Talking with Oompa Loompas: A Novel Framework for Evaluating Linguistic Acquisition of LLM Agents

Evaluating large language models (LLMs) on their ability to acquire new languages offers insights into their potential for adaptive learning beyond static fine-tuning and prompt engineering. Existing evaluation studies on linguistic competence of LLM agents have focused primarily on vocabulary learning, morphological rule induction, syntactic generalization, pragmatic inference, and cross-linguistic transfer. However, none assess whether LLM agents can acquire a language through pattern recognition and interactive feedback, a central feature of human language acquisition. This raises a fundamental research question: *Can LLM agents develop proficiency in a new constructed language through iterative interaction, by recognizing patterns and adapting with feedback?*

We propose a novel experimental framework in which an LLM agent is evaluated on its ability to acquire and use a synthetically generated language (`Tinkatongue`) in conversation with a bot (`Oompa Loompa`) that understands only `Tinkatongue`. The objective of the LLM agent is to minimize the number of turns required to complete a conversation under `Oompa Loompa`'s feedback mechanism. The interaction begins with `Oompa Loompa` producing an initial utterance and alternates until the LLM completes a full conversation or a configured round limit is reached. The `Oompa Loompa` reciprocates with a positive feedback (`"koro"` + response) whenever the LLM agent speaks a valid sentence and replies with a confused statement (`"moko lira bani"`) when the LLM agent is unable to do so. (refer to Figure 1)

We evaluated GPT-4o-mini, Gemini-2.5-flash and Claude-3.5-haiku in this experimental setup across various metrics: TVR (fraction of valid responses among all turns), AC (fractions of turns that adhere to adjacency compliance rules), FR (ability of a model to recover from feedback) and TTFK (number of turns taken by the LLM agent to speak the first valid sentence). All models show near-perfect ability to recover immediately after a negative signal, yet they largely fail to internalize the adjacency constraints that are crucial for sustained, coherent conversation. The experiment was replicated using another synthetically constructed language (`Zingaloom`) designed to preserve the syntactic framework of `Tinkatongue` while eliminating any lexical overlap. The results presented in Table 1 show closely aligned means and variances across both formal languages, indicating that the evaluation is robust to lexicon changes.

These findings suggest that LLM agents adapt to novel linguistic environments through strategies that mirror human language acquisition processes. This underscores the value of interactive synthetic language as a framework for probing the mechanisms of LLM agent adaptation. Further work involves a more comprehensive evaluation of this task by considering more variations of the language specification and performing ablation studies over language parameters.

Figure 1: Sample Conversation of `Oompa Loompa` with an LLM

		TVR			**AC**		
Lang	**Trial**	GPT	Gemini	Claude	GPT	Gemini	Claude
A	mean	0.01	0.06	0.34	0.10	0.08	0.08
	std dev	0.02	0.08	0.22	0.32	0.11	0.10
B	mean	0.02	0.07	0.33	0.00	0.00	0.07
	std dev	0.05	0.08	0.25	0.00	0.00	0.11
		FR			**TTFK**		
Lang	**Trial**	GPT	Gemini	Claude	GPT	Gemini	Claude
A	mean	1.00	1.00	1.00	12.90	9.90	4.90
	std dev	0.00	0.00	0.00	16.81	12.03	7.80
B	mean	1.00	1.00	1.00	3.20	10.90	6.30
	std dev	0.00	0.00	0.00	11.04	12.87	9.55

Table 1: Results (A: `Tinkatongue`, B: `Zingaloom`)