3L: Language, Linguistics, Literature® The Southeast Asian Journal of English Language Studies
Vol 29(1), March 2023 http://doi.org/10.17576/3L-2023-2901-09

Comparison of Lexical Complexity in L2 Speaking and Writing and Factors predicting English Speaking Proficiency

PARK, SHINJAE
Kookmin University, South Korea
muhando@kookmin.ac.kr

ABSTRACT

This study is to measure the three lexical complexity factors, i.e., lexical density, lexical diversity, and lexical sophistication using spoken and written data in a large-scale learner corpus, so that the lexical complexity can be observed in multidimensional terms. The study goes beyond most previous work on lexical complexity in SLA by examining the effects of all three measures of lexical complexity on predicting speaking proficiency of Korean EFL learners. Logistic regression analysis reveals that the verbal element of the lexical density factor was the most appropriate predictor of English proficiency (odds ratio: 2.259). In other words, the more frequently English learners use verbal elements, the higher their English-speaking ability is likely to improve 2.259 times. The results of this study will help to understand the correlation and difference compared to the results of other literatures in this field. Also, it will help educators to understand that the relationship between lexical complexity factors and proficiency is different in English speaking and writing, and to select more useful indicators when evaluating L2 learners. In addition, it is expected that it will serve as an evaluation method that can be used as an alternative in the field of pedagogy because it provides L2 discriminant function to evaluate learners.

Keywords: Lexis; Lexical complexity; Learner corpus; Monologue; Logistic regression

INTRODUCTION

Researchers in the areas of applied linguistics and second language acquisition have always been interested in measuring second language performance. One way to conceptualise the language performance is to measure the components of complexity, accuracy, and fluency (Housen, Kuiken, & Vedder, 2012). These three constructs, hereinafter abbreviated as CAF, have proved useful measures of second language performance in numerous studies (Skehan, 2009). Despite the wide use of CAF, some issues remain unresolved.

First, the focus on linguistic complexity is mainly on syntax aspects in task studies, so there is a tendency to lack lexical aspects strikingly (Skehan, 2009: 510-532). Skehan argues that one of the major omissions in performance measures is that of lexis (Skehan, 2009), therefore CAF competencies need to be supplemented by the characteristics of lexical use. Not only because empirical evidence suggests that the latter is a separate aspect of overall performance, but also because lexical access and retrieval figure prominently in all models of language production. Therefore, it is necessary to expand the range of measurement to the lexical competence evaluation. Second, among the lexical diversity, density, and sophistication known as the components of lexical complexity, most of the previous studies have been focused on only one or two components. Therefore, it is necessary to design a research that can capture the characteristics of these three lexical complexity factors. Third, it is important to investigate the difference in lexical complexity depending on modes, based on the existing claim that the representation of linguistic complexity differs according to speaking and writing tasks (Halliday, 1994: 61). Especially, in case of spoken data, it needs to go through not only time-consuming process for converting voice recordings into

texts, but also more complex process including handling disfluencies such as fragments and fillers (Lu, 2012: 194-95). Therefore, it is meaningful to examine lexical complexity using spoken data through systematic converting phases by several trained researchers. Finally, it is reported that errors may occur when measuring the indices only by relying on automatic calculating tools, so there is a demand for a process to increase reliability by adding the final manual inspection step of researchers.

The purpose of this study is to measure the lexical complexity factors that have not been dealt much in Korean English learners' speaking and writing tasks until recently, and to explore the relationship with English proficiency through reliable analysis. In this study, the lexis of the second language learners was observed in two modes in multidimensional manner, so other researchers will easily understand the characteristics of lexical competence by comparing the present results with their follow-up studies. In addition, it is expected that these results will serve as a useful information in EFL education context, because it will help educators understand the difference in lexical use in English speaking and writing, broaden their understanding of the different use of lexical complexity according to the modes, and help them select useful indicators when evaluating L2 proficiency.

LITERATURE REVIEW

Linguistic complexity is commonly defined as “the ability to use a wide and varied range of sophisticated structures and vocabulary in the L2” (Housen et al., 2012, p. 2). Linguistic complexity encompasses two dimensions: lexis and syntax. According to Skehan (2009: 514), lexis has been strikingly absent in task research, which is a serious omission, since the lexis–syntax connection is vital in performance models (Skehan, 2009), and lexis represents a form of complexity that has to be assessed in second language speech performance if any sort of complete picture is to be achieved.

In addition, Lu (2012: 190) empirically demonstrated that lexical complexity has been recognised as an important construct in L2 teaching and research, as it is directly related to the learner′s ability to communicate effectively in both spoken and written form. This study measures lexical complexity consisting of lexical density, lexical diversity, and lexical sophistication. In this section, I provide a brief overview of the lexical complexity measures that have been referred in this study. I also review empirical studies of the relationship of lexical complexity with L2 speaking proficiency and the different features of lexical complexity according to learners′ writing or speaking task.

