Humanities and Educational Sciences Journal

مجلـــة العلـــوم التربـــوية والدراســـات الإنســـانيـــة

ISSN: 2617-5908 (print)

ISSN: 2709-0302 (online)

DeepSeek vs. ChatGPT: Comparative Efficacy in Reasoning for Adults' Second Language Acquisition Analysis(*)

Mohammad Mahyoob Albuhairy
Department of Languages and Translation
Taibah University, Madina, KSA

Jeehaan Algaraady
Department of Languages and Translation
Taiz University, Yemen

*) موقع المجلة: http://hesj.org/ojs/index.php/hesj/index

DeepSeek vs. ChatGPT: Comparative Efficacy in Reasoning for Adults' Second Language Acquisition Analysis

Mohammad Mahyoob Albuhairy
Department of Languages and Translation,
Taibah University, Madina, KSA

Jeehaan Algaraady
Department of Languages and Translation,
Taiz University, Yemen

Abstract

The advent of new generation language models has revolutionized the field of natural language processing (NLP) due to their exceptional understanding and human language generation capabilities. ChatGPT emerged as an essential model with remarkable strengths for various applications. DeepSeek has recently emerged as the latest advancement in NLP, showing great potential in pure text-generation jobs, semantic analysis, and context-dependent language modeling capabilities. The study investigates and compares the performance of DeepSeek and ChatGPT in assessing adult L2 (second language) acquisition errors applied primarily to South Asian Arabic learners. With this premise, we aim to evaluate their efficacy in detecting linguistic inaccuracies (morphology, syntax, semantics) and diagnosing cases of L1 (first language). Methods include error analysis of non-native Arabic sentences, comparative evaluation of the two models, and contrasting assessment of depth of reasoning. Results show that DeepSeek was significantly better at context-driven error detection (for instance, in detecting SOV word-order transfer), and ChatGPT presented more instructively relevant feedback. However, both needed fine-tuned prompts to bring in feedback related to semantic/pragmatic errors, such as missing articles and dialectal mismatches. Contributions include a proposal for integrating AI tools into L2 pedagogy, emphasizing contrastive drill and sociolinguistic awareness, and recommendations for training AI concerning L1-targeted error profiles. This research pushes the integration of AI into language instruction for scalable solutions for adult L2 learners while pointing at improvements needed in the models.

Keywords: DeepSeek, ChatGPT, LLMs, Mother Tongue Influence (MTI), Second Language Acquisition (SLA), AI-Assisted Error Detection, Contrastive Linguistics

Introduction:

In the last couple of years, the emerging large language models (LLMs) have made notable progress in the field of natural language processing (Chai et al., 2024; Kumar et al., 2024; Hariri, 2023; Chowdhury & Haque; Le et al., 2023). These models tremendously understand, talk, and reason with human languages such as ChatGPT, Gemini, Copilot, DeepSeek, and others. They can offer exceptional capabilities as they help analyze and facilitate language acquisition. Language Acquisition is a rather intricate phenomenon involving cognitive linguistic and sociocultural dimensions (Chubaryan & Vardanyan, 2024; Mustafa et al., 2018). Acquire a language for adults is a challenging process as it does not only require the processing and generating of the language; instead, it requires reasoning about how acquiring is achieved that includes morphological aspects, syntactic structures, semantic variation, and pragmatic use (Krashen, 1982; Ellis, 2008). The rationale for choosing these models is their broader range of uses and unique capabilities in different fields of investigation. DeepSeek is designed to provide an in-depth analysis of specific linguistic phenomena, such as the intricacies of deep thinking and syntactic structures. This makes it one of the ideal solutions for understanding Language Acquisition, which is essential for educators, researchers, and linguists.

On a positive note, ChatGPT is known for its versatility, broad applicability, and interesting and relevant answers to users' prompts (Lo, 2023; Fraiwan & Khasawneh, 2023; Al-Garaady & Mahyoob, 2023). Many researchers say it can effectively aid learners by providing active communication and feedback during a language exchange (Masoudi, 2024; Prasetya & Syarif, 2023; Mahyoob et al., 2023; Brown et al., 2020, Mahyoob, 2020).

Adult second language acquisition (SLA) presents various theoretical notions; dissemination of the Critical Period Hypothesis (Lenneberg, 1967) claims that language acquisition fundamentally increases in difficulty from puberty onwards, owing to lesser neuroplasticity. The Input Hypothesis (Krashen, 1985) states that language is acquired through comprehensible input slightly above the learner's current level. Cognitive Theories (Ellis, 2008) view SLA as involving memory, pattern recognition, and problem-solving. Sociocultural Theory (Vygotsky, 1978) addresses the need for interaction and social context for learning.

