Talking in Sync: How Linguistic Synchrony Shapes Teacher-Student Conversation in English as a Second Language Tutoring Environment

Anna Pauline Aguinalde
University of Florida
Gainesville, USA
aaguinalde@ufl.edu

Jinnie Shin
University of Florida
Gainesville, USA
jinnie.shin@coe.ufl.edu

Abstract

Linguistic synchrony, or alignment, has been shown to be critical for student learning, particularly for L2 students (second language learners), whose patterns of synchrony often differ from fluent speakers due to proficiency constraints. While many studies have explored various dimensions of synchrony in global language tutoring contexts, there is a gap in understanding how linguistic synchrony evolves dynamically over the course of a tutoring session and how tutors' pedagogical strategies influence this process. This study incorporates three dimensions of synchrony—lexical, syntactic, and semantic—along with tutors' dialogue acts to evaluate their association with student performance using multivariate time-series analysis. Results indicate that lower-performing L2 students tend to lexically align with their tutor more consistently in the long term and with higher intensity in the short term. In contrast, higher-performing students demonstrate greater alignment with the tutor in syntactic and semantic dimensions. Furthermore, the dialogue acts of eliciting, scaffolding, and enquiry were found to play the strongest roles in influencing synchrony and impacting learning outcomes.

CCS Concepts

• **Computing methodologies → Natural Language Processing**; **Discourse, dialogue and pragmatics**; **Machine learning approaches**.

Keywords

Linguistic Synchrony, Dialogue acts, Language learner, Tutoring Conversation, Time-series analysis

1 Introduction

Linguistic synchrony, or linguistic alignment, has been identified as one of the key factors that are positively correlated with cognitive and affective dimensions essential to student learning. In classroom learning environments, students and teachers often engage in linguistic alignment both intentionally and unintentionally, by mirroring each other's use of vocabulary, syntax, and intonation patterns [46]. This alignment, encompassing lexical, syntactic, and semantic synchrony, enables students to more effectively grasp complex concepts through clearer communication [61]. High levels of linguistic synchrony have been shown to positively affect student task performance [51] and foster a sense of connection between tutors and students, serving as an indicator of improvement in engagement and rapport [49].

Linguistic synchrony assumes an even more critical role in informal educational settings such as peer-assisted learning, and in AI-powered or online tutoring systems [14, 25, 51]. Unlike traditional classroom environments where a teacher addresses a group, these one-to-one tutoring formats offer enhanced opportunities for dynamic, back-and-forth interactions. These interactions often facilitate the development of synchronous conversations crucial for effective learning [54]. Moreover, non-traditional environments introduce unique communication dynamics, including the absence of physical presence, delayed responses, and a reliance on technology, all of which can significantly affect the effectiveness of linguistic alignment [13, 22, 38]. For instance, in virtual tutoring sessions, the lack of face-to-face interaction may impede natural rapport building, making linguistic synchrony a critical bridge to close the communication gap and ensure students remain engaged and comprehend the material.

In second language (L2) tutoring, students must navigate not only comprehension but also the production of target language features [9, 47]. In such settings, linguistic synchrony facilitates not only understanding but aids in acquiring nuanced aspects of the target language [30]. Studies indicate that linguistic synchrony between tutors and students in L2 learning differs markedly from alignment in fluent conversational dialogues [9, 47]. This is because students' ability to align their language with the tutor may be constrained by their proficiency and familiarity with the target language [41, 53]. Additionally, although tutors aim to guide students toward synchronous language production, they must also balance this with fostering independent language generation, which may temporarily reduce synchrony [48]. This dynamic tension between alignment and independent language use makes the study of linguistic synchrony in online L2 tutoring especially critical for

optimizing teaching strategies and improving student outcomes [28].

Despite its significance, research specifically focusing on linguistic synchrony within language tutoring for second language learners remains limited in scope [28]. Existing studies often examine synchrony in a global context [14, 51], evaluating how overall or average synchrony manifests across an entire learning process. However, they seldom investigate how synchrony evolves dynamically as a moving variable throughout the session or how it interacts with the learning environment, including the tutor's pedagogical strategies. A few studies have highlighted the potential impact of tutors' strategies that could influence students' linguistic alignment (e.g., [53]), such as the complexity of their word choices [35, 62]. However, there is limited evidence to understand how tutors' pedagogical strategies, such as their dialogue acts, associate with the dynamics of linguistic synchrony over the course of a learning session (e.g., [48]).

Furthermore, many previous studies have focused on evaluating specific aspects of linguistic synchrony independently. Linguistic synchrony is multifaceted, meaning that the alignment between student and tutor can occur across different language dimensions [46], such as word choices (lexical [47]), sentence structure or grammatical patterns (syntactic), overall topics, themes, and concepts (semantic [14]), as well as timing and interaction patterns (temporal [51]). Recent advancements in computational linguistics and natural language processing (NLP) have significantly enhanced the capacity to efficiently and accurately capture varying dimensions of linguistic synchrony in dialogues [15, 46, 60].

Hence, this study aims to evaluate the dynamics of linguistic synchrony in second language virtual tutoring environments, while focusing on the relationship between the evolving dynamics of the multifaceted characteristics of linguistic synchrony—namely, semantic, syntactic, and lexical synchrony—between the tutor and the students. We also aim to understand the relationship between the specific pedagogical strategies (i.e., pedagogical dialogue acts) employed by tutors and their association with synchrony in encounters with higher- and lower-performing students. The following research questions were posed to navigate the study:

- **RQ1:** What is the relationship between semantic, syntactic, and lexical synchrony and the pedagogical dialogue acts (e.g., scaffolding, eliciting) introduced by the tutor?
- **RQ2:** How do the dynamics of semantic, syntactic, and lexical synchrony differ in tutoring conversations based on student performance?

In addressing these questions, we aim to enhance the empirical evidence through detailed, step-by-step process involving the computational modeling of the multifaceted aspects of linguistic synchrony using multivariate time-series analysis (e.g., [8]). By examining synchrony as a dynamic process, our research provides a more detailed analysis of student-tutor interactions, with potential applications for improving personalized learning environments and tailoring pedagogical strategies in virtual second language tutoring environments.

2 BACKGROUND

2.1 Linguistic Synchrony in Tutoring Conversation

Linguistic synchrony refers to the degree of shared linguistic features between participants in a dialogue [32, 41]. This concept is multifaceted, as interlocutors—often students and tutors—can align their language across various dimensions, including word choices (lexical), grammar and sentence structures (syntactic), timing and patterns (temporal), as well as the topics or concepts (semantic) of their conversation. Specifically, lexical synchrony focuses on alignment at the word-level, known to indicate shared communication amongst participants. Syntactic synchrony deals with structural alignment, such as grammar, sentence type, or parts of speech distributions [4]. Semantic synchrony is concerned with the alignment of the contextual meanings and captures the degree in which the common understanding are aligned [46].

Such patterns emerge naturally over the course of communication and have a significant impact on the flow and effectiveness of interaction [30]. In educational conversations, such as in classroom and tutoring environments, the presence of linguistic synchrony has been found to be positively correlated with students' learning outcomes across various subject domains [14, 49, 51]. When tutors, teachers, and students align their language, it helps create a shared understanding of instructional content by fostering common knowledge building [46], collaboration [14, 20], and rapport building [49]. Particularly, high levels of lexical synchrony with temporal patterns have been found to lead to positive student outcomes in math, as measured through students' online discussion-board conversations with instructors [51]. High levels of linguistic synchrony in lexical choices have been identified to have a positive association with collaborative problem-solving tasks (e.g., [20, 37]), such as outcomes in programming tasks where students' language becomes more synchronized as they work together as a team [49].

