From competence to performance: Do LLMs understand Cantonese sentence-final particles?

Yunqing Zhao, independent researcher, ayzhao06@gmail.com
Lok Yiu Yeung, independent researcher, odiliayeung@gmail.com

Cantonese sentence-final particles (SFPs) are frequently used in everyday discourse, appearing alone or in clusters to serve a broad range of communicative functions. Despite the well-established complexity of SFPs, Cantonese remains low-resource in natural language processing (NLP), including for large language models (LLMs) (Cheng et al., 2025). Their capacity to demonstrate genuine pragmatic competence in Cantonese, particularly in interpreting and producing SFPs, is not yet clear. This study addresses that gap by testing three LLMs, GPT by OpenAI, Gemini by Google, Claude by Anthropic. The test consists of two parts. The first part employs a multiple-choice section to assess whether models can select the most contextually appropriate SFP from a set of options, providing a controlled measure of receptive knowledge. The second part is a generation task comprising two sub-types: models are presented with a conversational scenario and instructed to produce a natural response from the perspective of a specified speaker; models are also required to translate Mandarin source sentences into colloquial Cantonese, where the illocutionary force of the original calls for particular SFPs in the target output. All generated outputs are evaluated according to established functional criteria, with acceptability further rated by native speakers. Findings from this study offer empirical insight into the strengths and limitations of three widely used LLMs. Gemini demonstrates the strongest overall performance in all tasks. Its responses are the most natural with the fewest errors in SFP usage, and it exhibits strong ability to differentiate the functions of individual particles as well as clusters. Claude achieves a perfect score in the identification task but shows less precision in its selection of SFPs during generation. OpenAI performs the weakest overall with evident errors in both parts. A number of register mismatches are observed in the output, which may reflect the most limited understanding of SFPs among the three models. The results carry implications for future training strategies aimed at improving model performance for Cantonese, supporting the goal of developing technologies that serve linguistically diverse communities. With over 85 million speakers worldwide, Cantonese may become more accessible to learners as LLM continues to advance. The results may serve as a reference for both learners and educators regarding the pragmatic competence of LLMs in the case of daily conversations, while also exposing deficiencies that should be noticed, such as the misuse of certain SFPs and incorrect clusters.

References

Cheng, T. C., Cheng, C. S., Lau, C. M., Lam, E., Yat, W. C., Yu, H. O., & Chong, C. H. (2025). Hkcanto-eval: A benchmark for evaluating cantonese language understanding and cultural comprehension in llms. In *Proceedings of the 29th Conference on Computational Natural Language Learning* (pp. 1-11).