

Exploring Active Learning in Papua: A Corpus and Conversation Analysis of EFL Classroom Discourse

¹*Rizdika Mardiana, ²Nurrahma Restia Fatkhiyati, ³Mukhamdanah, ¹Bambang Nur Alamsyah Lubis, ¹Sudarya Permana

¹Master Program of English Language Education, Faculty of Language and Arts, Universitas Negeri Jakarta, Jl. Rawamangun Muka, RT.11/RW.14, Rawamangun, East Jakarta, Indonesia

²English Literature Study Program, Faculty of Language and Arts, Universitas Negeri Jakarta, Jl. Rawamangun Muka, RT.11/RW.14, Rawamangun, East Jakarta, Indonesia

³Pusat Riset Bahasa, Sastra, dan Komunitas, Badan Riset dan Inovasi Nasional, Jakarta, Indonesia

*Corresponding Author e-mail: rizdikamardiana@unj.ac.id

Received: September 2025; Revised: December 2025; Accepted: April 2026; Published: June 2026

Abstract

This study investigates how active learning is conducted in EFL classroom discourse in Papua by combining Corpus Linguistics and Conversation Analysis (CA) framework. It specifically examines how lexical patterns and interactional organization jointly shape students' communicative development from five datasets. A mini spoken corpus was constructed from the video-recorded lessons and analyzed for lexical frequency, collocations, and formulaic language. In terms of CA, turn-taking, repair practices, and teacher scaffolding of the classroom conversations were identified. The findings reveal systematic variation across task types. Genre-focused lessons are dominated by metalanguage and definitional bundles, reinforcing classification routines within teacher-controlled IRF sequences. While performance-based and narrative tasks generate longer student turns supported by formulaic expressions and temporal markers. This facilitates discourse expansion and storytelling competence. Grammar-driven descriptive tasks prioritize structural templates, promoting accuracy but limiting lexical diversity. Across datasets, formulaic language functions as an interactional scaffold that enables participation despite limited grammatical complexity. However, repair practices operate as learning opportunities. The study demonstrates that communicative growth in active EFL classrooms is strongly shaped by task design and pedagogical orientation. Narrative and pragmatic performance tasks foster greater interactional competence than taxonomy-driven instruction. As for the methodology, the combination of corpus linguistics and CA framework provides a comprehensive account of how lexical recycling, scaffolding, and sequential organization provide learning opportunities in real-time classroom interaction. The findings offer pedagogical implications for designing tasks that expand learner autonomy and promote sustained student talk in multilingual and under-researched contexts such as Papua.

Keywords: Active learning; Conversation analysis; Corpus linguistics; Classroom discourse

How to Cite: Mardiana, R., Fatkhiyati, N.R., Mukhamdanah, Lubis, B. N. A. & Permana, S. (2026). Exploring Active Learning in Papua: A Corpus and Conversation Analysis of EFL Classroom Discourse. *Journal of Language and Literature Studies*, 6(2), 533-546. doi: <https://doi.org/10.36312/jolls.v6i2.3460>



<https://doi.org/10.36312/jolls.v6i2.3460>

Copyright © 2026, Mardiana et al.

This is an open-access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) License.



INTRODUCTION

Limited exposure to English, irregular access to locally relevant materials, and a persistent tendency toward teacher-faced communication contribute to a complex educational ecosystem in EFL classrooms, especially in Papua. Although current curriculum directions in Indonesia promote active learning and communicative competence, classroom talk in many settings still relies heavily on structured questioning, short student responses, and teacher confirmation or correction. Based on the previous

studies on Indonesian classroom discourse, the dominant organizational pattern of students-teacher interaction is Initiation–Response–Feedback/Evaluation (IRF/IRE), even in lessons designed to be interactive (Morton & Llinares, 2021; Wati et al., 2025). It shows that greater learner autonomy or extended student production is not automatically the result of increasing “activity”.

Within genre-based pedagogy, some previous studies have investigated the dominance of definitional talk and metalanguage recycling. These studies are focused on SFL-informed genre instruction demonstrating the classification of classroom discourse frequently conducted such as naming genres, identifying generic structures, and defining social functions, as part of routinized pedagogical cycles (Zhang & Liu, 2021). In interaction, studies on classroom metalanguage further show that definitional sequences outline what counts as acceptable knowledge and acceptable responses (Tang, 2021). In Indonesian contexts, genre-based teaching often emphasizes explicit metalanguage, which may strengthen conceptual understanding. However, this teaching sometimes limits opportunities for discourse expansion (Nurlaelawati et al., 2020).

Despite growing scholarship on classroom discourse, two methodological limitations remain. Corpus-based studies effectively identify lexical frequency, collocations, and recurrent bundles, yet they often do not explain how these patterns function sequentially within real-time interaction. In contrast, Conversation Analysis (CA) provides detailed accounts of turn-taking, repair, and pedagogical action, but it rarely integrates quantitative lexical evidence across task types. Classroom-based CA studies have consistently shown that learning opportunities emerge through the sequential organization of interaction, including teacher allocation of turns and scaffolding practices (Kunitz et al., 2021; Zhou & Li, 2023). At the same time, corpus studies demonstrate that repeated lexical bundle structure pedagogical explanations and participation frameworks in classroom discourse (Yoon & Römer, 2024). However, these two perspectives are seldom combined within a single empirical study.