In second language learning environments, in particular, where the student's ability to alignment may be constrained by their language proficiency, how the linguistic synchrony occur and and correlates to student learning is more complex [9, 16, 30]. A few studies indicated that lexical alignment [35, 62] and structural (syntactic) alignment [11] are positively correlated with students' task performance. However, many studies identified that varying factors could influence the extent and effectiveness of linguistic alignment in L2 tutoring. For instance, task difficulty plays a significant role (e.g., [39, 62]). As tasks become more complex, students may struggle to maintain alignment due to cognitive overload, prompting tutors to simplify their language and use more repetition and scaffolding for better comprehension.

Similarly, the nature of the interaction—whether the task involves open-ended conversation or more structured, close-ended activities—can also impact linguistic synchrony (e.g., [11]). Open-ended tasks may allow for more natural alignment, whereas structured tasks might limit the opportunities for synchrony [34]. Moreover, student proficiency levels can affect how easily alignment is achieved (e.g., [29, 47]). Lower-proficiency students might rely heavily on tutors to model language structures, while higher-proficiency students may engage in more dynamic back-and-forth interactions. Finally, technological factors in online L2 tutoring environments

(e.g., [30, 33], such as the absence of non-verbal cues, can reduce opportunities for real-time alignment, making it more challenging for both tutor and student to synchronize their language use effectively. By investigating such factors and their impact on the dynamics of linguistic synchrony in L2 tutoring, we can gain deeper insights into supporting and designing more personalized and effective online tutoring environments for L2 learners. One of the under-studied dimensions that could potentially impact this dynamic includes the pedagogical strategies present in online L2 tutoring environments.

2.2 Pedagogical Dialogue Act in Tutoring Conversation

Pedagogical discourse, or the language used in educational contexts, plays a crucial role in learning by enabling students to engage in discussions and share knowledge with one another [1, 3, 50]. Research has shown that rich dialogic interactions between teachers and students are linked to more effective learning outcomes, particularly when students actively contribute to the conversation [27, 55]. Additionally, collaborative interactions between students working on shared tasks have been shown to enhance their ability to construct meaning [55].

In tutoring environments, where direct interaction between the tutor and student is key, the dialogue often evolves into a collaborative effort to achieve shared learning goals [23]. These goals are realized through pedagogical dialogue acts, which reflect the tutor's instructional intentions and strategies within the conversation [56]. Pedagogical dialogue acts in the tutoring setting play a vital role in providing formal structure to the learning environment [23, 45] and fostering positive learning gains [7, 56]. For instance, pedagogical dialogue acts such as *positive feedback* and *direct procedural instruction* have been identified as positively associated with students' learning gains in tutoring dialogues [7]. Similarly, frequent *confirmation questions* by the tutor, followed by positive feedback, were also positively correlated with learning gains [56]. Conversely, several pedagogical dialogue acts have been associated with negative learning gains, particularly sequences of specific dialogue acts such as information provision and observation, which tend to provide a "lecturing" experience for the student. Additionally, the tutor's frequent use of evaluative questions followed by explanations were also identified as negatively associated with learning gains [56].

While generalized tutoring pedagogical frameworks are valuable for outlining strategies and interactions that support student learning, the pedagogical dialogue act in L2 (second language) tutoring pertains interesting dynamics [31]. Dialogue acts such as clarification and enquiry by L2 students bring insight into their competence for tutors to adapt the scaffolds they offer [12]. Studies show that the tutor's role in developing students' language proficiency consists of providing support in sequencing, structuring, and repairing their use of language [44, 59]. These nuances in pedagogical dialogue in the language learning context is distinct from native or fluent speakers [47], highlighting context-specific variations in instructional discourse. Due to the importance of these pedagogical dialogue acts in understanding L2 student performance, it is vital to understand it in conjunction with synchrony.

2.3 METHODOLOGY

2.3.1 Data Collection and Data Preparation. The language tutoring dataset in this study comes from the second version of the Teacher-Student Chat Room Corpus (TSCC, Caines et al. [6]). This dataset includes 260 unique dyadic online chats spanning about an hour each between a tutor and a student, with participants ranging in age from 12 to 40 years old. Each student's English proficiency level was determined by subject matter experts (SMEs) to be between B1 and C2 on the CEFR scale, as shown in Table 1. Each conversation in the dataset consists of an average of 99.38 utterances (SD=29.35), with each utterance containing an average of 14.03 tokens (SD=15.51). These utterances further break down into an average of 12.63 unique tokens each (SD=11.93). Analysis of the utterances reveals that tutors' talk averaged 18.45 tokens (SD=18.34) per utterance, while students' talk were shorter, averaging 9.6 tokens (SD=10.23) per utterance. In this study, we used the original, uncorrected chat lines to maintain the authenticity of the interactions.

Furthermore, pedagogical dialogue acts within the dataset were categorized by SMEs in English using a conversational analysis approach based on Sacks et al. [43]. This classification focused particularly on pedagogical sequence types that identify subtler shifts in the dynamics of the conversations, providing insights into the evolving interactions between tutors and students as shown in Table 2. To prepare the data for analysis, we conducted preprocessing techniques to equalize the dataset for synchrony calculation. Namely, we applied RegEx functions to expand contractions and remove punctuation marks. Each utterance is adjusted to alternate one at a time between the tutor and the student (i.e., multiple utterances by the same speaker in a row are combined). Furthermore, stop words were removed, and the rest of the words were stemmed and lemmatized before being tokenized to be utilized by each of the synchrony methods.

Table 1: CEFR Levels and Frequency Distribution

CEFR	Frequency	%
B1	35	13.46
B2	144	55.38
C1	29	11.15
C2	52	20.00

2.4 Analysis Procedures

This study consisted of three main phases: linguistic synchrony modeling, multivariate time-series feature extraction, and classification model building and evaluation. We detail the specific processes of each phase in this section.

2.5 Linguistic Synchrony Modeling

The linguistic synchrony methods used in this study are informed by a systematic review conducted by (Authors, blinded), which compares recent computational techniques for evaluating linguistic synchrony in dyadic educational conversations. We selected methods capable of measuring synchrony at the local or utterance level, enabling us to assess synchrony scores turn by turn and illustrate how synchrony evolves as a variable throughout the conversation

Table 2: Description of the Pedagogical Dialogue Acts

Talk Move	N	Description	Example
Clarification	37	Clarifying a previous utterance	Which one is more formal of the three?
Eliciting	614	Continuing to extract a response from the student	OK yes! Tell me a sentence, can you please? I mean with that exact example.
Enquiry	304	Asking for more information on a specific component	What does 'dissimilar' mean??
Opening	257	Marks the beginning of the conversation (e.g., greetings)	hi <student>
Redirection	10	Switching the topic	Ok, a bit of new grammar before we have to go!
Reference	3	Referring to an external source	I really recommend her Instagram account
Repair	277	Correcting a previous utterance, either to oneself or another	It's 'grey hair' in English
Revision	54	Returning to a previous topic	Let's quickly revise what we've done before.
Scaffolding	1256	Providing support	That's a good question! I would say 'the admin'
Topic Develop	350	Expanding on the current topic	Was there a recipe you liked in particular?
Topic Opening	191	Introducing a new topic	Anyway, can we talk about Lego?

or tutoring session. Our selection was aimed at capturing the multifaceted nature of linguistic synchrony, divided into three primary components: lexical synchrony, involving the alignment of word choices between conversational partners, assessed using cosine similarity and lexical overlap measures; syntactic synchrony, involving alignment in grammatical structures, measured through parse tree similarity and syntactic dependency patterns; and semantic synchrony, involving the alignment of meaning and topic coherence, assessed using various semantic representation methods. These techniques allow us to track adaptations in language use—lexical, syntactic, and semantic—across the interaction, providing insights into the dynamic nature of linguistic synchrony in educational settings.

