Available Online: 8 July 2026 DOI: 10.54254/2753-7080/2026.35237

The naturalness of AI tutoring in English speaking practice: a conversation analysis of turn-taking in human-AI interaction

Wenting Li

Graduate School of Education, University of Pennsylvania, Philadelphia, USA

li48@upenn.edu

Abstract. Interactional naturalness of human-AI dialogue is a key factor in fostering meaningful communicative engagement. This study addresses this gap by adopting Conversation Analysis (CA) as a methodological framework to examine recorded speaking practice sessions between English as a Second or Foreign Language (ESL/EFL) learners and an AI tutoring system. Drawing on the principles of emergence, indexicality, and recipient design, the study investigates why and how interactional naturalness is violated in AI-mediated conversation. These findings reveal that while the AI tutor produces grammatically well-formed and topically relevant utterances, its conduct frequently diverges from the moment-by-moment, context-sensitive organization that characterizes natural human interaction. Particularly, the AI system fails to consistently align with emergent turn sequences, indexical references, and recipient-designed responses. The results indicate that current AI tutoring systems, despite advanced language models, have not yet achieved interactional naturalness as defined from a conversation-analytic perspective. Therefore, a shift in evaluation criteria from technical accuracy to interactional quality in the design and assessment of AI-based language tutors is called-for.

Keywords: Artificial Intelligence (AI) tutoring, EFL/ESL speaking practice, interactional naturalness, conversation analysis, human-AI interaction

1. Introduction

In recent years, the integration of Artificial Intelligence (AI) into language learning has significantly transformed the ways learners engage with English as a Second or Foreign Language (ESL/EFL). AI tutoring applications have become widely adopted, offering learners flexible opportunities to practice spoken English, receive feedback, and develop communicative competence outside traditional classroom settings [1]. Unlike conventional instruction, AI tutors can provide individualized support on demand, enabling repeated practice and scaffolding tailored to learners' linguistic needs [2].

Among these developments, AI-supported speaking practice has undergone a notable transformation. As interaction in the real world has increased the demand for meaningful communication, the needs of language learners are changing from form-focused production toward interaction-oriented practice [3]. Therefore, AI-based speaking practice has increasingly taken the form of simulated conversations rather than isolated drills or sentence-level exercises [4]. Within this transition from form to meaningful interaction, naturalness becomes a central concern, as it directly influences whether AI-mediated speaking practice resembles authentic communicative engagement, which may have an impact on affect learning outcomes. In this study, naturalness is not treated as a surface-level imitation of human speech, but as an interactional achievement emerging moment by moment in situated talk [5]. Drawing on conversation analytic traditions, the author conceptualizes naturalness as grounded in emergence, indexicality, and recipient design [5]. From a user-experience perspective, naturalness is therefore not merely fluency or grammaticality, but the extent to which the system can participate in the dynamic co-construction of talk.

Despite rapid technological advancement, a critical gap persists in research on AI-mediated speaking practice. Some scholars argue that recent large language model-based tutoring systems have largely overcome earlier limitations in grammar, pronunciation, and lexicon, thus approximating human-like interaction [4]. However, emerging work in human-AI interaction suggests that breakdowns often arise from a lack of interactional naturalness rather than from limitations in grammar or vocabulary [6]. Others also caution that surface-level fluency should not be equated with interactional competence [7, 8]. Much prior research on AI in speaking instruction has focused primarily on technical performance—grammar accuracy, vocabulary, syntactic complexity, pronunciation and so on [9]. Yet relatively little empirical work has examined the naturalness of interaction in AI tutoring. In other words, although technical problems are becoming less serious, it is still unclear whether AI can manage real conversations step by step as humans do. This means the key issue is not whether AI can produce correct sentences, but whether it can respond naturally to the flow of interaction.

To address this gap, this study adopts Conversation Analysis (CA) as its methodological framework. CA enables a detailed analysis of how turns are organized and managed in interaction, including how participants take turns and respond to each other [10, 11]. By analyzing recorded English speaking practice sessions between ESL/EFL learners and an AI tutoring system, this study investigates why and how interactional naturalness is violated in human-AI speaking practice, focusing on the ways in which the AI tutor's conduct diverges from the principles of emergence, indexicality, and recipient design. The findings indicate that while the AI tutor produces grammatically well-formed and contextually relevant utterances at a topical level, its interactional conduct does not consistently conform to the principles of emergence, indexicality, and recipient design. These patterns suggest that current AI tutoring systems, despite advanced technology, do not fully achieve interactional naturalness as defined in this study.