This study addresses that gap by integrating corpus linguistics and conversation analysis to examine EFL classroom discourse in Papua. The novelty lies in treating lexical patterns not only as frequency-based phenomena but also as interactional resources whose pedagogical functions become visible through sequential analysis. By linking corpus-derived evidence (e.g., frequent genre labels, definitional bundles, and formulaic sequences) with CA-informed analysis of turn-taking, repair, and scaffolding, the study provides a more comprehensive account of how vocabulary, grammar, and interaction co-construct learning opportunities. This integrated approach responds directly to calls for classroom study that connects linguistic patterning with interactional competence and participation structures (Kunitz et al., 2021; Zhou & Li, 2023).

Focusing on multiple instructional topics (invitation, news item, recount text, personal recount, and describing people) this study investigates: (1) the characteristics of students’ utterances across task types; (2) the extent to which these utterances reflect communicative and authentic English use; and (3) the corpus-informed pedagogical implications emerging from observed interactional patterns. By situating Papua within broader discussions of genre pedagogy, formulaic language, and classroom interaction, this study contributes empirical evidence to the field of EFL pedagogy while offering methodological innovation through the integration of corpus and conversation analytic approaches.

LITERATURE REVIEW

Classroom Discourse and Language Learning

Classroom conversation is really important because it helps students learn and use language. In English as a foreign language class, where students don’t get much chance to

use English outside school, the classroom is often the main place they can practice real communication (Walsh, 2020). Studies show that talking in class helps students get language input, understand each other better, and learn how to interact in real conversations (Hellermann, Eskildsen, & Pekarek Doehler, 2019). Also, the way teachers talk and how students interact in class shows the teaching methods and the rules of the class.

When teachers use active learning methods, which focus on students working together and taking part in discussions, the classroom talk changes from being mostly teacher-centered to more student-centered and back-and-forth (Bonwell & Eison, 1991; Freeman et al., 2014). In these classes, students get to share their thoughts, work together to understand things, and respond to each other, which helps them both speak better and interact more naturally. But for this to work, the way students talk needs to be organized well, and they need support to use language that feels real and natural.

Corpus Linguistics in Language Education

Corpus linguistics has become a key method for studying how language is used and improving language teaching. A corpus is a collection of real language data that shows how words are actually used. This helps researchers look at how often words are used, how they often go together, and common expressions used in interaction (McEnery & Brezina, 2019). In language education, corpora help teachers find the most useful words, phrases, and patterns for learners, which then guides the creation of teaching materials and classroom activities (Flowerdew, 2020).

One important area of corpus-based research is how often words are used. Studies show that learners do better when they are exposed to words that are used a lot, as these words are more likely to stick in their memory and be used correctly (Nation, 2013). In the classroom, looking at how often teachers and students use certain words helps to understand how much learners are practicing and using important vocabulary. Another focus is collocations, which are common word pairs or groups that often appear together. Collocations are important for speaking and writing naturally and not knowing them can make language sound strange or wrong (Nesselhauf, 2005). Study shows that paying attention to collocations helps learners speak more accurately and use expressions more naturally (Boers & Lindstromberg, 2012).

Formulaic sequences, like "I would like to..." or "Thank you for...", are also important for communication. These are fixed phrases that help learners speak more fluently and with less thinking. Recent studies show that learners in English as a foreign language often use these prepared phrases, especially when they are still learning the language (Römer & Wulff, 2021). Looking at classroom conversations through a corpus helps teachers see which formulaic expressions students are learning and how these expressions help them communicate better.

Conversation Analysis and Classroom Interaction

Conversation Analysis (CA) gives another way of looking at how people talk by looking at how conversations happen right then and there. In contrast to corpus linguistics, which looks at how frequent words and phrases are used, CA looks at the talking order. CA studies investigate who gets to speak next, pairs of actions that happen one after another, sequences where people fix mistakes, and ways people interact. Turn-taking is a basic part of talking, and in classrooms, it's usually controlled by the teacher through questions, naming students, or giving instructions. Study shows that when teachers control the time of students speak, it can stop students from talking freely, but it also gives them chances to practice specific language skills (Seedhouse, 2004). Another important part of classroom talk is action pairs that happen at the same time, like a greeting and a response,

or a question and an answer. These patterns help learners practice useful and real-life conversations. For example, pairs like invitation and an acceptance/refusal help students understand how to handle social situations outside the classroom (Markee, 2015).

Fixing mistakes, or repair sequences, is very important in language lessons. Studies show that both when students fix their own mistakes and when others help them fix mistakes, can help them learn by focusing on language and meaning (Sert & Walsh, 2020). These moments are not about failure, but about improving how people use language. Also, ways people keep conversations going, like repeating what someone said, asking for clarification, or nodding, help maintain communication. In English as a Foreign Language (EFL) lesson, these strategies help students stay involved in conversations even when they don't have enough language skills (Walsh, 2020).

Integrating Corpus Linguistics and CA in Classroom Studies

While corpus linguistics and conversation analysis have usually been used on their own, recent studies suggest combining them to get a better understanding of classroom conversations. Corpus linguistics can show big patterns in word use, while conversation analysis looks at the small details of people interaction (Csomay & Petrović, 2022). For example, a frequency check might show that "invitation" is a word that comes up regularly, but conversation analysis can explain how that word is used in turn-taking or in standard phrases. Several recent studies show the advantages of using these two methods in combination. McEnery and Brezina (2019) say that using corpora in discourse studies can show both how often certain language features appear and how they are used in real conversations. Similarly, Römer and Wulff (2021) point out that looking at formulaic language through a corpus can be more meaningful when we also see how learners use these phrases during actual interactions. A suggestion comes from Csomay and Petrović (2022) to combine corpus linguistics and conversation analysis in order to understand classroom talk better and make teaching more effective. In English as a Foreign Language teaching, this combined method is especially helpful because it connects the vocabulary students learn with their ability to interact. By looking at not just which words students use but also how they are used during interactions, teachers and researchers can gain a deeper understanding of how students develop their communication skills.