The integration of artificial intelligence in error detection for reducing mother tongue influence in adult second language acquisition stands for the intersection of applied linguistics, SLA theory, and computational innovation. Mother tongue interference transfers phonological, syntactic, and lexical structures from a learner's first language (L1) into their L2. More mature systematic errors become, at times, hindrances to communicative competence, especially in typologically dissimilar languages such as South Asian languages to Arabic. Most traditional error analyses do not scale well and lack the granularity needed to study the multifaceted nature assumed by MTI. However, the advent of NLP via deep learning models allows contextual and contrastive analysis for AI models to detect examples of MTI errors at unprecedented accuracy. The computational application of mapping divergences between L1 and L2 structures gives immediate, adaptive feedback that eludes traditional pedagogy. Through the successful amalgamation of theoretical linguistics and AI diagnostics, this study aligns itself with burgeoning research focusing on maximizing adult L2 learning through technology-enhanced, evidence-based interventions.

This paper outlines the pros and cons of both models concerning second language acquisition of Arabic among South Asian adults. This study is essential in evaluating the two models and revealing the gap between them to sensitively assess their effectiveness at spotting specific linguistic errors that can be useful for developing particular linguistic devices and strategies, giving feedback, and simulating environments where a language is mastered. In addition, the study contributes to the existing literature by revealing the role of LLMs in language acquisition and providing insights into how these models can be used to enhance the language learning experience for adults of non-Arabic languages. The study indicates the gaps where there is still a need for improvement in these tools and approaches for facilitating language acquisition and learning.

The rest of the paper is structured as follows: Section 2 details the methodology and data collection. Section 3 analyses, discusses, and contextualizes the results. Section 4 concludes the study, summarizes key findings, and provides future remarks for additional investigations.

Objects of the study:

The current investigation is conducted to achieve the following objectives:

1. Conduct a thorough evaluation comparing ChatGPT's language mastery explanation to DeepSeek, with special consideration given to adult non-native Arabic speakers.
2. Investigating the strengths and weaknesses of both models in dealing with the linguistics, cognitive, and pragmatic aspects presented by adult language learners.
3. Determining the specific contexts and tasks where DeepSeek or ChatGPT would perform better in facilitating language adoption.
4. Providing educators, researchers, developers, and all other stakeholders with information on how those models can benefit adult language acquisition.

Research questions:

We have listed four research questions to highlight our contribution and guide the proposed methodology.

1. How do DeepSeek and ChatGPT differ in their evaluation and reasoning toward the mechanisms of language acquisition in adult speakers of non-Arabic languages?
2. Which models, DeepSeek or ChatGPT, can efficiently identify and address the syntactic, semantic, and morphological patterns pertinent to adult language learners?
3. What are the possible strengths and weaknesses of DeepSeek and ChatGPT in developing linguistic aspects of a non-Arabic adult learner?

These questions thoroughly assess the two models' performance, confirming a focused and evidence-based comparison.

Methods:

A mixed methods research approach was employed: Error analysis was conducted on fifty Arabic sentences produced by South-Asian adult learners to identify errors in morphology, syntax, semantics, pragmatics, and sociolinguistics. The models DeepSeek and ChatGPT underwent testing in two stages. The first testing stage was the baseline analysis, in which input of sentences was made without contextual prompts to measure the ability of the model to detect errors independently. In contrast, the second was prompt-driven analysis using incrementally increasing prompts-e.g., the learner L1

background, contrastive linguistic cues, and deep-structural cues for dependent error detection. Performance was evaluated regarding reasoning depth (qualitative analysis of L1 interference diagnoses) and pedagogical utility (actionability of feedback). Human expert annotations were used as a benchmark against which to pit results, while thematic and statistical analyses revealed strengths (e.g., DeepSeek-framework's context-aware syntax detection; explanatory clarity of ChatGPT.) and weaknesses (e.g., missed the semantic/pragmatic) of each model. Table one below shows some sample sentences from the collected data used for the analysis.

Table (1): Arabic Sentence produced by L2 speakers with gloss and translation

Arabic Sentence produced by L2 speakers	Gloss	Literal Translation	Corrected Arabic Sentence	Correct Translation
أنا قبل ثلاثة يوم يجي	I before three day comes	"I before three day comes."	(أنا) جئت قبل ثلاثة أيام	"I came three days ago."
واحد شهر أنا روح سفر	One month I go travel	"One month I go travel."	سأسافر بعد شهر واحد	"I will travel after one month."
أنا شغل عند نفر ثاني	I work job with person another	"I work at another person."	(أنا) أعمل مع شخص آخر	"I work for someone else."
أنا يحب مشي صباح	I loves walking morning	"I loves walking morning."	(أنا) أحب المشي في الصباح	"I like walking in the morning."

