# Computerized Diagnostic Language Assessment of Oral Requesting-in-interaction: Proof-of-concept

Allan Nicholas
*Center for Language Research*
*University of Aizu*
Aizuwakamatsu, Japan
anich@u-aizu.ac.jp

Maxim Mozgovoy
*School of Computer Science and Engineering*
*University of Aizu*
Aizuwakamatsu, Japan
mozgovoy@u-aizu.ac.jp

John Blake
*Center for Language Research*
*University of Aizu*
Aizuwakamatsu, Japan
jblake@u-aizu.ac.jp

2025 13th International Conference on Information and Education Technology (ICIET) | 979-8-3315-3784-5/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICIET66371.2025.11046328

***Abstract*—In this paper, we provide a proof-of-concept for a computerized diagnostic language assessment (C-DLA) of spoken requesting among Japanese learners of English at a Japanese computer science university. The program focuses on the pragmatics aspect of spoken communication, in which the language choices we make are affected by socio-contextual factors such as relative social status and familiarity with the interlocutor. In so doing, the C-DLA aims to address a number of challenges. Requesting is an important, but challenging skill for many learners, and yet is also undertaught in the language classroom, due to time and resource constraints. Further, assessments typically evaluate the learner holistically, providing an overall score, without providing insight into specific aspects of the learner's performance. A C-DLA addresses these challenges by employing a three-stage process: i) assessment administration; ii) provision of immediate individualized feedback to the learner that promotes learning, and iii) further instruction provision, based on the assessment results. Computerization of the DLA widens access, allowing large numbers of learners to engage with the program simultaneously. Here, we provide a rationale for the C-DLA and an outline of the key challenges – namely, speech recognition of L2 English, identification of pragmatic inappropriateness in a learner's interactive speaking performance and automated feedback delivery. We further provide a proof-of-concept for the C-DLA, in which the program administers a number of interactive spoken requesting tasks to the learner, acts as an automated spoken dialogue interlocutor, and provides immediate, automated pragmatics-focused feedback when necessary.**

***Index Terms*—computer-assisted language learning, diagnostic language assessment, speaking assessment**

## I. Introduction

As part of an ongoing project, here, we describe a proof-of-concept for a computerized diagnostic language assessment (C-DLA) of spoken interactive English among Japanese L2 English learners at a Japanese computer science university. The C-DLA focuses on making requests in English and the pragmatic aspect of oral interaction – the role socio-contextual factors play in affecting our language choices. In this paper, we outline the importance of assessing speaking in an interactive context and the benefits and challenges of computerizing both oral interactions with an automated agent and the provision of immediate, automated feedback to the learner regarding their performance. We then describe a proof-of-concept for a C-DLA of spoken requesting-in-interaction, including the spoken dialogue system and feedback management system. We conclude with a discussion of implications and challenges.

JSPS Grant-in-aid for Scientific Research, Kakenhi, 23K00773

## II. Background

This multidisciplinary project draws on the theoretical background of diagnostic and dynamic language assessment (DLA), the pragmatics literature relating to spoken requests in English, and the assessment of requesting-in-interaction. Each of these is discussed in turn in the following subsections.

### A. Diagnostic Language Assessment

Language assessments to date have typically focused on the overall performance of a learner on a task, rather than providing feedback to learners and teachers on specific aspects of a performance that are positive or problematic [1], [2]. A DLA aims to address this by implementing three phases in an assessment: i) assessment administration; ii) providing feedback to the learner and/or teacher, and iii) providing further pedagogical instruction based on the results of the assessment. In this process, therefore, the aim is not only to evaluate a learner's abilities, but also to promote learning. One issue with a DLA approach to assessment relates to efficiency and the time and resource demands it places upon the teacher, or institution [3]. Computerising a DLA may offer one way to address this need to increase efficiency, allowing for administration to large groups of learners simultaneously. The automatized nature of such assessments, however, has typically led to the provision of feedback that is general in nature, not tailored to a specific learner's needs [4].

### B. English Spoken Requesting and Interaction

With approximately 40% of faculty members at the Japanese university in which this study is taking place being non-Japanese, English is the campus *lingua franca* [5]. As such, carrying out a spoken request in English is a common and important speech act for the students, who must regularly interact with both faculty and non-Japanese students. It is also a challenging speech act, however, due to the importance of the pragmatic aspect of communication, in which the learner must consider contextual factors when making their language choices. Such factors include *power* (P; similar

to relative social status), *social distance* (D; the degree of familiarity between interlocutors) and *rank of imposition* (R; how potentially burdensome a request might be to the hearer) [6].

