VNU Journal of Foreign Studies

Journal homepage: https://jfs.ulis.vnu.edu.vn/

PROMPTING LANGUAGE MATTERS: A STUDY OF CHATGPT'S GRAMMAR FEEDBACK IN VIETNAMESE AND ENGLISH

Doan Minh Mao*

FPT University, Hoa Lac Hi-tech Park, Km29, Thang Long Boulevard, Hoa Lac, Hanoi, Vietnam

Received 04 April 2025
Revised 03 June 2025; Accepted 20 August 2025

Abstract: This research investigates whether ChatGPT tends to give explanations that focus more on exam tips (Rule-Centric - RC) or communicative functions (Meaning-Enriched - ME) when it is asked to give explanations to grammar exercises in Vietnamese and in English. The researcher developed four sets of grammar-focused multiple-choice questions on tenses, each of which consisted of 10 items. Each question was then duplicated into two copies: one with the instruction in Vietnamese and one in English, with the rest identical. The questions were sent to ChatGPT five times, which produced a total of 200 responses per language. Two independent raters then rated the explanation for each question as either RC or ME. The analysis showed a strong tendency for the Vietnamese prompts to produce more rule-centric explanations, while English prompts often resulted in more meaning-enriched results, except for the Present Simple set with the Vietnamese prompts, where the percentage of ME responses was considerably higher than for the others. Overall, these findings suggest that the prompt language may influence ChatGPT's grammar explanations, and this raises concerns about the reinforcement of exam-oriented thinking in students and teachers. The study also discusses implications for teachers, parents, and learners on the use of ChatGPT and other Large Language Models (LLMs) in exam-driven contexts like in Vietnam.

Keywords: ChatGPT, large language models (LLMs), EFL in Vietnam, washback, exam preparation

* Corresponding author.
Email address: minhmaodoan@gmail.com
https://doi.org/10.63023/2525-2445/jfs.ulis.5485

SỰ QUAN TRỌNG CỦA NGÔN NGỮ CÂU LỆNH: NGHIÊN CỨU VỀ PHẢN HỒI GIẢI THÍCH NGỮ PHÁP CỦA CHATGPT BẰNG TIẾNG VIỆT VÀ TIẾNG ANH

Doãn Minh Mão

Trường Đại học FPT, Khu Công nghệ cao Hòa Lạc, Km29, Đại lộ Thăng Long, Xã Hòa Lạc, Hà Nội, Việt Nam

Nhận bài ngày 04 tháng 4 năm 2025
Chỉnh sửa ngày 03 tháng 6 năm 2025; Chấp nhận đăng ngày 20 tháng 8 năm 2025

Tóm tắt: Nghiên cứu này tìm hiểu xem liệu ChatGPT có xu hướng đưa ra lời giải thích thiên về mẹo làm bài thi (Rule-Centric–RC) hay chú trọng đến chức năng giao tiếp (Meaning-Enriched–ME) khi được yêu cầu giải thích các bài tập ngữ pháp bằng tiếng Việt và tiếng Anh. Tác giả xây dựng bốn bộ câu hỏi trắc nghiệm ngữ pháp liên quan đến các thì, mỗi bộ gồm 10 câu hỏi. Mỗi câu hỏi sau đó được tạo thành hai bản: một bản với hướng dẫn bằng tiếng Việt và một bằng tiếng Anh, nội dung còn lại giữ nguyên. Các câu hỏi được gửi cho ChatGPT năm lần, tạo ra tổng cộng 200 phản hồi cho mỗi ngôn ngữ. Hai người chấm độc lập sau đó phân loại từng phản hồi thành RC hoặc ME. Phân tích cho thấy các câu lệnh bằng tiếng Việt có xu hướng tạo ra các lời giải thích thiên về mẹo làm bài thi nhiều hơn, trong khi câu lệnh bằng tiếng Anh thường dẫn đến các phản hồi mang tính giao tiếp và gắn với ngữ nghĩa nhiều hơn, ngoại trừ ở bộ câu hỏi về thì Hiện tại đơn bằng tiếng Việt, với tỷ lệ phản hồi ME cao. Nhìn chung, kết quả cho thấy ngôn ngữ được dùng trong câu lệnh có thể ảnh hưởng đến cách ChatGPT giải thích ngữ pháp, từ đó đặt ra mối lo ngại về việc AI có thể củng cố tư duy học vì điểm số khi được sử dụng bởi giáo viên và học sinh. Nghiên cứu cũng thảo luận các hàm ý dành cho giáo viên, phụ huynh và người học khi sử dụng ChatGPT và các mô hình ngôn ngữ lớn (LLMs) khác trong bối cảnh học tập bị ảnh hưởng nhiều bởi các kỳ thi như ở Việt Nam.

