

Tracking Meaning Through Allophonic Variation: Cross-Situational Learning of Tonal Words and Underlying Phonological Rules

Xinbing Luo¹

¹*University of Cambridge*

Cross-situational learning (CSL) provides a mechanism for resolving referential ambiguity by tracking co-occurrences between auditory forms and potential referents across contexts (e.g., Yu & Smith, 2007). While CSL research has primarily examined vocabulary learning with stable and highly distinctive auditory forms, natural languages exhibit systematic phonological alternations that create context-conditioned surface variation. Tone languages involve extensive tonal allophony; tone sandhi can cause citation tones and their surface realisations to diverge, creating variations from a learner's perspective. Prior work shows that tone is a perceptually salient dimension for tone-language speakers (Yip, 2023), and that even complex alternations can be acquired incidentally (e.g., Huang & Do, 2021; Zou & Luo, 2025). The present study investigates whether adult Mandarin speakers can acquire both word–referent mappings and underlying tone-sandhi patterns in a CSL setting when auditory forms vary systematically due to eight tone sandhi rules.

An artificial tone language (PseuSH, short for “Pseudo-Shanghainese”) was constructed, inspired by Shanghai Wu. The language contained two citation tones (T53, T13) and eight disyllabic sandhi rules. These were paired with two visually and functionally distinct morphosyntactic structures: Verb–Noun combinations were depicted as dynamic action scenes and consistently mapped to leftward-reduction sandhi patterns, whereas Modifier–Noun combinations were presented as static container–object images and mapped to rightward-extension sandhi patterns. All auditory stimuli were recorded as monosyllables by a Shanghai Wu speaker, and then resynthesised and F0-normalised. Seventy-nine native Mandarin speakers completed four tasks online: (i) an AX discrimination task (112 trials) assessing sensitivity to phonemic and allophonic tonal contrasts; (ii) a CSL exposure phase (192 trials) in which each trial paired two ambiguous disyllabic phrases with two possible referents, requiring learners to resolve ambiguity across repeated co-occurrences; (iii) a 2AFC memory test (64 trials) on learned sound–referent mappings; and (iv) a 2AFC generalisation test (64 trials) in which learners heard two segmentally identical disyllables differing only in whether the tone pattern conformed to the sandhi rule appropriate for the displayed referent that did not appear in the exposure phase.

Participants showed high discrimination accuracy (Mean = 87.2%), with all tonal and allotonal contrasts discriminated above chance, confirming perceptual access to the exact tonal distinctions used in the CSL tasks. In the 2AFC memory test, participants performed robustly above chance (Mean = 67.3%), despite every disyllable undergoing a sandhi alternation (Fig. 1). In the 2AFC generalisation test, participants again performed reliably above chance (overall Mean = 61%, Fig. 2). They systematically preferred sandhi-correct forms over incorrect (or no-sandhi) alternatives, and GLMMs revealed a significant main effect of condition: accuracy was higher for Correct-vs.–No-Rule than for Correct-vs.–Incorrect pairs. Rule-specific effects were modest and visual-referent type (dynamic vs. static) did not modulate performance. These results show that CSL remains effective even when auditory forms are systematically altered by contextual phonological processes. Learners not only successfully formed mappings between variable surface forms and referents but also showed sensitivity to the underlying tonal alternations governing those surface forms.