LEXICAL DENSITY

Lexical density, originally coined by Ure (1971), refers to the ratio of the number of lexical (as opposed to grammatical) words to the total number of words in a text. According to Ure, items that do not have lexical properties can be described "purely in terms of grammar" (p. 445), meaning that such words (or items) possess a more grammatical-syntactic function than the lexical items. Lexical density is then defined as the total number of words with lexical properties divided by the total number of orthographic words. Spoken texts reportedly have a lower lexical density than written texts (Halliday, 1994; Ure, 1971). Ure also demonstrates that a large majority of the spoken

texts have a lexical density of under 40%, while a large majority of the written texts have a lexical density of 40% or higher. Engber (1995) uncovered nonsignificant correlations between lexical density and holistic ratings of L2 writing. In terms of the relationship between lexical density and lexical diversity, on the other hand, Johansson (2008) concluded that the two measures are interchangeable, and they show a similar developmental pattern independent of the measure used for describing lexical development.

It is theoretically possible that a text has high lexical diversity (i.e. contains many different word types), but low lexical density (i.e. contains many pronouns and auxiliaries rather than nouns and lexical verbs), or, vice versa, that a text has low lexical diversity (i.e. the same words or phrases are repeated over and over) but high lexical density (i.e. the words that are repeated are nouns, adjective or verbs).

Given that, Johansson (2008) asserted that a text with a high proportion of content words generally contains more information than a text with a high proportion of function words. Thus, lexical density values are obtained by dividing the number of lexical items (e.g., verbal elements and noun elements) by the total number of tokens in each text.

LEXICAL SOPHISTICATION

Lexical sophistication, also known as lexical rareness, measures "the proportion of relatively unusual or advanced words in the learner's text" (Read, 2000: 203). Lexical frequencies are measures of the types of words used in a language production, compared with frequency lists from native English speakers' corpora analysis, to determine from what level the majority of words contained within a spoken or written sample come (Skehan, 2009).

Laufer (1994) and Laufer and Nation (1995) proposed the Lexical Frequency Profile (LFP) that looks at the proportion of word types in a text covered by the list of the first 1,000 most frequent words (K1 words), the list of the second 1,000 most frequent words (K2 words), the university word list, and none of these lists, respectively. Laufer and Nation (1995) claimed that LFP is a valid and reliable measure of lexical sophistication, but Meara (2005) challenged this claim by showing that it may not be able to "pick up modest changes in vocabulary size" (p. 32) (see Laufer, 2005). Higginbotham and Reid (2019), however, examined the validity of Laufer and Nation's (1995) LFP indicating that second language learners who received low scores for writing relied on high-frequency words (and vice versa). Although one can examine the details of LFP in different ways, the model also provides a measure of lexical sophistication, computed as the ratio of the number of sophisticated word types (i.e., the "beyond 2,000" words) to the total number of word types in a text (Wolfe-Quintero et al., 1998). Lee (2001) analysed the distribution of vocabulary in various genres included in the British National Corpus (BNC). According to Lee's research, only a very limited number of individual vocabulary was used in spoken language data and most of these individual vocabulary were included in the basic core individual vocabulary list (i.e., top frequency 2000 words).

On the other hand, in terms of cognitive burden, Skehan (2009) hyphothesised that planning opportunities make a big difference to the speaker's capacity to mobilise less frequent words. Bayazidi, Ansarin, and Mohammadnia (2019) examined speech monologs of 35 Iranian high-intermediate learners of English on three different tasks (i.e., argumentation, description, and narration) and discovered that the highest mean lexical sophistication score was obtained for the narration task and the lowest mean score was obtained for the argumentation task. In their argumentation task, the participants were asked to respond to the question whether money can

make people happy. The description task required the participants to describe someone they enjoyed spending time with. In the narration task, the participants were first asked to inspect a series of cartoon pictures with no text and then to narrate a story based on the pictures.

In the present study, I counted words as sophisticated, following Lee′s (2001) suggestion, the ratio of the number of types beyond the 2,000 most frequent words to the total types based on the British National Corpus (BNC).

LEXICAL DIVERSITY

The more varied a vocabulary a text possesses, the higher lexical diversity. For a text to be highly lexically diverse, the speaker or writer has to use many different words, with little repetition of the words already used (Johansson, 2008). Lexical diversity refers to the range of a learner's vocabulary as displayed in his or her language use (Lu, 2012; Yu, 2010; McCarthy & Jarvis, 2007).

