QIN, S. et al. (Eds.) (2025). Proceedings of the 1st International Conference on Learning Evidence and Analytics.
Asia-Pacific Society for Computers in Education

Advanced Learner Model for Error Collection Using Japanese Honorifics as an Example

Shubin QIN[a*], Wen GU[b], Koichi OTA[c] & Shinobu HASEGAWA[d]

[acd]*Japan Advanced Institute of Science and Technology, Japan;*

[b]Nagoya Institute of Technology, Japan.

[a]s2320007@jaist.ac.jp, [b]wgu@nitech.ac.jp, [c]ota@jaist.ac.jp, [d]hasegawa@jaist.ac.jp

Abstract: Over-reliance on context-unaware translation tools erodes learners' ability to detect pragmatically inappropriate expressions. This study develops a learner model for advanced learners to systematically collect and classify Contextual Errors, using Japanese Honorifics (Keigo) as a case study. The model enhances metalinguistic awareness through error visualization and provides a framework adaptable to other languages.

Keywords: Learner model; Advanced learners; Error collection; Honorifics.

1. Introduction

With the advancement of information technology, excessive reliance on translation tools and AI has led many language learners to struggle with recognizing their errors during practical language use—especially contextual errors, which refer to expressions that are grammatically correct but situationally inappropriate(Shu, Wang, & Zhai, 2025). Using Japanese Honorifics (Keigo) as a case study, this research aims to develop a Learner Model capable of systematically collecting such Contextual Errors.

The following are some related studies on Keigo: ①Shirado et al. (2007, 2011) developed a rule-based system for Keigo that corrects grammatical misused Keigo words; ② Matsumoto et al. (2022) created a task converting ordinary words into Keigo expressions.

However, existing systems and tasks primarily focus on correcting grammatical errors, while insufficient attention has been paid to recording and visualizing contextual errors—precisely what advanced learners require in open-world environments lacking clear reference standards.

To address this gap, we develop an error-collection model for advanced learners using Keigo as our test case. The model features: ① Dynamic contextual parameters as continuous variables; ② Context-sensitive tags for common Keigo expressions; ③ Visualized error patterns to cultivate pragmatic competence. Keigo no shishin【Guidelines for Keigo Expressions】(Agency for Japan Cultural Affairs, 2007) makes it an optimal testing ground.

2. Keigo as a Case Study for Contextual Errors

Keigo errors are categorized into two types: Grammatical Errors (e.g., tense or spelling mistakes) and Contextual Errors (Grammatically correct but mismatched with situational context).

Guided by Japan's Agency for Cultural Affairs, this study focuses on three dimensions: Social Status, Formality, and Relational Proximity. For example, when expressing "You've worked hard", there are two sentences (Table 1): [ご苦労様] and [お疲れ様でした], both two sentences are grammatically valid, but using [ご苦労様] toward someone of higher social status will be perceived as serious cultural disrespect, in this case, constituting a Contextual Error.

Table 1. *Example of lexicon in this model*

You've worked hard	Social status	Formality	Relational Proximity
ご苦労様	-1~-1	-2~0	-1~+2
お疲れ様でした	-1~+3	0~+3	-2~+3

The parameters in Table 1 are defined in Section 3.1

3. Learner Model in this study

This study proposes a learner model centered on recording and collecting contextual errors, distinct from error-correction mechanisms. The architecture consists of three core components: the Information Input Module, the Error Detection Module, and the Error Recording Module.

3.1 Information Input Module

The Information Input Module serves as the initial interface for learners. Its workflow comprises two sequential actions, the Contextual Parameter input section and the Text input section.

About Contextual Parameter definition, learners configure three continuous dimensions via sliders: Social Status [-3 (Low) ↔ +3 (High)], Formality [-3 (Informal) ↔ +3 (Formal)], Relational Proximity [-3 (Distant) ↔ +3 (Close)]. In this section, users will input different contextual parameters based on different situations. Such as, talking with professor will be like [Status=+3, Formality=+2, Proximity=-1], and talking with younger schoolmate [Status=-1, Formality=0, Proximity=+3].

After inputting the contextual parameters, the system will proceed to the Text input section. In this section, users will input text or sentences for error detection. Upon completion, the system automatically advances to the Error Detection Module.

3.2 Error Detection Module

The second module is the Error Detection Module, which detects Keigo expressions in the input text and determine whether contextual errors exist based on the contextual parameters provided by the Information Input Module.

Firstly, since many Keigo words derive from general forms, the Keigo lexicon in this model includes both the general forms and their variants: Sonkeigo (Respectful Language), Kenjougo (Humble Language), and Teineigo (Polite Language).

Secondly, this study focuses on Contextual Errors. For example, Keigo can be translated as "You've worked hard" into [ご苦労様] or [お疲れ様でした], but these two Keigo are used based on different Social Status, Formality, and intimacy levels. As shown in Table 1.

[お疲れ様でした] is highly formal, indicating deference to the listener's authority, and is appropriate when addressing superiors or clients. On the other hand, [ご苦労様] is neutral and suitable for peers, friends, or other informal situations. When the social status parameter is -1, using the formal [お疲れ様でした] would be overly polite and stiff. In this case, [ご苦労様] is a better choice. Similar distinctions apply to Kenjougo and Teineigo.