2.5.1 Modeling Syntactic Synchrony. Shapira et al. [46] introduced a method for measuring syntactic synchrony in psychotherapy sessions using the Jensen-Shannon Distance across part-of-speech (POS) tag distributions (JSDuPos). This method converts spoken text from these sessions into POS tag distributions and calculates the Jensen-Shannon Distance (JSD) between them. JSDuPos quantifies the similarity in linguistic styles, with lower values close to 0 indicating greater synchrony, and higher values closer to or equal to 1 indicating linguistic divergence.

2.5.2 Modeling Lexical Synchrony. ALIGN [15] is an open-source natural language processing tool designed to measure multi-level linguistic alignment in conversational interactions. This tool computes the degree of linguistic alignment by specifically measuring the reuse of words and phrases between speakers. For lexical alignment, ALIGN processes tokenized text into n-grams, while syntactic alignment uses POS-tagged versions of n-grams, generated with the PennPOS tagger. This approach discards lexical units that are exactly the same, to prevent merely calculating lexical overlap [15]. Frequency vectors, based on each interlocutor's n-grams per turn, calculate the cosine similarity between lexical or syntactic structures. A higher score close to 1 indicates greater synchrony.

2.5.3 Modeling Semantic Synchrony. Several studies have introduced methods for evaluating semantic synchrony using various

types of contextualized embeddings and semantic distance measures. These methods vary from word embeddings such as GloVe and Word2Vec to transformer-based embeddings like BERT and ELMO, to accurately represent the meaning of conversations. Similarity between the embeddings is measured using distance metrics (e.g., cosine distance). In this study, we compared several methods that vary in their embedding representations to assess semantic synchrony. First, the Conversational Linguistic Distance (CLiD; [36]) method utilizes Word2Vec embeddings, and the minimum distance for each conversation pair is recorded as coordination. This accounts for a lag window between utterances using the Word Mover's Distance approach, where lower distance value indicates a greater synchrony.

Second, SemDist [60] adopts various other pre-trained embeddings such as BERT, ELMO, FastText, and GloVe. Utterances are transformed into embedding vectors based on these pre-trained models, and the cosine distance (i.e., 1 minus the cosine similarity) between each utterance's representations is calculated to represent the local alignment, with a lower distance indicating higher synchrony. Differences in embedding techniques—from traditional word embeddings like Word2Vec to more context-sensitive models like BERT—can lead to varying interpretations of meaning within conversations. Therefore, it is important to examine whether these methods produce similar synchrony patterns or if their results diverge based on the specific nuances captured by each embedding type.

2.5.4 Classification Model Development with the Time-series Synchrony Dynamics. The varying dimensions—lexical, syntactic, and semantic—of synchrony features were used to investigate the dynamics of synchrony throughout a tutoring session, treated as multivariate time-series data points. Specifically, we employed the Tsfresh [8] feature extraction method to analyze the time-series dynamics of synchrony, observing how different types of synchrony fluctuated throughout the tutoring conversation. Tsfresh offers a total of 63 time-series characterization methods, enabling the computation of up to 794 different time-series features for individual variables. These variables include descriptive indices such as average length, duration, or standard deviation of the features, as well

as more advanced features as shown in Table 3. These features were paired with an examination of the pedagogical strategies used by the tutor to understand their impact on student performance.

Once the set of features were extracted, we explored the connection between these features and the performance outcome (i.e., CEFR Levels) of the student participating in the tutoring session. We adopted various classification models, including the Support Vector Machines (SVM, [52]), Random Forests (RF, [5]), Logistic Regression (LR), K-nearest Neighbors (KNN, [24]), and Gradient Boosting Machines (GBM, [18]). We trained the models using features extracted from the Tsfresh algorithms. Since many of these features are often highly correlated, we removed those with Pearson correlations greater than 0.95, as well as variables with negligible variance, specifically those smaller than 0.01.. The evaluation of the model was conducted using 5-fold CV and the fine-tuning of each model was conducted using a grid-search approach, where the final SVM (C=0.1, linear kernel), RF (max depth=10, minimum sample split=2, n estimators=300), LR (C=1, L1 penalty), KNN (n neighbors=7, p=1, uniform weights), and GBM (learning rate=0.01 and 0.2, max depth=5, n estimator=0.01) as the final models.

3 Results

3.1 RQ1. Linguistic Synchrony Measures and Pedagogical Dialogue Act

To address RQ1, we investigated synchrony methods that represent the lexical, syntactic, and semantic dimensions of the tutoring sessions. The average synchrony measures for syntactic synchrony were extracted from JSDuPOS and ALIGN (Penn POS distribution features). Lexical synchrony was evaluated using the ALIGN method (lexical token overlap). Lastly, semantic synchrony was calculated in relation to the pedagogical dialogue acts using CLiD (Word2Vec+WMD) and variants of semDist (using BERT, FastText, and GloVe embeddings with cosine distance).

Among the semantic synchrony measures, the *reference* and *clarification* dialogue act showed the notably lowest average values across all four methods of semantic synchrony, as illustrated in Figure 2 and Table 4. Since these metrics are distance-based, where a lower value indicates a closer distance, this suggests that making references and clarification is typically related to higher semantic synchrony on average. However, it is important to note that both categories appeared significantly less number of times as discourse act, limiting its representation in the dataset. This infrequent occurrence could lead to higher variability in the synchrony metrics and may not fully capture the role of reference acts in fostering semantic alignment. Conversely, the *revision* and *opening* dialogue acts indicated the lowest semantic synchrony, or highest semantic divergence.

Due to the non-normal distribution of the data, a non-parametric statistical test, the Kruskal-Wallis H test, was conducted to evaluate whether any dialogue acts exhibited significantly different synchrony values. Additionally, Dunn's post-hoc test with a Bonferroni correction was applied to pairwise comparisons to determine which dialogue acts significantly differed from each other. In summary, *opening* was often identified as the most semantically and lexically divergent compared to the other dialogue acts. For instance, across the semantic measures, the dialogue acts of opening showed

significant differences in semantic values when compared to enquiry (p<.001) and topic development (p<.001). Similarly, the lexical measure again indicated that opening was significantly different from other categories, including eliciting (p<.001), enquiry (p<.001), repair (p<.001), revision (p=.012), scaffolding (p<.001), topic development (p < .001), and topic opening (p = .008). Syntactically, the *'topic'-related* dialogue acts often showed significantly high synchrony. The JSDuPOS score revealed significant differences in syntactic synchrony between topic development and eliciting (p=.011), opening (p=.036), repair (p=.001), revision (p=.013), and scaffolding (p<.001). Significant differences were found between topic opening and repair (p=.035), as well as between topic opening and scaffolding (p=.015).

3.2 RQ2. Student Performance based on Synchrony Patterns in the Sessions

A total of 320 Tsfresh features were extracted after feature filtering (out of 396) was applied as described in section *2.5.4*. Separately, a total of 11 features that are dummy coded to identify the total frequency of pedagogical dialogue acts were extracted. Figure 2 provides an overview of how the final Tsfresh features were extracted to model the time-series patterns of the linguistic synchrony in the three dimensions–semantic, syntactic, and lexical. Table 5 showcases the performance results of the five classification models, where the left side of the table showcases the classification model accuracy when no pedagogical dialogue act features were introduced, hence solely based on the time-series characteristics of the semantic, syntactic, and lexical linguistic synchrony features. By contrast, the right side showcases when the pedagogical dialogue acts frequency features were introduced. We noticed that the overall performance across all five models improved. The performance improvement was most drastic in GBM, which achieved the accuracy of 0.808, precision of 0.814, recall of 0.808, and F1 of 0.810 when both the time-series features and the pedagogical dialogue acts were presented in the model training. This represents 17.4% higher accuracy and 17.1% higher F1 score in GBM performance when the pedagogical dialogue acts were introduced, indicating the significant role that pedagogical dialogue acts have in predicting student performance based on their tutor-student interaction over the period of tutoring sessions.