Research Gap in Indonesian and Papuan Contexts

Although classroom talk, corpus linguistics, and conversation analysis have been widely investigated, important gaps remain in secondary-school EFL contexts in Indonesia, particularly in understated regions such as Papua (Walsh, 2020; Widodo & Emilia, 2021; Sukirlan, 2022). Recent studies highlight the growing value of corpus-informed and interactional approaches for understanding classroom discourse, yet these methods are still rarely integrated to examine how lexical patterning and participation structures jointly shape learning opportunities in real-time classroom interaction (Csomay & Petrović, 2022; Römer & Wulff, 2021; Sert & Walsh, 2022). Consequently, there is still limited empirical evidence on how active learning tasks in Papua influence students' communicative development and interactional competence (Ellis & Shintani, 2021; Sert & Kunitz, 2020).

Also, not many studies have tried to combine corpus linguistics with conversation analysis in Indonesian classrooms. Usually, study either looks at common words and phrases using corpus methods or studies how people talk with each other using conversation analysis. By using combined approaches, we can get a better picture of what happens in class, and connect how students learn vocabulary with how they develop their communication skills.

RESEARCH METHOD

Research Design

This study employed a qualitative case study design complemented by descriptive quantitative corpus analysis. The qualitative orientation was chosen to capture the richness and complexity of classroom interaction in an active learning setting, while the corpus perspective provided systematic insights into the frequency and distribution of lexical and formulaic patterns. A case study approach was appropriate because it allowed the phenomenon of classroom discourse to be examined in depth within its natural context (Yin, 2018). The integration of corpus linguistics and Conversation Analysis (CA) reflects a growing trend in applied linguistics to combine macro-level lexical analysis with micro-level interactional insights (Csomay & Petrović, 2022).

Participants and Setting

The study was conducted at a public senior high school in Papua, Indonesia, which was purposely selected for its adoption of active learning strategies in English classes. Another reason of selecting this school was because the students are from different regions of center Papua province. Two English teachers were recruited based on the following criteria: 1) They had at least three years of teaching experience; 2) They explicitly implemented active learning strategies (e.g., group discussion, role-play, student-led questioning); 3) They consented to classroom recording and interviews. These two English teachers (initially with E and M) were from East Indonesia regions.

Two eleventh-grade classes taught by these teachers were included in the study. Each class consisted of approximately twenty students (total $\approx$ 40 students). The classes were selected because they represented typical academic-track classes within the school. The learning processes of these two classes were recorded resulting in five datasets. Invitation, news items and recount text datasets came from class with teacher E. While personal recount and describing people datasets came from class with teacher M. The local context of Papua is significant because it represents a linguistically diverse and under-researched region where English learning often takes place with limited resources. By focusing on this setting, the study contributes to expanding the geographical and sociolinguistic scope of EFL classroom discourse research, which has often been dominated by urban and tertiary contexts (Walsh, 2020).

Data Collection

Data were collected over two recorded classroom sessions (invitation, news items and recount text topics) in the classroom taught by teacher M and three recorded classroom sessions (personal recount and describing people topics) in the classroom taught by teacher E (five sessions in total). Multiple data sources were used to ensure triangulation and richness: 1) Video and audio recordings of teacher–student interaction; 2) Semi-structured interviews with teachers and selected students; and 3) Instructional documents, including lesson plans and worksheets. Teacher–student interactions were recorded using audio and video devices to capture both verbal and paralinguistic features of communication.

The recordings were transcribed using a simplified CA-informed convention that marked pauses, overlaps, and speaker turns (Seedhouse, 2004). Speaker codes (e.g., T = Teacher; S1, S2 = Students) were used for clarity. Supplementary materials such as lesson plans and student worksheets were collected to contextualize the discourse. The teacher and selected students were interviewed to elicit their perspectives on classroom interaction and learning. The combination of these methods provided a holistic dataset, enabling both lexical-corpus and interactional analyses (McEnery & Brezina, 2019).

Corpus Construction

The transcriptions were compiled into a mini spoken corpus. Following established procedures in classroom corpus research (Flowerdew, 2020), the corpus was processed through the following stages. First, data cleaning and standardization include standardizing spelling inconsistencies, retaining non lexical fillers (e.g. ‘uhm’, ‘eh’), and removing CA symbols. Second, corpus segmentation where data were organized according to activity type (e.g., group discussions, role-play, teacher–student Q&A). Third, linguistics annotation by conducting manual tagging for speech functions (requests, responses, clarifications), turn type (teacher-initiated, student-initiated) and repair episodes. Fourth, uploading cleaned corpus files into analysis tools called *AdTAT* (*Adelaide Text Analysis Tools*) to enable frequency analysis, keyword identification, concordance searches, and collocational analysis. This approach enabled the systematic identification of high-frequency vocabulary, collocational patterns, and formulaic sequences that emerged in classroom discourse.