Results and Discussion:

The study assesses the effectiveness of two LLMs, ChatGPT and DeepSeek, in detecting, classifying, and analyzing the linguistic errors produced by adult speakers of Arabic as a second language. The morphological, syntactic, pragmatic, and sociolinguistic angles, plus a systematic error analysis, help us understand the hindrances of Arabic on the part of South Asian Speakers. In this section, we thoroughly introduce two analysis versions of linguistic issues by ChatGPT4o and Deepseak and provide comparative statistics of their identification performance.

ChatGPT4o vs Deepseek (R1) performances:

The sentences were initially submitted to the two models with contextual information-based prompts as follows:

– **Contextualization**: Both Models were informed that South Asian speakers of Arabic produced the sentences.

- **Linguistic Contrast**: Explicit guidance on Arabic-South Asian language differences (e.g., word order, pro-drop phenomena) was provided.
- **Deep-Structure Analysis**: Prompts targeted advanced linguistic features (e.g., article omission, pluralization rules, pro-drop usage).

This incremental prompt significantly improved both models' diagnostic accuracy. However, without explicit contextual and linguistic guidance, their ability to autonomously detect L1 interference and deep-structural errors is supposed to remain limited.

ChatGPT analysis of various Linguistic issues:

The study's findings highlighted a significant presence of errors influenced by mother tongue transfer in the linguistic areas of Morphology, Syntax, Semantics, Pragmatics, and Sociolinguistics. Morphological errors are among the most common, and identification was noted particularly concerning verb inflection and article use. As shown in Table (2), respondents often failed to mark verbs for the correct person, tense, and gender, offering answers such as "أنا يحب مشي صباح / I like walking in the morning." instead of the correct " أنا أحب المشي في الصباح". The omission of definite articles (e.g., "ال/ the" in the word "صباح/morning") is also widespread, mirroring the absence of articles in South Asian languages. Such errors indicate a need for pedagogical focus on Arabic morphology when using gendered and tense-specific verb forms and marking definiteness in nouns. Syntax is the most problematic area, with many misfortunes, including incorrect word order and incorrect preposition use. Many speakers of Hindi, Urdu, and Bengali used SOV word order instead of Arabic SVO/VSO; for example, the sentence, "أنا قبل ثلاثة يوم يجي / I came three days ago" and sentence like "واحد شهر أنا روح سفر/ I will travel in one month", have an immediate transfer of L1 syntax, placing the time phrase at the beginning with the verb incorrectly at the end. Furthermore, participants tended to omit required prepositions (e.g., " to/إلى" or " in/ في"). They included unnecessary pronouns, as in "أنا قبل ثلاثة يوم يجي." These errors indicate that learners extensively resort to L1 syntactic patterns throughout the acquisition process, resulting in mismatching sentence construction in Arabic. The overall implications call for syntactic instruction, wherein sentence reordering exercises and explicit practice with Arabic word order can be undertaken to foster internalization of

the structural differences between their L1 and Arabic. These syntactic errors illustrate speakers' difficulty with the structural rules of Arabic, where verbs feature prominently in the ordering of information in sentence construction, consistent with Showalter & Hayes-Harb (2015).

The expressions classified as semantic errors were incorrect or vague word choices, where participants often used informal or contextually inappropriate vocabulary. For instance, "نفر" (a slang expression denoting "person") was utilized instead of the appropriate standard Arabic "شخص, in the sentences " أنا شغل عند نفر ثاني/ I work a job with a person another." Thus, learners may achieve communicative effectiveness in informal settings; their vocabulary usage in formal contexts seems imprecise. Therefore, it becomes essential to teach contextually appropriate vocabulary.

Table (2): A sample of ChatGPT analysis of adults' sentences

Sentence	Morphology Errors	Syntax Errors	Semantics Errors	Pragmatics Issues	Sociolinguistics Influence
أنا قبل ثلاثة يوم يجي	1. Verb conjugation incorrect (يجي → جئت). 2. Noun agreement mismatch (ثلاثة يوم → ثلاثة أيام).	1. Incorrect tense for verb relative to temporal phrase. 2. Misplaced verb (يجي placed after time phrase).	1. Temporal conflict: past time phrase with present-tense verb. 2. Indefiniteness in time phrase.	Misunderstood but partially communicative in informal settings (e.g., Gulf Arabic pidgin).	Explicit "أنا" due to South Asian language transfer. Overgeneralization of يجي as default.
واحد شهر أنا روح سفر	1. Misordered numeral and noun (واحد شهر → شهر واحد). 2. Base verb (روح) instead of conjugated form (سأسافر).	1. No future marker (e.g., سوف). 2. Verb placement reflects SOV transfer. 3. Missing preposition (إلى).	1. Ambiguity in time phrase (واحد شهر lacks بعد/في). 2. سفر without preposition affects meaning.	Simplified sentence works only in pidgin Arabic. Inaccurate for formal contexts.	SOV word order from South Asian languages. Use of "روح" from informal Gulf Arabic. Overuse of أنا.