The traditional lexical diversity measure is the ratio of different words (types) to the total number of words (tokens), the so-called type-token ratio, or TTR. However, this measure has been criticised for its sensitivity to sample size, as the ratio tends to decrease as the size of the sample increases (Hess, Sefton, & Landry, 1986).

In order to compare texts of different lengths, the D measure was presented by Brian Richards and David Malvern (Richards & Malvern, 1997; MacWhinney, 2000). The D measure is based on the predicted decline of the TTR, as the sample size increases (Johansson, 2008). Put differently, it is calculated through the TTR, which is corrected through calculating the D formula (Skehan, 2009). A computer program called vocd in CLAN (MacWhinney, 2000) provides a standardised procedure for measuring D (see Richards & Malvern, 1997). In essence, a higher value of D indicates higher lexical diversity, and thus a richer vocabulary (Gregori-Signes & Clavel-Arroitia, 2015).

Using D, i.e., VOCD, as a measure of lexical diversity, for instance, Yu (2010) reported that data from the Michigan English Language Assessment Battery (MELAB) showed significant positive correlations between the D measure and test takers' general proficiency, as well as the quality of their writing and speaking task performance. Lu (2012) found lexical diversity correlated most strongly with the raters' judgments of the quality of ESL learners' oral narratives and no effect for lexical density emerged, and a very small effect was found for lexical sophistication. Kormos (2004) examined 16 Hungarian learners of English and correlated measures of lexical diversity with fluency scores awarded by three experienced native and non-native speaking teachers. Strömqvist, Johansson, Kriz, Ragnarsdottir, Aisenman, and Ravid (2002) also used D to compare spoken and written expository and narrative texts produced by adults from four countries. The results show strong differences between speaking and writing, where writing has a much higher lexical diversity.

Another measure is MTLD (i.e., Measure of Textual Lexical Diversity) which is the average length of sequential word strings in a text which maintain a given TTR value (Kalantari & Gholami, 2017). McCarthy and Jarvis (2007) examined the validity of MTLD for lexical diversity assessment. To validate the MTLD approach, they compared it against the performances of the primary competing indices in the field, which include vocd-D and TTR. The comparisons involved assessments of convergent validity, divergent validity, internal validity, and incremental validity. The results of their assessments of these indices across two separate corpora suggest that MTLD and VOCD capture unique lexical information. Therefore, researchers are advised to use them in combination, rather than using any single index (McCarthy & Jarvis, 2007), noting that

lexical diversity can be assessed in many ways and each approach may be informative as to the construct under investigation.

One of the examples applied to SLA field is Hwang (2013)' study. She investigated the degree of lexical complexity in 319 Korean EFL college students' narrative writing. The four proficiency according to writing scores were compared with the level of lexical complexity which was assessed using the L2 Lexical Complexity Analyzer. As a result, only lexical diversity (i.e., variation) showed significant progress tendencies as L2 writing proficiency increased. In addition, Park (2013) examined whether different writing task (i.e., narrative and argumentative) elicit different lexical features using 75 university students' writing. The results revealed that there were significant differences between the two writing tasks and lexical diversity (TTR used in her study) was the best indicator of the writing proficiency, as was the result of Hwang (2013). On the other hand, Schnur and Rubio (2021) investigating the effect of task type on lexical complexity using writing from K-12 Spanish second language (L2) students. Results showed that all three measures (i.e., lexical density, lexical sophistication, lexical diversity) increase at each proficiency score between Novice High and Advanced. Diversity and sophistication were both shown to increase rapidly after the mid-point, indicating that a broad and deep lexical repertoire is a key feature of more advanced proficiency levels. Results for the different task modes indicate that text mode impacts learners' lexical density, while tasks that are more complex elicit higher lexical sophistication.

METHOD

This study focuses on the differences of lexical complexity between the spoken and written samples and the relationship of lexical complexity to the L2 proficiency of oral narratives produced by Korean learners of English. Specifically, we addressed the following main research questions.

> RQ1. Are there differences between spoken and written production of Korean EFL learners in terms of lexical complexity?
> RQ2. Which lexical complexity factors predict learners' speaking proficiency most reliably?

CORPUS

The data analysed in the present study included 139 writings and 224 monologues in Multi-language Learner Corpus (hereafter, MULC) of Korean university students (Park, 2021; 2022). The participants could join one or both tasks and choose one of the provided daily topics for each task. The writing task was assigned 30 minutes, and the monologue task was assigned 2 minutes. Writing was conducted using a Note program on a desktop computer so that internet search was not available, whereas monologue was conducted in a soundproof lab, and all data were recorded digitally in real time under the present author's supervision.