3.3 RQ2. Student Performance based on Dynamic Synchrony and Pedagogical Patterns

We identified the top 10% of those with the highest importance based on from the Tsfresh and dialogue act features. The importance of each feature in the Random Forest model was determined by the average decrease in impurity, calculated as the sum of the Gini impurity reduction each feature provided across all trees in the forest. Features that resulted in larger reductions in impurity were assigned higher importance scores. The total importance was normalized so that the sum of all feature importance values equaled 1. For the Gradient Boosting algorithm, feature importance was determined by the cumulative error reduction associated with each feature, normalized so that the total importance across all features

Table 3: Example Tsfresh Feature Descriptions and Equations

Feature	Equation	Description
CWT coefficient	$cwt = \frac{2}{\sqrt{3}a\pi^{14}}(1 - \frac{x^2}{a^2})(-\frac{x^2}{2a^2})$	The frequency of synchrony at a specific point in time
Absolute Sum of Change	$n = \sum_{i=1}^{n-1}|x_{i+1} - x_i|$	The extent to which synchrony fluctuates
Absolute Energy	$E = \sum_{i=1}^{n}x_i^2$	Intensity of synchrony measures across the session
Lempel-Ziv Complexity	$C(s) = k + \frac{n-P(k)}{P(k)}$	Randomness and Regularity of synchrony
Mean Change Quartile	$M_1 = \frac{1}{n_1}\sum_{i=1}^{n_1}x_i, x_i \in Q_1 \ \Delta M_{12} = M_2 - M_1$	Change in the mean between quartiles
FFT Coefficient	$A_k = \sum_{m=0}^{n-1}a_m \exp(-2\pi i\frac{mk}{n})$	The strength and prominence of the synchrony
Approximate Entropy	$ApEn(m, r, N) = \phi_m(r) - \phi_{m+1}(r)$	Regularity and unpredictability of fluctuations

Table 4: Average Synchrony by Pedagogical Dialogue Acts

	Semantic				Lexical	Syntactic	
	Word2Vec +WMD	BERT +Cosine	FastText +Cosine	Glove +Cosine	Token	JSDuPOS	PennPOS
Clarification	0.398	0.255	0.500	0.222	0.045	0.585	0.498
Eliciting	0.437	0.295	0.766	0.309	0.038	0.655	0.444
Enquiry	0.400	0.277	0.648	0.249	0.034	0.591	0.447
Opening	0.645	0.209	0.747	0.422	0.030	0.589	0.477
Redirection	0.406	0.326	0.680	0.352	0.039	0.645	0.373
Reference	0.302	0.191	0.022	0.070	0.040	0.415	0.468
Repair	0.445	0.280	0.710	0.290	0.036	0.664	0.436
Revision	0.516	0.337	0.792	0.504	0.056	0.751	0.422
Scaffolding	0.407	0.294	0.667	0.285	0.036	0.647	0.465
Topic Development	0.394	0.265	0.666	0.212	0.028	0.569	0.470
Topic Opening	0.413	0.259	0.705	0.255	0.032	0.585	0.467
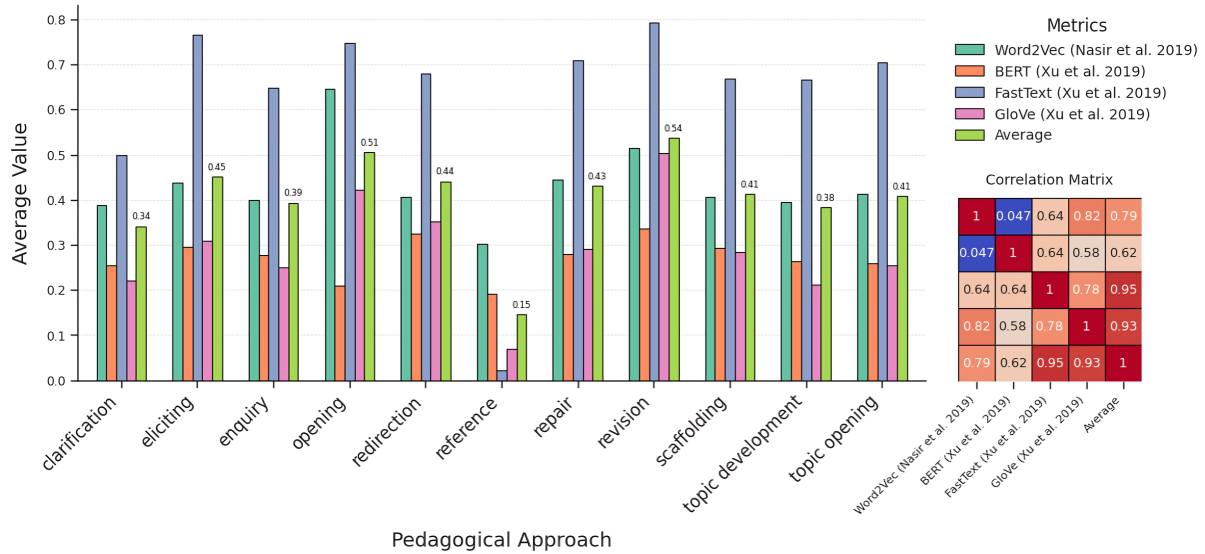
Figure 1: An Average of Semantic Synchrony Metrics by Pedagogical Dialogue Acts

summed to 1. Lastly, for the KNN algorithm, we used a permutation-based approach to compute feature importance. As shown in Figure 3, the importance of the features is ordered by each method, with the top categories representing the dialogue acts, lexical, semantic, and syntactic features.

In terms of pedagogical dialogue acts, the frequency of scaffolding, eliciting, and enquiry were found to have the highest importance in the GBM model, while the frequency of revision dialogue acts in the SVM model were identified as key pedagogical strategies predictive of student performance. Scaffolding and enquiry dialogue
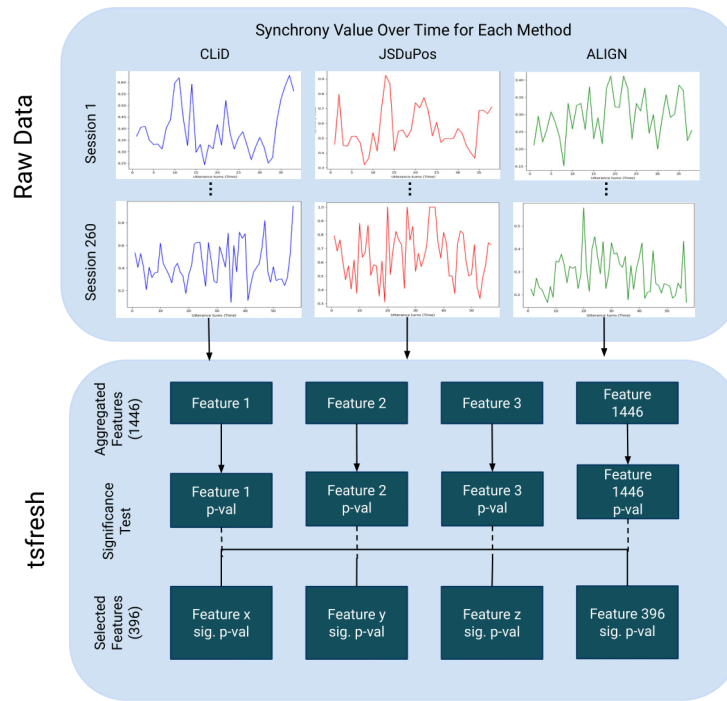
Figure 2: Feature Extraction using Tsfresh Algorithm

Table 5: Classification Model Performance with Tsfresh Features and Dialogue Acts

Methods	Without Pedagogical Dialogue Acts				With Pedagogical Dialogue Acts			
	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
SVM	0.707	0.777	0.712	0.723	0.731	0.742	0.673	0.682
RF	0.711	0.794	0.769	0.774	0.827	0.820	0.808	0.812
LR	0.669	0.754	0.673	0.673	0.788	0.820	0.750	0.756
KNN	0.707	0.803	0.788	0.788	0.788	0.797	0.788	0.786
GBM	0.688	0.773	0.692	0.692	0.808	0.814	0.808	0.810

acts occurred most frequently in higher performance groups, while eliciting and revision dialogue acts appeared more frequently in the other groups. Additionally, the frequency of eliciting dialogue act was also identified as a key predictive feature of student performance in the RF model (0.05).