Data Analysis

Data analysis in this study followed a sequential explanatory triangulation model, in which corpus findings informed and guided deeper CA analysis. This combined perspective is increasingly recommended in applied linguistics research as it provides a more holistic understanding of how language is used in classroom contexts (McEnery & Brezina, 2019; Walsh, 2020). From a corpus linguistics perspective, corpus driven pattern identification was conducted. The transcribed classroom interactions were compiled into a mini spoken corpus and analyzed using *AdTAT* (*Adelaide Text Analysis Tools*). The corpus analysis focused on three dimensions: (1) lexical frequency, to identify recurrent vocabulary produced by teachers and students; (2) collocational patterns, to highlight common word associations that signal emerging discourse structures; (3) formulaic sequences or chunks, to capture recurrent expressions such as *I would like...*, *Thank you for...*, or *I hope you are coming...*; and (4) distribution of teacher vs. student lexical contributions. Previous studies have shown that such corpus-informed analyses can reveal patterns of formulaic language that underpin classroom communication and learning (Flowerdew, 2020; Römer & Wulff, 2021).

From a Conversation Analysis perspective, the study explored the fine-grained organization of talk-in-interaction. Attention was given to (1) turn-taking mechanisms, including shifts, overlaps, and interruptions; (2) adjacency pairs, such as question–answer or invitation–acceptance/rejection sequences; (3) repair strategies, distinguishing between self-initiated and other-initiated repairs of errors or misunderstandings; and (4) interactional strategies, including repetition, clarification requests, *code-switching*, and *backchanneling*. CA has proven particularly effective for understanding the sequential and emergent nature of classroom talk in EFL/ESL contexts (Hellermann et al., 2019; Sert & Walsh, 2020).

Finally, the results from corpus analysis and CA were interpreted through the lens of classroom discourse analysis to link linguistic findings with pedagogical practice. This stage examined how interactional features supported or constrained active learning, such as how teachers scaffolded student participation and how learners negotiated meaning collaboratively. Integrating corpus linguistics with CA has been shown to provide a more comprehensive account of classroom discourse, capturing both the macro-patterns of language use and the micro-details of interaction (Csomay & Petrović, 2022).

Validity and Reliability

To ensure the trustworthiness of the study, several strategies were employed. Data from recordings, interviews, and documents were compared to validate findings.

Analytical interpretations were reviewed by colleagues with expertise in applied linguistics to reduce bias. Portions of the transcripts were independently coded by a second researcher to check coding consistency. Preliminary interpretations were shared with the teacher and students to confirm alignment with their lived classroom experiences. These strategies align with recommended practices in qualitative educational research for enhancing rigor and credibility (Creswell & Poth, 2018).

RESULTS AND DISCUSSION

There are five EFL topics namely invitation, news item, recount text, personal recount, and describing people, that were investigated in this study using an integrated corpus linguistics (CL) and conversation analysis (CA) framework. This study examined EFL classroom discourse in Papua through an integrated Corpus Linguistics (CL) and Conversation Analysis (CA) approach to understand how lexical patterns and interactional practices support active learning. Comparative analysis has been done to reveal systematic variation in lexical patterns, formulaic language use, and interactional organization across topics. The study also describes a consistent pedagogical structure in which teacher-led scaffolding and lexical recycling create participation opportunities for learners. A more comprehensive understanding of how language learning emerges through daily classroom interaction was enabled by the integration of corpus-based frequency analysis with micro-level CA. The findings are presented in detail in the following table (Table 1).

Table 1. Corpus and Conversation Analysis Across All Datasets

Dimension	Invitation	News Item	Recount Text	Personal Recount	Describing People
Corpus Size (tokens)	3,436	3,261	3,121	6,922	1,578
Dominant Lexical Focus	Event vocabulary (<i>invitation, party, time, place</i>)	Genre meta language (<i>news, item, structure, function</i>)	Sub-genre distinctions (<i>recount, personal, factual, biographical</i>)	Narrative time & past tense verbs (<i>went, last week, then</i>)	Descriptive grammar (<i>he, she, has, is, hair, eyes</i>)
Most Frequent Keywords	invitation (12), time (8), purpose (8)	news (33), item (29), generic (8), structure (8)	recount (26), text (19), next (19)	recount (39), went (22), then (12)	has (12), he (12), is (11), beautiful (5)
Core Formulaic Bundles	<i>I would like...; I hope you are coming...; Thank you for...</i>	<i>news item text; generic structure; written public information</i>	<i>recount text; personal recount; past tense</i>	<i>last week; and then; after that</i>	<i>he is...; has got...; red hair</i>
Collocational Strength	Moderate (<i>birthday party, invite you</i>)	Very strong (<i>news item ×29</i>)	Strong (recount text ×8)	Strong (last month, past tense)	Moderate (<i>has got, red hair</i>)
Dominant Interaction Pattern (CA)	Teacher-led pragmatic performance sequence	Definition–response–evaluation cycle	Classification–confirmation sequence	Narrative scaffolding sequence	Attribute listing with repair
Turn-taking Organization	Teacher allocates turns	Strong teacher control	Strong teacher control	Teacher-managed but longer student turns	Teacher-managed drills
Repair Pattern	Scaffolding repair for pronunciation & pragmatics	Concept clarification repair	Genre clarification repair	Grammar + narrative repair	Form-focused repair