The topics are shown in Table 1. The collected monologue recordings were manually transcribed by dozens of trained researchers and finally confirmed by English native linguistic experts. Furthermore, prior to the actual evaluation, the linguistic experts went through a pilot test for 5% of the transcriptions, and all discrepancies in evaluation were solved through discussion. After a series of convergence process, they conducted an evaluation of learners' speaking ability

based on Common European Framework of Reference for Languages (CEFR) standard. The CEFR describes foreign language proficiency at six levels: A1 and A2, B1 and B2, C1 and C2 (from A1 for beginners up to C2 for those who have mastered a language).

TABLE 1. Topics Provided in Each Task

Monologue	Writing
1. What do you usually do in your free time? Hobby, etc. 2. What is your favorite genre of movies? 3. Do you think there can be friendship between opposite genders? 4. Is it better to have a dog than cat?	1. Should everyone get married? 2. Is it essential to wear school uniforms in middle and high schools? 3.Should elementary, middle, and high school students be allowed to carry phones in class? 4. Should any college student join a club?

Table 2 represents information about the students who participated in the monologue task. Most students majored in English (78, 35%), followed by engineering colleges (42, 19%), natural science colleges (24, 11%), social science colleges (23, 10%), and other foreign language majors (22, 10%). The reason why most students majored in English was that the data were collected by conducting public advertisements, and predominantly those students who were relatively confident in their English production volunteered to participate. The sample had also a balanced gender distribution; the mean age of the participants was 20.9 years old.

The English-speaking proficiency of the participants as measured by CEFR standard was 2.81 (SD=0.870), which is close to B1 (based on a 6-point scale with A1=1 and C2=6; A1, A2, B1, B2, C1, and C2, from the lowest to the highest level). None of the participants had C2 proficiency, which is a native speaker's production level. In particular, a large number of students were in the mid- and low-level proficiency groups, i.e., B1 and A2 (B1: 86 (38.6%); A2: 82 (36.8%)), and followed by B2 (42, 18.8%).

TABLE 2. Participant Information

Majors								
Natural Science	Business Admin.	Engineering	Education	Law	Social Science	English	Other Languages	Arts & Physics
24	15	42	9	8	23	78	22	3
11%	7%	19%	4%	3%	10%	35%	10%	1%
Male: 112, Female: 112								
Age: 20.9								
A1 (7, 3%), A2 (82, 37%), B1 (86, 38%), B2 (42, 19%), C1 (7, 3%), C2 (0, 0%)								
Total: 224 (100 %)								

MEASUREMENT

All text files in the dataset were pre-processed using NotePad++ before being analysed. Each script was removed the followings: the header; fillers including “ah,” “eh,” “er,” “mm,” “oh,” and “um”; pause period in seconds. In the original text files, a word form containing an obvious pronunciation error (e.g., “work” mispronounced as “walk”) was enclosed in a pair of angle brackets and

preceded by the correct form provided by the transcriber (e.g., "the sun has risen <rised>"). The cleaned text files were saved separately.

This process entailed the following steps. A cleaned text file was firstly part-of-speech (POS) tagged using the Stanford tagger, which assigns every token in the language sample a label that indicates its part-of-speech category, for example, noun, verb, adjective, adverb, and so on. Then, the final step of manually checking by researchers was added to reduce errors that may occur due to the automatic tagging process, it is expected to be significantly increased the accuracy of measurement using this corpus than the existing studies that rely only on the results obtained using one of the automatic taggers. Next, the cleaned sample was processed by Text Inspector, i.e., a text complexity analyzer which computed the values of the D value and lexical sophistication was initially proposed and prototyped by Professor Stephen Bax and further developed by the software team at Versantus (Youssef, 2019). In supplementation, AntConc3.2.5 tool was also used to increase the accuracy in calculating content words using the tagged sample.

ANALYSIS

Data in each mode was analysed using the Text Inspector, which is an online lexical profiling tool that analyses the vocabulary of texts and aligns them with CEFR. For each sample, Text Inspector produces not only frequent counts for various indices related to lexis but also native-like percentage ranging from 0% to 100% in terms of the use of lexis.

After the lexical complexity indices have obtained for each sample, a set of independent samples t-test were run to compare differences between spoken and written data for each of the lexical complexity measures. In order to find the most predicting factors on L2 proficiency in speaking English, it is necessary to determine whether to analyse it by discriminant analysis or binary logistic regression. Normality Assumption Test is performed to find the right one among the two methods. If a multivariate normal distribution is not available, Binary logistic regression is performed. In this case, the dependent variable is inevitable to be reduced to two dimensions (Low (A1, A2): 89, High (B1, B2, C1): 135).

Table 3 shows basic information about samples of speaking and writing; there were more words per sentence in writing (Speaking: 14.384 vs. Writing: 14.701), especially the native-like level, which is automatically evaluated based on CEFR standards through Text Inspector, is also higher in writing (Speaking: 45.226 vs. 49.162).