For lexical features, the GBM model assigned the highest importance to a CWT coefficient with a larger window (n=14), indicating that changes in lexical synchrony over a longer conversational context were the most significant variable. Similarly, the GBM highlighted the FFT absolute coefficient for a larger context (n=26), emphasizing the importance of lexical synchrony over a broader window of synchrony changes. We found that the lower-performance group tends to have higher values for the CWT coefficient (B1:0.132 and B2:0.087) and higher FFT coefficient (B1:0.314, B2:0.347) compared to the higher-performance group (C1:0.080, C2:0.051 and C1:0.296, C2:0.292). Higher values in these coefficients indicate stronger long-term lexical synchrony that is persistent and smooth between the tutor and student, while lower values generally signify weaker or less consistent lexical synchrony over a longer window, suggesting rapid and short-term changes in lexical choices. The SVM model also emphasized the importance of FFT and CWT coefficients, and it also assigned the highest importance to the mean change quartile and Lempel-Ziv complexity (n=3), which identifies the repetitiveness of the lexical synchrony pattern over a short-term window of conversational turns. This pattern also showed a decrease in value as student performance level increased (B1:0.356 vs. B2:0.306 vs. C1:0.300 vs. C2:0.289), indicating that, in terms of short-term lexical alignment, lower-performance students exhibited higher lexical mimicry of the tutors than others. The RF model showed similar behavior, where the C3 lag score was identified as the most predictive feature. To summarize, both long-term and short-term lexical synchrony intensity and repetitiveness were predictive of student performance, with lower-performance students demonstrating higher synchrony in lexical use to the tutors in both contexts.

Gradient Boosting Importance		SVM Permutation Importance	
Freq of Scaffolding	0.06	0.02	Freq of Revision
Freq of Eliciting	0.02	0.03	Mean Change Quantiles (0–1.0, abs)
Freq of Enquiry	0.01	0.03	Lempel–Ziv Complexity (3 bins)
CWT 14 w 10 (2, 5, 10, 20)	0.04	0.03	Number of Peaks (10)
FFT Absolute Coefficient 25	0.03	0.03	Quantile 80%
FFT Absolute Coefficient 4	0.03	0.02	Approximate Entropy (m=2, r=0.1)
Max Linear Trend Intercept (chunk 5)	0.02	0.02	FFT Absolute Coefficient 14
Lempel–Ziv Complexity (3 bins)	0.01	0.02	FFT Absolute Coefficient 19
Linear Trend Intercept	0.01	0.02	Number of Peaks (3)
Number of CWT Peaks (5)	0.01	0.04	Absolute Energy
Quantile 80%	0.07	0.04	Mean Change Quantiles (0–0.2, abs)
C3 Lag 1	0.06	0.04	CWT Coefficients (2, w2): (2, 5, 10, 20)
Quantile 60%	0.04	0.04	Number of Peaks (5)
Lempel–Ziv Complexity (5 bins)	0.02	0.04	Sum of Values
Min Linear Trend Intercept (chunk 5)	0.02	0.03	CID CE (not normalized)
Approximate Entropy (m=2, r=0.5)	0.02	0.03	Fourier Entropy (100 bins)
Mean Change Quantiles (0.4–0.8, abs)	0.02	0.03	Quantile 80%
Variance Change Quantiles (0–0.8)	0.02	0.02	Ratio Beyond 1 Sigma
CWT Coefficients (2, w2): (2, 5, 10, 20)	0.02	0.05	Approximate Entropy (m=2, r=0.3)
skewness	0.02	0.05	Approximate Entropy (m=2, r=0.9)
Coefficient of Variation	0.01	0.04	Absolute Sum of Changes
Ratio of Value Number to Series Length	0.05	0.04	CWT Coefficients (14, w10): (2, 5, 10, 20)
Mean Change Quantiles (0–0.8, abs)	0.03	0.04	Has Duplicate Maximum
Mean Linear Trend Stderr (chunk 5)	0.02	0.04	Number of Peaks (1)
CWT Coefficients (14, w10): (2, 5, 10, 20)	0.02	0.03	Mean Change Quantiles (0–0.8, abs)
Max Linear Trend Stderr (chunk 5)	0.01	0.03	CID CE (normalized)
Variance Change Quantiles (0–0.8)	0.01	0.03	FFT Skew Aggregated
CWT Coefficients (3, w2): (2, 5, 10, 20)	0.01	0.03	Lempel–Ziv Complexity (100 bins)
FFT Absolute Coefficient 8	0.01	0.03	Minimum Value
Last Location of Maximum	0.01	0.03	Quantile 90%
Quantile 90%	0.01	0.02	Approximate Entropy (m=2, r=0.1)
Sum of Reoccurring Data Points	0.01	0.02	Variance Change Quantiles (0–0.8)

Figure 3: Average synchrony score shape obtained by averaging the sessions per pedagogical dialogue acts

In the case of semantic features, the GBM model identified the overall value of semantic synchrony across the session as a key feature, as evidenced by the high importance scores assigned to the quantile values of semantic synchrony. This pattern was identical in the RF model. The C3 lag score reflects how previous semantic patterns (such as topics, meanings, or concepts) immediately resurface in the dialogue. In the SVM model, important predictors of student performance included absolute energy (the overall intensity of the semantic score) and mean change (the intensity of semantic score fluctuations over the tutoring sessions). Interestingly, all of the coefficients showed lower semantic synchrony (with higher coefficient values) in the lower-performance group and higher synchrony (with lower coefficient values) in the higher-performance group. This indicates that the overall intensity (e.g., quantile, absolute energy) and short-term recurrence and repetitiveness of semantic topics (e.g., C3 lag score, mean) were much stronger in the higher-performance group.

Lastly, for syntactic features, the GBM model assigned the highest importance to the ratio of syntactic synchrony values, which indicates the density of syntactic synchrony occurrences. We observed that the higher-performance group showcased a higher ratio (B1:0.845, B2:0.859, C1:0.940, C2: 0.947). The SVM model highlighted approximate entropy indices as the most important features, signifying the predictability and variability of syntactic synchrony.

Again, higher performance was associated with higher entropy scores (B1:0.325, B2:0.422, C1:0.429, C2:0.452). These results indicate that the frequent occurrence of high syntactic alignment, which is repetitive and predictable, is predictive of students in the higher-performance group compared to those in the lower-performance group.

3.3.1 Time-series Pattern Comparisons. Comparison of patterns indicate differences in synchrony levels by the pedagogical dialogue acts as shown in Figure 4. We specifically focused on comparing the dialogue acts that were assigned as important features in the classification model decisions. At the syntactic (JSD) and semantic (WMD) level, lower values refer to higher levels of synchrony, whereas at the lexical level (ALIGN), higher values reflect higher synchrony. At the syntactic and semantic level, higher performing groups show lower distance values overall, displaying higher levels of synchrony throughout the session. Furthermore, the patterns of the dialogue acts align with our findings from the models' feature importance, where levels of consistency in long-term lexical synchrony contribute to predicting student performance.