Từ khóa: ChatGPT, mô hình ngôn ngữ lớn, học tiếng Anh như ngoại ngữ tại Việt Nam, hiệu ứng washback, luyện thi

1. Introduction

1.1. Background and Rationale

Grammar instruction has been essential in English as a Foreign Language (EFL) education, but the best approach to teaching grammar remains debated. Traditional methods focusing on rules often dominate exam-oriented settings, especially in contexts such as Vietnam, where high-stakes national English exams emphasize discrete linguistic components, such as grammar and vocabulary more than communicative skills due to the lack of resources. This tendency has reinforced test-oriented learning, encouraging students to memorize rules and engage in intensive drilling. Nguyen (2020) pointed out that "Excessive reliance on tricks-based performances and other test-taking strategies for achievement of good scores made them stuck at the recognition level and confronted with struggle in generation of ideas in conversational situations and written communication." (p. 53). Such practices may lead to quick gains on the targeted tests, but they can also create negative washback (Alderson & Wall, 1993; Bailey, 1996), and instruction tends to reflect exam content rather than authentic language use.

In contrast with the purely form-focused approach, Second Language Acquisition (SLA) research has increasingly argued for integrating meaning in grammar instruction. Researchers suggest that grammar should not be treated as a fixated, static list of grammatical

rules, but rather as communicative resources (Ellis, 2006; Larsen-Freeman, 2015). For example, Larsen-Freeman (2001) proposes a three-dimensional framework of form, meaning, and use. This illustrates that learners benefit more when grammar instruction focuses on how structures function in the real communicative contexts.

Another crucial aspect for consideration in EFL contexts is the choice of instructional language. Although teachers were once discouraged from using learners' first language (L1) in class, many teachers now recognize the value of L1 for explaining complex grammar rules. It has been shown that "L1 use reduces students' anxiety levels" (Alshehri, 2017, p. 28). In the same study, one teacher shared: "I try my best to [help] them to comprehend in L2 … but the thing is sometimes we feel the need … I say what do you call it in Arabic and all of a sudden she said something in Arabic and they get the idea…" (p. 26).

2.2. Problem Statement

Recent developments in Artificial Intelligence (AI) have opened up new possibilities for delivering grammar instructions. Large Language Models (LLMs) such as ChatGPT can generate increasingly natural text on demand, potentially serving as "always online" tutors. However, there is one important issue: because ChatGPT depends on the text input from local websites and books in the local language as its training data, it raises the issue of the extent to which ChatGPT is influenced by the local tradition of language teaching, in this case the exam-oriented practices. If it is heavily influenced, and the outputs focus on tricks and shortcuts for quick gains on tests, they may reinforce narrow test-oriented norms rather than support meaningful language use.

2.3. Research Questions

Building on these concerns, the present study addresses two primary questions:

1. *Do Vietnamese-language prompts elicit more rule-centric (RC) grammar explanations from ChatGPT compared to English-language prompts?*
2. *To what extent do the grammar topics (Present Perfect, Present Simple, Present Continuous, Past Simple) affect the distribution of RC vs. ME explanations?*

2.4. Significance of the Study

This study reveals how prompt language can influence ChatGPT's grammar explanations, often reflecting exam-oriented norms in Vietnamese EFL contexts. The findings inform educators, parents, and policymakers on how to guide AI use toward communicative learning, not just test-taking. As secondary students increasingly use AI for self-study, understanding these patterns becomes crucial. The study also offers a replicable coding method for future research on cross-linguistic AI behavior.

3. Literature Review

3.1. Washback Effects in Vietnamese EFL Contexts

There is an overall strong agreement on the term washback. It is defined as the influence that testing has on teaching and learning (Alderson & Wall, 1993). Brown (2005) similarly defines, "washback is the degree to which a test affects the curriculum that is related to it" (p. 242). High-stakes tests can exert a very powerful washback because teachers often "tailor" their instruction to the content and format of the exam test items, and learners adjust their strategies accordingly. According to these authors, washback can be positive if the test encourages

positive teaching practices or negative if it causes narrowing of the curriculum to focus on the test. In the context of Vietnam's EFL education, washback has long been a subject of considerable research because of the exam-driven educational culture here (Pham, 2007).

Several studies have reported the washback effects of Vietnam's English exams on classroom practices and learner behavior. Nguyen (2020), for example, investigated the washback of the 2018 National English Exam on first-year English majors' learning strategies. The findings were telling: because the exam was an objective test focusing on discrete language points, it "resulted in intense memorization of language knowledge to perform test-preparation exercises rather than concentrating on developing non-tested skills (listening, speaking, and writing)" (p. 1). Students spent a large amount of time drilling grammar rules, practicing isolated sentences, and taking mock tests. The study reported that, while these practices may have helped them succeed on the exam, they led to a neglect of untested skills; "only under 20% regularly did the speaking and writing prompts that were provided as part of the English textbook" (p. 45).

3.2. Grammar Explanation Styles

Different approaches to grammar explanation in language teaching have long been debated, which often revolve around the continuum of explicit rule-based instruction and the implicit, meaning-focused instruction. An extensive body of SLA research has investigated how learners best acquire grammatical competence and what role explicit explanations play in that process (Ellis, 2006; Norris & Ortega, 2000). Here, we review key insights from this literature, focusing on how various explanation styles impact L2 learning. These include deductive rule teaching, inductive discovery learning, form-focused instruction, and meaning-enriched explanation.