2. Literature review

2.1. AI tutoring in EFL speaking

With the rapid development of artificial intelligence, AI tutoring systems have become widely used in English as a Foreign Language (EFL) and English as a Second Language (ESL) contexts. Many studies have examined how AI tools support language learning, especially in speaking practice. Early research mainly focused on Automatic Speech Recognition (ASR), pronunciation scoring, and feedback accuracy. For example, scholars discussed how AI systems can detect pronunciation errors and provide immediate corrective feedback [2]. These studies show that AI tutoring can help learners improve pronunciation, fluency, and vocabulary use through repeated practice.

In addition, researchers have found that AI tutoring increases learner autonomy. Because AI systems are available anytime and anywhere, learners can practice without waiting for a teacher. This flexibility is especially helpful for adult learners who have busy schedules. Studies also report that learners often feel less anxious when speaking to a machine than to a human teacher. As a result, they may be more willing to speak and take risks [12]. This suggests that AI tutoring can create a low-pressure environment that supports oral practice.

More recent research has moved beyond pronunciation and grammar. With the development of large language models, AI tutoring systems can now simulate longer conversations instead of short drills. According to Godwin-Jones [4, 13], many AI applications have shifted from sentence-level exercises to dialogue-based interaction. These systems present themselves as conversational partners that can ask questions, respond to answers, and give feedback. This shift reflects a broader change in language teaching, from focusing only on correct forms to focusing on interaction [7].

As AI becomes more interactive, scholars pay more attention to the quality of interaction between learners and machines. Some studies suggest that interaction plays an important role in language development. In sociocultural and interaction-based theories, learning happens through meaningful communication. Therefore, the way AI responds to learners may influence learning outcomes. Research shows that learners value AI systems not only for correct answers, but also for how natural and responsive the conversation feels. When the AI reacts quickly and appropriately, learners report higher satisfaction and stronger engagement.

However, much of the existing research still measures AI performance mainly through technical indicators. These include speech recognition accuracy, error detection rate, lexical richness, and grammatical correctness. From this perspective, if the AI can produce correct sentences and understand learner input, it is considered successful. Many recent studies argue that technological limitations are becoming smaller because AI systems are more advanced than before [4]. As a result, some researchers believe that AI tutoring is already close to human-like performance in speaking practice.

In summary, previous research has shown that AI tutoring can improve pronunciation, increase practice opportunities, reduce anxiety, and simulate conversation. It has made important progress in technical performance and feedback design. At the same time, the growing use of AI in education has led scholars to examine human-AI interaction more closely. Researchers now recognize that interaction is central to speaking development, and that the quality of this interaction deserves deeper study.

2.2. Naturalness

Although AI tutoring has improved in many technical aspects, the question of naturalness remains debated. In speaking practice, naturalness does not only mean correct grammar or fluent speech. It also means whether the conversation feels smooth, connected, and responsive.

Some scholars argue that modern AI systems are already very natural. They claim that with advanced language models, AI can generate contextually relevant and grammatically accurate responses. From this view, problems such as incorrect grammar or inappropriate vocabulary have largely been solved [4]. If the AI can produce long, coherent sentences and stay on topic, it is seen as performing well.

However, other researchers disagree. Scholars such as Seedhouse [7], Kasper and Wagner [8] emphasize that interactional competence is different from grammatical competence. A speaker may produce correct sentences, but still fail to respond appropriately in a conversation. Conversation is organized turn by turn. Each turn depends on the previous one and shapes the next. If a response does not align with the prior turn, the interaction may feel unnatural, even if the sentence itself is correct.

Many past studies on AI speaking tools focus mainly on surface features. They test whether the system recognizes speech correctly, whether it gives accurate pronunciation scores, or whether it produces complex syntax. These are important factors, but they do not fully explain whether learners can engage in smooth, real-time dialogue with the AI. In other words, technical success does not always mean interactional success.

Research in human-AI interaction provides further insight. Moore and Arar [14] suggest that breakdowns in human-machine conversation often happen at the level of timing, alignment, and relevance. For example, the AI may respond too slowly, ignore part of the learner's meaning, or continue with a pre-set script without adapting to unexpected answers. In such cases, the conversation may feel mechanical or disconnected. These problems are not about grammar, but about how turns are organized and how participants respond to each other.

Therefore, a gap exists in the literature. While many studies show that AI systems are technically advanced, fewer studies examine whether they follow the principles of natural conversation. There is limited research that uses detailed interactional analysis to study turn-taking, response design, and context-sensitive reference in AI-mediated speaking practice. Most evaluations rely on quantitative scores rather than a close analysis of real talk.

This study addresses this gap by focusing on interactional naturalness. Instead of asking whether the AI produces correct sentences, it asks whether the AI can participate in the emergent, indexical, and recipient-oriented organization of conversation. By examining emergence, indexicality, and recipient design in real learner-AI dialogues, this study contributes a new perspective to the discussion of AI tutoring in EFL speaking. It shifts the focus from technical accuracy to interactional quality, and from product-based evaluation to process-based analysis.