Traditionally, a request speech act was conceptualized as a single pair of turns in speech as shown below [7]:

John: Can I borrow your textbook?
Jane: Sure, here you are.

In recent years, however, the interactive nature of making a request (requesting-in-interaction) has been increasingly emphasized, in which it is recognized that a request typically occurs within a conversation, and across extended sequences of talk [7], [8]. Conversation analysis studies seek to describe such sequences, identifying typical ways in which requesting interactions are organized by the interlocutors. Such features include a *pre-request*, in which, the speaker signals to the hearer that they intend to make a request ("Are you free next Thursday morning?"), and a *post-request*, in which the speaker may comment on the hearer's granting or refusing of the request [9], [10]. Additionally, there may be other features not specific to a request, such as a conversation *opening*, a *pre-closing*, in which a space in the conversation is opened, allowing for the interaction to end (*"I have to go, sorry."*) and a *closing* (*"Bye."*) [9].

### C. Requesting-in-interaction and Assessment

Assessing requesting-in-interaction has typically involved employing role-play-type tasks, allowing for elicitation of extended sequences of talk [8], [11]. Such assessments have commonly been administered in-person; as with DLAs, this may lead to efficiency challenges. Further, assessments have typically been holistic, evaluating the overall performance of a learner, rather than identifying specific positive or negative aspects of the performance [1]. A computerized DLA has the potential to address both efficiency and feedback issues; however, there are a number of challenges in developing a C-DLA of spoken requesting-in-interaction. The first relates to the conversation itself – a spoken dialogue system must be developed and employed that allows for the eliciting of extended sequences of talk between the learner and automated agent, with a reasonable approximation of authenticity. In addition, the C-DLA must be able to recognise specific instances of pragmatic inappropriateness by the learner in the interaction, and must comprise a second system implementing immediate, automated and individualized feedback to the learner on the specific identified problems arising during the interaction. In this paper, we outline a proof-of-concept for how such a C-DLA of spoken requesting may be implemented.

## III. Key Challenges

The C-DLA system for spoken requesting aims to address the unique challenges associated with evaluating and enhancing the pragmatic competence of Japanese learners of English. There are three key challenges to be overcome, namely speech recognition of learner English, identification of instances of pragmatic inappropriatness, and the development of a feedback delivery mechanism.

### A. Speech Recognition of the English of Japanese L2 English Learners

Speech recognition for non-native speakers, particularly Japanese learners of English, presents distinct challenges. Learner-specific phonological influences, such as vowel epenthesis (e.g. pronouncing *cream* as *kurimu*, following the Japanese syllabary) or consonant substitutions (e.g. substituting *d* for *th*), can reduce recognition accuracy. Additionally, speech variability arising from differing proficiency levels necessitates robust automatic speech recognition (ASR) models adapted to Japanese learners. Code-switching between languages further exacerbates the difficulty of recognition since the model not only has to identify the sound of the words, but distinguish between languages.

### B. Identification of Pragmatic Inappropriateness

Pragmatic competence involves understanding and producing language adapted to varying contexts. The language used when speaking with a professor, for example, may differ from that used with a close friend. The system identifies specific instances of pragmatic inappropriateness based on the typical pragmatic norms and conventions of the target community the learners are members of [8], [12]. This complexity is compounded by the variability and unpredictability in the interlanguage of learners.

### C. Feedback Delivery Mechanism

Effective feedback should be both specific, identifying precise aspects of the interaction deemed pragmatically inappropriate, and delivered in a manner that promotes learning [3], [8], [12]. Regarding the latter, a graduated prompt approach [12], [13] aims to promote learning by providing the least amount of feedback to the learner that allows them to successfully resolve the issue and continue with the task. Initially, highly implicit feedback is provided; if this does not help the learner to resolve the problem, further feedback is given, gradually increasing in explicitness, until the learner successfully resolves the issue. This approach, in which the learner takes maximal responsibility for resolving a problem, has been found to reliably promote learning and improvement in task performance [1], [14], [15].

