

Usefulness of prosodic cues in parsing: Evidence from a novel cross-modal maze task

Zeping Liu¹, Xin Chen², Chien-Jer Charles Lin¹

¹Indiana University Bloomington, ²Fudan University

Introduction Previous studies suggest that prosodic information guides syntactic parsing and can be immediately integrated during speech comprehension [1-2]. Evidence indicates that listeners can draw on cues such as prosodic boundaries to resolve structural ambiguity, as their placement within phrases often aligns with syntactic structures [3-4]. However, most of the existing evidence focuses on attachment ambiguity (e.g., PP- or RC-attachment), little is understood about how the resolution of structural ambiguity beyond attachment can benefit from prosodic boundaries.

The aim of this study is to investigate how listeners use prosodic cues to predict head-final structures and alternative parses, and thus facilitate the resolution of structurally ambiguous fragments. The ambiguous sequence *V+N1+DE+N2* in Chinese is compatible with two parses: a complement clause (CC) parse and a relative clause (RC) parse. Notably, the two parses are associated with prosodic boundaries (e.g., pauses) placed at different locations within the fragment [5-6] (see **Fig. 1**). The semantics of N2 (patient vs. agent) can serve to disambiguate between the two parses (CC: patient vs. RC: agent). Note that overall the RC parse is favored over the CC parse due to syntactic constraints [7-8]. We aim to see whether listeners can make different predictions for N2 based on the location of pause within the fragment.

Method We employed a novel *cross-modal maze task* (N=37), where participants first listen to a lead-in sentence and fragment (with pause placed at different locations), and then continuously make choices between two visually presented words to finish the sentence. We focus on choices and reaction times (RTs) on the target screen (see **Fig. 2**), which inform us of the structures pursued based on the placement of pause. We included a training phrase to help listeners establish the association between the prosodic cue and a syntactic parse, during which visual choices on the target screen are unambiguous, including a distractor word and a word that legitimately continues the sentence (i.e., a *patient* noun aligning with the CC parse in early pause and an *agent* noun aligning with the RC parse in late pause). In the testing phase, choices on the target screen are ambiguous between an agent and a patient noun. Our hypothesis is, if participants successfully established the prosody-syntax relationship in the training phase, they should be able to choose corresponding nouns based on the prosodic cue in the testing phase.

Results & Discussion In the training phase, the early pause condition (aligning with the CC parse) resulted in significantly longer RTs than the late pause condition (RC parse) (Est. = 0.25, $t = 8.47$, $p < .001$). This confirms the previous findings that the CC parse is generally dispreferred, suggesting that the presence of prosodic cues that align with the less frequent parse leads to parsing difficulty. In the testing phase, however, no effect of prosodic conditions was observed in either choices or RTs: the presence of early pause (88%) or late pause (89%) invariably led to the RC parse (taking N2 as the agent), and there is no significant difference in RTs (Est. = -0.01, $t = -0.16$, $p = .87$) (see **Fig. 3**). These findings suggest that the association between prosodic cues and syntactic parses established in the training phase may be overshadowed by the strong preference for the RC parse in the *V+N1+DE+N2* segment. In other words, even though an early pause in the CC parse signals a constituent boundary between V and N1, listeners still actively attempt to fill the missing NP before V and therefore tend to adopt an RC parse in predictive processing [8]. The current research thus provides two notable contributions: (i) While prosodic cues are effective in disambiguating prepositional phrase attachments in English, they may prove insufficient in countering parsing preferences, such as those observed in Chinese RCs, where a strong preference exists for filling a relativized gap. (ii) We introduce an innovative experimental paradigm, the "cross-modal" maze tasks, to investigate the impact of prosody on parsing decisions in the context of ambiguous structures.