Therefore, it is necessary to identify the applicable range of each Keigo expression under different contextual parameters and assign appropriate contextual labels. These ranges are defined through a three-stage methodology:

(1) Rule extraction from authoritative guidelines: Baseline rules are extracted from the Keigo no shishin (Agency for Cultural Affairs, 2007), focusing on standardized Keigo usage patterns across Social status, Formality and Relational Proximity;

(2) Threshold validation through LLM-powered surveys: Parameter thresholds are validated using large language models (LLMs) to analyze authentic conversational corpora, identifying boundary values where honorific appropriateness shifts;

(3) Boundary calibration via multimodal analysis: Boundary values are calibrated using LLM-based contextual analysis that incorporates both linguistic features and sociopragmatic cues.

This multi-source approach ensures linguistically grounded and pragmatically validated range definitions.This study focuses on these three contextual parameters.

3.3 Error Recording Module

This module is divided into two sections visualized in Figure 1: the Real-Time Recording section (left half of Figure 1) and the Total Recording section (right half of Figure 1). Both sections contain a coordinate system and a table.

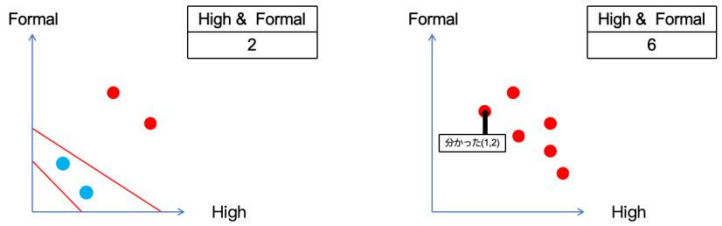

Figure 1. Example of Recording Module

In the Real-Time Recording section, a region is formed at coordinates based on contextual parameters from the Information Input Module, where each Keigo expression becomes a dot. Correct Keigo appears within this area, marked in blue. Conversely, Keigo outside the area is incorrect and marked in red. The table records errors occurring under these contextual parameters.

After the Real-Time Recording section, the model proceeds to the Total Recording section, as illustrated in the right half of Figure 1. In the Total Recording section, blue dots are eliminated after each session, while red dots remain. Moving the mouse over retained red dots in the Total Recording section displays error messages. Numerical values from the Real-Time section's table are automatically accumulated in the Total Recording section.

4. Conclusion

This study proposes a learner model for detecting contextual errors in Keigo by integrating dynamic parameters. The framework identifies grammatically correct but pragmatically inappropriate expressions, aiming to enhance advanced learners' metalinguistic awareness through error recording and visualization. Current limitations include subjective parameter scaling and limited adaptability to dynamic conversational shifts. Future work will focus on automating parameter calibration using sociolinguistic metrics, incorporating multimodal cues for context adaptation, and extending the model to other languages. This research highlights the potential to bridge grammatical accuracy and pragmatic competence in second language education, fostering culturally appropriate communication.

References

Agency for Cultural Affairs. (2007). Keigo no shishin: Kokugo ni okeru keigo no shiyō ni tsuite [Guidelines for honorific expressions: On the use of honorifics in the Japanese language]. Ministry of Education, Culture, Sports, Science and Technology, Japan.

Shirado, T. (2007). Development of a system to point out misuse of Japanese honorific. In Natural Language Processing, 1-31.

Shirado, T., Marumoto, S., Murata, M., Kanzaki, K., & Isahara, H. (2011). A system for judging the misuse of honorifics in Japanese: Quantifying the degree of validity of judgment rules for flexibility. In Natural Language Processing, 1-31.

Asahi, D., Okuhara, S., & Ito, T. (2015). Prototype and evaluation of a keigo learning support system. Proceedings of the 77th National Convention, 2015(1), 843-844. (2015 in Japanese)

Matsumoto, Y., Hayashizaki, Y., Kitayama, K., Funayama, H., Mita, M., & Inui, K. (2022). Proposal of a keigo conversion task for supporting Japanese language learners. Proceedings of the 36th Annual Conference of the Japanese Society for Artificial Intelligence, 3Yin239-3Yin239. (2022 in Japanese)

Kobayashi, I. (2022). Efforts to construct a Japanese keigo corpus considering social relationships. Proceedings of the 36th Annual Conference of the Japanese Society for Artificial Intelligence, 2C1GS601-2C1GS601. (2022 in Japanese)

Sekizawa, R., & Yanaka, H. (2023). Analyzing syntactic generalization capacity of pre-trained language models on Japanese honorific conversion. arXiv preprint arXiv:2306.03055.

Shu, Q., Wang, W. Z., & Zhai, J. (2025). The impact of artificial intelligence aids on students' reading and writing skills: An exploration of the balance between self-directed learning and dependency. Overseas English Testing Teaching and Research, 7(1), 25–30. https://doi.org/10.12677/oetpr.2025.71003 (2025 in Chinese)