Explicit instruction teaches grammar rules directly, often using metalanguage, while implicit instruction relies on learners inferring rules from examples. Though useful for natural acquisition, implicit methods alone may not ensure reliable classroom outcomes. Norris and Ortega (2000) showed that explicit instruction leads to significantly greater short-term gains. However, as Ellis (2016) argues, effective explicit instruction should still be meaning-focused, using tasks or examples that connect form with real-world use. Within explicit instruction, grammar can be taught deductively or inductively (Widodo, 2006). In the deductive approach, rules are presented first, followed by examples–common in many EFL textbooks. This method offers clarity and efficiency but may encourage passive learning. In contrast, the inductive approach begins with examples, allowing learners to infer the rules themselves, which can foster deeper engagement.

In our study, *rule-centric explanations* refer to rigid, shortcut-style grammar tips that prioritize test score over students' actual understanding and usage. These explanations are common in exam-oriented contexts, which reduce grammatical usage to decontextualized cues. A typical example can be, "if you see *since*, choose Present Perfect," or "with *will* in the main clause, use Present Simple in the *if*-clause." While they may originate from deductive instruction, such rules have gradually become oversimplified tricks that bypass any semantic or contextual understanding or reasoning.

In contrast, for *meaning-enriched explanations,* the researcher refers to explanations that aim to link grammatical structures to their communicative purposes and semantic relationship. Rather than offering test-taking formulas, these explanations emphasize *why* a grammatical structure is used in a particular context. For example, instead of saying *since* equals Present Perfect, an ME explanation tends to clarify that "since 2010" signals a starting point in

the past with relevance to the present, thus the use of Present Perfect is appropriate.

3.3. Bias in Large Language Models in Educational Contexts

Bias in machine learning refers to systematic patterns in AI outputs shaped by the data they are trained on. O'Brien et al. (2023) note that models like ChatGPT "learn how to 'write' like us, complete with all of our best and worst qualities" (p. 1), meaning they can reproduce human biases embedded in the training data. Informational bias, for instance, arises when certain groups or viewpoints are overrepresented, leading to skewed or unfair outputs, such as image models generating more white than non-white figures. Solaiman et al. (2021) suggest mitigation strategies like diversifying training data and applying filters, while Kasneci et al. (2023) emphasize the need for AI literacy among educators and students to critically engage with LLMs and their limitations in educational contexts.

4. Research Methodology

4.1. Research Design

This study adopted a comparative content analysis design, which combined quantitative methods of frequency statistics and chi-square tests and qualitative methods with content analysis of representative explanations. The primary objective was to observe how ChatGPT (the free version) explained English grammar questions when prompted in two linguistic contexts: Vietnamese (VN) and English (EN).

Rationale for the research design: The study did not interfere with ChatGPT's internal processing but controlled the question format and prompt language to ensure systematic data collection. This design allowed direct comparison between the two conditions (prompting in VN vs. EN), leading to conclusions about tendencies toward Rule-Centric (RC) or Meaning-Enriched (ME) explanations.

The researchers analyzed all the responses from ChatGPT to 40 questions (10 per topic × 4 topics), each prompted 5 times in each language, giving a total of 400 coded responses.

4.2. Grammar Topics and Question Construction

Four grammar points were selected for this study: *Present Simple, Present Continuous, Present Perfect*, and *Past Simple.* These topics were chosen because they were foundational tenses that appeared consistently across all official secondary-level English curricula issued by the Vietnamese Ministry of Education and Training (MOET) and were frequently tested in national exams and classroom assessments.

This age group of learners (typically 11–15 years old) also represented a critical stage in self-directed learning: students began to access the internet more independently, including using tools like ChatGPT for homework support. As such, the way AI models explained these grammar points might exert a significant influence on their learning strategies and grammatical understanding. Therefore, these tenses were not only pedagogically central but also highly relevant for assessing potential AI-induced instructional biases.

Each topic consisted of 10 multiple-choice questions requiring the respondent to select the correct verb form from a list of 4 options (or appropriate tense structure).

Each question set was written in English, but made into two copies with the instruction in Vietnamese - "*Chọn thì ngữ pháp đúng."* - and English - *"Choose the correct verb tense."*. Thus, two content-identical versions were created, differing only in the prompt language.

When submitted to ChatGPT in Vietnamese, this phrase was added before the question set: *"Đáp án là gì, tại sao?"*

For the English version, this phrase was added: *"What's the correct answer, and why?"*

4.3. Data Collection and Analysis

A total of 80 grammar questions (40 in Vietnamese and 40 in English), covering four tenses–Present Simple, Present Continuous, Present Perfect, and Past Simple–were created. Each question was submitted to the free version of ChatGPT five times, resulting in 400 responses per language. The questions were entered manually without providing user profiles, simulating typical usage by students, teachers, and parents. During data collection, responses were categorized into two types: *Specific-Only* (just the explanation for the selected answer) and *Specific+General* (which included a broader explanation about the tense). All responses were saved and labeled with prompt language, tense type, run number, and response type.