TABLE 3. Summary of Data

	Monologues	Writings
Number of samples	224	139
Total Words	28,149	32,581
Words per sample	156.67	234.40
Words per sentence (SD)	14.384 (6.369)	14.701 (7.740)
Percentage of native-likeness (SD)	45.226 (11.494)	49.162 (7.548)

RESULTS

DIFFERENCES BETWEEN THE SPOKEN AND WRITTEN DATA

In the first part of the analysis, I aimed to determine whether significant differences existed in the lexical complexity in spoken and written data and if yes, what they were. As Table 3 shows, the written data had higher mean values than the spoken data for 3 out of the 5 lexical complexity indices. Independent-samples t-test were run to determine which differences between them were statistically significant. The results suggest that the two groups differed significantly in all indices (VOCD: t=-13.327; verbal density: t=4.474; noun density: t=4.355; BNC(2K): t=9.132 (df=361, p<.05)). In detail, the scores in writing were higher in diversity (i.e., VOCD and MTLD), verbal density, but, in noun density and sophistication, the scores in speaking were higher. However, in general, the scores of lexical complexity in writing were higher than in speaking according to the evaluation of comprehensive lexical complexity through Text Inspector (See Table 4: Speaking 45.226 vs. Writing 49.162; Figure 1).

TABLE 4. Measures of Lexical diversity, Lexical density and Lexical Sophistication

Lexical Complexity		Monologue Mean (SD)	Writing Mean (SD)	*t*
Diversity	VOCD	48.802 (17.076)	74.910 (19.744)	-13.327
	MTLD	37.368 (15.942)	65.512 (17.894)	-15.593
Density	verbal E./S.	1.670 (0.787)	2.138 (1.208)	4.474
	noun E./S.	2.640 (1.113)	2.073 (1.339)	4.355
Sophistication	beyond BNC(2K)	0.36 (6.716)	0.30 (5.074)	9.132

(*df*: 361, *p*=.000)

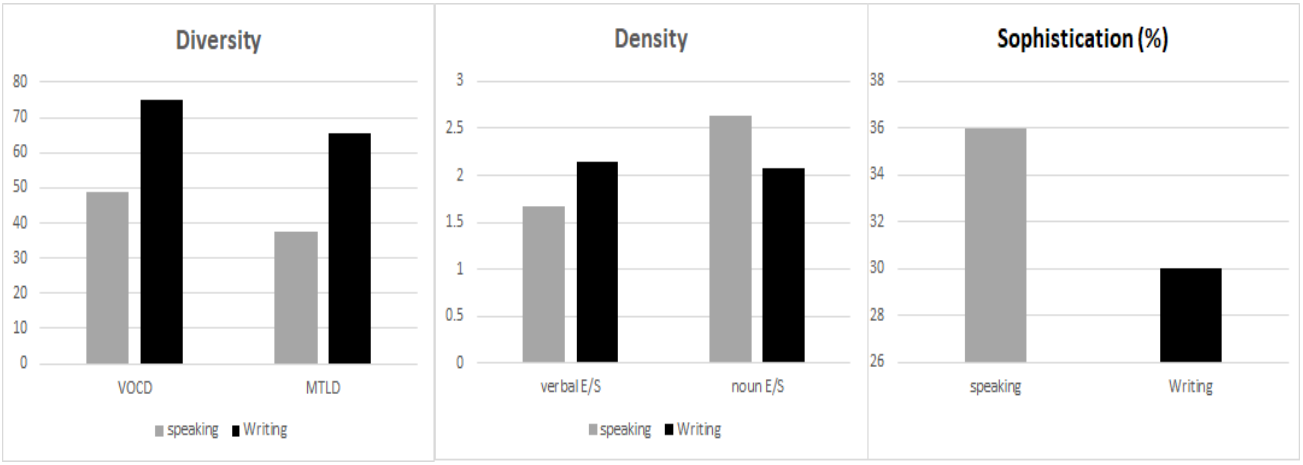

FIGURE 1. Lexical Complexity by modes

The discriminant analysis and the logistic regression analysis are two statistical techniques used for predicting group membership from a set of predictors (Abdulqader, 2015). Discriminant Analysis is applied to the situation when dependent variable has two or more categories and they are mutually exclusive. Discriminant analysis assumes that the data are normally distributed and there is an equal variance among all independent variables, but logistic regression does not assume such things (Park, 2022). In logistic regression analysis, dependent variables should be dichotomous, and its probability or odds will be modeled based on the combination of predictors. That is, it is well known that if the populations are normal with identical covariance matrices, discriminant analysis estimators are preferred to logistic regression estimators, but under non-normality, it is preferred the logistic regression model free from such assumptions. Then, the logistic regression has emerged as a robust alternative to discriminant analysis.