أنا شغل عند نفر ثاني	1. Use of noun (شغل) instead of verb (أعمل). 2. Informal نفر instead of شخص.	1. Missing verb in predicate. 2. Redundant pronoun أنا. 3. No agreement between subject and predicate.	1. نفر ثاني informal and less precise. 2. Vague meaning due to noun-based predicate.	Informal and works only in specific dialectical contexts. Not suitable for formal communication.	Influence of Gulf Arabic workplace pidgin. Borrowing of نفر as "person." Redundant أنا from L1.
أنا يحب مشي صباح	1. Verb conjugation incorrect (يحب أحب →). 2. Missing ال in nouns (المشي, الصباح).	1. Misaligned subject-verb agreement (أنا vs. يحب). 2. Missing preposition for time (في).	1. Generic nouns lead to unclear meaning. 2. Temporal expression lacks preposition for clarity.	Communicative in informal settings but grammatically inappropriate.	Influence of bare noun usage from L1. Overgeneralization of subject pronouns. Verb mismatch from L1 syntax.

The pragmatic and sociolinguistic errors pose challenges to understanding language variation across registers. Many participants resorted to informal structures like ما في معلوم for "I don't know." These constructions are somewhat functional for day-to-day communication but inappropriate for some formal or academic environments. The informality suggests socio-cultural adaptations made by non-Arabic Native learners in particular settings while highlighting their inability to switch to formal Arabic registers.

Based on quantitative data, Figure (1) below shows the percentage of the ChatGPT error analysis. Syntax (33%) and Morphology (26%) were the top representing errors, with Semantics (20%), Pragmatics (14%), and Sociolinguistics (7%) being less frequent but equally important in understanding the difficulties faced by learners. AI tools could detect and classify errors. However, insights on context-dependent issues, especially those regarding Pragmatics and Sociolinguistics, were more apparent when errors were fed into human annotation. Thus, these findings reflect the multidimensional way L1 influences interactions and highlight the need to focus on structural and contextual aspects when teaching Arabic to adult

learners. Therefore, the study offers many pedagogical insights to language instructors, focusing on teaching error patterns for syntax and morphology while fostering vocabulary growth and register awareness.

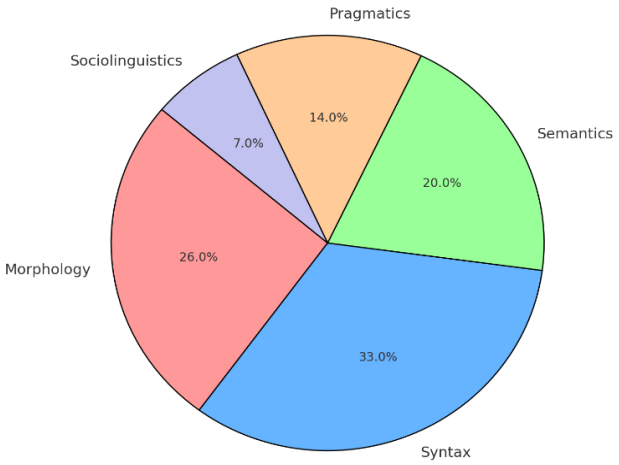

Figure (1): The percentage of ChatGPT error detection for different linguist fields

DeepSeek analysis of various Linguistic issues:

To assess the performance of ChatGPT and DeepSeek comparatively, we fed both systems with identical sentences. Further analysis of DeepSeek's output-error detection capacity, linguistic feature specification, and responsiveness to non-native speaker behavior was strung on the L1 syntactic transfer paradigm (SOV word order). Still, it revealed limitations in autonomously detecting nuanced semantic and pragmatic errors without targeted prompts, as discussed in this subsection. The deeper linguistic analysis reveals additional layers of complexity in second language acquisition. By addressing morphological, syntactic, pragmatic, and sociolinguistic factors and conducting a systematic error analysis, we can better understand the challenges South Asian learners face when learning Arabic. In addition, if L1 interference and contextual gaps are systematically addressed, learners can achieve greater grammatical accuracy and communicative competence in Arabic.