The *eliciting* dialogue act at the semantic level shows that the tutor and student in the higher performing groups have greater levels of synchrony in comparison to the lower performing groups. As the tutor elicits information during the session, responses of students in the lower performance group appear not to align well

with the tutors' intended topic. At the lexical level, synchrony of the eliciting dialogue act improves for the lower performing groups towards the latter part of the session. However, at the syntactic and semantic level, initial synchrony values are maintained and worsen. This suggests that students from this group tend to replicate the vocabulary associated with the elicited topic; however, the use of these terms do not necessarily reflect full comprehension. For the *scaffolding* dialogue act, the higher performing group shows steady improvement in synchrony over time at all three dimensions of synchrony, indicating that the tutors' supports are effective in improving understanding. On the other hand, synchrony levels for the lower performing group remain lower and are steadily decreasing, suggesting that the content discussed in the session presents substantial challenge in their understanding. Alternatively, the enquiry dialogue act shows similar patterns between higher and lower performance groups. A prominent distinction in synchrony levels across the three dimensions is that higher performance groups exhibit prominently low levels of synchrony at the end of the tutoring session. This is depicted by high values at the syntactic and semantic level, and a flat line towards the 60-turn mark of the lexical dimension. This suggests that during the latter stage of the session, the questions raised refer to challenging content resulting in the disconnect of their shared understanding.

4 Discussion

4.1 Interpretation of the Results

4.1.1 RQ1 The Relationship between Pedagogical Dialogue Act and Linguistic Synchrony. We aim to identify the relationships between pedagogical dialogue acts and the various dimensions of synchrony. At the lexical level, there was notably less alignment at the beginning of the session, as indicated by significant differences in the synchrony values of the *opening* dialogue act that marks the start of the conversation, compared to most other dialogue acts (e.g., eliciting, repair). This finding supports the interactive alignment model [40], suggesting that certain aspects of language, such as lexical choices, tend to automatically align as conversations progress between participants. At the semantic level, *opening* significantly differed from *enquiry* and *topic development*, aligning with the findings at the lexical level, where alignment improves over time. Syntactically, *topic development* and *topic opening* showed significant differences from *repair* and *scaffolding*. These topic-related acts indicate shifts in content and potential changes in syntax. Cumulative structural priming, where repeated exposure leads to alignment in syntax, aids comprehension [17]. However, topic changes may disrupt this alignment, requiring the tutor and student to adapt to new syntactic structures.

4.1.2 RQ2. The Dynamics of Linguistic Synchrony based on Student Performance. We focused on evaluating how patterns of linguistic synchrony across the dimensions vary along with student performance. By extracting time series features of lexical, syntactic, and semantic synchrony at the local level, in conjunction with pedagogical dialogue acts, we gain insight on the nuances of linguistic synchrony between high and low performing groups. Thus, we blend synchrony research that individually evaluates insights from each dimension, presented in [30]. Our analysis demonstrates that

features of lexical, semantic, and syntactic synchrony, along with pedagogical dialogue acts– particularly the eliciting, scaffolding, and enquiry– play a key role in predicting student performance.

Results on lexical synchrony revealed that lower-performing students exhibited higher long-term lexical alignment, evidenced by the high values in CWT and FFT coefficients. In the short-term, Lempel-Ziv complexity showed that frequent lexical synchrony changes predicted performance. This can be explained by the task-based nature of tutoring dialogue, which is cognitively challenging; aligning language helps reduce cognitive load [42]. When the task is perceived as difficult, lexical synchrony tends to increase especially for language learners [41]. Our study supports this, revealing that lower-performing students, likely facing more challenges, tended to closely mimic tutors' lexical choices.

Higher-performing students demonstrated frequent patterns of high syntactic synchrony with tutors, indicated by greater importance on entropy, reflecting better adaptation to syntactic structure. This contradicts Healey et al. [26]'s claim that interlocutors consistently diverge in syntax. Given the online, text-based nature of the sessions, students' ability to construct sentences with correct syntax may indicate their level of comprehension, aligning with findings from Wang and Wang [58]'s study on written structural priming in language learning. The predictability of syntactic features in higher-performing groups mirrors the findings of Branigan et al. [4], which showed that syntactic priming occurs as interlocutors incidentally mirror sentence structures, reducing cognitive load.

Additionally, improved levels of semantic synchrony in the higher performance group are more reflective of fluent speakers' comprehension ability (i.e., more advanced language ability). This was evident in the intensity of quantile and absolute energy features, as well as stronger short term recurrence and repetitiveness of topics in the higher-performing groups, supporting the claim by Garrod and Anderson [21] in that semantic coordination is attained locally with the joint effort of the interlocutors. This echoes the notion that as language learners' ability approaches that of a fluent speaker, they begin to utilize similar, but not identical language– indicating nuanced understanding [10]. Similarly, tutors also align their language with students of higher proficiency [48], contributing to their shared understanding.

The significance of specific dialogue acts coincides with Barnes et al. [2]'s study, which revealed that *elicitation* and *enquiry* dialogue acts support explanatory rigor in drawing out students' ideas revealing their learning. That is, elicitation and enquiry uncovers the quality of one's understanding, also exhibited by the patterns shown in Figure 4 in which higher performing groups had higher levels of synchrony. In this study, students from higher performing groups consistently had stronger levels of synchrony based on the *scaffolding* dialogue act compared to lower performing groups. Scaffolding has been established as a pedagogical move that allows tutors to provide academically challenging instruction for language learners [57]. While scaffolding is intended to provide support to students of any level, higher performing students may be better able to internalize these to assist their understanding. In contrast, lower performing students may face more challenges in integrating this support with their understanding despite the scaffolding provided.
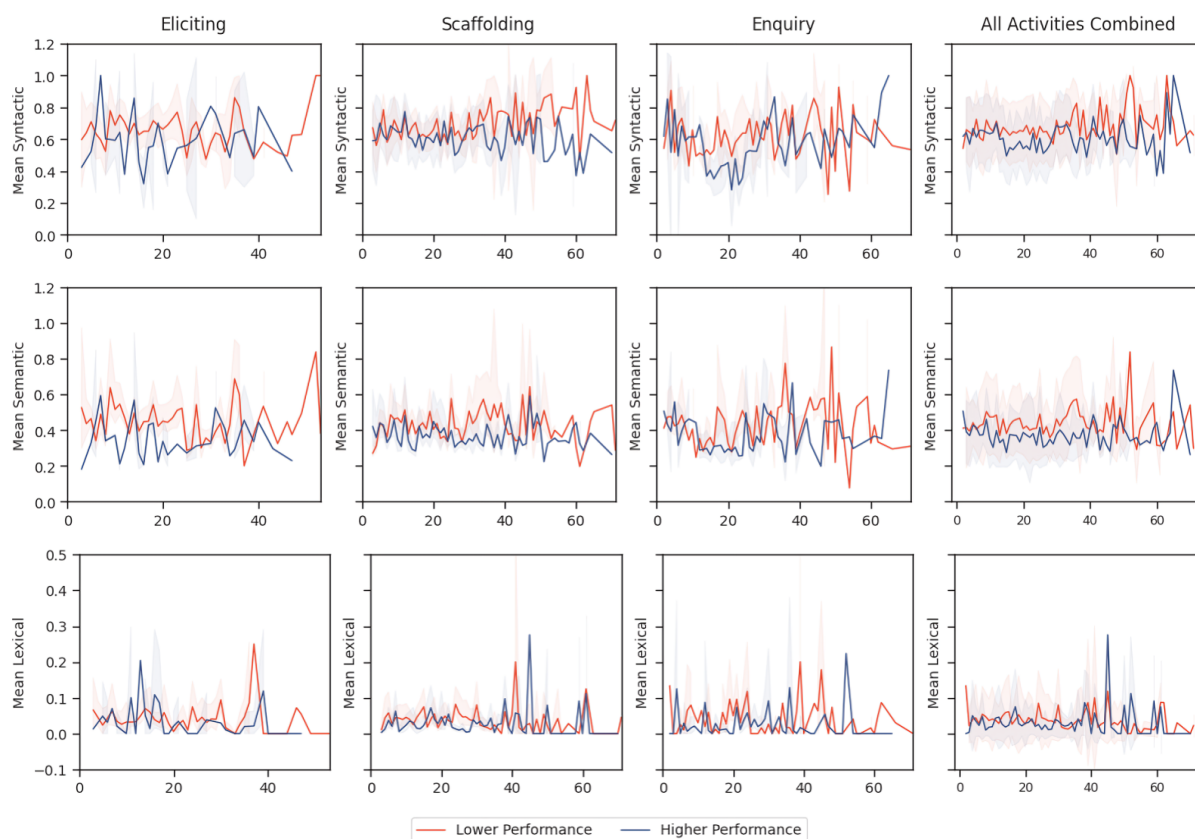
Figure 4: Average synchrony score shape obtained by averaging the sessions per pedagogical dialogue acts for three different conditions.