## IV. Dialog Flow

Users of the system participate in a number of tasks, in which they produce spoken dialogues incorporating a request. Conversation analysis studies have identified common patterns in the way interlocutors typically co-construct request-based spoken interactions, known as *sequential organisation*. A request-based interaction's sequential organisation, for example, typically includes an opening, a pre-request, the actual request, a post-request, a pre-closing, and a closing [9]. The steps in a request-based conversation can be represented using a dialogue flow. Pragmatic inappropriateness can

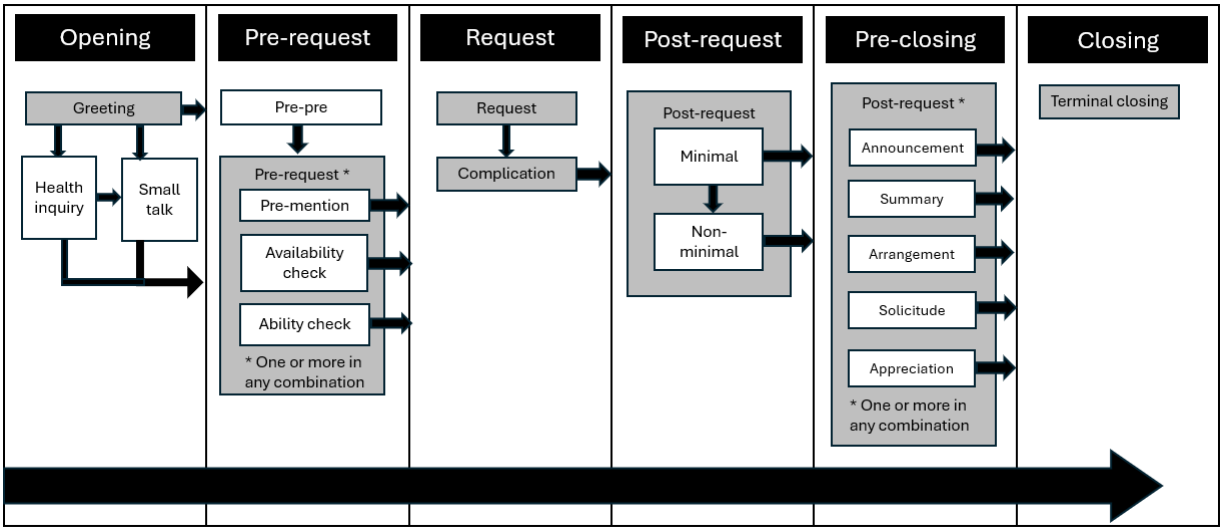


Fig. 1. Six-move dialogue flow

be identified within the dialogue flow based on pinpointing the stage the dialogue is currently in (such as the opening stage), identifying the absence of obligatory expressions or determining the presence of inappropriate expressions. Fig. 1 provides a visual representation of the flow of a dialogue that incorporates a request. This flow can be thought of a series of changes of states with each state mapping onto a sequential organisation stage in the dialogue structure. There are six stages, comprising obligatory components (coloured grey) and optional components (coloured white). Obligatory components relate to those stages of a request conversation deemed to be pragmatically important, and typically present in studies of request conversations. The system needs to be able to ascertain the users' current state to be able to identify pragmatic inappropriateness and deliver appropriate feedback if or when inappropriateness is detected. Taking the opening stage as an example, it can be seen there are three likely components, among which the greeting is obligatory while health enquiries (*"How are you?"*) or small talk are optional. Should a conversation begin with small talk (*"Cold today, isn't it?"*), the system should recognize that the obligatory greeting has been omitted, identifying pragmatic inappropriateness and triggering feedback.

Instances of pragmatic inappropriateness are flagged based on the mismatch between learner utterances and expected conversation stages or linguistic patterns. Both contextual appropriateness and linguistic accuracy are evaluated. Instances of pragmatic inappropriateness are detected using a two-pronged approach. The system monitors the stage of the request conversation, identifying transitions between stages (e.g., pre-request to request), and determines whether the utterance aligns with expected conversational moves (e.g., pre-request, request, or post-request). Further, specific lexical or grammatical patterns indicative of pragmatic inappropriateness can be identified, such as overly direct requests (e.g., *"I want you to... ."*) when more indirect requests (e.g., *"Would you mind ... ?"*) would be more appropriate.

## V. System Architecture

The system comprises three functional layers: a spoken dialogue layer, pragmatic inappropriateness identification layer and feedback layer as shown in Fig. 2.

