HNUE JOURNAL OF SCIENCE
Educational Sciences 2026, Volume 71, Issue 2, pp. 3-16
This paper is available online at https://hnuejs.edu.vn/es
DOI: 10.18173/2354-1075.2026-0020

ANALYZING LEARNER INTERLANGUAGE DISCOURSE WITH AI: FACILITATING NOTICING IN THE EFL CLASSROOM

Anthony G. Ryan
Nanzan University, Japan
*Corresponding author: Anthony G. Ryan, e-mail: ryan@nanzan-u.ac.jp

Received January 12, 2026. Revised February 22, 2026. Accepted March 10, 2026.

Abstract. Schmidt's (1990, 2001) Noticing Hypothesis proposes that conscious awareness of gaps in output is necessary for second language development. This classroom-based action research study operationalizes that principle in spoken EFL by integrating structured discourse analysis with AI-supported feedback. Twenty-two Japanese university students recorded five-minute English conversations at the beginning and midpoint of a seven-week applied linguistics course. Using predefined analytic definitions, students manually analyzed transcripts for macro-structure, speech-act "interacts," exchange patterns, and embedded story sub-genres before comparing their analyses with AI-generated categorizations using standardized ChatGPT prompts. Quantitative summaries served as descriptive indicators of change, while structured reflections provided primary evidence of noticing. Across the instructional cycle, conversations showed a modest increase in total word production (6,951 → 7,287 words overall), increased follow-up information and clarification moves, and more frequent and more deliberately structured storytelling (7 of 22 → 20 of 22 students). Reflections indicated heightened awareness of pre-closings, turn balance, limited elaboration, and narrative sequencing. A post-course questionnaire (N = 22) further indicated that learners perceived AI as a useful analytic mediator, though requiring critical oversight. Rather than claiming causal effectiveness, the study documents how AI-mediated discourse analysis can make conversational architecture visible and support structured noticing within communicative classrooms. The findings highlight the pedagogical potential and practical limitations of integrating generative AI into discourse-focused EFL instruction.

Keywords: Noticing Hypothesis, interlanguage discourse, EFL speaking, discourse analysis, AI-mediated feedback, ChatGPT in education

1. Introduction

Conversation remains the most dynamic and interactive form of language use, yet in many EFL classrooms it is also the least systematically taught. Learners are often encouraged to "speak freely," but such open practice rarely leads to development unless students become aware of the internal architecture of talk. Schmidt (1990) emphasized that noticing is a necessary condition for converting input into intake; without it, repeated experience may not translate into learning. This study applies Schmidt's principle to conversational structure, proposing that explicit awareness of how talk is organized enables learners to notice gaps in their performance and consciously adjust their interaction.

In typical Japanese university communication classes, students are competent in sentence production but struggle to sustain interaction. Many conversations stop after a few question answer pairs or end without closure. Ventola (1979) showed that English casual conversation follows a predictable progression of phases: Greeting, Approach, Centering, and Leave-taking. Each stage serves a social purpose and provides cohesion. Ryan (2014) extended this framework in the Interlanguage Conversation (ILC) model, describing learner talk as a system of nested structures ranging from single speech acts to whole conversations. The ILC model is both descriptive and pedagogical it allows teachers to diagnose what is missing in a learner's performance and gives learners a vocabulary to discuss how conversations work.

At the same time, advances in AI now make discourse visible in ways previously impossible in ordinary classrooms. Automatic transcription and AI-based labeling reveal interactional features instantly, providing concrete data for reflection. When used carefully, these tools can make Schmidt's concept of noticing operational: learners can literally "see" how their conversations unfold. The present study explores whether guided noticing, informed by the ILC model and supported by AI, is associated with positive changes in conversational performance among Japanese university EFL learners.

2. Content

2.1. Literature Review

2.1.1. Noticing and Interactional Awareness

Schmidt's (1990) Noticing Hypothesis argued that learners must consciously attend to linguistic forms for them to become part of the developing system. Subsequent research (Ellis, 2012; Gass & Mackey, 2007) extended this idea beyond grammar, suggesting that noticing of pragmatic and discourse features contributes to interactional competence. Noticing occurs when learners compare their output with models or feedback, recognize discrepancies, and modify their behavior. However, the Noticing Hypothesis has been criticized for its limited explanatory power regarding complex discourse-level development, particularly in spontaneous interaction. In classroom contexts, tasks that draw attention to form–function relationships such as analyzing transcripts are effective in promoting noticing and uptake.

2.1.2. Conversation as a Genre

Systemic-functional linguistics (SFL) and genre theory view conversation not as random talk but as a recognizable social process with an underlying structure. Ventola (1979) proposed the *Generic Structure Potential* (GSP) of conversation, comprising phases that commonly appear in English casual talk: Greeting, Approach, Centering, and Leave-taking. These phases correspond to social moves such as establishing contact, exchanging information, sharing experiences, and closing interaction. Understanding these elements helps learners organize their talk more coherently. However, applying genre-based models of conversation in EFL classrooms remains challenging, as learners often lack awareness of how these structures operate in real-time interaction.