3. Theory

This study draws on Conversation Analysis (CA) as its main theoretical framework. CA was first developed by Harvey Sacks, Emanuel Schegloff, and Gail Jefferson in the 1960s and 1970s. It studies how people organize talk in everyday life. Instead of focusing on grammar or vocabulary alone, CA looks at how speakers take turns, respond to each other, repair problems, and build meaning together step by step. From this view, conversation is not random. It has an orderly structure, and this order is created by participants in real time.

One key concept in CA is turn-taking organization. In their early work, Sacks et al. [10] showed that speakers follow certain rules when they speak. Usually, one person talks at a time. Speakers also show when a turn is ending, and the next speaker responds at the right moment. Smooth timing and relevant responses make the conversation feel natural. If turns are too long, too delayed, or not clearly connected to the previous turn, the interaction may feel strange or awkward.

This study also builds on three related ideas: emergence, indexicality, and recipient design. Emergence means that each turn is shaped by the previous turn. Speakers do not simply follow a fixed script. They adjust their words according to what just happened. Indexicality, a concept influenced by Harold Garfinkel, means that speakers use context-dependent expressions to refer to earlier talk. The meaning of words such as pronouns depends on the local conversational context. Recipient design means that speakers design their talk for a specific listener. They choose words, tone, and level of detail based on what the listener knows or needs.

In language education research, CA has been widely used to study classroom interaction. For example, scholars have analyzed how teachers ask questions, how students respond, and how feedback is given in ESL/EFL classrooms. Researchers such as Paul Seedhouse [7] have shown that classroom talk has its own interactional patterns, and that learning opportunities are closely connected to turn-taking and repair practices. More recently, CA has been applied to technology-mediated interaction, including online learning and humancomputer communication. These studies show that even when technology works well at a technical level, interactional problems may still occur.

Based on this theoretical tradition, the present study treats naturalness not as fluency or correctness, but as an interactional achievement. By examining emergence, indexicality and recipient design in the turn-taking between learners and AI tutoring, this study aims to show whether the AI system can follow the same interactional principles that guide human talk.

4. Method

This study uses a qualitative case study design based on Conversation Analysis (CA). The aim is to examine how interactional naturalness is achieved or violated in AI-mediated English speaking practice. The focus is on detailed analysis of real conversations between learners and an AI tutor.

4.1. Participants

A total of 20 adult ESL/EFL learners were recruited through university mailing lists and social media groups. All volunteers completed an initial screening questionnaire.

To be eligible for the study, participants had to meet three criteria: (1) be at least 18 years old; (2) be currently learning English as a second or foreign language; and (3) have prior experience using AI tutoring systems for English speaking practice. Eight volunteers were excluded at this stage because they did not meet one or more of these requirements, leaving 12 eligible participants.

The remaining 12 participants completed a background questionnaire and a short speaking task. The questionnaire collected information on participants' English proficiency, first language, age, learning goals, and prior experience with AI tutoring systems.

Because conversation analysis requires detailed examination of naturally occurring interaction, it was not feasible to analyze data from all 12 participants. Therefore, a purposive maximum-variation sampling strategy was adopted. The 12 eligible participants included several learners with similar proficiency levels, linguistic backgrounds, and learning goals. To obtain a broader range of interactional patterns, five participants were selected who represented different proficiency levels (A1-C1), first-language backgrounds, age groups, and learning purposes. The remaining seven participants were not included because their profiles substantially overlapped with those already represented in the final sample.

Table 1. Participant characteristics

Participant	Age	L1 / Background	ESL / EFL	CEFR Level	Learning Context	Purpose of AI Tutoring
P1	50	Chinese	EFL	A1	Family visit	Basic spoken communication and pronunciation
P2	43	Japanese	ESL	B2	Academic study	Academic interaction and spoken fluency
P3	32	Russian	EFL	B1	Workplace	Workplace communication and fluency
P4	20	Korean	EFL	B1	Test preparation	IELTS speaking practice
P5	36	Indian	ESL	C1	Sociocultural adaptation	Culture-bound expressions

The final sample (Table 1) therefore consisted of five participants who provided the greatest diversity in learner characteristics and AI interaction contexts.

The study received institutional ethical approval. All participants signed informed consent forms before data collection. Their names were replaced with codes (P1-P5) to protect privacy.

4.2. Data collection

Data were collected over one month. Each participant took part in several 10-15 minute speaking sessions with the same AI tutor, ELSA Speak (version 8.0.6), which is an APP developed by Vu Van and Dr. Xavier Anguera. This is a mobile-assisted language learning application originally developed for English pronunciation training. For this study, data were collected using ELSA's AI Conversations module, which provides scenario-based spoken interactions through a Large Language Model (LLM)-supported conversational interface. The module allows learners to engage in simulated dialogues on designated topics and receive automated feedback on their spoken performance. The AI conversation module employed a voice-plus-text interaction format. Learners communicated with the AI through spoken language, while transcriptions of both learner and AI utterances were simultaneously displayed on the screen.