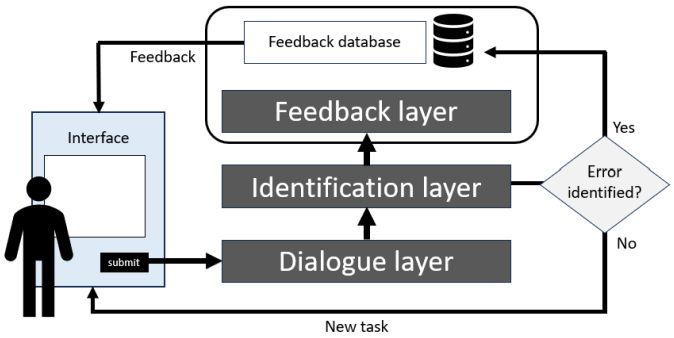


Fig. 2. Three-layer architecture

### A. Dialogue Layer

At the core of the system lies the dialogue layer, powered by a large language model (LLM) operating as an automated conversational interlocutor agent. Prompt engineering is employed to ensure the output of the LLM adheres to the expected conversational stages and supports the pedagogical goals of the system. This layer facilitates interactions, enabling learners to practice making requests in predetermined task scenarios.

### B. Pragmatic Inappropriateness Identification Layer

This layer is responsible for identifying the occurrence of specific instances of pragmatic inappropriateness. Inappropriateness may be identified through the presence or absence

of inappropriate language forms, e.g. overly direct or indirect language, or through the absence of an expected conversation stage, such as an opening greeting to initiate a dialogue.

### C. Feedback Layer

The feedback layer is responsible for delivering feedback based on identified inappropriateness. When an instance of pragmatic inappropriateness is detected, the system generates a feedback message tailored to the specific type of identified inappropriatess (for example, feedback will differ for an opening-related issue compared with an issue with the closing of the conversation). Graduated feedback ensures that a learner takes maximal responsibility for resolving the problem. Should implicit feedback fail to help the learner resolve an issue, the feedback gradually increases in explicitness until the problem is resolved. In this way, the feedback system adapts to a learner's specific needs.

## VI. Proof of Concept

This section provides a proof-of-concept for different components within the system, namely the request dialogue, state identification and pragmatic inappropriateness detection.

### A. General Design Considerations

The target system is aimed to be cross-platform and lightweight, allowing the users to access it from a variety of devices with minimal effort. In turn, the researchers should have easy access to user data to evaluate ongoing student progress.

These goals motivated us to design a system as a web application, accessible from a browser. The user can log in to a personal account, and continue work from the last checkpoint.

The server-side component of the system relies on the Open AI Whisper speech recognition system, a locally stored LLaMA large language model, and a number of handcrafted pragmatic inappropriateness detection rules, based on pattern matching. Open AI Whisper performs better on native speaker accents but demonstrates good performance on non-native speaker accents, including Japanese [16]. The accuracy is sufficient to recognise the speech of a single interlocutor with little background noise, resolving the first key challenge.

Generally, we envision a typical work session with the system as a dialogue between the user and the LLM-based "Interlocutor". This dialogue is relatively flexible in structure, but must include certain "trigger" utterances advancing the conversation. The automated interlocutor assumes a somewhat passive role: it replies to user utterances, but never takes the lead, forcing the user to produce trigger events.

All user utterances are evaluated by the "Tutor" assistant subsystem, responsible for feedback generation. It is the task of the tutor to detect instances of pragmatic inappropriateness and to make sure the dialogue is progressing (by suggesting to advance the conversation when the user is "stuck" in a lengthy sequence of turns of conversation without producing trigger utterances).

### B. User Interface

The current system supports both text and voice input, with the plan to focus on the latter. Similarly, Interlocutor's utterances are shown as text, but can be played using text-to-speech capabilities.

User and Interlocutor are shown as participants of a conventional messenger chatroom (see Fig. 3 and Fig. 4), while the Tutor assumes the role of an observing assistant. When Tutor detects an instance of pragmatic inappropriateness, the user is provided with feedback and asked to revise the utterance before it is sent to the Interlocutor again. When Tutor detects long sequences of conversation turns without trigger events, it urges the user to progress, but not directly intervening into the conversation.