2.1.3. The Interlanguage Conversation (ILC) Model

Building on SFL and interlanguage analysis, Ryan (2014) formulated the Interlanguage Conversation (ILC) model to describe learner conversations as hierarchical systems of nested structures. The model identifies a *rank scale* of six elements, each subsumed within the next:

1. **Interact** – the smallest unit of meaning or speech act (e.g., ask, inform, comment, agree).
2. **Move** – one or more interacts forming a coherent communicative function.
3. **Exchange** – a sequence of moves (e.g., question–answer–follow-up) or a branch into a stage of a sub-genre such as an event in a recount or narrative.
4. **Transaction** – a complex of exchanges with optional story sub-genres forming a topical segment of discourse.

5. **Element of the GSP** – a macro-phase such as greeting, well-being exchange, story sub-genre, pre-closing, or leave-taking.
6. **Conversation** – the complete interaction comprising all the above levels.

This scale allows teachers and researchers to analyze learner talk from micro to macro perspectives and provides learners with a concrete framework for self-analysis. Ryan (2016) later applied the model to Japanese EFL data, showing that learners' conversations often lack centering stages or narrative sub-genres, resulting in short, fragmented talk. While the ILC model provides a detailed analytic framework, its pedagogical effectiveness in classroom settings has not been extensively examined, particularly in relation to learner noticing and interactional development. Teaching students to recognize these missing elements can significantly improve discourse coherence.

2.1.4. AI and Noticing in Language Learning

AI technologies have recently become tools for awareness raising in language classrooms. Chen, Wei, and Zou (2025) demonstrated that generative AI tools, when used to provide automated feedback, increase learners' motivation and reflection. However, AI's role should be mediational rather than evaluative it facilitates noticing but cannot replace teacher interpretation. Used in combination with explicit models like the ILC, AI can accelerate reflection by providing immediate visual data of interactional performance.

2.2. Methodology

2.2.1. Research Questions

The following research questions were formulated for this study:

RQ1: What descriptive changes in learners' conversational organization and interactional features were observed during an instructor-guided, AI-supported noticing cycle?

RQ2: How do learners perceive and evaluate the role of AI-assisted discourse analysis as a support for noticing within an instructor-guided analytic cycle?

These questions reflect the dual focus of the study: (a) documenting descriptive changes in conversational performance, and (b) examining learner perceptions of AI as a mediational tool in the analytic process.

2.2.2 Research Design

This study adopts a classroom-based action research design with an embedded pre–post descriptive comparison. The purpose was not to establish causal effectiveness through experimental control or comparison groups, but to document observable changes occurring within a structured instructional cycle in which noticing was explicitly scaffolded.

The instructional model integrated three mediational elements: (1) learner self-analysis, (2) instructor-guided modeling and discussion, and (3) AI-assisted categorization using standardized prompts. Quantitative indicators (e.g., word counts, frequencies of speech-act and exchange types, and presence of story structures) were used descriptively to illustrate shifts between BEFORE and AFTER conversations. Qualitative reflections and questionnaire responses were treated as central evidence of noticing and perceived development.

Because the study was embedded within a regularly scheduled university course, it was neither feasible nor pedagogically appropriate to introduce a separate control group receiving different instructional content under the same course designation. While full validation was not conducted, targeted checks indicated broad alignment between AI-generated categorizations and instructional definitions. Maintaining curricular consistency across cohorts was necessary for program coherence and equity. Consequently, the design prioritised within-group developmental comparison and pedagogical documentation over between-group experimental contrast.

The study therefore positions AI not as an independent evaluator, but as a mediational tool operating within the analytic cycle.

2.2.3 Participants and Context

The study was conducted in an elective applied linguistics course for English majors at a Japanese university.

Twenty-two undergraduate students (L1 Japanese) completed both the pre- and post-conversation tasks and submitted written reflection reports. These 22 participants constitute the primary dataset for quantitative and qualitative analysis.

A voluntary post-course questionnaire focusing on perceptions of AI-assisted analysis was completed by 12 students.

The course spanned seven weeks within a quarter system and focused on developing awareness of conversational organization through structured discourse analysis supported by AI tools.

2.2.4 Instructional Procedure

2.2.4.1 Baseline Conversation (Week 1)

Students were paired and instructed to conduct a natural, unstructured five-minute conversation in English (BEFORE conversation). Conversations were audio-recorded using smartphones and transcribed using Microsoft Word's speech-to-text function.