Dimension	Invitation	News Item	Recount Text	Personal Recount	Describing People
Student Autonomy Level	Low–Moderate	Low	Low	Moderate (more extended turns)	Low
Lexical Diversity Level	Moderate	Low–Moderate (metalinguage heavy)	Low (taxonomy-focused)	Moderate–High (storytelling verbs)	Low (limited adjective set)
Pedagogical Orientation	Pragmatic communicative action	Genre literacy	Genre taxonomy	Narrative production	Grammar-based description
Interactional Competence Evidence	Invitation performance routines	Academic discourse routines	Genre labeling competence	Emerging storytelling competence	Early descriptive competence

Dominant Lexical Focus across Topics: From Genre Labels to Narrative Action

Clear thematic differentiation across the five datasets is revealed by corpus analysis (Table 1). Based on the data of most frequent keywords presented in table above, the high frequency items found are *news* (33) and *item* (29) in news item dataset; *recount* (26) in recount text dataset; *recount* (39) and *went* (22) in personal recount dataset. These findings show that classification and definitional explanation is mostly identified in classroom interaction. As a result, in relation to dominant lexical focus, genre metalanguage and sub-genre distinctions were mostly found in the classroom discourse across all datasets. The dominance of genre metalanguage across datasets aligns with recent studies (Lee & Deakin, 2022; Nguyen & Marlina, 2023). These studies find that genre-based pedagogy often produces repeated definitional bundles and classification sequences in EFL classrooms. Similar patterns have been reported in Indonesian EFL contexts, where genre teaching often emphasizes metalanguage such as *generic structure* and *social function* as part of classroom discourse routines (Widodo & Emilia, 2021).

On the other hand, findings from the personal recount dataset show a shift from classification toward narrative production. Past tense verbs such as *went* (22), temporal markers such as *last week* (8) and *last month* (9), and sequencing markers such as *then* (12) and *after that* (8) are frequently found indicating that the students are engaged in constructing temporally ordered narratives. This finding consistent with research showing that storytelling tasks promote extended turns and sequencing markers in L2 development (Eskildsen, 2020; Li & Qian, 2022; Sari & Rido, 2022).

Specifically, in terms of dominant lexical focus, findings from invitation dataset show event vocabulary (e.g. *invitation, party, place, time*); findings from news item dataset show genre meta language (e.g. *news, item, structure, function*); findings from recount text dataset focus heavily on sub-genre distinctions (e.g. *personal, recount, factual, and biographical*); findings from personal recount dataset demonstrates narrative time and past tense verbs (*went, last week, then*) using in more sustained storytelling; and findings from describing people dataset are mostly in the form of descriptive grammar patterns rather than thematic lexical expansion. Lexical items such as *he* (12), *has* (12), and *is* (11) are highly found which indicate reliance on pronoun–copula and possessive constructions. Evaluative adjectives in describing people dataset (e.g. *beautiful* and *handsome*) suggest limited lexical diversity compared to findings from personal recount dataset that are rich in lexical diversity.

In general, there is a developmental pattern across tasks shown from these lexical contrasts. The pattern goes from genre labeling (news item, recount text) to pragmatic

performance (invitation), narrative construction (personal recount), and finally to attribute-based description (describing people). This pattern highlights the importance of task types in shaping lexical and interactional development. In EFL pedagogy, the role of task type in shaping both lexical and interactional development emphasizes task design as a driver of communicative growth (Ellis & Shintani, 2021). Study conducted by Sert & Walsh (2022) supports this finding from a classroom interaction perspective. It reflects how interactional competence emerges through structured participation and the gradual expansion of learner roles within classroom discourse.

Formulaic Language as an Interactional Resource

Formulaic language is found to be a key technique that encourages student involvement and serves as interactional resource based on the results from all datasets. In the invitation dataset, formulaic language serves clear interactional functions in the following sequences: *I would like to invite you...* functions as a turn-entry device; *I hope you are coming...* plans the invitation action; and *'Thank you for the time given to me'* closes a presentation. These chunks reduce cognitive burden and enable students to perform social actions within structured sequences. This supports recent work on L2 pragmatics showing that formulaic language are interactional resources for learners' pragmatic development in instructed settings (Taguchi, 2021). Moreover, formulaic language has been widely recognized as a main factor in promoting fluency and interactional competence, particularly when learners have lack of grammatical control (Wray & Perkins, 2021).

Findings from the news item and recount text datasets show that formulaic language appears mainly in definitional bundles such as *news item text*, *generic structure*, *written public information*, *recount text*, *past tense*. These lexical bundles are constantly reused by teachers and students, creating stable interactional frames. Although such repetition may appear mechanical, CA analysis shows that these bundles function as anchors within teacher-initiated question–response sequences. From a corpus perspective, definitional bundles are typical in pedagogical discourse and play a key role in structuring classroom explanations. These bundles enable learner participation in teacher-led question–response routines (Yoon & Römer, 2021). Moreover, conversation analytic findings show that these bundles enable learners to reduce uncertainty, make expected responses more predictable, and provide learners with interactional confidence in contexts where spontaneous L2 production is still limited (Sert & Kunitz, 2020).

The personal recount dataset shows a different kind of formulaic language. Instead of definitional bundles, students rely on temporal and sequencing expressions such as *last week*, *and then*, and *after that*. These bundles enable students to organize multi-event narratives and sustain longer turns. Here, formulaic language supports discourse expansion rather than mere participation. This supports research showing that temporal markers function as key resources for narrative coherence in L2 oral production and facilitate extended discourse development (Li & Qian, 2022; Pallotti, 2020).