In total, five topics were included in the study, and each participant was assigned one topic. The topics were related to the learners' real goals, such as daily communication, academic speaking, workplace English, test preparation, or cultural adaptation. Every participant was assigned to answer one topic that was closest to their life.

The conversations were audio-recorded in a quiet room using a digital recorder. In total, the dataset includes 75 minutes of learner-AI interaction. All recordings of AI tutor and learners were transcribed word by word following basic CA conventions, based on Jefferson transcription conventions [15].

4.3. Data analysis

The analysis focuses on three key concepts: emergence, indexicality, and recipient design.

First, to examine emergence, the study analyzes how each AI turn responds to the immediately previous turn. Special attention is paid to timing, relevance, and whether the AI adjusts its response to unexpected learner actions.

Second, to examine indexicality, the analysis looks at whether the AI uses pronouns or other referring expressions (such as it, they, this, that) to refer back to earlier talk, instead of simply repeating the same words.

Third, to examine recipient design, the study investigates whether the AI adapts its language, feedback, and level of detail to the learner's proficiency level and displayed needs.

Through turn-by-turn analysis, the study identifies patterns where naturalness is maintained or disrupted. This study focuses on selected excerpts of 3-6 conversational turns as analytic units for close CA analysis.

5. Results

This study draws on three interactional principles of natural conversation—emergence, indexicality, and recipient design. Conversation is understood as emergent in real time, constructed through variable turn formats and subject to ongoing modification via self-repair and action negotiation. It is deeply indexical, relying on prior talk, shared knowledge, and the immediate interactional environment to achieve coherence. Moreover, each turn is recipient-designed, tailored to the particular interlocutor and moment, requiring continuous monitoring of the unfolding interaction.

Together, these principles provide an interactional lens for evaluating the naturalness of AI tutoring dialogue, highlighting whether AI-generated turns display sensitivity to the context.

5.1. Descriptive data

Table 2 shows how many times each principle is violated in the five participants' conversations. Emergence is violated the most, with 19 times (55.9%). Recipient design is the second most violated, with 12 times (35.3%). Indexicality is violated the least, only 3 times (8.8%). Overall, violations of emergence happen much more often than the other two principles.

Table 2. Frequency of violations of three interactional principles

	Participants					Total	Proportion
	P1	P2	P3	P4	P5		
Emergence	5	3	5	4	2	19	55.9%
Indexicality	0	1	0	1	1	3	8.8%
Recipient Design	4	1	2	3	2	12	35.3%
Total	9	4	7	8	5		

5.2. Emergence

Violations of emergence are the most common pattern in this section, as the AI tutor often produces long, pre-structured turns that limit the learner's possible responses and hinder the co-construction of conversation.

Table 3. Excerpt of emergence violation

Turn	Participant	Utterance Content	Interactional Function	CA Notes
T1	AI	Teaching Chinese and English is a great choice. What part of teaching are you most excited about?	Positive assessment + Initiating question	Question projects learner response
T2	Learner	Um:: (.) I'm <mostly> excited about °teaching°	Answer	Off-target response; trouble source
T3	AI	That's great to hear! Teaching can be so rewarding…What kind of activities would you like to do in your classes?	Evaluation + Question	trouble unsolved; sequence expansion
T4	Learner	Uh:: I am:: (.) uh:: I'd like to de-design <some games>	Extended response with hesitation	incremental formulation
T5	AI	That sounds like a fun idea! Games can make learning more enjoyable…	Positive evaluation	No repair of earlier off-target response
T6	Learner	Thank you.	Show gratitude	Sequence closure

As shown in Table 3, the learner's response in T2 is a vague and mismatched response. Although the AI's question in T1 asks which part of teaching the learner is most excited about, the learner only repeats the general term "teaching" without giving any specific details. The hesitation marker "Um" also shows that the learner is unsure and still in the process of forming an answer. However, the AI tutor in T3 treats this response as fully adequate and gives an immediate positive evaluation ("That's great to hear!"). It does not notice the lack of specificity, nor does it try to ask a follow-up question or help the learner elaborate. Instead, it closes the sequence too early.

The problem here is not only that its turn is designed as if the next relevant action is already pre-determined, but that the AI's response is verbose. In other words, the AI does not leave space for the learner to shape the direction of the interaction. From the perspective of emergence, conversation should develop step by step, with meaning being co-constructed through ongoing interaction [11]. However, the AI stabilizes the interaction too early and fails to build meaning in real time. As a result, learners tend to produce shorter and less elaborated responses. Over time, this may reduce opportunities for active participation and restrict the co-construction of meaning in interaction, with fewer occasions for extended and more flexible language use.