2.2.4.2 Guided Analytic Cycle (Weeks 2–5)

Students were introduced to layered discourse analysis based on the Interlanguage Conversation (ILC) model. Each week focused on a different analytic lens:

- **Macro-structure** (greeting, centering, pre-closing, closing)
- **Speech-act interacts** (e.g., Information Seek, Information Provide, Comment, Clarification)
- **Exchange moves** (e.g., Initiation, Response, Pay-Back Initiation, Follow-up Information, Feedback, Clearing-Up, Staller)
- **Embedded story sub-genres** (e.g., Habitual Recount, Recount, Anecdote, Narrative, Dilemma, Foretell)

Students first manually labeled their own transcripts using shared operational definitions. Only after completing their own analysis did they enter standardized prompts into ChatGPT to generate AI-assisted categorizations.

The manual → AI comparison sequence was intentional. AI functioned as a mediational contrastive tool designed to stimulate noticing rather than replace learner analysis.

2.2.4.3 Post-Instruction Conversation (Week 6)

Students recorded a second five-minute conversation under similar conditions (AFTER conversation). These conversations were transcribed and analyzed using the same multi-layer procedure.

2.2.4.4 Reflection and Questionnaire (Week 7)

Students submitted written reflection reports that quantitatively and qualitatively compared their BEFORE and AFTER conversations. These quantitative data sets (N = 22) and qualitative reflections (N = 22) form the primary qualitative dataset for examining noticing.

A voluntary post-course questionnaire (N = 12) gathered perceptions regarding the usefulness, limitations, and pedagogical value of AI-assisted discourse analysis.

2.2.5 Data Collection and Analysis

2.2.5.1 Quantitative Descriptive Indicators (N = 22)

Student-generated quantitative summaries included:

- Total and individual word counts

- Frequency of speech-act interacts
- Frequency of exchange moves
- Presence and type of embedded stories

Simple descriptive statistics (sums, averages, ranges) were used to compare BEFORE and AFTER conversations. These indicators are interpreted as descriptive evidence of change rather than as experimentally validated outcome measures.

2.2.5.2 Qualitative Reflections (N = 22)

Reflection reports were anonymized (S1, S2, etc.). Analysis focused on identifying recurring themes in what learners reported noticing (e.g., lack of pre-closings, limited follow-up moves, overuse of minimal responses, absence of storytelling) and how they described adjusting their interactional behavior. Qualitative noticing data are treated as evidence of learners' perceived changes within the instructional process.

2.2.5.3 Post-Course Questionnaire (N = 22)

A post-course questionnaire (N = 22) was administered in Japanese at the end of the instructional cycle. The instrument consisted of five thematic sections:

1. Awareness of Conversation Structure and Interaction
2. Changes in Confidence and Speaking Behavior
3. Noticing and Analytic Development
4. Use of AI Tools in the Course
5. Overall Evaluation of the Course

The questionnaire included Likert-scale items and open-ended comments across these domains.

For the purposes of the present paper, only five Likert-scale items from Section 4 (Items 18–22), together with one open-ended prompt (Comment C), are analyzed and reported. These items specifically address learners' perceptions of AI-assisted noticing, comparative analysis, critical evaluation of AI output, and intentions for future AI use. Findings from the remaining sections will be examined separately in a subsequent study.

Responses are reported descriptively and are used to contextualize learners' perceptions of AI within the instructor-guided analytic cycle.

2.2.5.4 Role and Validation of AI

All AI-assisted categorizations were generated using identical prompt templates and shared operational definitions supplied by the instructor. It was neither feasible nor pedagogically aligned with course aims for the instructor-researcher to validate every AI-generated label across all transcripts. Instead, reliability was supported through:

- Shared category definitions
- Uniform prompt structures
- Instructor modeling
- Targeted spot-checking of selected transcripts

AI functioned as a mediational analytic aid rather than as an independent research coder. Reported frequency patterns should therefore be interpreted as outcomes of a structured classroom analytic process rather than as fully independently coded corpus data.

2.3. Results

This section presents findings organized according to the two research questions. Sections 4.1–4.5 address RQ1 by reporting descriptive changes observed between BEFORE and AFTER conversations following the instructor-guided, AI-supported noticing cycle. These findings draw on student-generated quantitative summaries (N = 22) and reflective analyses completed as part of the course (N = 22).

Section 4.6 addresses RQ2, examining learner perceptions of instructor-guided and AI-assisted discourse analysis. These findings draw on reflections (N = 22) and the post-course questionnaire (N = 12).

Across instruments, the emphasis remains on the noticing process and how structured analytic comparison shaped learners' awareness of conversational organization, interactional balance, and narrative development.

2.3.1 Macro-Structure of the Conversation

Analysis indicated increased macro-structural completeness in the AFTER conversations. BEFORE recordings often omitted identifiable genre stages and moved abruptly from greeting to short question–answer sequences, with many dialogues ending without a recognizable pre-closing. AFTER conversations more frequently included a staged progression: brief small talk, one or two sustained topics (often supported by a personal narrative), and an explicit winding down leading to leave-taking. One learner captured this shift succinctly: "In the BEFORE conversation, we did not have Pre-closing. In the AFTER conversation, we had time for Pre-closing. Thanks to Pre-closing, our AFTER conversation was more natural than BEFORE." (S1)

Students consistently reported that including short pre-closings (e.g., "Anyway, it was nice talking with you") made endings smoother and more socially appropriate.

