

Prominence and Phrasing in Spoken Discourse Processing

Shari R. Speer

Department of Linguistics
Ohio State University
Columbus, Ohio, USA
speer@ling.osu.edu

Kiwako Ito

Department of Linguistics
Ohio State University
Columbus, Ohio, USA
ito@ling.osu.edu

Abstract

We review psycholinguistic research on the use of intonation in dialogue, focusing on our own recent work. In experiments using complex real-world tasks and naïve speakers and listeners, we show that speakers reliably specific prosodic cues to signal their intentions, and that listeners use these cues to recognize syntactic and pragmatic aspects of discourse meaning.

1 Introduction

The intonation of an utterance conveys a great deal of information about a speaker's intended message. Recent research has addressed whether, when, and how speakers use intonation to transmit linguistic and paralinguistic meaning. Speakers use intonation for a broad range of functions in communication, such as: to mark the difference between immediately relevant vs. background information; to express contrast, contradiction, and correction; and to indicate the intended syntax of ambiguous utterances.

In this paper, we will review recent experimental studies of naïve speakers' and listeners' use of intonation during production and comprehension. We focus primarily on our own work, where we have used naturalistic, relatively complex real-world tasks to elicit speakers' intonation in dialogue, and to examine listeners' ability to use intonation during comprehension.

1.1 Structure of the Review

The review includes a very brief overview of the ToBI (Tone and Break Indices: Beckman, Hirschberg & Shattuck Hufnagel, 2005; Brugos, Shattuck Hufnagel, & Veilleux, 2006), a system widely used to annotate the locations and types of intonational entities present in speech. We then turn to the correspondence between intonational phrasing and syntactic constituency, examining whether and how speakers use the location and size of prosodic breaks as they produce syntactic structure, and testing listeners' ability to use this correspondence. Next, we will discuss the role of intonational prominence in discourse structuring. In particular, we will examine the production and comprehension of accentual prominence to express referential contrast. Our experimental findings suggest that both phrasing and prominence cues are robust for understanding spoken messages, and infelicitous use of those prosodic cues may mislead the listener's comprehension processes. The closing discussion emphasizes the contribution of intonation to spoken language processing, and its essential role in discourse modeling.

2 Prosodic phrasing and syntactic structure

Production experiments investigating the correspondence of prosodic and syntactic phrasing in language production were conducted using a partially-scripted two-player board game task (Speer, Warren, & Schafer, 2003, Schafer, Speer, & Warren, 2005). The game was constructed to elicit particular syntactic contrasts as the players spoke to

each other to exchange information and move game pieces from start to goal, avoiding hazards and collecting bonuses. Speakers were required to use a fixed set of sentence frames and game piece names to construct instructions, requests and acknowledgements. Board layouts were carefully constructed to create situational contexts for the resulting dialogues produced by the speakers, so that potentially syntactically ambiguous utterances sometimes remained ambiguous given the game context, but sometimes were contextually disambiguated. The task allowed recording of multiple renditions of a sentence from the same speaker, as well as across speakers.

Phonetic analyses and phonological transcription of speech from the game task experiments demonstrates that speakers reliably used prosodic cues to signal syntactic structure, and to convey the intended meaning of syntactically ambiguous sentences. Although speakers produced a wide variety of prosodic patterns for the same sentence, they consistently placed the strongest prosodic break in the syntactic location that indicated their intended meaning. This effect held for different syntactic types, and was robust across various types of potentially disambiguating situational contexts.

Companion comprehension experiments used the naïve speakers' productions from the game task as stimuli in forced-choice experiments. Results show that listeners make reliable use of prosodic regularities to accurately recover the speakers' intended meanings.

3 Intonational prominence and contrast

To investigate naïve speakers' use of intonational prominence to mark contrast, Ito, Speer & Beckman (2004; Ito & Speer, 2006) elicited unscripted speech using a holiday tree decoration task. Naïve speakers were asked to be 'directors,' giving instructions about how to hang ornaments on a tree to a confederate 'decorator.' Photos of ornaments and the tree were used to give 'directors' the sequence of ornaments and their intended locations on the tree. By using a set of common color terms and object names for the ornaments, we collected multiple spontaneous productions of target adjective-noun pairs (e.g. "*green bell*") in a natural conversational setting. The sequence of decoration was constructed to create contrastive discourse contexts (green bell -> blue bell). Prosodic annota-

tion of the spontaneous speech and supporting phonetic measurements indicated that speakers used prominent pitch accent more frequently when the decoration sequence was contrastive than when it was not (e.g. "green bell.... → Now, find a BLUE bell": where the adjective 'blue' is accented and the repeated noun 'bell' is deaccented).

To test how prominent accent guides referential resolution during comprehension, a set of experiments was conducted using eye movement monitoring technique (Ito & Speer, 2008). Spoken stimuli modeled on the productions from the previous production study were recorded by a trained phonetician. This time, naïve listeners followed instructions to locate ornaments and hang them as specified on a small tree.

Results showed the immediate use of intonational cues by the listeners. For example, when a prominent accent felicitously marked contrast on a color adjective ("First hang the red ball. → Next, hang the BLUE ball"), fixation proportions to target cells increased more quickly than when intonation did not cue the contrast ("Hang the red ball. → Next, hang the blue ball."). The timing of fixations indicates that listeners moved their eyes to the correct ornament before processing the segmental information of the noun. In addition, prominent accent misled listeners when the decoration sequence did not prompt contrast (blue ball → GREEN drum: here, participants looked at balls before they redirected fixations to the drum). These incorrect fixations increased toward the end of the adjective and continued to rise halfway into the noun with conflicting segmental information. Thus, listeners were executing their initial visual search based on the intonational information of the adjective, demonstrating the use of prosody for making predictions during comprehension of running speech.

Acknowledgments

To our colleagues and co-investigators, especially Ping Bai, Mary Beckman and Laurie Maynell, and to NSF BCS-0617609 & REU-0617609 and NIH DC007090.

References

- Beckman, M.E., Hirschberg, J., & Shattuck-Huffnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In Sun-Ah Jun (ed.), *Prosodic models and transcription: Towards prosodic typology*. Oxford University Press.
- Brugos, A., Shattuck-Hufnagel, S., & Vielleux, N. (2006). *Transcribing Prosodic Structure of Spoken Utterances with ToBI*, MIT Open Courseware, <http://ocw.mit.edu/OcwWeb/Electrical-Engineering-and-Computer-Science/6-911January--IAP--2006>.
- Ito & Speer, S.R. (2006). Using interactive tasks to elicit natural dialogue. In P. Augurzky & D. Lener-tova (Eds.), *Methods in Empirical Prosody Research*, Mouton de Gruyter.
- Ito, K., & Speer, S.R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language, Special issue on Language-Vision interaction*.
- Ito, K., Speer, S.R., & Beckman, M. (2004). Informational status and pitch accent distribution in spontaneous dialogues in English. *Proceedings of the International Conference on Speech Prosody*. Nara, Japan.
- Schafer, A.J., Speer, S.R., & Warren, P. (2005). Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task. In M. Tanenhaus & J. Trueswell (Eds.) *Approaches to Studying World Situated Language Use: Psycholinguistic, Linguistic and Computational Perspectives on Bridging the Product and Action Tradition*, Cambridge: MIT Press.
- Speer, S.R., Warren, P., & Schafer, A.J. (2003). Intonation and sentence processing. Proceedings of the Fifteenth International Congress of Phonetic Sciences, Barcelona, Spain.