As shown in Table (3), morphological errors include improper verb conjugations, i.e., "يحب / he loves" instead of "أحب / I love", in the sentence, " أنا يحب مشي صباح/ I like walking in the morning.", or improper pluralization, i.e., يوم / a day" instead of أيام / days" in the sentence, "أنا قبل ثلاثة يوم يجي / I came three days ago". Errors in Syntax mainly involve the incorrect ordering of words, especially the use of SOV structures, for example, in the sentences," واحد شهر أنا روح سفر/ I will travel in one month" instead of Arabic's SVO/VSO structures. Errors of Semantics, representing 20% of the total, included the wrong omission of prepositions (e.g., في / in) and articles (e.g., الـ /al-), rendering the sentences ungrammatical and unnatural. Infrequently (5% each), errors in Pragmatics and Sociolinguistics relate to inappropriate word choices (e.g., "نفر / person" rather than "شخص / person") and lack of an appreciation for dialectal variations. Like ChatGPT, the Deapseek analysis demonstrates a significant influence of the speakers' L1 on their Arabic production. The absence of articles and the relatively simple verb conjugation pattern in L1 lead to devastating morphological and syntactic errors. These cases fit into Interlanguage Theory, which illustrates how speakers access their L1 structures while targeting the full development of their transitional L2 grammar. Furthermore, insufficient exposure to the sociolinguistic diversity of Arabic (i.e., the difference between Modern Standard Arabic and dialects) creates pragmatic mismatches. Because of that, errors here need specific instruction in Arabic grammar, particularly in conjugation, order of words, and articles and prepositions, and more exposure to natural contexts in Arabic to enhance pragmatic and sociolinguistic competence. Proficiency in Arabic.

Table (3): A sample of DeepSeek analysis of adults' sentences

Sentence	Intended Meaning	Errors	Linguistic Features	Corrected Sentence
أنا قبل ثلاثة يوم يجي	"I came three days ago."	- **Verb tense**: Present يجي **(yiji)** instead of past جئت **(ji'tu)**. - **Pluralization**: يوم **(yawm)** should be أيام **(ayyam)**. - **Word order**: SOV influence from L1.	- **Morphology**: Incorrect verb conjugation. - **Syntax**: Incorrect word order. - **Semantics**: Time expression error.	أنا جئت قبل ثلاثة أيام. (*Ana ji'tu qabl thalatha ayyam.*)
واحد شهر أنا روح سفر	"I will travel after one month."	- **Verb tense**: Present روح **(rooh)** instead of future سأسافر **(sa'usafir)**. - **Word order**: SOV influence from L1. - **Preposition**: Missing بعد **(ba'da)**.	- **Syntax**: Incorrect word order. - **Morphology**: Incorrect verb tense. - **Semantics**: Missing time expression.	سأسافر بعد شهر واحد. (*Sa'usafir ba'da shahr wahid.*)
أنا شغل عند نفر ثاني	"I work for someone else."	- **Noun-verb confusion**: شغل **(shughl)** instead of أعمل **(a'mal)**. - **Word choice**: نفر **(nafar)** instead of شخص **(shakhs)**. - **Word order**: SOV influence from L1.	- **Lexicon**: Incorrect word choice. - **Syntax**: Incorrect word order. - **Morphology**: Noun used instead of verb.	أنا أعمل مع شخص آخر. (*Ana a'mal ma'a shakhs akhar.*)
أنا يحب مشي صباح	"I like walking in the morning."	- **Verb conjugation**: يحب **(yuhibbu)** instead of أحب **(uhibbu)**. - **Article omission**: Missing الـ **(al-)** for مشي **(mashi)**. - **Preposition**: Missing في **(fi)**.	- **Morphology**: Incorrect verb conjugation. - **Syntax**: Missing articles and prepositions. - **Semantics**: Incomplete time expression.	أنا أحب المشي في الصباح. (*Ana uhibbu al-mashi fi al-sabah.*)

As illustrated in Figure (2), the most frequent errors were found to occur mainly in Morphology and Syntax, which stand at 40% and 30% of the total errors, while the Semantics errors reach 20%, Sociolinguistics and Pragmatics show 5% of the total.