A second macro-structural change concerned topic development. Reflections suggest that BEFORE conversations involved more rapid topic cycling, whereas AFTER conversations often centered around fewer topics developed through storytelling and follow-up questions. As S2 noted:"In BEFORE conversation, the topic did not last and we changed topic three times in five minutes. In AFTER conversation, we had three stories… our conversation became more consistent" (S2).

By course end, students demonstrated explicit genre awareness. Reflections frequently listed conversation stages (e.g., "Greeting, Well-being, Initial Topic, Story, Redirection, Pre-Closing, Closing") and described intentionally including them in the second recording. Several also commented that the AFTER conversations "flowed more easily" (S4), suggesting that structural awareness contributed to perceived naturalness.

Overall, AFTER conversations more frequently exhibited (a) identifiable openings and closings, (b) sustained centering around one or two substantive topics, and (c) deliberate pre-closings preceding leave-taking. These descriptive patterns are consistent with uptake of the genre model introduced during the instructional cycle.

2.3.2 Conversation Length and Word Counts

Under equivalent time limits (approximately five minutes per recording), the AFTER conversations showed a modest increase in overall language production. Total words across all pairs rose from 6,951 (BEFORE) to 7,287 (AFTER). The mean words per student increased from 315 to 331, and the minimum individual contribution rose from 160 to 182 words (see Table 1).

Table 1. Word count metrics (5-minute conversations)

ITEM	BEFORE	AFTER
Total words	6,951	7,287
Mean per student (N = 22)	315	331
Range (min-max)	160-655	182-681

Although the aggregate increase at the class level was modest, individual variation was notable. For example, one student (S5) increased from 266 to 380 words (+114 words), while another (S8) increased from 401 to 681 words (+280 words). In their reflections, these learners attributed the increase not to filler or repetition, but to greater elaboration and sustained topic development. As S5 explained, "I tried to add more information and reasons instead of finishing with short answers."

Importantly, increased word count did not necessarily reflect dominance of the conversational floor. Several students reported consciously adjusting participation patterns. One learner, S6, whose output rose only slightly, stated that she had spoken disproportionately in the first conversation and deliberately moderated her contribution in the second, aiming for more balanced interaction: "In the BEFORE conversation, I talked too much. In the AFTER conversation, I tried to make the amount more equal."

Conversely, initially reticent students reported making deliberate efforts to initiate topics or extend responses. For example, S7, whose word count rose from 182 to 276 words, described shifting from passive responding to more active initiation: "In the BEFORE talk, I was more passive. In the AFTER talk, I made an effort to speak more and ask questions."

Students frequently linked increased output to structural awareness gained through the analytic cycle. Rather than simply "speaking more," they described asking follow-up questions, adding contextual detail, and incorporating short narrative segments. In this respect, word-count gains appear to reflect expanded interactional work (e.g., follow-up information and re-initiations) rather than simple verbosity.

Taken together, the quantitative data indicate a modest increase in production, while reflections suggest learners perceived greater engagement and balance. Within the scope of this classroom comparison, the findings point toward small but observable modifications in participation behavior.

2.3.3 Use of Speech "Interacts" (Speech-Act Functions)

Students' coded "interacts" provide a more detailed picture of how conversational functions shifted between the two recordings (see Table 2). Unlike the word-count data, these patterns show a mixture of increases and decreases across categories rather than a uniform directional change.

Table 2. Interacts

ITEM	BEFORE	AFTER
Information Seek	134	133
Information Provide	397	423
Clarification Seek	31	39
Clarification Provide	19	19
Comment	48	36
Feedback (backchannel)	186	193
Greeting	18	19
Opinion Seek	41	12
Opinion Provide	56	51
Leave-Taking	8	14

Information Seek remained essentially stable (134 → 133), suggesting that students continued to ask questions at similar rates. By contrast, Information Provide increased moderately (397 → 423), indicating slightly greater elaboration in responses. In their reflections, some learners described intentionally expanding answers rather than replying minimally. For example, S5 noted that she tried to "add more information and reasons instead of finishing with short answers," which aligns with the rise in informational provision.

Clarification Seek increased from 31 to 39 instances, while Clarification Provide remained stable (19 → 19). Although the numerical increase is modest, several students interpreted this shift as evidence of greater willingness to negotiate meaning. As S9 wrote, clarification moves reflected "making sure what my partner said." This suggests that learners may have become somewhat more attentive to repair and mutual understanding.

Not all categories increased. Comment moves declined (48 → 36), and both Opinion Seek and Opinion Provide decreased (41 → 12; 56 → 51, respectively). Reflective comments suggest that in the second conversation learners relied more on storytelling and extended information exchange than on explicit opinion prompting. In the initial "getting-to-know-you" task, questions about likes and dislikes naturally generated Opinion Seek moves. By contrast, as familiarity and rapport developed, evaluative stances appeared more frequently embedded within Information Provide turns rather than structured as separate opinion questions and responses. This pattern suggests a shift from discrete preference exchanges toward more integrated, narrative-based interaction.