In contrast, describing people dataset shows different formulaic language, that is grammar-driven: *he is*, *she has*, and *has got* functioning as structural templates for listing attributes. While lexical diversity remains limited, these templates allow learners to produce grammatically structured utterances within a controlled environment. These findings are consistent with studies showing that form-focused classroom interaction often promotes controlled structural routines rather than lexical expansion (Nassaji & Kartchava, 2020; Sato & Loewen, 2020).

Thus, across all topics, formulaic language serves as an interactional scaffold. However, its function varies: interactional function in invitation topic, definitional function in news item and recount text topic, narrative sequencing function in personal recount topic, and descriptive structuring function in describing people topic. This

variation demonstrates that formulaic language serves as an interactional scaffold, but its pedagogical value depends on how it is embedded in classroom interaction and task design (Sert & Walsh, 2022; Wray & Perkins, 2021; Yoon & Römer, 2021).

Turn-taking Organization and Teacher Scaffolding

Conversation Analysis reveals strong teacher control across datasets. Turn-taking is predominantly teacher-allocated, with minimal student self-selection. This pattern was documented widely in EFL classroom interaction research (Seedhouse, 2004; Walsh, 2020; Sukirlan, 2022). In the invitation dataset, teacher control remains present, but interaction shifts toward performance-oriented sequences. Students are invited to present invitations in front of the class, using formulaic openings and closings. Although turns are still allocated by the teacher, students produce longer stretches of talk.

In news item and recount text datasets, interaction typically follows an Initiation–Response–Evaluation (IRE) pattern: the teacher poses a question, a student provides a short answer, and the teacher confirms or corrects. In other words, this study shows that the pattern reinforces genre metalanguage, but limits extended student talk. This findings in line with other studies showing such similar pattern that has been shown to reinforce conceptual understanding but often restrict extended learner talk (Morton & Llinares, 2021; Sukirlan, 2022).

In contrast, personal recount dataset shows the highest degree of extended student production. While the teacher continues to manage turn allocation, students produce multi-clause narratives supported by temporal markers, as it is in line with Li and Qian (2022) and Sari & Rido (2022) studying in Indonesian context. However, describing people dataset reflects a drill-like pattern. Repetition, confirmation, and repair dominate interaction, reinforcing pronoun–verb agreement and descriptive adjectives. This interactional scaffolding prioritizes form over narrative content, as it is also described by Eskildsen (2020).

According to CA study, teacher-controlled discourse is not always restricted. Teacher control is dynamic and can constrain or enable learning (Walsh and Li, 2021). By giving scaffolding, teachers can expand learner participation when aligned with task design (Sert and Walsh, 2022). The task design has a relatively education impact especially in shaping interaction (Kunitz and Markee, 2021). Teacher domination could inhibit student growth in lessons that rely much on classification according to findings from news items and recount text datasets. However, teacher scaffolding could facilitate longer turns in narrative and performance tasks based on findings in invitation and personal recount datasets.

Repair Practices and Learning Opportunities

Repair sequences appear across all datasets but differ in function. In invitation dataset, repair focuses on pronunciation, as its focused repair in Indonesian context has been identified by Syafryadin & Salniwati (2021), and pragmatic appropriateness, as its repair and negotiation in classroom interaction has also been discussed by Kasper and Ross (2020). Saito & Plonsky (2021) also found that pronunciation repair and corrective feedback play role as central in classroom talk. While Taguchi (2021) shows that repair is linked to pragmatic appropriateness in speech acts.

In news item and recount text datasets, repair often targets conceptual accuracy (e.g., correcting genre definitions). This finding in line with Morton and Llinares (2021) that conducted a study on conceptual repair inside IRF cycles. Based on Lee and Deakin (2022), genre metalanguage, which the dominant lexical focus of news item dataset, is being corrected and recycled through teacher repair. This finding is also supported by

Widodo and Emilia (2021) that investigating conceptual repair in genre-based classroom instruction.

In personal recount dataset, repair frequently addresses verb tense consistency and narrative clarity. This finding is aligning with evidence that tense and temporal markers are key resources for coherence in L2 storytelling (Li & Qian, 2022; Sert & Kunitz, 2020). In describing people dataset, repair emphasizes grammatical agreement, which is typical of form-focused interaction in controlled descriptive tasks (Nassaji & Kartchava, 2020; Sato & Loewen, 2020). Importantly, repair rarely stops interaction. Instead, teachers reformulate or prompt students to self-correct, allowing talk to continue. This supports the CA perspective that repair can function as a learning opportunity embedded within interaction rather than as a disruptive correction mechanism (Sert & Walsh, 2022; Waring, 2021).

Cross-Dataset Developmental Implications

When viewed comparatively, the datasets suggest a progression in interactional complexity that is strongly shaped by task design and pedagogical orientation. Genre-labeling lessons (findings from news item and recount text datasets) prioritize metalanguage, definitional explanation, and classification, a pattern widely reported in genre-based EFL instruction. In this kind of instruction, learners' participation is often constrained to short IRF responses that recycle genre labels rather than extended meaning-making (Lee & Deakin, 2022; Nguyen & Marlina, 2023; Widodo & Emilia, 2021).

Performance-based lessons (findings from invitation dataset) prioritize formulaic action sequences. The tasks in this kind of lesson can expand interactional space even within teacher-managed turn allocation, as learners draw on formulaic sequences to perform social actions and sustain longer talk (Taguchi, 2021; Wray & Perkins, 2021).