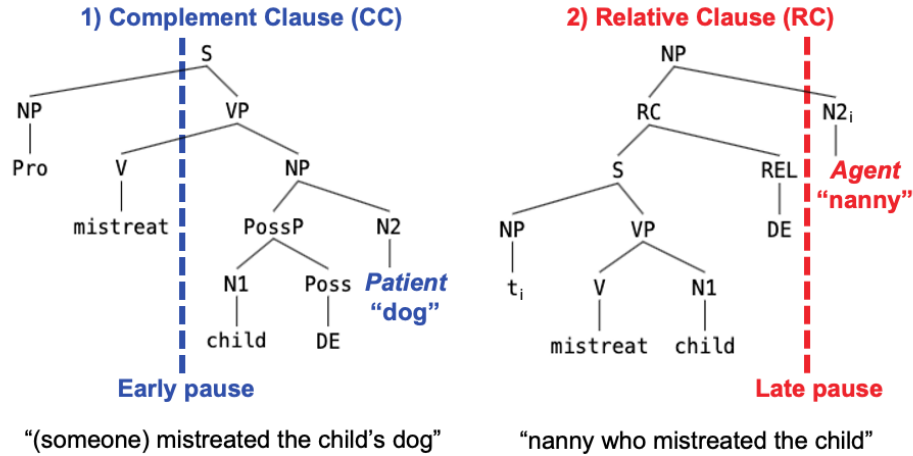


Fig. 1 Chinese fragment $V+N1+DE+N2$ aligns with both CC and RC analyses. Based on the location of pause, listeners have different prediction for N2. Early pause is associated with the CC parse so that N2 should be a *patient* noun. In comparison, late pause is associated with the RC parse so that N2 should be an *agent* noun.

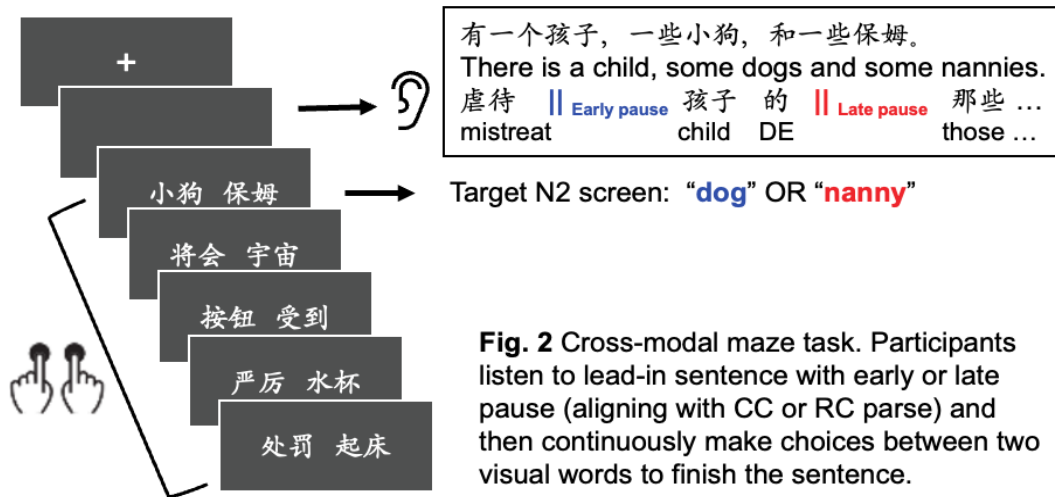


Fig. 2 Cross-modal maze task. Participants listen to lead-in sentence with early or late pause (aligning with CC or RC parse) and then continuously make choices between two visual words to finish the sentence.

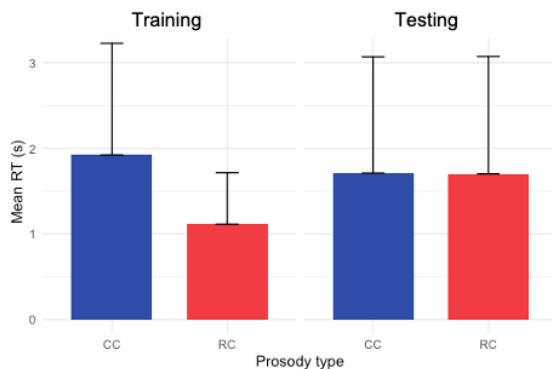


Fig. 3 Mean RTs of CC and RC prosody types in the training and testing phases

References. [1] Marslen-Wilson et al., 1992. *The Quarterly Journal of Experimental Psychology*. [2] Snedeker & Casserly, 2010. *Language and Cognitive Processes*. [3] Nakamura et al., 2022. *Language, Cognition, and Neuroscience*. [4] Kraljic & Brennan, 2005. *Cognitive Psychology*. [5] Li et al., 2010. *Biological Psychology*. [6] Yu et al., 2019. *Frontiers in Psychology*. [7] Hsieh & Boland, 2015. *Journal of Psycholinguistic Research*. [8] Ng & Wicha, 2014. *Journal of Memory and Language*.