Feedback (backchannel) moves rose slightly (186 → 193), remaining a prominent feature of interaction in both rounds. Some students reported becoming more aware of when to insert feedback and when to allow longer turns. S10 reflected, "I respond way too much… I don't have to do it by cutting [in] while someone's talking," suggesting that metapragmatic awareness of backchannel timing increased even if overall frequency did not decrease.

Leave-Taking moves increased from 8 to 14, consistent with the macro-structural findings reported in Section 4.1 that AFTER conversations more frequently included explicit closings.

Overall, the interact data indicate selective rather than uniform change across categories, including modest increases in informational provision and clarification seeking, relatively stable question frequency, and reduced reliance on explicit opinion prompts; when considered alongside learner reflections, these patterns suggest targeted adjustments in elaboration, repair practices, and structural closure within the instructional cycle.

2.3.4 Exchange Structure (Initiation–Response Patterns)

Exchange-move analysis provides further insight into how conversational sequences were organized across the two recordings (see Table 3).

Table 3. Exchange moves

ITEM	BEFORE	AFTER
Initiation (I)	82	91
Response (R)	214	186
Follow-up Information (FI)	217	226
Clarification (C)	50	58
Pay-Back Initiation (PBI)	59	106

Initiations increased modestly (82 → 91), suggesting slightly greater topic introduction or question-posing in the AFTER conversations. Follow-up Information (FI) also rose (217 → 226), indicating somewhat more elaboration following initial responses.

The most notable numerical shift occurred in Pay-Back Initiations (PBI), which increased from 59 to 106 instances. PBI moves questions that return the topic to the interlocutor (e.g., "How about you?") represent a mechanism for sustaining reciprocal exchange. Several students commented explicitly on attempting to maintain conversational momentum through such moves. For example, S14 wrote, "I increased Initiations (I), Response (R), and Pay-Back Initiations (PBI)… I gave more topics and some questions related to these topics."

Clarification sequences (C) increased moderately (50 → 58), aligning with the clarification-seek findings reported in Section 4.3. As S9 noted, clarification reflected "making sure what my partner said," suggesting greater attention to mutual understanding.

Responses (R), by contrast, decreased (214 → 186). One possible interpretation is that some simple response tokens were replaced by longer Follow-up Information or PBI sequences, resulting in more extended multi-move exchanges rather than short, closed adjacency pairs.

Students' reflections cautiously support this interpretation. S15 contrasted the first conversation's "long monologues plus short acknowledgments" with what she perceived as more evenly paced turn-taking in the second conversation. Other learners described deliberately avoiding rapid "question–answer–next question" patterns in favor of adding explanation or inviting reciprocal sharing.

Taken together, the exchange data suggest selective adjustments rather than uniform expansion across categories. The increase in PBIs and modest growth in FI moves point toward somewhat more sustained reciprocal development, while the reduction in simple Response moves may reflect redistribution across more complex exchange sequences. Similarly to word counts and interacts, these patterns are consistent with incremental changes in how students managed turn-taking and topic continuation during the instructor-assisted, AI-supported analytic cycle.

2.3.5 Storytelling in Conversation

A central instructional focus was the development of embedded storytelling during the centering phase of conversation. The presence of identifiable story segments increased markedly between recordings. In the BEFORE conversations, 7 of 22 students (32%) produced at least one identifiable story. In the AFTER conversations, this figure rose **to** 20 of 22 students (91%)**.**

Beyond frequency, there were observable differences in story type and organization. In the AFTER conversations, students produced a range of sub-genres, including Recount (10 instances), Anecdote (4), Habitual Recount (3), Foretell (4), Narrative (1), and Dilemma (2). Several conversations contained more than one story segment.

Additionally, learners increasingly applied narrative sequencing (e.g., setting → event → result). S16: "In the BEFORE conversation, I didn't tell the story in order… In the AFTER conversation, I could tell the story in the right order." For some, simply adding stories represented a breakthrough: S17 moved from no stories to talking about summer vacation and family.

Other students emphasized depth and engagement with the interlocutor rather than mere inclusion. S2 reported that their AFTER conversation included "three stories," which they felt made the interaction "more consistent." S19 similarly described "going deep into each story" rather than shifting rapidly between topics. Students also linked storytelling to engagement. S18 noted that adding stories allowed them to "deepen our story by asking about reasons or personal preferences." In some cases, reciprocal storytelling emerged, with one narrative prompting another, creating extended centering sequences rather than isolated question–answer exchanges.

Several learners explicitly referred to temporal expansion (past–present–future framing) as a strategy for enriching conversation. S21 described deliberately incorporating past and future references to avoid superficial present-tense exchanges. This awareness reflects the instructional framing used during the course, in which key story sub-genres were organized temporally: Recount, Anecdote, and Narrative were introduced as past-event genres**;** Habitual Recount and Dilemma as present-oriented genres; and Foretell as future-oriented events. Students were encouraged to use this temporal framework strategically to diversify centering sequences and extend conversational development across time horizons.