For analysis, responses were coded into two categories: *Rule-Centric (RC)*, focusing on formulas and signal words without contextual meaning–and *Meaning-Enriched (ME)*, highlighting communicative intent and contextual usage. Two trained raters, each with over ten years of EFL experience, independently coded the responses, achieving high inter-rater reliability (Cohen's Kappa > 0.85). Discrepancies were resolved through discussion.

Quantitative analysis included frequency comparisons of RC vs. ME responses across languages and tenses, and between Specific-Only and Specific+General formats. Inferential statistics were conducted using 2×2 chi-square tests ($\alpha = 0.05$, $p < 0.001$ for significance). In addition, qualitative analysis was used to explore patterns in selected responses, with particular attention given to the unexpectedly high ME rate in the Vietnamese Present Simple group.

5. Results

This section presents the key findings of the study, organized into descriptive statistics, inferential analysis, and qualitative insights. The results aim to show how prompt language influences the explanatory style of ChatGPT across different grammar tenses and response formats.

5.1. Descriptive Statistics

A total of 800 explanations given by ChatGPT were collected, which reflected two prompt languages - Vietnamese and English - and two response formats: Specific-Only and Specific+General. The summary in Table 1 clearly shows differences in style depending on the prompt input.

Table 1

Overall Summary

Prompt Lang	Data Version	N (responses)	RC (n)	RC (%)	ME (n)	ME (%)
VN	Specific-Only	200	91	45.5	109	54.5
VN	Specific+General	200	71	35.5	129	64.5
EN	Specific-Only	200	28	14.0	172	86.0
EN	Specific+General	200	19	9.5	181	90.5

For the Specific-Only type, when prompted in Vietnamese, ChatGPT gave 45.5% rule-centric (RC) explanations and 54.5% meaning-enriched (ME) explanations. In contrast, when the prompt was in English, RC responses accounted for only 14%, while ME responses made up 86%.

When a general explanation was added (Specific+General), the tendency toward ME became even more substantial: ME accounted for 64.5% of responses in Vietnamese and up to 90.5% in English.

These results suggest that prompt language has strong influences on the model's approach: Vietnamese prompts tend to elicit more RC explanations, whereas English prompts are more likely to result in ME explanations. The inclusion of a general explanation increased the proportion of ME responses, especially in the English group.

However, a notable exception was observed in the Present Simple tense for the Vietnamese prompt: the ME rate was unexpectedly high. This suggests that, beyond prompt language, ChatGPT's exploratory tendencies may vary depending on the topic.

5.2. Detailed Analysis Based on Descriptive Tables

Table 2

Tense-Based Comparison in the Specific-Only Condition

Language	Tense	ME (n)	RC (n)	Total	RC (%)	ME (%)
EN	Present Continuous	46	4	50	8	92
EN	Past Simple	39	11	50	22	78
EN	Present Perfect	40	10	50	20	80
EN	Present Simple	47	3	50	6	94
VN	Present Continuous	26	24	50	48	52
VN	Past Simple	21	29	50	58	42
VN	Present Perfect	25	25	50	50	50
VN	Present Simple	37	13	50	26	74

Table 2 shows a detailed breakdown of RC and ME proportions by tense and prompt language. For the Vietnamese prompts, Past Simple and Present Continuous showed the highest RC rates, at 58% and 48% respectively. This suggests that the model leaned more on rule-based or test-taking tips when dealing with past tenses, those contexts that often involve fixed grammatical formulas.

In contrast, under English prompts, the same tenses produced far lower RC rates: only 22% for Past Simple and 8% for Present Continuous. This indicates a stronger tendency toward meaning-oriented explanations when the prompt is in English.

An interesting result appears in the Present Simple tense under Vietnamese prompts, where ME reached 74%. This was contrary to the initial expectation that explanations would remain rule-centric. This anomaly may come from the nature of the input data or the LLMs' training data.

Table 3

The Impact of General Explanations (Specific+General Condition)

Language	Tense	ME (n)	RC (n)	Total	RC (%)	ME (%)
EN	Present Continuous	49	1	50	2	98
EN	Past Simple	44	6	50	12	88
EN	Present Perfect	40	10	50	20	80
EN	Present Simple	48	2	50	4	96
VN	Present Continuous	26	24	50	48	52
VN	Past Simple	22	28	50	56	44
VN	Present Perfect	32	18	50	36	64
VN	Present Simple	49	1	50	2	98

When general explanations were added to the table, ME proportions increased significantly in both languages, though the effect was stronger for the responses with English prompts. For instance, in tenses like Present Simple and Present Continuous, the ME rate rose to 98%, suggesting that after individual question responses, ChatGPT tended to "expand" its reasoning toward semantic and communicative interpretations.

For the Vietnamese prompts, the general explanation had a more limited impact on the explanatory pattern. For example, Past Simple still showed a high RC rate of 56%, nearly unchanged from the version without the general explanation. One exception was Present Perfect, where the ME rate increased from 50% in Table 2 to 64% in Table 3.

