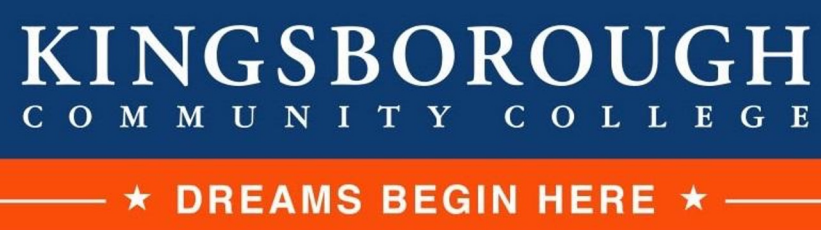


Prompting Across Languages: Are Bilingual Speakers More Effective AI Prompt Engineers?

Janet Brito, Denisse Ramos, Anastasiia Myslyk, Laura Spinu



INTRODUCTION

Bilingualism

Enhances:

- Metalinguistic awareness.
- Strategic language use.

Leads to:

- Greater structure, explicitness, and precision.
- May support more effective LLM interaction (Vatsal et al., 2025).

Prompt Engineering

- Crafting inputs for AI systems.

Requires:

- Clarity & Structure.
- Explicit constraints.
- Applied metalinguistic competence.

Research Gap

We know multilingual prompting affects outputs (Wang et al., 2025) but we don't know if the user's own linguistic experience shapes prompt quality.

Study Aim

- Compare bilingual vs. monolingual speakers.
- Prompt construction
- AI output quality.

Hypothesis

Bilinguals produce more:

- Structured, explicit, constrained prompts.
- Better task-aligned outputs.

Relevance

- Prompt design impacts risks like identity misrepresentation.
- Highlights importance of human input (Wang et al., 2025).
- If prompt skill varies by linguistic background, AI may amplify existing inequalities.

RESEARCH QUESTION

Do bilingual speakers produce more explicit and structurally complex prompts than monolingual speakers during AI-based tasks?

METHODOLOGY

Participants

A total of **40 adult participants** will be recruited:

- **half MONLINGUAL speakers;**
- **half BILINGUAL speakers.**

Language background will be collected via self-report (language use, proficiency, age of acquisition).

Materials

4 prompt generation tasks will assess constraint use and prompt effectiveness in human-AI chats:

Task 1: Cartoon replication

Generate a cartoon that's as close as possible to a specific image.

Task 2: Parameter Specification (Code Generation)

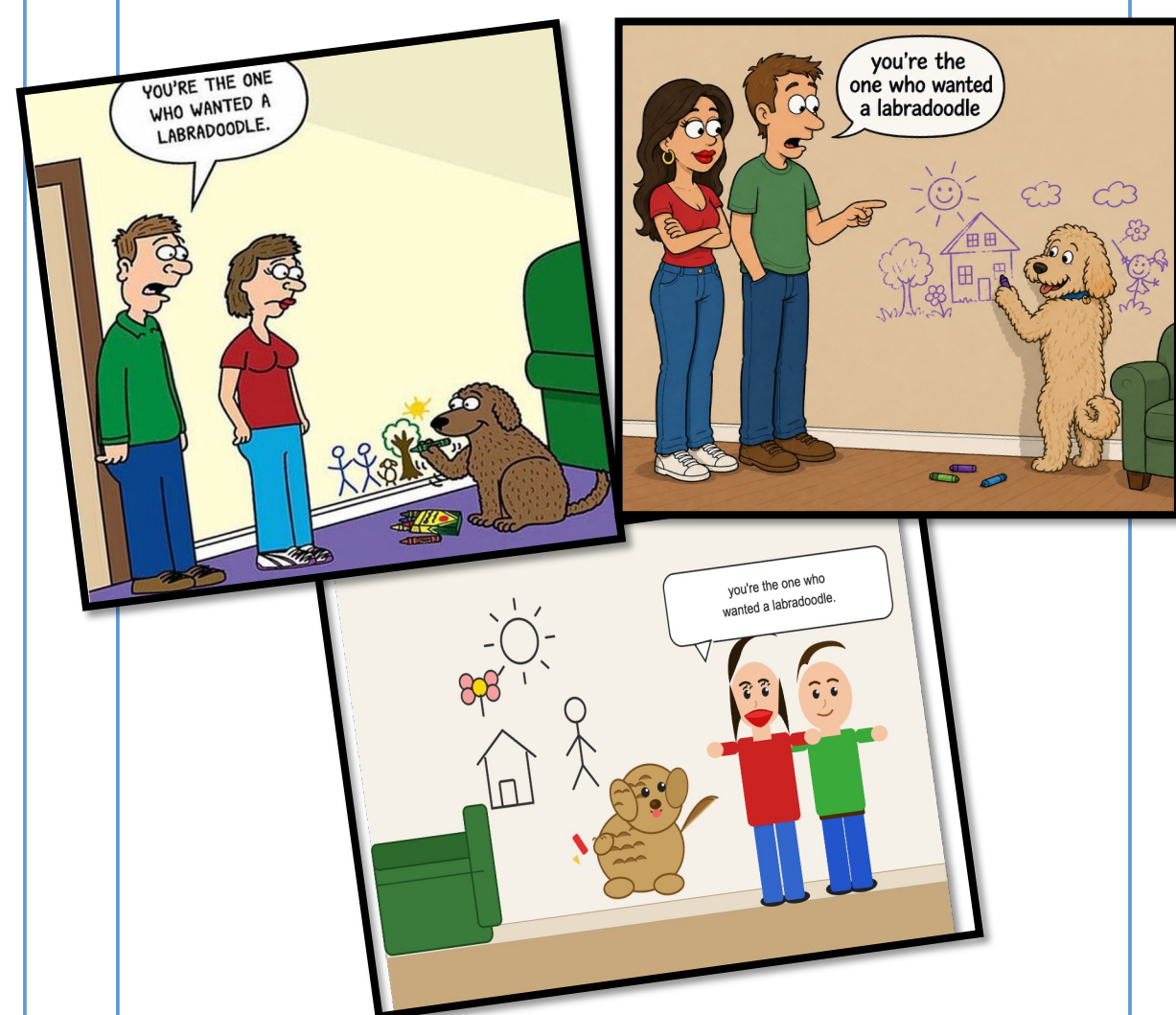
Generate Python code with specific formatting constraints (e.g., font sizes).