Taken together, the data indicate a marked increase in the frequency of storytelling and suggest qualitative adjustments in narrative organization. Within the context of the instructional cycle, these patterns are consistent with greater awareness of storytelling as a conversational resource and its role in sustaining centering phases.

2.3.6 AI Performance and Assistance in Analysis (RQ2)

A secondary aim of the study (RQ2) was to examine how learners perceived the role of instructor-guided and AI-assisted discourse analysis within the instructional cycle. Across reflections (N = 22) and the post-course questionnaire (N = 22), students described AI tools as useful analytic supports, while also identifying technical and interpretive limitations.

2.3.6.1 Transcription Phase: Benefits and Constraints

Students initially used Microsoft Word's speech-to-text function to generate transcripts of their recorded conversations. Many reported that the speed of automated transcription allowed them to "see" their conversations in written form more quickly than manual transcription alone. One student wrote, "It was helpful to see our conversation immediately in text because I could notice what we actually said" (S3). Another commented that transcription "made the structure visible" and allowed her to identify missing pre-closings (S11).

However, limitations were frequently mentioned. The program did not distinguish between speakers, producing a continuous stream of text that required manual segmentation. Several learners also noted transcription inaccuracies linked to pronunciation. In cases where Katakana-influenced English forms were used, the software sometimes appeared to generate its "best guess" based on contextual inference rather than phonological accuracy. One student observed that "some words were not correct because of our pronunciation," requiring manual correction (S8). Another commented that "it sometimes changed what we said into something similar but different" (Q4).

Students also reported that automated transcription did not adequately capture interactional features such as hesitations, false starts, or Japanese backchannels. As a result, many pairs supplemented the process with cellphone recordings and manual turn segmentation in Excel. Although this added time, several reflections indicated that the process of correcting transcripts itself supported noticing. As S15 explained, "When I checked the transcript carefully, I realized how many short answers I gave."

2.3.6.2 ChatGPT as an AI-Mediated Analytic Tool

After completing manual coding, students used standardized prompts to generate AI-assisted categorizations via ChatGPT. Questionnaire results provide quantitative insight into learner perceptions.

For Item 18 ("AI tools helped me notice features I missed on my own"), 18 of 22 respondents (81.8%; M = 4.09) selected Agree or Strongly Agree. Similarly, for Item 19 ("Comparing my own analysis with AI output deepened my understanding"), 18 of 22 (81.8%; M = 4.18) expressed agreement. These results suggest that most learners perceived the AI-supported comparison process as enhancing noticing and conceptual understanding.

Item 20 ("I became more critical rather than trusting AI blindly") showed 16 of 22 agreements (72.7%; M = 3.95). Item 21 ("I now understand better how AI tools can and cannot support language learning") yielded 15 of 22 agreements (68.2%; M = 4.00), though 27.3% selected Neutral. These findings indicate that while most learners perceived increased critical awareness, variability remained in how confidently students evaluated AI's limitations.

Finally, Item 22 ("I would like to use AI tools again") received the strongest endorsement, with 19 of 22 respondents (86.4%; M = 4.18) selecting Agree or Strongly Agree. Despite recognizing limitations, the majority expressed willingness to continue using AI as a learning support.

Open-ended responses (Comment C; 22 responses) reinforce these patterns. Recurring themes included:

- Increased awareness of subtle grammatical errors (e.g., prepositions, tense use)
- Recognition of conversational habits and clarity of expression
- Efficiency gains in identifying and revising errors
- Explicit acknowledgment that AI output was "not always correct" and required verification
- Discovery of alternative vocabulary and phrasing

Taken together, these responses suggest that learners perceived AI not as an infallible evaluator but as a comparative analytic aid within an instructor-guided noticing cycle.

2.3.6.3 Perceived Pedagogical Value

Across reflection and survey data, learners often linked AI-supported analysis to increased awareness rather than automatic improvement. Several described the experience as "seeing our conversation from outside" (S14) or "realizing my habits" (S6). The combination of high agreement for noticing (Items 18–19) and continued critical stance (Items 20–21) suggests that AI functioned as a mediational support rather than a replacement for instructor guidance.

Although responses were generally positive, interpretation must remain cautious. The sample reflects a single cohort (N = 22) within one instructional context. The survey therefore documents perceived impact rather than generalizable attitudinal trends.Within the instructor-guided analytic cycle, AI appears to have functioned less as an evaluator and more as a catalyst for structured comparison and reflective noticing.