A similar pattern can be found in Table 4. It presents an instance in which the AI responds to learner uncertainty by bypassing the trouble source and initiating a topic shift through evaluation and instruction.

Table 4. Another excerpt of emergence violation

Turn	Participant	Utterance Content	Interactional Function	CA Notes
T1	AI	Today let's talk about what you did yesterday. What did you do?	Initiating question	Question projects learner response
T2	Learner	Um:: I:: (.) don't know. Maybe it's- it's hard to say.	Answer	Response
T3	AI	That's great to hear! It sounds like you're making good progress so far. Now, in order to further improve your use of the past tense, we are going to do a short practice activity together.	Positive evaluation / Topic shift	Ignores learner trouble
T4	Learner	Alright.	Sequence closure	Accepts topic shift without pursuing the original question

Across the interactions of the five participants, violations of emergence are the most frequent issue. This suggests that the AI tutor tends to rely on pre-structured responses and fixed interactional patterns, rather than responding flexibly to the unfolding interaction. As a result, the conversation is less collaboratively constructed, and the learner has limited opportunities to shape the direction and develop meaning through interaction.

5.3. Indexicality

Violations of indexicality are relatively rare. The AI tutor occasionally fails to use context-dependent referential expressions, relying instead on repetition that weakens inter-turn cohesion.

Table 5. Excerpt of indexicality violation

Turn	Participant	Utterance Content	Interactional Function	CA Notes
T1	AI	Great. In a job interview in an international company, it is important to communicate your work experience in a clear way. You should explain your work experience in the job interview step by step.	Explanation / Instruction	Repetition of full noun phrase; absence of indexicals such as "it" or "there"
T2	Learner	(.) Okay.	Acknowledge	Minimal response; sequence continuation without elaboration
T3	AI	Now, let us practice a job interview in an international company. Please describe your work experience in the job interview.	Task instruction	Continued reliance on explicit noun phrases instead of indexical reference
T4	Learner	Uh:: should I- um:: talk about (.) my current job or (.) my °previous° job	Clarification request with hesitation	learner seeks referential clarification
T5	AI	In the job interview in an international company, you can start with your current job, then move on to your previous job. This job interview practice will HELP you improve your confidence in workplace communication. Would you like to continue this job interview practice?	Guidance + Closing question	Repetitive formulation; indexicals avoided

As shown in Table 5, the learner's response in T6 reveals a moment of interactional trouble. After the AI repeatedly frames the task as "a job interview in an international company" and instructs the learner to describe "your work experience in the job interview in an international company", the learner asks whether he should talk about his current or previous job. This turn shows difficulty in identifying the relevant focus of the task, and the clarification request is a relevant attempt to resolve referential ambiguity. However, the AI does not treat this as a problem of understanding. Instead, it repeats the same explicit noun phrases without narrowing the focus or clarifying which type of experience is relevant, leaving the learner's question only partially addressed.

This interaction illustrates a violation of indexicality. Although the learner signals confusion, the AI does not use context-dependent referential expressions (such as it, this, or that) to refer back to the job interview topic. Instead, it repeatedly uses the same full noun phrases ("a job interview in an international company", "your work experience in the job interview"), which creates referential ambiguity and weakens inter-turn cohesion. As a result, the learner's clarification request does not lead to any modification of instruction, and the interaction fails to progress toward shared understanding. This may reduce opportunities for negotiated meaning and limit effective learning.

A similar pattern can be found in Table 6. It presents a three-turn interaction in which an AI provides task instructions and follow-up guidance to a learner in a classroom discussion context, while consistently relying on repeated full noun phrases instead of indexical expressions.

Table 6. Another excerpt of indexicality violation

Turn	Participant	Utterance Content	Interactional Function	CA Notes
T1	AI	In a classroom discussion, a student should share her opinions clearly.	Instruction / Framing	Establishes task context
T2	Learner	Should- um:: should the student give personal examples >as well<.	Clarification request	Asks for task requirements
T3	AI	In a classroom discussion, a student can also support the opinion with personal examples to strengthen it.	Guidance / Expansion	Reformulates advice

Across the dataset, however, such violations are infrequent. As shown in Table 2, indexicality is violated only three times (8.8%), suggesting that the AI generally maintains referential cohesion between turns. This indicates that the AI tutor is relatively capable of tracking prior discourse at a surface level, even if occasional lapses in naturalness still occur.

5.4. Recipient design

Violations of recipient design are the second most frequent in this section, as the AI tutor does not consistently adjust its responses to learners' proficiency level or interactional needs, resulting in limited tailoring of interaction.