4.2 Implications and Future Directions

Our study evaluated the evolution of linguistic synchrony dynamics in a dyadic, online, text-based tutoring context, revealing distinctive patterns based on students' performance levels. The findings confirm that linguistic synchrony is dynamic across multiple dimensions, teacher and student adapting to each others' speech [19]. *Methodologically*, this study showcases the pipeline for analyzing diverse dimensions of linguistic synchrony in a dynamic manner by incorporating multivariate time-series feature analysis using Tsfresh. This approach allowed us to highlight specific patterns, often used in signal processing, to understand how synchrony differs across dimensions and between groups. Moreover, our methodological pipeline points to the potential for technological tools that track synchrony in real time, enabling tutors to adjust their instructional strategies during sessions to optimize learning outcomes. *Theoretically*, the pedagogical implications of this study highlight the critical role of linguistic synchrony in online L2 tutoring environments. In line with previous research, our results demonstrate that novice language learners exhibit different synchrony patterns compared to learners approaching fluency [9], shedding light on specific features that contribute to these differences and the effective pedagogical practices. For lower-performing students, stronger lexical synchrony suggests that tutors should focus on vocabulary

alignment, helping students mirror the tutor's word choices to enhance language acquisition. The identification of eliciting, scaffolding, and enquiry as key dialogue acts influencing synchrony underscores the importance of dynamic and adaptive instruction, where tutors actively encourage student responses and guide their understanding. Additionally, the weaker syntactic and semantic synchrony observed in lower-performing students, compared to higher-performing ones, indicates that appropriate scaffolding is needed to help students stay on topic. This involves providing tasks that encourage students to mimic not only the tutor's lexical choices but also the syntactic structure (grammar) of their speech to maximize learning. Furthermore, the ability of synchrony to reduce cognitive load suggests that tutors must carefully manage task complexity, particularly in online settings, where students may rely more on aligning with their tutor's language patterns to alleviate cognitive demands.

References

[1] Maha H. Alsoraihi. 2019. Bridging the Gap between Discourse Analysis and Language Classroom Practice. *English Language Teaching* 12, 8 (2019), 79–88.
[2] Evan R. Barnes, Ron Gray, and Anna S. Grinath. 2023. Talk moves as pedagogical tools for eliciting and working with student ideas in an undergraduate general biology laboratory. *Science Education* 107, 1 (2023), 89–123.
[3] G. Boulet. 2007. How does language impact the learning of mathematics? Let me count the ways. *Journal of Teaching and Learning* 5, 1 (2007).

[4] Holly P Branigan, Martin J Pickering, and Alexandra A Cleland. 2000. Syntactic co-ordination in dialogue. *Cognition* 75, 2 (2000), B13–B25.

[5] Leo Breiman. 2001. Random forests. *Machine learning* 45 (2001), 5–32.

[6] Andrew Caines, Helen Yannakoudakis, Helen Allen, Pascual Pérez-Paredes, Bill Byrne, and Paula Buttery. 2022. The teacher-student chatroom corpus version 2: more lessons, new annotation, automatic detection of sequence shifts. In *Swedish Language Technology Conference and NLP4CALL*. 23–35.

[7] Lin Chen, Barbara Di Eugenio, Davide Fossati, Stellan Ohlsson, and David Cosejo. 2011. Exploring effective dialogue act sequences in one-on-one computer science tutoring dialogues. In *Proceedings of the Sixth Workshop on Innovative Use of NLP for Building Educational Applications*. 65–75.

[8] Maximilian Christ, Nils Braun, Julius Neuffer, and Andreas W Kempa-Liehr. 2018. Time series feature extraction on basis of scalable hypothesis tests (tsfresh–a python package). *Neurocomputing* 307 (2018), 72–77.

[9] Albert Costa, Martin J Pickering, and Antonella Sorace. 2008. Alignment in second language dialogue. *Language and cognitive processes* 23, 4 (2008), 528–556.

[10] Scott A. Crossley. 2013. Assessing automatic processing of hypernymic relations in first language speakers and advanced second language learners: A semantic priming approach. *The Mental Lexicon* 8, 1 (2013), 96–116.

[11] Phung Dao, Pavel Trofimovich, and Sara Kennedy. 2018. Structural alignment in L2 task-based interaction. *ITL-International Journal of Applied Linguistics* 169, 2 (2018), 293–320.

[12] Simona Pekarek Doehler. 2019. On the nature and the development of L2 interactional competence: State of the art and implications for praxis. In *Teaching and Testing L2 Interactional Competence* (1st ed.). Routledge, 25–59. First Published 2019.

[13] Bruce P Doré and Robert R Morris. 2018. Linguistic synchrony predicts the immediate and lasting impact of text-based emotional support. *Psychological Science* 29, 10 (2018), 1716–1723.

[14] Nicholas D Duran, Amie Paige, and Sidney K D'Mello. 2024. Multi-Level Linguistic Alignment in a Dynamic Collaborative Problem-Solving Task. *Cognitive Science* 48, 1 (2024), e13398.

[15] Nicholas D Duran, Alexandra Paxton, and Riccardo Fusaroli. 2019. ALIGN: Analyzing linguistic interactions with generalizable techNiques—A Python library. *Psychological methods* 24, 4 (2019), 419.

[16] María Fernández-Toro and Concha Furnborough. 2018. Evaluating alignment of student and tutor perspectives on feedback on language learning assignments. *Distance Education* 39, 4 (2018), 548–567.

[17] Alex B. Fine and T. Florian Jaeger. 2016. The role of verb repetition in cumulative structural priming in comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 42, 9 (2016), 1362.

[18] Jerome H Friedman. 2002. Stochastic gradient boosting. *Computational statistics & data analysis* 38, 4 (2002), 367–378.

[19] Riccardo Fusaroli, Bahador Bahrami, Karsten Olsen, Andreas Roepstorff, Geraint Rees, Chris Frith, and Kristian Tylén. 2012. Coming to terms: Quantifying the benefits of linguistic coordination. *Psychological Science* 23, 8 (2012), 931–939.

[20] Riccardo Fusaroli and Kristian Tylén. 2016. Investigating conversational dynamics: Interactive alignment, interpersonal synergy, and collective task performance. *Cognitive science* 40, 1 (2016), 145–171.

[21] Simon Garrod and Anthony Anderson. 1987. Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition* 27, 2 (1987), 181–218.

[22] Jayne Goode and James D Robinson. 2013. Linguistic synchrony in parasocial interaction. *Communication Studies* 64, 4 (2013), 453–466.

[23] Grzegorz Grzegorczyk. 2018. Dialogue and language as factors contributing to transformative learning in academic tutoring. *Psychology of Language and Communication* 22, 1 (2018), 164–186.

[24] Gongde Guo, Hui Wang, David Bell, Yaxin Bi, and Kieran Greer. 2003. KNN model-based approach in classification. In *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings*. Springer, 986–996.

[25] Yugo Hayashi. 2023. Modeling Synchronization for Detecting Collaborative Learning Process Using a Pedagogical Conversational Agent: Investigation Using Recurrent Indicators of Gaze, Language, and Facial Expression. *International Journal of Artificial Intelligence in Education* (2023), 1–42.