2.3.7 Summary of results

In regard to **RQ1,** descriptive comparisons between BEFORE and AFTER conversations indicate selective structural and interactional adjustments rather than uniform growth. These include more consistent inclusion of openings and closings, modest increases in overall language production, expanded use of Pay-Back Initiations, and a marked increase in identifiable storytelling. Macro-structural analysis suggests clearer centering sequences and more deliberate pre-closings, while interact and exchange patterns reflect incremental shifts in elaboration and clarification practices. Reflective reports indicate that learners noticed and intentionally adjusted specific conversational features.

In relation to **RQ2,** reflection data (N = 22) and questionnaire responses (N = 12) suggest that students generally perceived AI-assisted analysis as a useful comparative tool, while recognizing technical and interpretive limitations. AI tools contributed to the visibility of conversational structure and facilitated structured comparison with self-analysis, though instructor mediation remained central.

Within the constraints of a classroom-based pre–post design and a limited perception sample, the combined findings are consistent with heightened metapragmatic awareness across the instructional cycle.

2.4. Discussion

The findings of this classroom-based study suggest that structured attention to conversational organization, combined with instructor-guided and AI-supported analytic comparison, may be associated with observable shifts in learners' interactional behavior. Rather than establishing causal effectiveness, the present study documents patterns of change within a pedagogically designed instructional cycle. This section situates those patterns in relation to Schmidt's Noticing Hypothesis, Ventola's model of conversational structure, and sociocultural perspectives on mediated learning. It also considers the role of AI as a comparative analytic tool within instructor-supported discourse awareness.

2.4.1 Noticing Interactional Structures and Schmidt's Hypothesis

The descriptive patterns observed in Sections 2.3.2 – 2.3.5 are broadly consistent with Schmidt's

(1990) claim that conscious noticing is a necessary condition for changes in language use. Although Schmidt focused primarily on morphosyntactic forms, the present findings suggest that discourse-level and pragmatic features may also be amenable to noticing-based adjustment.

Students' reflections frequently documented awareness of specific interactional behaviors (e.g., interrupting, failing to use pre-closings, limited clarification). In many cases, learners described intentional adjustments in the second recording. While the study design does not permit causal attribution, the alignment between reflective noticing and descriptive shifts in exchange moves, storytelling frequency, and macro-structure suggests that conscious comparison may have contributed to incremental behavioral modification.

Noticing in this context was not solely individual. It was scaffolded through instructor modeling, peer discussion, and AI-generated categorizations. This layered mediation resonates with sociocultural accounts of learning (e.g., Swain, 2000), in which awareness emerges through dialogic engagement. Although students often framed insights individually ("I noticed…"), these insights developed within a collaborative analytic environment.

Importantly, not all categories showed uniform increase. Instead, the data suggest selective adjustment (e.g., increased PBI use, expanded storytelling, more consistent closings). This pattern aligns more closely with strategic adaptation than with generalized proficiency growth. Within the limits of a pre–post classroom comparison, the findings are therefore best interpreted as evidence of heightened metapragmatic awareness rather than definitive gains in communicative competence.

2.4.2 Ventola's GSP and the Pedagogical Operationalization of Centering

Ventola's (1979) generic structure potential (GSP) of casual conversation describes macro-phases such as openings, centering, and closings, but was developed primarily from native-speaker corpora. Learner interactions, particularly between unfamiliar partners, often deviate from this pattern. The present instructional cycle operationalized these macro-phases explicitly, making conversational structure visible and teachable.

The most marked descriptive shift occurred in the centering phase, particularly through the inclusion of identifiable story genres. By framing sub-genres temporally Recount, Anecdote, and Narrative as past-oriented; Habitual Recount and Dilemma as present-oriented; Foretell as future-oriented students were provided with tangible strategies for sustaining topic development. Reflective comments indicate that learners consciously applied this temporal framework when constructing their second conversations.

Rather than implying that learners achieved Ventola-consistent performance, the data suggest increased structural awareness. AFTER conversations more frequently included explicit pre-closings and sustained centering segments. This pattern is consistent with the claim that conversational genre knowledge can be scaffolded pedagogically.

The CHAT/CHUNK distinction functioned as accessible metalanguage, helping students conceptualize turn distribution and development. While metaphorical, this framing appears to have supported reflective evaluation of interactional balance. In this sense, the ILC model may be viewed not as replacing descriptive frameworks such as Ventola's, but as providing a pedagogical interface through which learners engage with them.

2.4.3 AI as a Mediational Tool in Discourse Awareness

In this study, AI tools did not function as autonomous evaluators but as comparative analytic aids within instructor-mediated instruction. ChatGPT was used to apply predefined coding categories after students had completed their own analyses. This sequencing ensured that AI did not replace human judgment but instead provided a second interpretive lens.

Students' responses suggest that AI's value lay less in accuracy than in visibility and speed. Rapid categorization enabled learners to compare their own coding with an alternative output, prompting discussion when discrepancies occurred. In several cases, disagreements required students to justify their analytic decisions using shared definitions, potentially deepening conceptual understanding.