The strongest evidence of discourse expansion appears in narrative lessons (findings from personal recount dataset) encourage temporal organization and extended turns. In this lesson, students produce multi-clause storytelling supported by temporal markers and sequencing bundles. This finding is consistent with research showing that narrative tasks facilitate coherence, extended turns, and the development of interactional competence through repeated participation in temporally organized talk (Eskildsen, 2020; Li & Qian, 2022; Pallotti, 2020; Sari & Rido, 2022). Grammar-based descriptive lessons (findings from describing people dataset) reinforce structural accuracy but limit lexical expansion. In this lesson, students rely on structural templates that prioritize accuracy and controlled production. In contrast, students have limit lexical diversity and thematic elaboration which is in line with findings from (Nassaji & Kartchava, 2020). The study highlighted that form-focused interaction can limit opportunities for discourse expansion.

These cross-topic contrasts indicate that task design significantly shapes interactional competence. In terms of student autonomy level, students demonstrate greater autonomy and discourse expansion in tasks that require storytelling (personal recount) than in tasks focused primarily on classification (recount text, news item). Therefore, active learning in EFL classrooms may benefit from increased emphasis on narrative and pragmatic performance tasks that encourage extended student production.

Conclusion of Comparative Analysis

By integrating corpus-based lexical analysis with conversation analysis of interactional organization, this study reveals that learning in EFL classrooms emerges through repeated and structured participation. Formulaic language, lexical recycling, and teacher-led scaffolding consistently support student participation across topics. However, the extent to which these resources foster extended discourse depends on task design.

Narrative and performance-oriented tasks appear to create more opportunities for interactional growth than taxonomy-driven instruction.

This comparative perspective strengthens the argument that corpus-informed conversation analysis provides a powerful methodological framework for understanding classroom discourse rather than treating vocabulary, grammar, and interaction as separate domains. The integrated approach shows how lexical patterns and sequential organization jointly construct learning opportunities in real-time classroom interaction.

CONCLUSION

This study set out to analyze classroom interaction in an active English learning context through a combined corpus linguistics and Conversation Analysis (CA) approach. This study contributes to EFL pedagogy by showing how students' communicative development in active learning classrooms is shaped by the interaction between task design, formulaic language, and teacher scaffolding. Addressing the first research question, corpus analysis across five datasets demonstrates clear differences in students' utterance characteristics as follows. Genre-focused lessons (news item and recount text) are dominated by metalanguage and definitional bundles. In contrast, performance-based and narrative task lessons (invitation and personal recount) elicit longer turns supported by formulaic language, temporal markers, and sequencing expressions. Meanwhile, grammar-driven descriptive lessons (describing people) prioritize structural accuracy through repeated templates but show more limited lexical diversity and thematic elaboration.

Regarding to the second research question, these patterns indicate that learners' English becomes more communicative and locally authentic when tasks require pragmatic action or storytelling rather than classification. In such tasks, learners produce extended multi-clause talk and manage discourse continuity. Further conversation analysis shows that teacher-managed turn allocation is common, yet not always restrictive. It limits expansion in genre-focused lessons but facilitates participation and repair-as-learning in performance-based and narrative task lessons. As a result of the third research question, the study recommends designing classroom tasks that prioritize narrative and pragmatic performance, recycling formulaic language as interactional scaffolds while gradually expanding learner autonomy. Methodologically, the study demonstrates the value of integrating corpus linguistics and Conversation Analysis to connect lexical patterns with the sequential organization of learning opportunities in real-time classroom discourse.

ACKNOWLEDGMENT

This study was funded by Faculty of Language and Arts, Universitas Negeri Jakarta.

REFERENCES

- Boers, F., & Lindstromberg, S. (2012). *Experimental and intervention studies on formulaic sequences in a second language*. *Annual Review of Applied Linguistics*, 32, 83–110.
- Bonwell, C. C., & Eison, J. A. (1991). *Active learning: Creating excitement in the classroom*. Washington, DC: ASHE-ERIC Higher Education Report.
- Brezina, V., & Pallotti, G. (2021). Lexical diversity measures in L2 spoken production: A corpus-based comparison. *International Journal of Corpus Linguistics*, 26(3), 1–25.
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). Thousand Oaks: Sage.
- Csomay, E., & Petrović, A. (2022). *Integrating corpus linguistics and classroom discourse studies: A methodological synthesis*. *Language Teaching Research*, 26(5), 633–655.