5.3. Inferential Analysis

To determine whether the differences in RC ME explanation rates across the two prompt language conditions - Vietnamese vs. English - were statistically significant, two independent **Chi-square tests** were conducted: corresponding to the two response types: Specific-Only and Specific+General.

a. Specific-Only: Significant Differences in Explanatory Style

Table 4

ChiSquare - Specific Explanation

Language	RC (n)	ME (n)	Total
EN	28	172	200
VN	91	109	200
Total	119	281	400

Statistic	Value
Chi-square	45.9822
p-value	< 0.001
Degrees of Freedom	1

Table 4 shows that under the Specific-Only condition, RC explanations accounted for **45.5%** of responses in the Vietnamese prompt group but it made up only 14% in the English prompt group. A Chi-square test performed on a 2×2 frequency table produced a highly significant result:

$\chi^2(1, N = 400) = 45.98, p < .001$

The expected frequencies indicate that if the prompt language had had no effect, the number of RC responses in each group should have been around 59.5. However, the actual values deviated substantially from this expectation. This confirms that prompt language had a strong influence on ChatGPT's explanatory style in the Specific-Only condition.

b. Specific+General: Effect Persists with General Explanations

Table 5

ChiSquare - Specific+General Explanation

Language	RC (n)	ME (n)	Total
EN	19	181	200
VN	71	129	200
Total	90	310	400

Statistic	Value
Chi-square	37.2903
p-value	0.00000001
Degrees of Freedom	1

Table 5 shows that even with the addition of general explanations, the influence of prompt language remained strong. RC responses made up 35.5% in the Vietnamese group but only 9.5% in the English group. The Chi-square test yielded:

$\chi^2(1, N = 400) = 37.29, p < .000000001$

Again, the expected frequency under the null hypothesis was 45 per group, but the observed frequencies differed significantly. This result reflects a clear bias toward rule-based explanations under Vietnamese prompts, even when general comments were included.

c. Coding Reliability: High Inter-Rater Consistency

Table 6

Rater Agreement - Specific Explanation

Rater1 \ Rater2	RC	ME	Total
ME	14	269	283
RC	106	11	117
Total	120	280	400

Cohen's Kappa: 0.8501

Table 7

Rater Agreement - Specific+General Explanation

Rater1 \ Rater2	RC	ME	Total
ME	7	304	311
RC	84	5	89
Total	91	309	400

Cohen's Kappa: 0.914

To ensure the reliability of the RC–ME classification, Cohen's Kappa coefficients were calculated. Results presented in Tables 6 and 7 demonstrate high inter-rater agreement:

Specific-Only version: $\kappa = 0.8501$

Specific+General version: $\kappa = 0.914$

According to Landis & Koch (1977), both Kappa values fall within the range of "almost perfect agreement", ensuring that the coding process was stable and objective across raters.

5.4. Illustrative Examples and Qualitative Analysis

5.4.1. The Simple Present with Vietnamese Prompts

To clarify the unexpectedly high ME rate in the Present Simple tense under Vietnamese prompts, the researcher conducted a targeted qualitative content analysis. Three grammar questions (Questions 1, 3, and 7 from the Present Simple set) were randomly selected from the main dataset. For each question, all the five responses from different runs were analyzed.

Table 8a

Responses in Vietnamese for Selected Questions on the Simple Present Tenses - Specific Explanation

Question	Run_Order	Specific Explanation
Question 1	1	1. A. goes → She usually goes to school by bike. (Cô ấy thường đi học bằng xe đạp.)
Question 1	2	1. A → Vì "usually" chỉ thói quen, dùng hiện tại đơn.
Question 1	3	1. A. goes → Thì hiện tại đơn (Present Simple) vì "usually" diễn tả thói quen.
Question 3	1	3. A. play → They play soccer on Sundays. (Họ chơi bóng đá vào Chủ nhật.)
Question 3	2	3. A → "on Sundays" chỉ thói quen, dùng hiện tại đơn.;
Question 3	3	3. A. play → Hiện tại đơn, vì "on Sundays" diễn tả hành động thường xuyên.
Question 7	1	7. B. does → She rarely does her homework after dinner. (Cô ấy hiếm khi làm bài tập sau bữa tối.)
Question 7	2	7. B → "rarely" chỉ thói quen, dùng hiện tại đơn, "does" đúng với he/she/it.
Question 7	3	7. B. does → Hiện tại đơn, vì "rarely" là trạng từ chỉ tần suất.
Question 7	4	B. does → Thì hiện tại đơn, vì "rarely" (hiếm khi) là trạng từ tần suất.;
Question 7	5	7. B. does → "She" là số ít, động từ "do" chuyển thành does ở thì hiện tại đơn.