Technical limitations such as transcription inaccuracies linked to pronunciation variation or inconsistent grammatical repairs across runs reinforced the need for human oversight. Rather than undermining learning, these inconsistencies sometimes stimulated critical reflection about prompt specificity and category boundaries.

From a sociocultural perspective, AI in this context functioned as a mediational artifact rather than as a "More Knowledgeable Other." The instructor remained central in modeling prompt precision, validating interpretations, and contextualizing discrepancies. Thus, AI's contribution appears to lie in facilitating iterative noticing rather than in delivering authoritative evaluation.

2.4.4 Pedagogical Implications

Within communicative curricula, discourse structure is often assumed to develop implicitly through practice. The present findings suggest that structured analytic comparison may support learners in attending to macro- and micro-interactional features that otherwise remain unnoticed.

Integrating short record-and-analyze cycles whether AI-supported or manually conducted may provide opportunities for metapragmatic awareness. Providing learners with explicit meta-vocabulary (e.g., centering, PBI, clarification) appears to support reflective evaluation of interaction.

However, these implications should be interpreted cautiously. The study was conducted within a single elective course with motivated participants. Replication in different institutional contexts and proficiency levels would be necessary before generalization.

2.4.5 Limitations and Future Directions

Several limitations constrain interpretation. The design was classroom-based and lacked a comparison group; therefore, changes cannot be attributed exclusively to the analytic intervention. Observed shifts may partially reflect increased familiarity with partners or task repetition effects.

The analytic focus emphasized structural and pragmatic features rather than prosodic or temporal dimensions (e.g., pause length, intonation contour), which also influence perceived fluency. Future research could integrate multimodal analysis to provide a more comprehensive account of conversational development.

Quantitative indicators such as word counts and move frequencies provide only coarse-grained evidence of change. Independent expert coding or inter-rater reliability measures would strengthen future studies. Additionally, longitudinal follow-up could assess whether heightened awareness transfers to spontaneous interaction beyond structured classroom tasks.

Finally, while AI tools showed potential as mediational supports, their effectiveness depends on instructor scaffolding and prompt precision. The findings do not suggest that AI replaces human guidance; rather, they indicate that AI may serve as a supplementary analytic resource within carefully designed pedagogical frameworks.

3. Conclusions

This classroom-based study examined how structured attention to conversational organization, supported by instructor guidance and AI-assisted comparison, may be associated with shifts in learners' interactional behavior. By integrating the ILC model with Schmidt's Noticing Hypothesis and Ventola's genre framework, students engaged in systematic reflection on the macro- and micro-structure of their own talk.

Descriptive comparisons indicated selective changes in areas such as storytelling frequency, inclusion of pre-closings, reciprocal exchange moves, and modest increases in language production. Reflective data suggest that learners became more aware of turn balance, clarification practices, and topic development. Within the limits of a pre–post classroom design, these findings are consistent with heightened metapragmatic awareness rather than definitive gains in proficiency.

AI tools functioned as mediational supports, making conversational structures visible and comparable, while instructor scaffolding remained central. The results suggest that explicit attention to conversational genre and exchange patterns can be incorporated into communicative teaching through structured analytic cycles, with or without technological support.

Future research should examine durability of these shifts and transfer to spontaneous interaction beyond structured classroom tasks.

REFERENCES

[1] Chen, Z., Wei, W., & Zou, D, (2025). *Generative AI technology and language learning: Global language learners' responses to ChatGPT videos in social media*. Interactive Learning Environments. Advance online publication.

[2] Ellis, N. C, (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics*, 32, 17–44.

[3] Gass, S. M., & Mackey, A, (2007). *Input, interaction, and output in second language acquisition*. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition* (pp. 175–199). Routledge.

[4] Hall, E. T, (1976). *Beyond culture*. Anchor Books.

[5] Johns, T, (1991). Should you be persuaded: Two samples of data-driven learning materials. *ELR Journal,* 4, 1-16.

[6] Ryan, A. G, (2014). *Analysing and assessing the interlanguage conversation of Japanese EFL learners (Unpublished doctoral dissertation).* Monash University.

[7] Ryan, A, (2016). Analysis of an interlanguage English conversation: Macrostructure and internal genres. Gaikokugo Kenkyu (*Journal of Foreign Language Studies*), 49, 127–165.

[8] Schmidt, R. W, (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158.

[9] Schmidt, R, (2001). *Attention*. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge University Press.

[10] Sinclair, J., & Coulthard, M, (1975). *Towards an analysis of discourse: The English used by teachers and pupils*. Oxford University Press.

[11] Swain, M, (2000). *The output hypothesis and beyond: Mediating acquisition through collaborative dialogue*. In J. P. Lantolf, (Ed.), *Sociocultural theory and second language learning,* (pp. 97-114). Oxford University Press.

[12] Ventola, E, (1979). The structure of casual conversation in English. *Journal of Pragmatics*, 3(3–4), 267–298.

[13] Vygotsky, L. S, (1930/1978). *Mind in society: The development of higher psychological processes.* Harvard University Press.