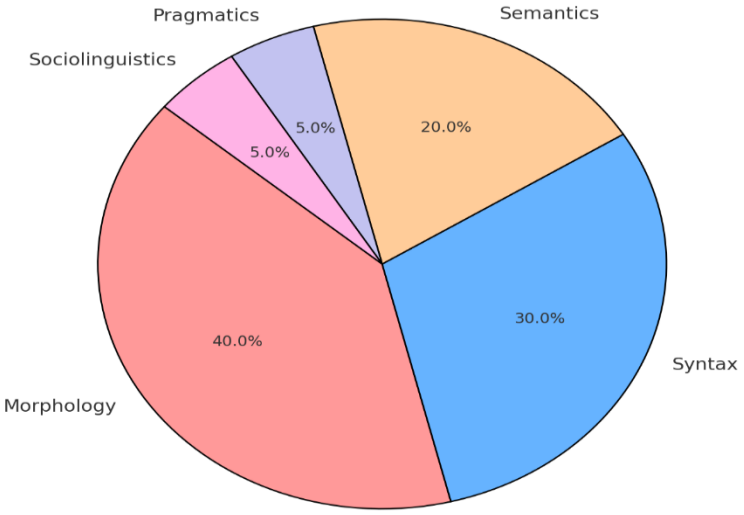

Figure (2): The distribution of errors across different linguistic fields detected by DeepSeek

ChatGPT and DeepSeek are useful AI tools that help identify and categorize errors pertinent to the different linguistic domains. In Syntax and Morphology, AI could measure high-frequency mistakes, whereas in Semantics, AI could identify problematic use of vague or incorrect word choices. However, it failed to comprehend the subtle distinctions violated across specific context cases in Pragmatics and Sociolinguistics. The findings demonstrate that AI could provide some automation and assist in error detection with the promise of scalability to support any post-processing of massive datasets. In contrast, human individual effort is crucial in interpreting context-dependent errors.

Table (4) features the contrasting study of ChatGPT and DeepSeek on the detection and evaluation of the linguistic characteristics involved in L2 speech by Arabic non-native speakers. The table presents the measurable performance ratings of the two models based on salient linguistic areas

(morphology, syntax, semantics). It compares the two models regarding errors originating from L1 interference and structural complexity.

Table (4): Linguistic error detection statistics for ChatGPT4o and DeepSeek

Linguistic Field	Percentage of error detection		Statistics		Description of Errors with examples	
	Chat GPT (4o)	DeepSeek (R1)	T-test	P-value		
Morphology	24%	30%	0.0	1.0	ChatGPT	Verb conjugation errors, incorrect pluralization, and omission of gender/number agreement.
					DeepSeek	Verb conjugation (يحب → أحب), pluralization (يوم → أيام), noun-verb confusion (شغل → أعمل).
Syntax	30%	40%			ChatGPT	Word order errors (SOV influence), omission of prepositions, and redundant pronoun usage.
					DeepSeek	SOV word order (أنا روح سفر), misplaced time phrases (قبل ثلاثة يوم → قبل ثلاثة أيام).

Linguistic Field	Percentage of error detection		Statistics		Description of Errors with examples	
	Chat GPT (4o)	DeepSeek (R1)	T-test	P-value		
Semantics	16%	20%			ChatGPT	Misuse of lexical items due to native language interference, leading to unintended meanings.
					DeepSeek	Missing prepositions (صباح → في الصباح), missing articles (مشي → المشي)
Pragmatics	20%	5%				Overuse of informal expressions, culturally inappropriate structures, and miscommunication.
					DeepSeek	Inappropriate word choice (نفر → شخص).
Sociolinguistics	10%	5%			ChatGPT	Influence of social norms from L1 (e.g., overgeneralization of formal/informal speech).
					DeepSeek	6.7% Colloquial vs. formal mismatch (مافي → لا أعرف معلوم).

The analysis indicated that both models were initially confronted with identifying deeper syntactic and semantic structures of Arabic, such as

phenomena related to pro-drop and slight differences in word order. However, whenever the interaction of L1 syntactic transfer (i.e., word order, influenced by the mother tongue) was at stake in L2 Arabic production, DeepSeek was able to identify it. At the same time, ChatGPT paid more attention to surface grammatical errors than recognizing more significant linguistic or pragmatic issues. Interestingly, the underlying diagnostics of both models improved their task performance when provided specific prompting to attend to these linguistic errors (e.g., L1 interference or semantic mismatches); however, outside of these prompts, their independent ability to fallibility in identifying these deeper issues was limited.