As shown in Table 7, the learner's response in T4 is an off-target answer. The learner shows difficulty understanding the key term "stand out" from the AI's previous question. This is indicated by hesitation markers ("uh", "um"), repetition, and the explicit uncertainty ("I'm not sure"). Instead of addressing the meaning of the question, the learner shifts to describing personal experience, suggesting a breakdown in comprehension at the semantic level. However, the AI does not treat this as a misunderstanding. It neither reformulates the question nor checks understanding. Instead, it focuses on a minor grammatical issue in the learner's previous sentence, leaving the original problem unresolved.

This pattern shows a lack of recipient design. Although the learner displays misunderstanding, the AI does not adjust its question to the learner's level or reformulate it in simpler terms. From a conversation-analytic perspective, recipient design involves tailoring talk to the interlocutor's displayed understanding [10]. Here, the AI focuses on correction and task progression rather than alignment. As a result, the interaction becomes less supportive, with limited space for clarification or elaboration, reducing opportunities for shared understanding. This pattern also violates the principle of emergence. The AI's insistence on its original formulation, despite clear signals of misunderstanding, treats the learner's response as a mere obstacle to be bypassed rather than a resource for redesigning the next turn. Consequently, the sequence lacks emergent sensitivity because it has no repair initiation and no attempt to respond based on the learner's displayed competence.

Table 7. Excerpt of recipient design violation

Turn	Participant	Utterance Content	Interactional Function	CA Notes
T1	AI	Alright! Let's dive into healthy cooking. What is one healthy dish that you absolutely LOVE to cook?	Initiating question	Question projects learner response
T2	Learner	I- I:: like to cook vegetable soup it is <easy and healthy>. I put °carrot° and the potato sometimes I- I:: eat egg.	Answer	Response to prior question
T3	AI	That sounds delicious! I really like how you mentioned that it's easy and healthy. What do you think makes your vegetable soup stand OUT compared to others?	Evaluation + Question	Assessment + sequence expansion
T4	Learner	Uh, stand out (0.4) uh:: I'm not sure it is hot. I cook it long time um:: my daughter likes it so, I- I:: think it is GOOD	Response with hesitation	Off-target response; trouble source
T5	AI	You probably meant to say "I cook it for a long time" instead of "I cook it long time". It's great that your daughter enjoys it! It really shows how special your soup is. What other ingredients do you think make it even better?	Correction + Follow-up question	Other-initiated form-focused repair; meaning-level trouble unsolved + inserted expansion

A similar pattern can be found in Table 8. It shows an interaction where the AI fails to adjust its explanation despite the learner expressing confusion and requesting clarification, continuing to use highly technical language.

Table 8. Another excerpt of recipient design violation

Turn	Participant	Utterance Content	Interactional Function	CA Notes
T1	AI	In thermodynamics, entropy is a state function defined as the logarithmic measure of the number of microstates consistent with a macrostate.	Instruction / Explanation	technical explanation
T2	Learner	Um:: Sorry I can't understand (.) can-can:: you (.) explain °more°?	Clarification request	Displays trouble understanding

Table 8. Continued

T3	AI	Entropy is the logarithmic measure of the number of microstates consistent with a macrostate in thermodynamics, and it is calculated using statistical mechanics principles.	Reformulated explanation (but not adjusted)	does not simplify despite request
T4	Learner	OK.	Sequence closure	Does not signal understanding

Across the interactions of the five participants, violations of recipient design are the second most frequent issue. This suggests that the AI tutor does not consistently adapt its responses to the learner's displayed understanding or interactional needs. As a result, the interaction is not sufficiently tailored to the recipient, and opportunities for scaffolding understanding are reduced, limiting the development of shared meaning through interaction.

6. Discussion

6.1. Interactional patterns

The findings of this study reveal that current AI tutoring systems still have important limitations in interactional competence. Although the AI tutor can usually produce grammatically correct and topically relevant responses, many interactions still sound unnatural at a deeper interactional level. The results show that the AI often fails to respond flexibly to the learner's ongoing talk, especially in terms of emergence and recipient design.

The most important finding is that violations of emergence are the most frequent interactional problem. In many cases, the AI tutor follows a pre-structured conversational script instead of adapting to the learner's actual response. Even when learners show hesitation, confusion, or misunderstanding, the AI often continues the conversation as if no problem exists. For example, in several excerpts, the learner gives unclear or off-target answers, but the AI still provides positive evaluation and quickly moves to the next question. This suggests that the AI mainly treats conversation as a sequence of fixed slots to complete, rather than an interaction that should develop collaboratively step by step. This finding supports previous research arguing that AI interaction is often script-driven instead of context-driven [14]. It also aligns with Conversation Analysis research which emphasizes that conversation is emergent and jointly constructed turn by turn [11]. However, this study further shows that even advanced AI tutoring systems still struggle with this interactional flexibility, despite their high level of grammatical fluency. In this sense, the findings partly challenge optimistic views that modern AI tutoring has already become close to human-like conversation [4]. The problem is no longer mainly grammatical accuracy, but interactional responsiveness.

