

Introduction: Mandarin has a rich system of sentence-final particles (SFPs), conventionally associated with interactional functions, as they are concerned with marking the epistemic states of either the speaker or the addressee. This study explores their use in a non-canonical interactional context, i.e. self-talk, where the speaker and addressee coincide and epistemic negotiation should, in principle, be unnecessary. However, preliminary fieldwork suggests that SFPs do occur in self-talk and that their distribution is systematically constrained. To further explore the interactional nature of these constraints, the study compares human judgements with those of large language models (LLMs), which lack genuine epistemic state management.

Background: SFPs have been analyzed as occupying different syntactic layers, which give rise to their strict linear order (Zhu 1982; Paul & Pan 2017; Paul & Yan 2021; Fang & Hengeveld 2022; Xu 2023). Following Wiltschko's (2021) Interactional Spine Hypothesis, where an addressee-oriented layer (Ground-Adr) dominates a speaker-oriented layer (Ground-Spr), I argue *ne* is speaker-oriented, while *ma* is addressee-oriented, a distinction reflected in their linear order (*ne ma* is grammatical, but *ma ne* is not). Self-talk, however, requires the speaker to simultaneously undertake both roles. As observed in Holmberg (2010), two types of self-talk can be distinguished: I-centered self-talk, where the speaker refers to themselves with the 1st person pronoun and you-centered self-talk, where the speaker refers to themselves with the 2nd person pronoun. Following Wiltschko (2025), I assume in I-centered self-talk the interactional structure is deficient in that it lacks Ground-Adr, while in you-centered self-talk both Speaker- and Addressee-oriented Grounding layers are present. These distinctions make self-talk an ideal environment for testing SFPs' interactional orientations.

Research question: how do *ne* and *ma* behave in different types of self-talk?

Hypothesis: *ne* is compatible with both I-centered and you-centered self-talk, whereas *ma* is restricted to you-centered self-talk.

Methodology: I employ a storyboard methodology designing illustrated scenarios that visually cue self-talk contexts (e.g. talking to oneself in the mirror), manipulating both person reference (1st vs. 2nd person) and SFP choice (*ne* or *ma*) in the utterances. Mandarin native speakers judge the naturalness of these utterances on a five-point Likert scale. GPT-4o judgments are also collected as a supplementary diagnostic. If the two SFPs differ in interactional orientation, human judgements should reflect this contrast, whereas LLMs, lacking representations of interlocutor's epistemic states, are not expected to do so. This design addresses two methodological challenges: naturalistic self-talk data containing target SFPs are rare, and even state-of-the-art LLMs do not spontaneously generate self-talk data containing target SFPs without explicit prompting.

Results: Preliminary observations reveal a clear divergence between human and LLM judgements. Human ratings align with my hypothesis: *ne* is judged as compatible with both I-centered and you-centered self-talk, while *ma* is judged as only compatible with you-centered self-talk. In contrast, GPT-4o judges both SFPs as similarly natural across both self-talk types. This divergence provides experimental support for analyzing *ma* as addressee-oriented and *ne* as speaker-oriented, while also suggesting a key deficiency in LLMs' ability to parse interactional language.