To test the assumption that the dependent variables follow a multivariate normal distribution and the variance-covariance matrices are equal across groups, Box's M test was conducted (SPSS 14 Help-Tutorial). Table 5 presents the results of the test of equality of covariance matrices. A significance value of 0.000 indicates that the data differ significantly from multivariate normality. This would mean that the discriminant analysis could not proceed further because its essential assumptions were violated. Under non-normality as in this case, the logistic regression is used (Park, 2021; 2022; Sio & Ismail, 2019: 29).

TABLE 5. Normality Assumption Test

Box's M		140.647
F	Approx.	1.847
	*df*1	60
	*df*2	1477.163
	Sig.	0.000

For the binary logistic regression, the dependent variable for English proficiency was categorised into two groups: LP and HP. The following analysis investigates the predictive validity of independent variables to confirm the predictive quality of lexical complexity indices. Table 6 presents the two methods (i.e. Cox & Snell R2 and Nagelkerke R2) to calculate the explained variation. The explained variation in the dependent variables in this model was 11.5 % according to Nagelkerke R2 values (Nagelkerke R2 = .115).

TABLE 6. Model summary

Step	-2 Log likelihood	Cox & Snell R-square	Nagelkerke R-square
1	286.719	0.062	0.084

Logistic regression estimates the probability of the event occurring, that is, the observed number of students in each proficiency group and the predicted number according to logistic regression ($p < .05$). Table 7 presents the assessment of the effectiveness of the predicted classification as opposed to the actual classification. It indicates that 25.8 % of the total 89 cases in the lower group, or 23 cases, were correctly classified into the lower group, and 85.2 % of the

total 135 cases in the upper group, or 115 cases, were correctly classified into the upper group. The overall classification accuracy was 61.6 %. This logistic analysis model was more suitable for predicting the upper group.

TABLE 7. Category prediction

Observed		Predicted		
		L2 Proficiency		Correct %
		Low	High	
Step 1	Low	23	66	25.8
	High	20	115	85.2
	Overall Percentage			61.6

Table 8 presents a summary of the stepwise logistic regression results used to determine the influence of predictors of L2 proficiency in a spoken corpus. Given the odds ratio (i.e. Exp(B)), it shows that the strongest predictor of L2 proficiency is Verbal density, i.e., Verbal elements per sentence, with an odds ratio of 2.285. This indicates that L2 learners are over 2.285 times more likely to achieve high proficiency when they use more verbal elements per sentence. The Wald statistics are also generated to show the significant levels and to establish which specific predictor variable is influencing the target variable.

TABLE 8. Variables in the equation

		B	S.E.	Wald	*df*	*p*	Exp(B)	95% C.I. for Odds Ratio	
								Lower	Upper
Step 1	Verbal density	0.826	0.248	11.100	1	0.001	2.285	1.405	3.714
	Constant	-0.906	0.407	4.966	1	0.026	0.404		

In assessing the influence of lexical complexity on L2 speaking proficiency, the independent variable (i.e., Verbal density) makes a statistically significant contribution to the model. As shown in Table 8, the strongest predictor of student achievement in proficiency is the frequency of verbal elements per sentence, recording an odds ratio of 2.285. This indicates that L2 learners are 2.285 times more likely to achieve high speaking proficiency when they use verbal elements per sentence more frequently.

DISCUSSION

This study measured lexical complexity, a relatively less studied rather than syntactic complexity in linguistic complexity, which was mainly studied as an indicator of performance measurement and evaluation in the language development stage. In addition, lexical complexity was conceptualised in both speaking and writing. In particular, the spoken data, which was difficult to use in existing studies due to the hassle of data collection and preprocessing, was included in the present investigation. Therefore, the results of this study will be readily exploited to understand the similarities and differences when comparing and analyzing other literatures in this field.

Considering the method and range of the study and the characteristics of the data set, this study can be said to have more reliable and objective results than the existing studies. It is expected that the results of the study will be more advanced than the previous studies in terms of index selection, measurement, data processing, sample type, and corpus amount for vocabulary ability evaluation.

First, as a result of comparing the scores of the lexical complexity factors, according to the task mode, Diversity and Sophistication scores were higher in writing than speaking. That is, the scores of Diversity's two indices, VOCD and MTLD, were all higher in writing, as were previous studies. Sophistication was reported to be higher in writing in the use of those over 1000 words commonly used in BNC, but the use of those beyond 2000 words was higher in speaking. However, the complicated results were presented in the case of Density, verbal elements per sentence and noun elements per sentence. In other words, the verbal density was measured higher in writing as predicted, but the noun density was higher in speaking. The supporting grounds for the results are found in previous studies (Johansson, 2008; Buhr et al., 2015).