Table 8b

Responses in Vietnamese for Selected Questions on the Simple Present Tenses - Specific+General Explanation

Question	Run_Order	General + Specific Explanation
Question 1	1	1. A. goes → She usually goes to school by bike. (Cô ấy thường đi học bằng xe đạp.); Thì hiện tại đơn dùng cho thói quen, sự kiện thường xuyên xảy ra, lịch trình cố định. Chủ ngữ số ít (she, my brother, my cat, the train) → động từ thêm -s hoặc -es (goes, watches, leaves, does, sleeps). Chủ ngữ số nhiều hoặc I/you/we/they → dùng động từ nguyên mẫu (play, drink, have, go, live).
Question 1	2	1. A → Vì "usually" chỉ thói quen, dùng hiện tại đơn.; Tất cả các câu trên đều dùng thì hiện tại đơn vì diễn tả thói quen, sự thật hiển nhiên hoặc lịch trình cố định.
Question 1	3	1. A. goes → Thì hiện tại đơn (Present Simple) vì "usually" diễn tả thói quen.; Thì hiện tại đơn vì diễn tả thói quen, sự thật hiển nhiên hoặc lịch trình cố định.
Question 1	4	A. goes → Thì hiện tại đơn, vì "usually" (thường xuyên) diễn tả thói quen.;
Question 1	5	1. A. goes → "She usually" diễn tả thói quen, dùng thì hiện tại đơn.; Tất cả các câu trên dùng thì hiện tại đơn vì diễn tả thói quen, sự thật hiển nhiên hoặc lịch trình cố định.

Question 3	1	3. A. play → They play soccer on Sundays. (Họ chơi bóng đá vào Chủ nhật.); Thì hiện tại đơn dùng cho thói quen, sự kiện thường xuyên xảy ra, lịch trình cố định. Chủ ngữ số ít (she, my brother, my cat, the train) → động từ thêm -s hoặc -es (goes, watches, leaves, does, sleeps). Chủ ngữ số nhiều hoặc I/you/we/they → dùng động từ nguyên mẫu (play, drink, have, go, live).
Question 3	2	3. A → "on Sundays" chỉ thói quen, dùng hiện tại đơn.; Tất cả các câu trên đều dùng thì hiện tại đơn vì diễn tả thói quen, sự thật hiển nhiên hoặc lịch trình cố định.
Question 3	3	3. A. play → Hiện tại đơn, vì "on Sundays" diễn tả hành động thường xuyên.; Thì hiện tại đơn vì diễn tả thói quen, sự thật hiển nhiên hoặc lịch trình cố định.
Question 3	4	A. play → Thì hiện tại đơn, vì "on Sundays" (vào Chủ Nhật) diễn tả thói quen.;
Question 3	5	3. A. play → "They" là chủ ngữ số nhiều, dùng động từ nguyên mẫu trong thì hiện tại đơn.; Tất cả các câu trên dùng thì hiện tại đơn vì diễn tả thói quen, sự thật hiển nhiên hoặc lịch trình cố định.
Question 7	1	7. B. does → She rarely does her homework after dinner. (Cô ấy hiếm khi làm bài tập sau bữa tối.); Thì hiện tại đơn dùng cho thói quen, sự kiện thường xuyên xảy ra, lịch trình cố định. Chủ ngữ số ít (she, my brother, my cat, the train) → động từ thêm -s hoặc -es (goes, watches, leaves, does, sleeps). Chủ ngữ số nhiều hoặc I/you/we/they → dùng động từ nguyên mẫu (play, drink, have, go, live).
Question 7	2	7. B → "rarely" chỉ thói quen, dùng hiện tại đơn, "does" đúng với he/she/it.; Tất cả các câu trên đều dùng thì hiện tại đơn vì diễn tả thói quen, sự thật hiển nhiên hoặc lịch trình cố định.
Question 7	3	7. B. does → Hiện tại đơn, vì "rarely" là trạng từ chỉ tần suất.; Thì hiện tại đơn vì diễn tả thói quen, sự thật hiển nhiên hoặc lịch trình cố định.
Question 7	4	B. does → Thì hiện tại đơn, vì "rarely" (hiếm khi) là trạng từ tần suất.;
Question 7	5	7. B. does → "She" là số ít, động từ "do" chuyển thành does ở thì hiện tại đơn.; Tất cả các câu trên dùng thì hiện tại đơn vì diễn tả thói quen, sự thật hiển nhiên hoặc lịch trình cố định.

a. ME Dominance, Even with Formal Markers Present

The analysis of 15 responses in the Specific-Only version showed that most explanations leaned toward ME, although explanations tended to be brief. Even for two questions that included clear formal "markers" such as *usually* (Question 1) and *rarely* (Question 7), ChatGPT still tended to provide semantic, functional explanations rather than rule-based formulas. Only two responses were categorized as Rule-Centric (RC), both from Question 7, which accounted for 20% of the 10 responses with formal markers, and just 13.3% of the total 15.

For example, in response to **Question 1**:

A. goes → She usually goes to school by bike.

→ Explanation: Present Simple is used to express habits, regular events, or fixed schedules.