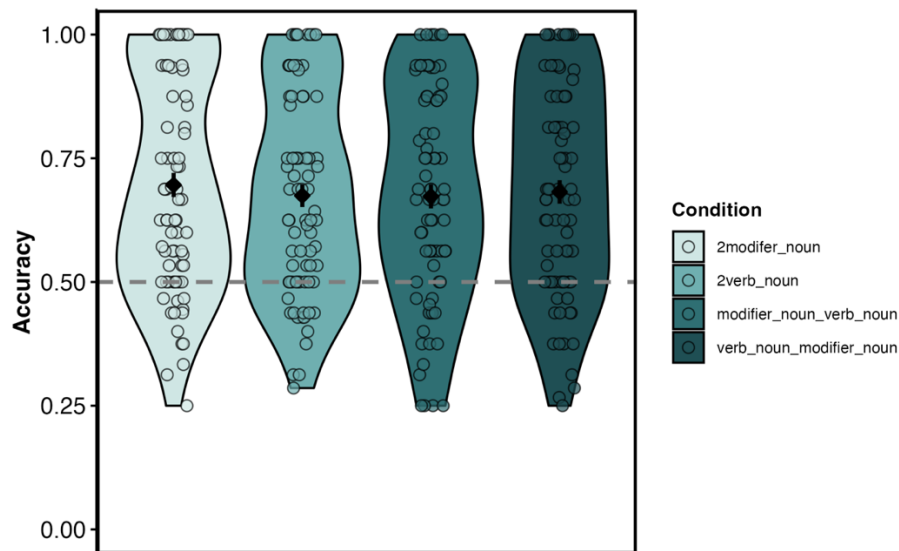


Fig. 1. Accuracy in the 2AFC memory test by visual referent condition. Dots represent individual participants' mean accuracy. Diamonds indicate the mean accuracy across participants for each condition, and error bars show the Mean $\pm$ SE. The dashed line marks the chance level (0.5).

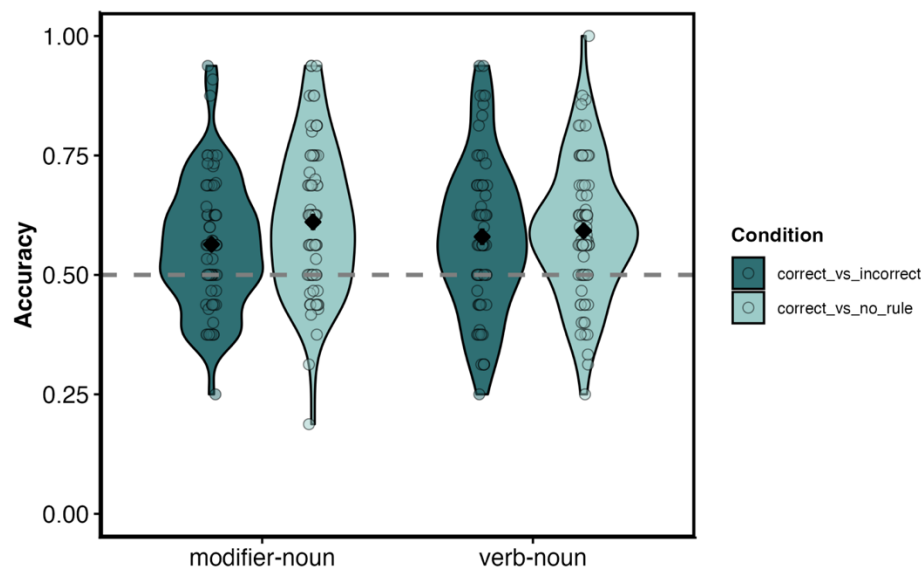


Fig. 2. 2AFC generalisation accuracy by visual reference type and condition. Violins show the distribution of per-participant mean accuracy; dots are individual participants. Diamonds indicate the condition mean, and the error bars indicate the Mean $\pm$ SE. The dashed line marks chance-level performance (0.5).

References

- [1] Huang, T., & Do, Y. (2021). Phonetically grounded structural Bias in learning tonal alternations. *Frontiers in Psychology*, 12, 705766.
- [2] Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, 18(5), 414-420.
- [3] Yip, M. C. (2023). Cross-situational word learning of Cantonese Chinese. *Psychonomic Bulletin & Review*, 30(3), 1074-1080.
- [4] Zou, T., & Luo, X. (2025). Implicit learning of unfamiliar tone sandhi patterns in lexical recognition. *Frontiers in Psychology*, 15, 1414732.