For instance, Johansson asserted that it is theoretically possible that a text has high lexical diversity (i.e. contains many different word types), but low lexical density (i.e. contains many pronouns and auxiliaries rather than nouns and lexical verbs), or, vice versa, that a text has low lexical diversity (i.e. the same words or phrases are repeated over and over) but high lexical density (i.e. the words that are repeated are nouns, adjective or verbs). In addition, speech errors are typically associated with content words such as nouns, verbs, and adjectives, and the errors were also related to disfluency such as self-correction due to the characteristics of speaking mode. L2 learners were often seen to re-pronounce the corrected content words again, and especially in the examples presented by Buhr et al (2015), the case was more frequently found in nouns than verbs.

According to the traditional model of language production (Levelt et al., 1999), speech errors can emerge during linguistic planning, as content words are assigned to their relevant slots within a syntactic structure (e.g., a noun phrase). During this process interference can occur between content words, resulting in anticipation, perseveration, or exchange errors. Such interference is apparent in a tongue twister such as “she sells seashells,” in which anticipation of sh in “shells” might result in the selection of sh at the syllable-initial position of the preceding word “sea”, resulting in an error. According to the traditional model, an interruption occurring within a content word would be assumed to originate at a phonological level. In contrast, function words such as conjunctions, prepositions, and determiners play a grammatical role in the serial ordering of content words, and are not thought to be associated with phonological errors (Garrett, 1975).

On the other hand, logistic regression analysis showed that the indices that significantly predict L2 speaking proficiency were verbal density and BNC (~2K). The former was the most significant, and the more verbs used per sentence, the level of reaching the upper proficiency was 2.259 times higher than when not. This result is also related to the study that the factor determining speaking proficiency among syntactic complexity indices is the use of subordinating clauses (Park, 2021). She examined Korean EFL undergraduates and shows that the factors affecting learners’ speaking proficiency were Subordination and Length of Production among 14 syntactic complexity indices, which were proposed by Lu (2011). Namely, the high proportion of verbs per sentence can be assumed to be the high proportion of subordinating clauses in a sentence. Therefore, this result provides the basis for the necessity of studying both areas of linguistic complexity because lexical density and syntactic complexity are related to the prediction of proficiency.

Second, the conclusions derived from a series of measurements and systematic analysis designed to enhance reliability and validity in this study are expected to contribute academically to the second/foreign language acquisition area of linguistics. It is expected that the accuracy of calculation would be significantly higher than that of existing studies because the manual inspection process by researchers was added after the calculation by several tools for tagging, calculation, and data cleaning work such as transcription and disfluency processing work. Moreover, the analyzing method proposed in this study would be meaningful in education because it can be useful for educators to analyse lexical complexity of students and track the developmental stage of the abilities.

Finally, the current research methodology can be flexibly changed and expanded to meet various needs in future studies. That is, the measurements and the methodology systematically demonstrated which were designed to be suitable for speaking and writing analysis and can be applied to verify topics related to the relationship between proficiency and evaluation factors in the research area. For example, test settings can be diversified by using the data collected from the other L1 or other backgrounds other than the Korean university students or other types of tasks. Therefore, further studies can explore the relationship with this study.

CONCLUSION

This study is to investigate the relationship between lexical complexity and L2 proficiency in speaking and writing in English, and thus, it helps us understand in depth how the lexical characteristics used by learners are related to their proficiency. It also provides evidence of the validity of lexical factors for L2 performance evaluation. In other words, this study selected five indices of lexical competence based on previous literature, calculated lexical competence in speaking and writing of Korean English learners using objectively verified measurement methods, and investigated the relationship with their proficiency level classified by the CEFR standards. D was used as a measure of lexical diversity, i.e., VODC and MTLD, verbal and noun elements per sentence were used as a measure of lexical density, and lexical sophistication was defined as the percentage of words beyond the 2000 most frequent words based on BNC frequency lists. These indices of lexical complexity of Korean EFL learners in speaking and writing tasks were compared, and a series of comprehensive analysis methods were conducted to examine what variables can predict speaking proficiency among the lexical factors. Moreover, it was also found that the derived logistic regression model predicts L2 proficiency to a certain extent.

As a result of comparing lexical competence in speaking and writing, the complexity of writing was more prominent than speaking in general (Percentage of native likeness: monologue: 45.226, writing: 49.162). First, the lexical diversity scores were obtained through the D and MTLD indices, and the scores were significantly higher than those of speaking in writing (D: 48.7802 vs. 74.910, MTLD: 37.368 vs. 65.512, $p<.05$). These are in the same line with many previous studies, which are predictable results.