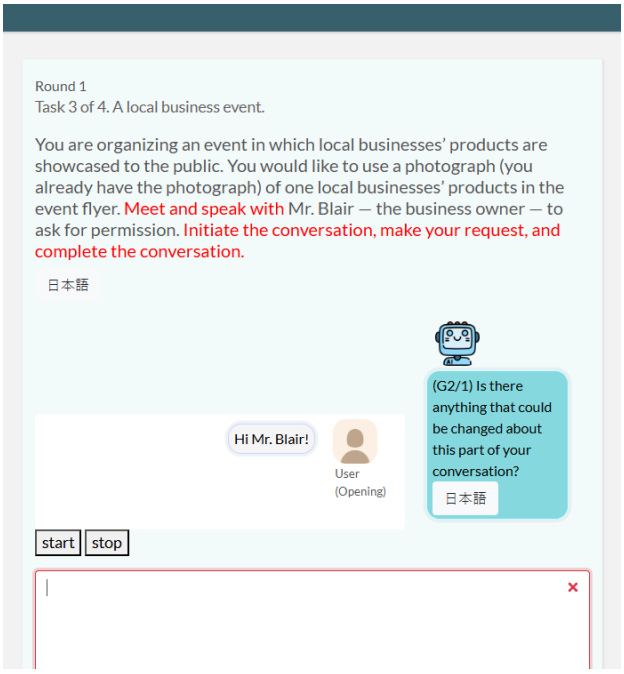


Fig. 3. Dialogue flow: Tutor responses to inappropriate greeting.

### C. State Identification

New dialogues start in the "opening" state (see Fig. 1), where the user is expected to greet the Interlocutor. Trigger utterances advance the dialogue into subsequent states, corresponding to typical elements of request-based interactions identified in the conversation analysis literature [9], and also aspects of spoken requesting in L2 English that have been identified as frequently being challenging for Japanese L2 English learners [8].

Detection is done via a combination of LLM-based and pattern matching-based approaches. Certain states, such as Opening, require the user to produce one of a closed set of possible phrases (e.g., *"Good morning, Mr. Smith."*), while any other utterance would be treated as pragmatically inappropriate. In such cases, state transition is controlled by a pattern-matching system, equipped with a repository of admissible set phrases.

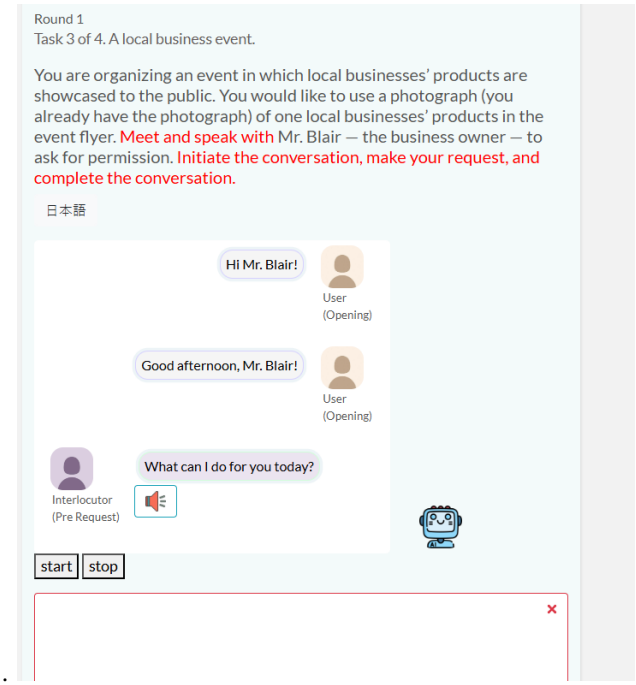


Fig. 4. Dialogue flow: Interlocutor responses to greeting.

Other states allow greater flexibility. For example, a Pre-request to Request transition can be accomplished in a variety of ways that are hard to reduce to keywords or set phrases. In these cases, we ask the LLM directly whether the desired condition is fulfilled by a user-supplied utterance.

### D. Pragmatic Inappropriateness Identification

Pragmatic inappropriateness identification is based primarily on pattern recognition. One of the primary goals of the system is to administer multiple request-based task scenarios, each with varying levels of power, social distance and rank of imposition, and provide feedback to the user based on their ability to appropriately adapt the directness and formality of their language to these changing factors. At our target educational level, a certain combination of these factors effectively translates into the requirement to use (or not use) specific set phrases, deemed appropriate for a given context. If the Tutor fails to detect a set phrase, it identifies an issue and generates feedback. For instance, overly direct phrases like *"I want you to…"* may be flagged based on power, social distance and formality factors.