- Ellis, R., & Shintani, N. (2021). Task-based language teaching and L2 interactional development. *Language Teaching Research*, 25(5), 1–18.
- Eskildsen, S. W. (2020). Usage-based development in L2 interaction: From formulaic sequences to narrative expansion. *Modern Language Journal*, 104(3), 557–575.
- Flowerdew, L. (2020). *Corpora and language education in a changing world*. Language Teaching, 53(2), 216–230.
- Freeman, S., et al. (2014). *Active learning increases student performance in science, engineering, and mathematics*. PNAS, 111(23), 8410–8415.
- Hellermann, J., Eskildsen, S. W., & Pekarek Doehler, S. (2019). *L2 interactional competence: Theoretical and methodological considerations*. Applied Linguistics, 40(2), 205–214.
- Kasper, G., & Ross, S. (2020). Classroom pragmatics: Interactional repair and appropriateness in instructed settings. *System*, 92, 102–246.
- Kunitz, S., & Markee, N. (2021). Task-as-workplan versus task-in-process: A conversation analytic perspective on classroom task implementation. *Language Teaching Research*, 25(4), 1–22.
- Kunitz, S., Markee, N., & Sert, O. (Eds.). (2021). *Classroom-based conversation analytic research: Theoretical and applied perspectives on pedagogy*. Springer.
- Lee, J., & Deakin, L. (2022). Genre-based pedagogy and classroom metalanguage in EFL contexts. *Journal of English for Academic Purposes*, 56, 101–194.
- Li, S., & Qian, D. (2022). Temporal markers and narrative coherence in L2 oral production. *Applied Linguistics*, 43(4), 725–748.
- Markee, N. (2015). *The handbook of classroom discourse and interaction*. Hoboken: Wiley Blackwell.
- McEnery, T., & Brezina, V. (2019). *Corpus linguistics and the description of English*. International Journal of Corpus Linguistics, 24(3), 261–289.
- Morton, T., & Llinares, A. (2021). Classroom interaction and IRF patterns in L2 subject teaching. *System*, 96, 102405.
- Nassaji, H., & Kartchava, E. (2020). Grammar-focused interaction and second language development. *Language Teaching Research*, 24(3), 1–20.
- Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge: Cambridge University Press.
- Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam: John Benjamins.
- Nguyen, T. T., & Marlina, R. (2023). Metalanguage and genre instruction in secondary EFL classrooms. *Linguistics and Education*, 73, 101120.
- Nurlaelawati, I., Lengkanawati, N. S., & Gunawan, W. (2020). Recontextualising genre-based pedagogy in Indonesian context: A case of preservice teachers. *International Journal of Education*, 12(2).
- Pallotti, G. (2020). Narrative structure and development in L2 spoken performance. *System*, 92, 102255.
- Rahman, A. (2023). Lexical diversity in Indonesian EFL learners' spoken production. *Journal of Language and Education*, 9(2), 155–169.
- Römer, U., & Wulff, S. (2021). *Applying corpus linguistics to TESOL: Innovations and challenges*. TESOL Quarterly, 55(2), 529–556.
- Saito, K., & Plonsky, L. (2021). Effects of corrective feedback on second language pronunciation development: A meta-analysis. *Studies in Second Language Acquisition*, 43(2), 1–27.
- Sari, N., & Rido, A. (2022). Narrative development in Indonesian EFL learners' spoken recounts. *TEFLIN Journal*, 33(1), 58–76.

- Sato, M., & Loewen, S. (2020). Do learners benefit from form-focused interaction? Evidence from classroom research. *Studies in Second Language Acquisition*, 42(2), 1–23.
- Seedhouse, P. (2004). *The interactional architecture of the language classroom: A conversation analysis perspective*. Oxford: Blackwell.
- Sert, O., & Kunitz, S. (2020). Classroom interaction and narrative development in L2 learning. *Classroom Discourse*, 11(2), 123–142.
- Sert, O., & Walsh, S. (2020). *Classroom interactional competence and teacher development*. RELC Journal, 51(3), 439–452.
- Sert, O., & Walsh, S. (2022). Interactional competence and classroom ecology: Expanding participation in L2 classrooms. *Modern Language Journal*, 106(4), 912–928.
- Sukirlan, M. (2022). IRF structure in Indonesian EFL classrooms: Implications for learner talk. *Studies in English Language and Education*, 9(1), 220–237.
- Syafradin, S., & Salniwati, S. (2021). Corrective feedback and pronunciation repair in Indonesian EFL classrooms. *Studies in English Language and Education*, 8(3), 1045–1062.
- Taguchi, N. (2021). Pragmatic development in instructed second language settings. *Annual Review of Applied Linguistics*, 41, 134–152.
- Tang, K.-S. (2021). The interconnections among metadiscourse, metalanguage, and metacognition: Manifestation and application in classroom discourse. *Linguistics and Education*, 65, 100977.
- Walsh, S. (2020). *The role of classroom discourse in teacher development and language learning*. Language Teaching Research, 24(1), 25–44.
- Waring, H. Z. (2021). Repair in classroom interaction: Managing learning opportunities in teacher–student talk. *Classroom Discourse*, 12(2), 145–166.
- Wati, S. O., Suzuasmisyah, A., Chair, H., & Efrini, M. (2025). Exploring classroom discourse in applied linguistics: Analysing IRF interaction patterns in offline and online English classrooms. *Journal of English Language and Education*.
- Widodo, H. P., & Emilia, E. (2021). Genre-based pedagogy in Indonesian EFL classrooms: Pedagogical affordances and challenges. *Indonesian Journal of Applied Linguistics*, 11(2), 389–401.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- Wray, A., & Perkins, M. (2021). Formulaic language and communicative competence revisited. *Language Teaching*, 54(3), 1–20.
- Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). Thousand Oaks: Sage.
- Yoon, H., & Römer, U. (2021). Lexical bundles in pedagogical discourse: Functions and classroom implications. *System*, 98, 102463.
- Yoon, H.-J., & Römer, U. (2024). Exploring the functions of lexical bundles in undergraduate learner writing for pedagogical use. *Language Learning in Higher Education*, 14(2), 337–356.
- Zhang, Y., & Liu, X. (2021). SFL pedagogies in language education: Special issue introduction. *System*.
- Zhou, S., & Li, W. (2023). Classroom interactional competence in an English-medium instruction classroom: Interactional resources and learning opportunities. *System*.