One possible reason why emergence violations are the most common is that emergence requires real-time adaptation to unpredictable learner behavior. Human conversation constantly changes according to hesitation, misunderstanding, repair, emotion, and topic development. However, AI tutoring systems are still largely designed around instructional goals and pre-programmed conversational flows. As a result, the AI tends to prioritize task completion and topic progression over interactional negotiation. This makes emergence especially difficult for AI systems to achieve.

The study also finds that violations of recipient design are relatively frequent. In many interactions, the AI does not sufficiently adjust its language to the learner's displayed level of understanding. For example, when learners explicitly ask for clarification or show confusion, the AI often repeats the same explanation using similarly difficult language. In some cases, the AI focuses on correcting grammar instead of solving the learner's meaning-level problem. This suggests that the AI pays more attention to linguistic form than to the learner's interactional needs. This finding is consistent with previous CA studies showing that effective interaction depends on recipient-oriented design [10]. In human interaction, speakers usually simplify, reformulate, or explain differently when the listener shows trouble understanding. However, the AI tutor in this study often fails to make such adjustments. As a result, opportunities for scaffolding and negotiated meaning become limited.

Importantly, emergence violations and recipient design violations often occur together. When the AI ignores learner hesitation or misunderstanding, it not only disrupts the emergent development of conversation, but also fails to design the response appropriately for the learner. In other words, insufficient attention to learner understanding may directly lead to reduced interactional emergence. This suggests that the three interactional principles are closely connected rather than completely separate.

Compared with emergence and recipient design, violations of indexicality are much less common. In most cases, the AI can maintain basic topical continuity and refer back to previous talk. This suggests that current AI systems are relatively capable of tracking surface-level discourse connections. However, occasional overuse of repeated noun phrases still weakens interactional cohesion and makes the conversation sound less natural.

Overall, the findings suggest that current AI tutoring systems have achieved considerable progress in grammatical fluency and topical relevance, but still face important challenges in interactional naturalness. The AI can produce correct language, yet it cannot always participate naturally in the dynamic co-construction of conversation.

6.2. Theoretical and pedagogical implications

This study has important theoretical implications for Conversation Analysis (CA) and AI-assisted language learning research.

First, this study extends CA theory into the field of human-AI interaction. Traditional CA research mainly focuses on interaction between humans in classrooms, daily conversations, or institutional settings [10, 11]. However, with the increasing use of AI in education, communication between humans and AI systems has become more common. This study shows that CA concepts such as emergence, indexicality, and recipient design can also be applied to AI-mediated interaction. In this way, the study expands the contemporary significance of CA and demonstrates that interactional organization is still important even when one participant is a machine.

Second, this study contributes to second language acquisition research by shifting attention from technical performance to interactional quality. Previous studies on AI tutoring mainly focused on grammar accuracy, pronunciation scoring, vocabulary richness, and speech recognition [2]. However, this study shows that grammatically correct language does not necessarily create natural interaction. Even when the AI produces fluent and topic-related responses, communication problems may still occur if the AI cannot respond flexibly to learners' understanding, hesitation, or clarification requests. Therefore, this study provides a new interactional framework for evaluating AI tutoring systems.

The findings also have important pedagogical implications. The study suggests that current AI tutoring systems are still largely script-driven rather than context-driven. In many cases, the AI treats conversation as a fixed sequence of question-and-answer slots instead of a collaboratively developed interaction. As a result, learners may have fewer opportunities to negotiate meaning, clarify misunderstandings, and develop extended responses. This may limit the effectiveness of speaking practice.

Based on the findings, future AI tutoring systems could be improved in three main areas. First, AI systems should become more interaction-sensitive. Instead of simply following pre-designed scripts, AI tutors should react more flexibly to learners' unexpected responses and interactional trouble. Second, AI tutors should improve recipient design by simplifying explanations, reformulating difficult questions, and checking learner understanding when confusion appears. Third, AI systems should improve indexicality by using more context-based references to make conversations smoother and more connected.

For language teaching, the findings also suggest that AI tutoring should be used as a supplementary tool rather than a complete replacement for human interaction. AI can provide convenient speaking practice and reduce learner anxiety, but human teachers still play an important role in scaffolding communication, managing repair, and responding to learners' emotional and interactional needs. Therefore, combining AI tutoring with human classroom interaction may provide more effective speaking practice for ESL/EFL learners.

Overall, this study suggests that interactional naturalness is an important part of effective AI-assisted language learning. Improving AI interactional competence may help learners participate more actively, produce longer responses, and experience more meaningful communication.