[26] Patrick GT Healey, Matthew Purver, and Christine Howes. 2014. Divergence in dialogue. *PloS one* 9, 6 (2014), e98598.

[27] Sara Hennessy, Elisa Calcagni, Angel Leung, and Neil Mercer. 2021. An analysis of the forms of teacher-student dialogue that are most productive for learning. *Language and Education* 37, 2 (2021), 186–211. https://doi.org/10.1080/09500782.2021.1956943

[28] Carrie N Jackson. 2018. Second language structural priming: A critical review and directions for future research. *Second Language Research* 34, 4 (2018), 539–552.

[29] YouJin Kim, YeonJoo Jung, and Stephen Skalicky. 2019. Linguistic alignment, learner characteristics, and the production of stranded prepositions in relative clauses: Comparing FTF and SCMC contexts. *Studies in Second Language Acquisition* 41, 5 (2019), 937–969.

[30] YouJin Kim and Marije Michel. 2023. Linguistic alignment in second language acquisition: A methodological review. *System* 115 (2023), 103007.

[31] Sungjin Lee, Hyungjong Noh, Jonghoon Lee, Kyusong Lee, and Gary Geunbae Lee. 2012. Foreign language tutoring in oral conversations using spoken dialog systems. *IEICE TRANSACTIONS on Information and Systems* 95, 5 (2012), 1216–1228.

[32] Wenda Liu, Holly P Branigan, Lifen Zheng, Yuhang Long, Xialu Bai, Kanyu Li, Hui Zhao, Siyuan Zhou, Martin J Pickering, and Chunming Lu. 2019. Shared neural representations of syntax during online dyadic communication. *NeuroImage* 198 (2019), 63–72.

[33] Kim McDonough and Pavel Trofimovich. 2011. *Using priming methods in second language research*. Routledge.

[34] Marije Michel and Marco Cappellini. 2019. Alignment during synchronous video versus written chat L2 interactions: A methodological exploration. *Annual Review of Applied Linguistics* 39 (2019), 189–216.

[35] Marije Michel and Bryan Smith. 2017. Measuring lexical alignment during L2 chat interaction: An eye-tracking study. In *Salience in second language acquisition*. Routledge, 244–268.

[36] Md Nasir, Sandeep Nallan Chakravarthula, Brian Baucom, David C Atkins, Panayiotis Georgiou, and Shrikanth Narayanan. 2019. Modeling interpersonal linguistic coordination in conversations using word mover's distance. In *Interspeech*, Vol. 2019. NIH Public Access, 1423.

[37] Utku Norman, Tanvi Dinkar, Barbara Bruno, and Chloé Clavel. 2021. Studying alignment in a collaborative learning activity via automatic methods: The link between what we say and do. *arXiv preprint arXiv:2104.04429* (2021).

[38] Trena Paulus, Amber Warren, and Jessica Nina Lester. 2016. Applying conversation analysis methods to online talk: A literature review. *Discourse, context & media* 12 (2016), 1–10.

[39] Jinfang Peng, Chuming Wang, and Xiaofei Lu. 2020. Effect of the linguistic complexity of the input text on alignment, writing fluency, and writing accuracy in the continuation task. *Language teaching research* 24, 3 (2020), 364–381.

[40] Martin J Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences* 27, 2 (2004), 169–190.

[41] Martin J Pickering and Simon Garrod. 2006. Alignment as the basis for successful communication. *Research on Language and Computation* 4 (2006), 203–228.

[42] David Reitter and Johanna D. Moore. 2006. Priming of syntactic rules in task-oriented dialogue and spontaneous conversation. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 28.

[43] Harvey Sacks, Emanuel A Schegloff, and Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *language* 50, 4 (1974), 696–735.

[44] Paul Seedhouse. 2005. Conversation analysis and language learning. *Language Teaching* 38, 4 (2005), 165–187.

[45] Mi-Suk Seo and Irene Koshik. 2010. A conversation analytic study of gestures that engender repair in ESL conversational tutoring. *Journal of pragmatics* 42, 8 (2010), 2219–2239.

[46] Natalie Shapira, Dana Atzil-Slonim, Rivka Tuval-Mashiach, and Ori Shapira. 2022. Measuring linguistic synchrony in psychotherapy. In *Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology*. 158–176.

[47] Arabella Sinclair, Adam Lopez, Christopher G Lucas, and Dragan Gasevic. 2018. Does ability affect alignment in second language tutorial dialogue?. In *Proceedings of the 19th annual SIGdial meeting on discourse and dialogue*. 41–50.

[48] Arabella Sinclair, Jon Oberlander, and Dragan Gašević. 2017. Finding the zone of proximal development: Student-tutor second language dialogue interactions. In *SemDial 2017: SaarDial*. University of Amsterdam, 134–142.

[49] Arabella J Sinclair and Bertrand Schneider. 2021. Linguistic and Gestural Co-ordination: Do Learners Converge in Collaborative Dialogue?. *International Educational Data Mining Society* (2021).

[50] David Skidmore. 2000. From pedagogical dialogue to dialogical pedagogy. *Language and Education* 14, 4 (2000), 283–296.

[51] Yukyeong Song, Wanli Xing, Chenglu Li, Xiaoyi Tian, and Yingbo Ma. 2024. Investigating the relationship between math literacy and linguistic synchrony in online mathematical discussions through large-scale data analytics. *British Journal of Educational Technology* (2024).

[52] Shan Suthaharan and Shan Suthaharan. 2016. Support vector machine. *Machine learning models and algorithms for big data classification: thinking with examples for effective learning* (2016), 207–235.

[53] Višnja Pavii Taka. 2008. Vocabulary learning strategies and foreign language acquisition. In *Vocabulary Learning Strategies and Foreign Language Acquisition*. Multilingual Matters.

[54] Seng-Chee Tan and Aik-Ling Tan. 2006. Conversational analysis as an analytical tool for face-to-face and online conversations. *Educational Media International* 43, 4 (2006), 347–361.

[55] Amy Tsui. 2008. Classroom discourse: Approaches and perspectives. In *Language education in Europe: The Common European Framework of Reference*, N. Van Deusen-Scholl and N. Hornberger (Eds.). Springer, 209–226. https://doi.org/10.1007/978-0-387-30424-3

[56] Alexandria Katarina Vail and Kristy Elizabeth Boyer. 2014. Identifying effective moves in tutoring: On the refinement of dialogue act annotation schemes. In *International conference on intelligent tutoring systems*. Springer, 199–209.

[57] Aida Walqui. 2006. Scaffolding instruction for English language learners: A conceptual framework. *International Journal of Bilingual Education and Bilingualism* 9, 2 (2006), 159–180.

[58] Chuming Wang and Min Wang. 2015. Effect of alignment on L2 written production. *Applied Linguistics* 36, 5 (2015), 503–526.

[59] Jean Wong and Hansun Zhang Waring. 2020. Conversation Analysis and Second Language Pedagogy: A Guide for ESL/EFL Teachers. In *Conversation Analysis and Second Language Pedagogy: A Guide for ESL/EFL Teachers* (2nd ed.). Routledge, Chapter 1. https://doi.org/10.4324/9780429488023

[60] Yang Xu. 2021. Global divergence and local convergence of utterance semantic representations in dialogue. *Society for Computation in Linguistics* 4, 1 (2021).

[61] Reiko Yamamoto. 2019. THE ROLE OF TEACHER-STUDENT RELATIONSHIPS TO CAUSE SYNCHRONY IN EFL CLASS. In *LANGKAWI INTERNATIONAL MULTIDISCIPLINARY ACADEMIC CONFERENCE*. 280.

[62] Xiao Zhou and Chuming Wang. 2024. Effects of interactive alignment on L2 vocabulary learning by Chinese EFL learners. *Language Teaching Research* 28, 2 (2024), 466–496.