Task 3: Formatting Restriction

Rewrite content while avoiding all list-based or styled formatting.

Task 4: Role-Based Simplification

Summarize a technical text for a child using restricted vocabulary and short sentences.



Do you have any ideas for us?

- 1)
- 2)
- 3)
- 4)
- 5)
- 6)

Procedure

Participants will:

- Complete all four tasks using an AI system (e.g., ChatGPT).
- Generate prompts for each task.
- Submit both **initial and revised prompts**, along with corresponding AI outputs.
- Use a new interaction for each task to ensure response independence.

Measures

Prompt-level measures:

- Similarity of key elements.
- Structural complexity.
- Explicitness of instructions.
- Constraint specification and enforcement.

Output-level measures:

- Accuracy of constraint adherence.
- Clarity and coherence.
- Task success (alignment with instructions).

Data Analysis

Planned analyses will compare **monolinguals and bilinguals** on:

- Prompt construction strategies.
- Use of explicit and structured constraints.
- Resulting AI output quality.

Study Status

Study in progress: task design completed and data collection is underway.

NEXT STEPS & DISCUSSION

Planned Experimental Implementation

The next phase will involve controlled data collection with $N = 40$ adult participants, stratified into monolingual and bilingual groups based on self-reported language background (proficiency, age of acquisition, and frequency of use).

Bilingualism → Metalinguistic awareness → Better prompt structure → Better AI output?

Planned Statistical Analysis

Inferential statistical methods:

- **ANOVAs** (group comparisons).
- **Correlation analyses** between aspects of language experience and output quality.

Theoretical Extensions and Future Research

- Future research will extend this paradigm to investigate **cognitive and linguistic mechanisms** in prompt engineering.
- Examine **bilingualism as a predictor of metalinguistic control** in human - AI interaction.
- Differentiate effects of **age of acquisition, proficiency, and language dominance**.
- Investigate the impact of **explicit prompt-engineering training** on performance.
- Expand to **multilingual populations** and diverse sociolinguistic backgrounds.

FOOD FOR THOUGHT

- Is prompt engineering just language skill in disguise?
- Are bilinguals better at "thinking for the listener" or in this case, the AI?
- Does AI reward explicitness because it lacks human inference?
- Could training monolinguals close the gap?

Limitations

- Self-reported bilingualism.
- Task type may influence results.
- AI variability across sessions.

SELECTED REFERENCES

- Wang, Q., Pan, S., Linzen, T., & Black, E. (2025). Multilingual prompting for improving LLM generation diversity. *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, 6367–6389. <https://aclanthology.org/2025.emnlp-main.324.pdf>
- Vatsal, S., Dubey, H., & Singh, A. (2025). *Multilingual prompt engineering in large language models: A survey across NLP tasks*. arXiv, 1–27. <https://doi.org/10.48550/arXiv.2505.11665>
- Wang, A., Morgenstern, J., & Dickerson, J. P. (2025). Large language models that replace human participants can harmfully misportray and flatten identity groups. *Nature Machine Intelligence*, 1–12.