

Using ChatGPT to Obtain Effective Feedback for Learning Mandarin Near-synonyms

The current study aims to explore the effectiveness of ChatGPT for providing feedback to Chinese as a second language (CSL) learners who are learning Mandarin near-synonyms. In recent years, researchers have explored the efficacy of integrating generative AI tools into second language curricula, highlighting their strengths, such as convenience and accessibility, as well as limitations, including mediocre quality and accuracy. However, most of these studies focus on English, while few involving ChatGPT in CSL education are limited in that they investigate broad language contents, such as correcting general grammatical errors, rather than specific areas of difficulty, such as near-synonyms. This study aims to bridge these gaps by utilizing ChatGPT to identify errors and provide corrective feedback for erroneous sentences involving Mandarin near-synonyms and having CSL instructors evaluate the quality of the responses.

The current study focuses on two near-synonym pairs differing in difficulty, *yizhi/yixiang* ‘always/consistently’ and *bimian/yimian* ‘to avoid/in case of’, and implements few-shot and chain-of-thought (CoT) prompt engineering methods to elicit responses from ChatGPT 4.0 aimed to support learners’ acquisition of Mandarin near-synonyms. Four responses were produced, focusing on ChatGPT’s error identification and corrective feedback aimed toward learners. The original prompts and corresponding responses were inserted into four surveys in which instructors were presented with each near-synonym pair and prompt engineering method once. Instructors then answered questions on 10-point Likert scales regarding the quality of the AI-generated responses and the accuracy of ChatGPT’s ability to identify errors in sentences and provide corrective feedback to learners.

Participants’ responses are analyzed via linear mixed-effects models (LMM), and the results indicate a significant main effect of “prompt-type” with CoT prompts receiving higher ratings and a significant interaction effect of “near-synonym pair*prompt-type” for items regarding error identification (all $ps < .01^{**}$). We explore the interaction effect via Tukey’s HSD test, and the results indicate that responses for *yizhi/yixiang* ‘always/consistently’ do not significantly differ across prompt-type ($p = .736$), but responses for *bimian/yifang* ‘to avoid/in case of’ in the CoT prompt are rated significantly higher than those in the few-shot prompt ($p < .001^{***}$). For items regarding corrective feedback, the LMM results show a significant main effect of “near-synonym pair” ($p < .001^{***}$), with the responses for *yizhi/yixiang* ‘always/consistently’ being rated significantly higher than those for *bimian/yifang* ‘to avoid/in case of’. Fixed effects explain 20.88%-23.83% of variance in participants’ ratings ($R^2_m = .2088-.2383$) across models. The results illustrate that CSL instructors consider CoT prompting more effective in having ChatGPT identify near-synonym errors. They also believe that corrective feedback provided for *yizhi/yixiang* ‘always/consistently’ is of higher quality and accuracy compared to that of *bimian/yifang* ‘to avoid/in case of’ across prompting methods, suggesting that ChatGPT struggles to analyze higher-level near-synonyms. The findings indicate that while CoT prompting can potentially support learners in learning Mandarin near-synonyms, CSL instructors should be wary of ChatGPT’s limitations in processing advanced language contents.

Keywords: AI-assisted learning; Mandarin near-synonyms; corrective feedback