work, we gathered naturalistic uses of this ambiguity from conversation recordings. Here, we compare interpretations of these naturalistic uses as text-only, audio-only, text-in-context, and audio-in-context. We find that both context and prosody contribute significant and partially-redundant information.

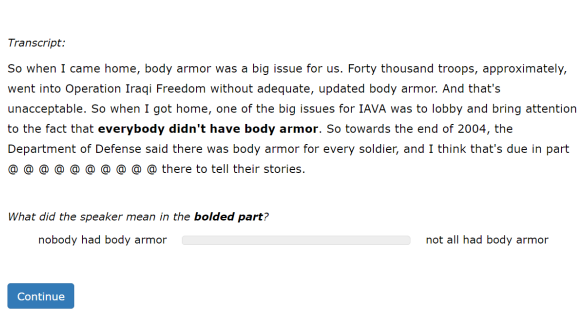
Background. Utterances like *Every vote doesn't count*, with a quantified subject and verb negation, are potentially ambiguous between a surface scope interpretation *every*$>$*not* (*No vote counts*) and an inverse scope interpretation *not*$>$*every* (*Not all votes count*). A striking facet of prior research on scope interpretation is both a strong expectation that prosody matters and a lack of clear evidence that it does (e.g., Halliday, 1967; Jackendoff, 1972; Liberman and Sag, 1974; Ladd, 1980; Ward and Hirschberg, 1985; Büring, 1997). A larger question emerges from this body of work about how redundant prosody is with context, since many describe the information provided by prosody as information that might also be provided by context. In one of the only experimental studies investigating prosody, Syrett et al. (2012) found a speaker-specific but no cross-speaker mapping between interpretation and prosody; conversely, listeners show a success rate between 53% and 77% at matching between what they hear and what the speaker intended (Syrett et al., 2014). This weak and variable mapping between prosody and interpretation may be due to many reasons, highlighting the value of understanding the extent of the disambiguating information in both context and prosody of naturalistic data.

Methods. We ran an experiment on Prolific (N=94 monolingual English speakers) to annotate the 63 conversational *every*-negation items collected in past work from radio and TV interviews. Participants judged the speaker's intended meaning on a sliding scale between paraphrases of the item's surface and inverse scope interpretations, in a 2x2 design with factors context (with or without context) and modality (text or audio): each item appeared in each of four conditions (text, audio, text-in-context, audio-in-context). Figure 1a shows an example trial. Each participant judged twenty items (5 randomly-selected items in each of the 4 conditions) in a random order. Between 2 and 15 judgments were collected per item in each condition.

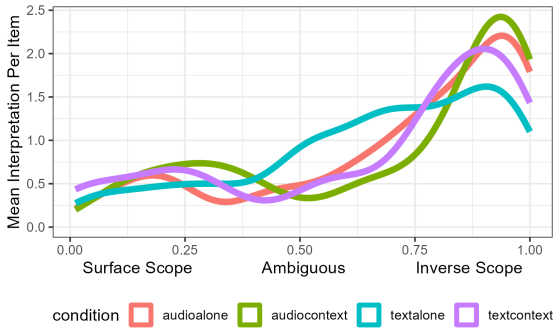
Results. To test the amount of additional information provided by context and prosody, we coded a variable (int-diff) for each item that encodes the absolute value difference in interpretations between the text-only condition and the three other conditions (e.g., for a hypothetical item that received an average interpretation of 0.6—60% inverse—in text-only, 0.8 in text-in-context, 0.9 in audio-only, and 0.9 in audio-in-context, the corresponding int-diff values would be 0, 0.2, 0.3, and 0.3). We then used a mixed effects model predicting int-diff by an interaction of context and modality, with random intercepts for item and participant, using the lme4 package in R (Bates et al., 2015). The main effects of context (β=-0.1455, SE=0.00524, p<2e-16) and modality (β=0.01765, SE=0.005232, p=0.000757) were significant, as was their interaction (β=0.1416, SE=0.007424, p<2e-16). As a measure of the confidence of interpretations in the different conditions, we compared the entropies of the mean interpretations in the four different conditions. We estimated the Shannon entropy, using the entropy package in R (Hausser et al., 2012), of each mean interpretation distribution, where mean interpretations were calculated using the non-parametric bootstrap method from the

Hmisc package in R (Harrell Jr and Harrell Jr, 2019). We found that entropy decreased between conditions in the following order: text-only (5.89) > audio-only (5.86) > text-context (5.83) > audio-context (5.69). Figure 1b shows the four distributions of mean responses per item.

Discussion. In spite of variation, we found that both context and prosody contribute significant information to the interpretations of naturalistic ambiguity, with context providing more confidence than audio, and with the audio information partially redundant with the contextual information. In future work, we investigate more specifically where the disambiguating aspects of context and prosody are redundant with each other. In a previous study that only considered text-in-context interpretations, we identified a specific contextual cue that predicts interpretations; in another study, we identified potential acoustic cues. Future work will test how these contextual and acoustic cues, alone and in interaction, predict interpretations of naturalistic items, using the experimental paradigm we introduce in this study and on the basis of a larger corpus of naturalistic items.



(a) Sample trial of a text-in-context condition.



(b) Distribution of responses.

Figure 1: Sample trial from the experimental task, and the distributions of mean interpretations per item in each of the four conditions (text-only, audio-only, text-in-context, and audio-in-context).

References

D. Bates, M. Mächler, B. Bolker, and S. Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015. doi: 10.18637/jss.v067.i01.

D. Büring. *The meaning of topic and focus: The 59th Street Bridge accent*, volume 3. Psychology Press, 1997.

M. A. K. Halliday. Intonation and grammar in british english. In *Intonation and grammar in British English*. De Gruyter Mouton, 1967.

F. E. Harrell Jr and M. F. E. Harrell Jr. Package 'hmisc'. *CRAN2018*, 2019:235–236, 2019.

J. Hausser, K. Strimmer, and M. K. Strimmer. Package 'entropy'. *R Foundation for Statistical Computing: Vienna, Austria*, 2012.

R. S. Jackendoff. Semantic interpretation in generative grammar. 1972.

D. R. Ladd. The structure of intonational meaning (bloomington), 1980.

M. Liberman and I. Sag. Prosodic form and discourse function. In *Chicago Linguistics Society*, volume 10, pages 416–427, 1974.

K. Syrett, G. Simon, and K. Nisula. Prosodic disambiguation of scopally ambiguous sentences. In *Proceedings of the Meeting of the North East Linguistic Society*, volume 43, pages 141—152. GLSA (University of Massachusetts), 2012.

K. Syrett, G. Simon, and K. Nisula. Prosodic disambiguation of scopally ambiguous quantificational sentences in a discourse context. *Journal of Linguistics*, pages 453–493, 2014.

G. Ward and J. Hirschberg. Implicating uncertainty: The pragmatics of fall-rise intonation. *Language*, pages 747–776, 1985.