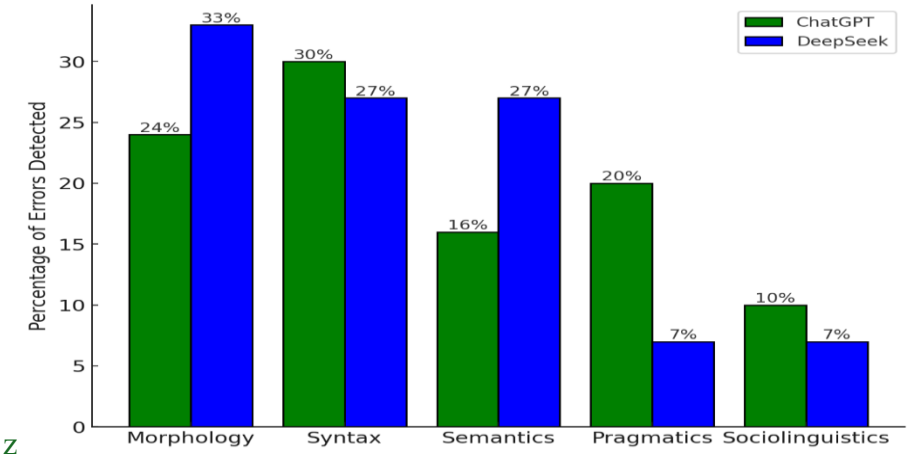

Figure (3): ChatGPT and DeepSeek error detection comparison

Figure (3) compares error detection in five domains of language ChatGPT's and DeepSeek's performance. DeepSeek did better in morphology (33% vs. 24%) and semantics (27% vs. 16%), meaning that, generally, DeepSeek seems to master the rules of grammar and word choice better. On the contrary, ChatGPT outperformed DeepSeek in Syntax (30% vs. 27%), Pragmatics (20% vs. 7%), and Sociolinguistics (10% vs. 7%); that is, DeepSeek can understand a broader context of sentence organization, moment relevance, and sociolinguistic meaning. So, the two models might possess complementing strengths: DeepSeek reaches for the rule-governed language errors, whereas ChatGPT thrives on contextual and communicative effects. These two may yield powerful results for AI-assisted testing and learning in Arabic as a second language.

Intervention teaching SLA to South Asian speakers of Arabic must focus specifically on morphology and syntax regarding verb conjugation, pluralization rules, and SVO/VSO word order in response to L1 interference. A contrastive analysis of Arabic and the learners' South Asian language(s), such as SOV vs. SVO structures, would enhance metalinguistic awareness. Input needs to be contextualized to incorporate actual texts and some dialect exposure; it also must embrace explicit instruction on articles, prepositions, and pragmatic rules to help bridge semantic and sociolinguistic gaps. Training models ChatGPT and DeepSeek to improve error detection would involve annotated learner L2 corpora focusing on common L1 transfer errors, while prompt engineering would also help insight deep structural analysis. A further development of learner autonomy at this stage, combined with immersive practice and formal feedback loops with human instructors and AI tools facilitating, can bridge a gap between grammatical accuracy and communicative competence.

The paired t-test obtained a t-value of 0.0 and a p-value of 1.0, indicating that the error detection percentage means of ChatGPT and DeepSeek are not significantly different in the different linguistic areas. The p-value is much more significant than the conventional level of significance (e.g. 0.05), thereby indicating that DeepSeek is relatively stronger in morphology, syntax, and semantics. In contrast, ChatGPT has a relative advantage in pragmatics and sociolinguistics.

Evidently, from an educational perspective, teachers must pay special attention to addressing structural and functional gaps in learning a second language. Syntax exercises that lay stress on word orders and verb placements and a little semantic training in the formality and informality of lexical items would significantly enhance learners' abilities. In addition, AI-based language learning tools will effectively offer students instant feedback on their errors while practicing. This study stands at the juncture of linguistic analysis and AI capabilities, as it aims to forward the second language acquisition research agenda with theoretically sound yet practically rich insights on instructional methods to realize this aim.

Conclusion:

This study emphasizes the significant role of L1 interference in causing errors among South Asian speakers of Arabic, with the bulk of errors being

morphological (e.g., verb conjugation, pluralization) and syntactic (e.g., SOV word order). Whereas Arabic and South Asian languages show mismatches in structures such that the errors may be most salient), DeepSeek was shown to be more capable of detecting L1-induced syntactic transfer errors. In contrast, neither ChatGPT nor DeepSeekc can detect deep-structural and pragmatic errors independently without specific prompting and context. This brings to light the necessity for context-aware training and linguistically annotated corpora. In enhancing Arabic L2 pedagogy, instruction should concentrate on contrastive grammar drills (e.g., SVO vs. SOV), the correct use of articles/prepositions, and awareness of sociolinguistics (e.g., dialect differences). L1-specific error classifications and prompt engineering would further fine-tune AI-based error detection, making a bridge between pure grammatical correction and communicative competence. Following this integrated dual approach of human-based pedagogy and AI scaffolding is a way to reduce L1 interference and promote Arabic competency.