Similarly, for **Question 3** (*They ___ soccer on Sundays*), here *on Sundays* could be interpreted less as a formulaic signal for the tense as adverbs of frequency *usually*, ChatGPT

stilled explained it in terms of frequency and repeated action:

A. play → Present Simple, because "on Sundays" indicates habitual action.

This stands in stark contrast to responses in the Present Continuous set with Vietnamese prompts, where the model frequently gave purely rule-based justifications such as:

B. is doing → 'right now' is a marker of the Present Continuous."

C. "are waiting → 'now' is a marker of the Present Continuous."

These responses lacked any mention of communicative function or usage context, exemplifying a distinct RC style.

b. General Explanations Reinforce the ME Style

In the Specific+General version, all 15 responses included general explanations added by ChatGPT after the specific answers. The analysis showed that these general explanations consistently mentioned semantic factors, such as habitual actions or frequency, as the reason for using the Present Simple. For example:

"All of the above sentences use the Present Simple tense to express habits, general truths, or fixed schedules."

Notably, in the three responses to Question 7, which previously showed RC tendencies in the Specific-Only version, the added general explanation changed the response toward ME. As a result, in all 15 Specific+General responses, none were classified as purely RC.

c. Hypotheses on Underlying Causes

This phenomenon suggests that in certain contexts, Vietnamese prompts do not necessarily result in RC explanations as originally assumed. Instead, factors such as the tense in question, the nature of training data, or the model's response generation mechanism, especially in generating general explanations, may moderate the explanatory style.

5.4.2. A Comparison of Responses in Vietnamese With Responses in English

Table 9a below shows the specific explanations for the first run of the Vietnamese prompts for the Present Continuous while Table 9b shows the corresponding results for the English version.

Table 9a

Specific Explanations for the First Run of Vietnamese Prompts for the Present Continuous

1. B. are playing → Dấu hiệu "Look!" chỉ một hành động đang diễn ra ngay lúc nói, dùng thì hiện tại tiếp diễn.
2. B. is doing → "right now" là dấu hiệu của thì hiện tại tiếp diễn.
3. A. is knocking → "Listen!" nhấn mạnh hành động đang diễn ra, dùng thì hiện tại tiếp diễn.
4. B. are having → "at the moment" là dấu hiệu của thì hiện tại tiếp diễn.
5. B. am listening → "now" là dấu hiệu của thì hiện tại tiếp diễn.
6. A. is cooking → "at the moment" là dấu hiệu của thì hiện tại tiếp diễn.
7. B. is raining → "Look!" chỉ hành động đang diễn ra, dùng thì hiện tại tiếp diễn.
8. B. are waiting → "now" là dấu hiệu của thì hiện tại tiếp diễn.

9. A. are taking → "at the moment" là dấu hiệu của thì hiện tại tiếp diễn.
10. A. is wearing → "today" có thể chỉ trạng thái tạm thời, nên dùng hiện tại tiếp diễn.

General Explanations: Tất cả các câu này đều có dấu hiệu nhận biết của thì hiện tại tiếp diễn (Present Continuous) nên cần dùng to be + V-ing.

Table 9b

Specific Explanations for the First Run of English Prompts for the Present Continuous

1. B. are playing → ("Look!" suggests something happening right now.)
2. B. is doing → ("right now" indicates an ongoing action.)
3. A. is knocking → ("Listen!" suggests something happening at this moment.)
4. B. are having → ("at the moment" signals an ongoing action.)
5. B. am listening → ("now" shows the action is in progress.)
6. A. is cooking → ("at the moment" indicates an action currently happening.)
7. B. is raining → ("Look!" suggests an action happening right now.)
8. B. are waiting → ("now" shows an ongoing action.)
9. A. are taking → ("at the moment" signals an action in progress.)
10. A. is wearing → ("today" suggests a temporary action happening now.)

An analysis of Table 9a shows that ChatGPT's explanations for the Vietnamese prompts were predominantly **rule-centric**. Only 4 out of 10 responses (Questions 1, 3, 7, and 10) mentioned semantic meaning or communicative context. The remaining responses relied entirely on identifying markers such as "now," "at the moment," or "right now," and applying formulas without requiring learners to understand the meaning. Furthermore, the general explanation added by ChatGPT after the specific answers reinforced this approach, stating: *"Tất cả các câu này đều có dấu hiệu nhận biết của thì hiện tại tiếp diễn, nên cần dùng to be + V-ing"*. This is a purely procedural strategy focused on test-taking tips rather than language use.

In contrast, the explanations provided under English prompts in Table 9b, though short in length, always connected the Present Continuous form with its meaning or communicative function. Every sentence included references such as "ongoing action," "currently happening," or "temporary action," reflecting meaning-enriched explanations. These explanations encouraged learners to interpret why the tense is used in each case rather than simply how to form it. Notably, ChatGPT did not add a general summary at the end of the English responses, and yet the individual explanations already demonstrated greater emphasis on meaning. This contrast highlights how prompt language may shape the depth and focus of the responses provided by AI.

