別記様式第4

<div align="center">論 文 の 要 旨</div>

ふ り が な	トウィン ミン ミン モウ ㊞
氏　　　名	THWIN MYINT MYINT MAW
論 文 題 目	Investigating Factors Affecting Lexical Diversity Measure Predictions of Writing and Speaking Proficiency: Word-Counting Criteria, L1 Background, Language Proficiency, and Text Length

論 文 の 要 旨

Lexical diversity (i.e., the different words) used in a written or spoken text is crucial in estimating L2 language proficiency. Various LD measures have been developed for vocabulary and language assessment. Because of the text length sensitivity of basic measures (*Types*; simple count of every word that occur once, and *Type-Token-Ratio*; the proportion of different words to total words), sophisticated measures with more complicated quantifications have been formulated. Previous research has validated existing LD measures' applicability in the L2 context and has found that LD measures can be reliable L2 general, writing, and speaking proficiency indicators.

Researchers (e.g., Treffers-Daller et al., 2018; Yu, 2010; Zenker and Kyle, 2021) have identified important factors that can influence the accuracy of LD measures in predicting these wider L2 proficiencies, namely: the analysis units used, L1 background, language proficiency, and text length. However, previous validation studies have been lacking in address and controlling and incorporating these four factors into L2 lexical diversity assessment because the studies have considered only one or two of these factors.

The current dissertation, therefore, addresses this important gap in LD research. The dissertation investigates whether LD measures predict inter- and intra-group writing variability under a controlled text length (200 words) and for a specific L1 background (Chinese). It also examines the extent to which LD measures predict speaking proficiency based on using different constant spoken-sample text lengths (200 to 450 words). It examines the extent to which LD measures predict writing and speaking using different-word counting techniques, focusing on the utility of the flemma count (a base word and its inflections under different word classes as the same types).

The dissertation comprises four experiments, the partial replications of Treffers-Daller et al.'s (2018) study. The first experiment was based on an entire population (N = 194 L2 English writers from mixed L1 backgrounds). It investigated the extent to which LD measures could discriminate between the three IELTS-based writing proficiency levels (6.5, 7, 7.5) under a controlled text length based with the different analysis units. The second experiment controlled the L1 background and so examined the extent to which LD measures predict the writing proficiency of an L1 Chinese L2 English learner group (N = 105).

The third experiment controlled both L1 background and language (writing) proficiency. It explored the extent to which LD measures to predict writing proficiency of L1 Chinese L2 English learners (N = 103) based on different writing proficiency levels (6.5, 7, 7.5). The fourth experiment analyzed the different participants (55 L2 English speakers from various L1 backgrounds). It examined whether LD measures were predictive of the IELTS-based speaking proficiency levels (6.5, 7, 7.5) based on the different analysis units and text lengths. It followed similar procedures to the first writing experiment to gain greater comparability of the findings of the LD measure predictions of two different language modes (L2 writing and speaking).

Overall, the four experimental studies' findings indicate that different analysis units influenced LD measure predictions of L2 language proficiency. Furthermore, LD measures were stronger L2 writing predictors than speaking predictors, and LD measures required longer constant text length for speaking than writing to achieve accurate predictability. This thesis concludes that LD measure predictions of L2 language proficiency is dependent on these four factors, so future LD research should consider and control them carefully.

備考　要旨は，日本語 4,000 字以内又は英語 1,500 ワード以内とする。