Limitations and Future Direction:

This study is limited to investigating morphology, syntax, and semantics, while phonology and phonetics were excluded mainly due to the focus on written language errors only. Demographic variables, such as age and gender, were not specified, and these could also affect second language acquisition patterns. It will be interesting for future studies to fill these gaps by considering phonology to investigate pronunciation and intonation in error production and detection. Broadening the scope to factor in these elements will offer a fuller picture of second language acquisition and increase the relevance of AI-enabled error detection tools.

References:

Al-Garaady, J., & Mahyoob, M. (2023). ChatGPT's capabilities in spotting and analyzing writing errors experienced by EFL learners. Arab World English Journals, Special Issue on CALL, (9).

Chai, Y., Jin, L., Feng, S., & Xin, Z. (2024). Evolution and advancements in deep learning models for Natural Language Processing. Applied and Computational Engineering. https://doi.org/10.54254/2755-2721/77/20240674.

Chowdhury, M., & Haque, A. (2023). ChatGPT: Its Applications and Limitations. 2023 3rd International Conference on Intelligent Technologies (CONIT), 1-7.

https://doi.org/10.1109/CONIT59222.2023.10205621.

Chubaryan, A., & Vardanyan, M. (2024). EXPLORING THE COGNITIVE DIMENSIONS OF LANGUAGE ACQUISITION. Armenian Folia Anglistika. https://doi.org/10.46991/afa/2024.20.1.13.

Ellis, R. (2008). *The Study of Second Language Acquisition*. Oxford University Press.

Fraiwan, M., & Khasawneh, N. (2023). A Review of ChatGPT Applications in Education, Marketing, Software Engineering, and Healthcare: Benefits, Drawbacks, and Research Directions. ArXiv, abs/2305.00237. https://doi.org/10.48550/arXiv.2305.00237.

Hariri, W. (2023). Unlocking the Potential of ChatGPT: A Comprehensive Exploration of its Applications, Advantages, Limitations, and Future Directions in Natural Language Processing. ArXiv, abs/2304.02017. https://doi.org/10.48550/arXiv.2304.02017.

Krashen, S. D. (1982). *Principles and Practice in Second Language Acquisition*. Pergamon.

Krashen, S. (1985). The input hypothesis: Issues and implications. Longman.

Kumar, R., J, P., S, G., Devadas, R., Ambreen, L., & Pal, R. (2024). Emerging Trends of Language Models and Their Significance in Digital Era. 2024 Second International Conference on Advances in Information Technology (ICAIT), 1, 1-4. https://doi.org/10.1109/ICAIT61638.2024.10690389.

Le, N., Hadiprodjo, N., El-Alfy, H., Kerimzhanov, A., & Teshebaev, A. (2023). The Recent Large Language Models in NLP. 2023 22nd International Symposium on Communications and Information Technologies (ISCIT), 1-6. https://doi.org/10.1109/ISCIT57293.2023.10376050.

Lenneberg, E. H. (1967). Biological foundations of language. Wiley.

Lo, C. (2023). What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. Education Sciences. https://doi.org/10.3390/educsci13040410.

Mahyoob, M., Al-Garaady, J., & Alblwi, A. (2023). A proposed framework for human-like language processing of ChatGPT in academic writing. *International Journal of Emerging Technologies in Learning (iJET)*, 18(14).

Mahyoob, M. (2020). Challenges of e-Learning during the COVID-19 Pandemic Experienced by EFL Learners Arab World English Journal (AWEJ).

Mustafa, M., Masnan, A., Alias, A., & Radzi, N. (2018). Sociocultural Theories in Second Language Acquisition. *The International Journal of Academic Research in Business and Social Sciences*, 7, 1168-1171. https://doi.org/10.6007/IJARBSS/V7-I12/3747.

Masoudi, H. (2024). Effectiveness of ChatGPT in Improving English Writing Proficiency among Non-native English Speakers. International Journal of Educational Sciences and Arts. https://doi.org/10.59992/ijesa.2023.v3n4p2.

Prasetya, R., & Syarif, A. (2023). ChatGPT as a Tool for Language Development: Investigating Its Impact on Proficiency and Self-Evaluation Accuracy in Indonesian Higher Education. VELES (Voices of English Language Education Society). https://doi.org/10.29408/veles.v7i3.19303.

Showalter, C., & Hayes-Harb, R. (2015). Native English speakers learning Arabic: The influence of novel orthographic information on second language phonological acquisition. Applied Psycholinguistics, 36, 23 - 42. https://doi.org/10.1017/S0142716414000411.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.