6. Discussion

6.1. General Interpretation of Results

The results show a clear pattern: Vietnamese prompts led to far more Rule-Centric (RC) explanations (45.5% vs. 14% in English). Even with general explanations added, RC responses remained higher in Vietnamese (35.5% vs. 9.5%). These differences were statistically

significant ($p < .001$), confirming that prompt language strongly influences ChatGPT's feedback style. While English prompts consistently triggered Meaning-Enriched (ME) explanations across all tenses, Vietnamese prompts leaned toward test-focused, formulaic responses–except in the Present Simple tense, where the model produced more meaning-based explanations, possibly due to the semantic clarity of adverbs like *usually* and *on Sundays*.

6.2. Pedagogical Implications for EFL Teaching and the Role of Parents and Students

The findings have clear pedagogical implications: teachers should consider training students to use English prompts or ask for usage and context, not just rules. Parents should also be aware that when students use ChatGPT in Vietnamese, they are more likely to receive shallow, rule-based help. Encouraging the use of English prompts, or follow-up questions about meaning, can support deeper, communicative learning.

6.3. Limitations and Directions for Future Research

This study focused only on multiple-choice questions about four verb tenses, limiting generalizability. Other grammar topics (e.g., conditionals, passives) or skills (e.g., writing) were not tested. Also, all data came from one free version of ChatGPT at a specific time, so future versions may behave differently. Findings may not apply to other LLMs like Claude or Gemini. In addition, the study did not control for cross-linguistic influence in bilingual users, which could affect interpretation. Future studies should examine broader grammar areas, more diverse tasks, and multiple LLMs to further investigate prompt-language bias and learner interaction.

6.4. Ethical Implications and Mitigation Strategies

This study highlights ethical concerns when LLMs produce biased outputs in under-resourced languages like Vietnamese. Rule-heavy explanations may reinforce exam-oriented learning while overlooking communicative competence. These biases likely stem from unbalanced training data and dominant language trends. To address this, educators should promote AI and prompt literacy, helping learners critically engage with AI outputs. Long-term solutions may include multilingual fine-tuning and diverse pedagogical datasets to ensure more equitable, context-sensitive support for all learners.

7. Conclusion

This study shows that prompt language significantly affects ChatGPT's grammar explanations. Vietnamese prompts tended to produce rule-based, test-oriented responses, while English prompts led to more meaning-focused explanations. An exception was the Present Simple tense in Vietnamese, where meaning-enriched feedback was more common–likely due to tense-specific features. These findings highlight the influence of language and educational culture on AI behavior. In Vietnam's exam-driven context, promoting thoughtful prompt use can help shift AI-assisted learning toward communicative competence. Integrating such insights into teaching practices may enhance the responsible and effective use of LLMs in language education.

References

Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics, 14*(2), 115–129. https://doi.org/10.1093/applin/14.2.115

Alshehri, E. (2017). Using learners' first language in EFL classrooms. *IAFOR Journal of Language Learning, 3*(1), 20–36. https://files.eric.ed.gov/fulltext/EJ1167274.pdf

Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, *13*(3), 257–279. https://doi.org/10.1177/026553229601300303

Brown, J. D. (2005). *Testing in language programs*. McGraw-Hill Companies.

Ellis, R. (2006). Current issues in the teaching of grammar: An SLA perspective. *TESOL Quarterly*, *40*(1), 83–107. https://doi.org/10.2307/40264512

Ellis, R. (2016). Focus on form: A critical review. *Language Teaching*, *49*(3), 341–374. https://doi.org/10.1177/1362168816628627

Kasneci, E., Sessler, K., Küchemenn, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Larsen-Freeman, D. (2001). Teaching grammar. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language* (3rd ed., pp. 251–266). Heinle & Heinle.

Nguyen, M. H. T. (2020). Beyond the washback of the English National Exam on learning: Subsequent impact on language learning strategies of first-year English majors. *Language Education & Assessment*, *3*(2), 38–57. https://doi.org/10.29140/lea.v3n2.356

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*(3), 417–528. https://doi.org/10.1111/0023-8333.00136

O'Brien, M. J., Alsmadi, I., & Bentley, R. A., Tuba, M. (2024). OpenAI ChatGPT and biased information in higher education. In M. Labby (Ed.), *Empowering digital education with ChatGPT* (pp. 223–234). Chapman and Hall/CRC.

Shaffer, C. (1989). A comparison of inductive and deductive approaches to teaching foreign languages. *Modern Language Journal*, *73*(4), 395–403. https://doi.org/10.2307/326874

Solaiman, I., & Dennison, C. (2021). Process for adapting language models to society (PALMS) with values-targeted datasets. In *Proceedings of the 35th International Conference on Neural Information Processing Systems* (pp. 5861–5873). https://proceedings.neurips.cc/paper/2021/file/2e855f9489df0712b4bd8ea9e2848c5a-Paper.pdf

Widodo, H. P. (2006). Approaches and procedures for teaching grammar. *English Teaching: Practice and Critique*, *5*(1), 122–141. https://edlinked.soe.waikato.ac.nz/journal/files/etpc/files/2006v5n1nar1.pdf