6.3. Limitations and future research

This study has several limitations. First, the sample size was very small. Only five people took part in the study. With such a small group, the results may not represent people from different cultural backgrounds. For example, the way people learn languages can be very different depending on their culture, so five people are not enough to draw general conclusions. Second, the observation time was only 75 minutes. Therefore, the study could not show the long-term effects of using the method.

To improve future research, several changes are needed. First, future studies should use a much larger sample with more participants. Ideally, researchers should include people from different countries, cultures, and language backgrounds. This would make the results more generalizable and reliable. Second, future studies should last much longer. Instead of just 1.25 hours, researchers should observe learners over weeks or even months. The main focus should be on long-term oral language learning. For example, researchers could track how learners improve their speaking skills over time and whether the method works well in daily practice. By doing this, future studies can provide stronger evidence and more useful advice for real-life language learning.

7. Conclusion

This study examined interactional naturalness in AI-mediated English speaking practice through the lens of Conversation Analysis. Drawing on the concepts of emergence, indexicality, and recipient design, the study analyzed recorded interactions between ESL/EFL learners and an AI tutoring system to investigate how naturalness is disrupted in human-AI conversation.

The findings show that although the AI tutor was generally able to produce grammatically accurate and topically relevant responses, its interactional conduct frequently diverged from the organization of naturally occurring conversation. Violations of emergence were the most common, indicating that the AI often relied on pre-structured conversational patterns rather than adapting to learners' unfolding contributions. Violations of recipient design were also frequent, as the AI did not consistently adjust its responses to learners' displayed understanding or communicative needs. By contrast, violations of indexicality occurred relatively rarely, suggesting that the system was generally capable of maintaining basic referential continuity across turns. Taken together, these findings indicate that current AI tutoring systems have made considerable progress in linguistic performance but have not yet achieved interactional naturalness as understood from a conversation-analytic perspective.

The study contributes to the growing literature on AI-assisted language learning by shifting attention from technical accuracy to interactional quality. Rather than evaluating AI tutors solely in terms of grammar, pronunciation, or vocabulary, it highlights the importance of examining how AI systems participate in the moment-by-moment co-construction of talk. The findings suggest that future AI language tutors should be designed not only to generate correct language, but also to demonstrate greater sensitivity to learners' actions, understanding, and interactional needs. As AI continues to play a larger role in language education, interactional competence may become an essential criterion for evaluating the effectiveness of AI-mediated speaking practice.

References

[1] Hao, C., Xu, W., Halim, H. B. A., & Hao, M. (2025). AI chatbot-assisted vocabulary learning: Relationships with self-regulation, motivation, and performance among Chinese private college students. *Language Teaching Research*, 13621688251352595.

[2] Godwin-Jones, R. (2019). Emerging technologies: Artificial intelligence and language learning. *Language Learning & Technology, 23*(3), 1-10.

[3] Zou, B., Guan, X., Shao, Y., & Chen, P. (2023). Supporting speaking practice by social network-based interaction in artificial intelligence (AI)-assisted language learning. *Sustainability, 15*(4), 2872. https://doi.org/10.3390/su15042872

[4] Godwin-Jones, R. (2022). Partnering with AI: Intelligent writing assistance and conversational agents. *Language Learning & Technology, 26*(1), 6-25.

[5] Mondada, L. (2014). The local constitution of multimodal resources for social interaction. *Journal of Pragmatics, 65*, 137-156. https://doi.org/10.1016/j.pragma.2014.04.004

[6] Voss, C., & Waring, H. Z. (2023). Human-AI interaction and the question of interactional competence. *Applied Linguistics Review, 14*(2), 245–267.

[7] Seedhouse, P. (2004). *The interactional architecture of the language classroom: A conversation analysis perspective*. Blackwell.

[8] Kasper, G., & Wagner, J. (2011). A conversation-analytic approach to second language acquisition. In D. Atkinson (Ed.), *Alternative approaches to second language acquisition*, 117-142. Routledge.

[9] Kristiawan, D., Bashar, K., & Pradana, D. A. (2024). Artificial intelligence in English language learning: A systematic review of AI tools, applications, and pedagogical outcomes. *The Art of Teaching English as a Foreign Language (TATEFL), 5*(2), 207-218.

[10] Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*(4), 696-735.

[11] Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge University Press.

[12] Fryer, L., & Carpenter, R. (2006). Bots as language learning tools. *Language Learning & Technology, 10*(3), 8–14.

[13] Godwin-Jones, R. (2018). Chasing the butterfly effect: Informal language learning online as a complex system. *Language Learning & Technology, 22*(2), 8–27.

[14] Moore, R. J., & Arar, R. (2019). *Conversational UX design: A practitioner's guide to the natural conversation framework*. ACM Books.

[15] Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. H. Lerner (Ed.), *Conversation analysis: Studies from the first generation*, 13–31. John Benjamins.