

## 論文の英文要旨

論文題目	Criteria Features in L2 English of Japanese Learners Based on Complexity, Accuracy, and Fluency
氏名	Xiaolin Hu

The Common European Framework of Reference for Languages: learning, teaching, assessment (CEFR) has gained widespread recognition as a highly influential standard for language assessment, curriculum design, and materials development. It provides a vertical framework, where second language (L2) learners are classified into six proficiency levels and a horizontal framework offers different types of communicative language activities in which illustrative descriptors describe what language learners as social agents can do with language at different levels. One of the features of the CEFR is its deliberate under-specification in terms of language-dependent points, which avoids language-specific details such as grammar and lexis. Theoretically, this makes it applicable to various foreign languages, but this feature poses challenges for stakeholders in implementing the CEFR to identify specific language features of L2 learners. To enhance the practicality and usability of CEFR scales in various contexts, there have been attempts to identify *criteria features* that can distinguish one CEFR level from the others. However, limited research has been conducted on criteria features of L2 oral proficiency in English, particularly with regards to the aspects of complexity, accuracy, and fluency (CAF).

To address this gap, this dissertation aims to examine the CAF features that can serve as criterial features for distinguishing CEFR levels of L2 learners' oral performance. Furthermore, this research aims to empirically investigate the feasibility of establishing a CEFR prediction model using selected CAF features, and to determine the accuracy of this model in predicting the CEFR levels of L2 oral performance in dialogic and monologic tasks, respectively.

Using a quantitative/corpus-based approach, this study employed the National Institute of Information and Communications Technology Japanese Learner English (NICT JLE) Corpus as the primary database. From this corpus, a total of 170 error tagged dialogic transcripts, each spanning 15 minutes, were utilized for analysis. Based on the analysis of this corpus, a CEFR prediction model was constructed using CAF features. Subsequently, the reliability and validity of this model was confirmed through verification with 45 transcripts from a distinct database, the Telephone Standard Speaking Test (TSST) corpus, which comprises ten monologic tasks in each transcript.

The basic data analysis was carried out as follow. First, 36 CAF features were extracted from the NICT JLE Corpus, using both automatic tools and manual calculation. These CAF features were then analyzed to identify variations across different CEFR levels. The findings revealed that the majority of the CAF features exhibited significant differences over CEFR levels, with higher-proficiency speakers demonstrating better performance by utilizing more complex language structures with less errors and displaying greater fluency compared to lower-proficiency learners.

Second, a CEFR prediction model was constructed using selected significant CAF features as variables. Based on the results of Kruskal Wallis test, nine CAF features that demonstrated no significant contribution to distinguishing CEFR levels were excluded from further analysis. The remaining 27 CAF features showed significant differences between at least one pair of the three CEFR groups (A1, A2 and B1). Subsequently, a discriminant analysis was conducted to determine the most optimal combination of variables for inclusion in the CEFR prediction model. Additionally, two more CAF features were excluded from the model to address

multicollinearity issues, based on bivariate correlations between variables. Ultimately, seven CAF features were used to construct the CEFR prediction model, including three complexity measures (characters/word, clauses/AS-unit, and mean length of clause), one global accuracy measure (error-free clause ratio), and three dysfluency measures (repetitions/100 words, short pauses/100 words, and end-clause long pauses/100 words). Among them, clauses/AS-unit played a major role when predicting the three CEFR levels of L2 learners (from A1 to B1). The results indicated that 86.7% of cross-validated cases were correctly classified using these seven CAF features.

Third, to verify the robustness of this CEFR prediction model, a subsequent analysis was conducted by reapplying the model to a new dataset comprising the TSST Corpus. Despite differences in task types and transcript lengths between the two corpora, the results showed that 73.3% of L2 learners were correctly classified into the appropriate proficiency groups. However, it should be noted that due to differences in task types, lower-proficiency speakers in the TSST Corpus exhibited higher fluency in monologic transcripts, resulting in an overestimation of their proficiency levels, with only 60% of A1 speakers accurately predicted. In contrast, the prediction accuracy for A2 and B1 levels was 80% and 100%, respectively.

The findings of this dissertation confirmed that there are distinct differences in the CAF features among L2 learners at different CEFR levels. Moreover, some of these CAF features could be identified as potential criterial features that can effectively distinguish CEFR levels. Using the CEFR prediction model established in this study, it was observed that the CEFR level (from A1 to B1) of L2 learners could be predicted based on the seven major CAF features. Furthermore, the robustness of this model was verified, and its application scope was extended to both dialogic and monologic speech. However, the decreased prediction accuracy observed in the TSST Corpus suggests that adjustments might be required when applying this model to data of different task types in order to achieve higher prediction accuracy.

Limitations of this study mainly include several aspects: the lack of comprehensive CAF features, the inconsistency of data types between the two corpora,

the time-consuming calculation of CAF features caused by inevitable manual counting, and the limited sample size. These problems call for further investigation in future research.

Notwithstanding the limitations inherent in the present study, the identification of CAF features as criterial features made a valuable contribution to the research into L2 learner profiling using the CEFR. The findings promote the understanding of how L2 learners' oral performance progresses in terms of CAF measures at different proficiency levels. Furthermore, the findings of this study bear notable theoretical, pedagogical, and methodological implications. Theoretically, this study delves into the description and characterization of the CEFR level criteria in terms of associated spoken language features based on CAF measures. Pedagogically speaking, the criterial features and the CEFR prediction model proposed in this study hold potential for practical application, particularly in the development of CEFR-based teaching materials and their implementation, as well as the enhancement of self-assessment practices and automated assessment of spoken proficiency. From a methodological perspective, this study contributes a new methodological path to connecting existing CAF feature research to the CEFR criterial feature studies using learner corpora.