While this approach is less flexible than LLM-based evaluation, it is consistent with our pedagogical goals. Our aim is to raise learner awareness of conventions and norms regarding the pragmatic aspect of their language choices within the academic community they are members of, and enable them to make informed language choices.

This way, Tutor evaluates incoming user utterances twice: with a pattern matching-based system designed to identify pragmatic inappropriateness, and with an LLM assessing whether the utterance triggers state transition. By harnessing pattern detection at lexical, grammatical and discourse levels; the system is able to detect pragmatic inappropriateness, resolving the second key challenge. The feedback delivery mechanism is triggered once inappropriateness is identified. This mechanism draws on the four-level graduated prompt approach described in [12], which resolves the third challenge.

## VII. Conclusion

A working prototype was implemented in Python using the Django framework, providing a robust and scalable foundation for the system. The Graphical User Interface integrates seamlessly with the LLaMA and Whisper API, while Django's modular architecture facilitates iterative development and testing of new features, such as expanded pragmatic inappropriateness categories and advanced state-tracking mechanisms.

We have shown that we were able to overcome the three challenges stated in Section III, namely speech recognition of learner English for Japanese speakers, identification of pragmatic inappropriateness and the development of a feedback delivery mechanism in our working prototype.

## References

[1] M. E. Poehner, J. Zhang, and X. Lu, "Computerized dynamic assessment (C-DA): Diagnosing l2 development according to learner responsiveness to mediation," *Language Testing*, vol. 32, no. 3, pp. 337–357, 2015.
[2] Y.-W. Lee, "Diagnosing diagnostic language assessment," *Language Testing*, vol. 32, no. 3, pp. 299–316, 2015.
[3] M. E. Poehner, *Dynamic assessment: A Vygotskian approach to understanding and promoting L2 development*. Springer Science & Business Media, 2008, vol. 9.
[4] L. Harding, J. C. Alderson, and T. Brunfaut, "Diagnostic assessment of reading and listening in a second or foreign language: Elaborating on diagnostic principles," *Language Testing*, vol. 32, no. 3, pp. 317–336, 2015.
[5] E. Kaneko, M. Park, I. Wilson, Y. Heo, D. Roy, T. Yasuta, A. Nicholas, and J. Blake, "English curriculum innovation for computer science majors in the japanese efl context: From needs to tasks," in *2018 IEEE International Professional Communication Conference (ProComm)*. IEEE, 2018, pp. 84–89.
[6] P. Brown and S. C. Levinson, *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press, 1987.
[7] G. Kasper, "Speech acts in interaction: Towards discursive pragmatics," *Pragmatics and language learning*, vol. 11, 2006.
[8] A. Nicholas, "Dynamic assessment and requesting: Assessing the development of Japanese EFL learners' oral requesting performance interactively," *Intercultural Pragmatics*, vol. 17, no. 5, pp. 545–575, 2020.
[9] J. Sidnell, *Conversation Analysis: An Introduction*. Wiley-Blackwell, 2010.
[10] A. Nicholas, "Profiling the development of a japanese efl learner's interactional competence: A dynamic assessment of requesting-in-interaction," *JLTA Journal*, vol. 21, pp. 42–64, 2018.
[11] C. Roever, *Teaching and testing second language pragmatics and interaction: A practical guide*. Routledge, 2022.
[12] A. Nicholas and J. Blake, "Profiling learner development with a computerized dynamic assessment of a Japanese learner's L2 email writing," *Research Methods in Applied Linguistics*, vol. 3, no. 3, p. 100164, 2024.
[13] M. E. Poehner and J. P. Lantolf, "Vygotsky's teaching-assessment dialectic and l2 education: The case for dynamic assessment," *Mind, Culture, and Activity*, vol. 17, no. 4, pp. 312–330, 2010.
[14] A. Aljaafreh and J. P. Lantolf, "Negative feedback as regulation and second language learning in the zone of proximal development," *The Modern Language Journal*, vol. 78, no. 4, pp. 465–483, 1994.
[15] A. Nicholas, J. Blake, J. Perkins, and M. Mozgovoy, "Evaluating the effectiveness of a computerised dynamic assessment of L2 English email requests," *Computer Assisted Language Learning*, pp. 1–33, 2024.
[16] C. Graham and N. Roll, "Evaluating openai's whisper asr: Performance analysis across diverse accents and speaker traits," *JASA Express Letters*, vol. 4, no. 2, 2024.