Second, to measure lexical density, content words per sentence were measured. The ratio of verbal elements and noun elements per sentence was examined. The results showed that the use of verbal elements was higher in writing (1.670 vs. 2.138), but the use of noun elements was higher in speaking (2.640 vs. 2.070). Lexical density is well known to be higher in writing (Lu, 2012), which is different from the current result. However, the result that speaking was higher than writing in the use of noun elements per sentence should be wary of assuming that lexical density was

higher in speaking than writing (Johansson, 2008; Buhr et al., 2015).

Third, in the sophistication, the use of words beyond BNC frequent 2K words was higher in speaking (0.36 vs. 0.30). The examples belonging to the BNC 2K group were 'place', 'alone', 'student', 'join', and 'deal', so this word group can also be assumed to be a combination that can be used easily by college students. According to Lee's research (2001) which examined the distribution of vocabulary in various genres included in BNC, only a limited number of individual vocabulary was used in spoken data and most of these individual vocabulary were included in the basic core individual vocabulary list (top frequency 2000 words).

Finally, as a result from the binary logistic analysis, the strongest predictor of student achievement in proficiency was the frequency of verbal elements per sentence, recording an odds ratio of 2.285. This indicated that L2 learners are 2.285 times more likely to achieve high speaking proficiency when they use verbal elements per sentence more frequently.

The results of this study will help educators understand that the relationship between lexical complexity and L2 proficiency is not equal in English speaking and writing, and select useful indicators when evaluating learners' proficiency. And it is expected that this study will play a role as pedagogically beneficial information because it has found the most relevant lexical indices in evaluating learners through the analysis of the relationship between the proficiency levels evaluated based on CEFR by well-trained native English linguistics experts.

ACKNOWLEDGEMENT

The author would like to thank the anonymous reviewers. This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2022S1A5A8054734).

REFERENCES

Abdulqader, Q. M. (2015). Comparison of discriminant analysis and logistic regression analysis: An application on caesarean births and natural births data. *Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 20*(1-2), 34-46.

Bayazidi, A., Ansarin, A. A., & Mohammadnia, Z. (2019). The relationship between syntactic and lexical complexity in speech monologues of EFL learners. *Applied Research on English Language, 8*(4), 473-488.

Buhr, A. P., Jones, R. M., Conture, E. G., & Kelly, E. M. (2016). The function of repeating: The relation between word class and repetition type in developmental stuttering. *International journal of language & communication disorders, 51*(2), 128-136.

Engber, C. A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of second language writing, 4*(2), 139-155.

Garrett, M. F. (1975). The analysis of sentence production. In *Psychology of learning and motivation* (Vol. 9, pp. 133-177). Academic Press.

Higginbotham, G., & Reid, J. (2019). The lexical sophistication of second language learners′ academic essays. *Journal of English for Academic Purposes, 37*, 127-140.

Hwang, E. (2013). Analysis of lexical complexity in Korean EFL students' narrative writing. Journal of Korea English Education, 12(3), 161-180.

Gregori-Signes, C., & Clavel-Arroitia, B. (2015). Analysing lexical density and lexical diversity in university students′ written discourse. *Procedia-Social and Behavioral Sciences, 198*, 546-556.

Halliday, M. A. K. (1994). Spoken and written modes of meaning. *Media texts: Authors and readers, 7*, 51-73.

Hess, C. W., Sefton, K. M., & Landry, R. G. (1986). Sample size and type-token ratios for oral language of preschool children. *Journal of Speech, Language, and Hearing Research, 29*(1), 129-134.

Housen, A., Kuiken, F., & Vedder, I. (Eds.). (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (Vol. 32). John Benjamins Publishing.

Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: A developmental perspective. *Working papers/Lund University, Department of Linguistics and Phonetics, 53*, 61-79.

Kalantari, R., & Gholami, J. (2017). Lexical Complexity Development from Dynamic Systems Theory Perspective: Lexical Density, Diversity, and Sophistication. *International Journal of Instruction, 10*(4), 1-18.

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System, 32*(2), 145-164.

Laufer, B. (1994). The lexical profile of second language writing: Does it change over time?. *RELC journal, 25*(2), 21-33.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied linguistics, 16*(3), 307-322.

Lee, D. Y. (2001). Defining core vocabulary and tracking its distribution across spoken and written genres: Evidence of a gradience of variation from the British National Corpus. *Journal of English Linguistics, 29*(3), 250-278.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences, 22*(1), 1-38.

Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners′ oral narratives. *The Modern Language Journal, 96*(2), 190-208.

McCarthy, P. M., & Jarvis, S. (2007). vocd: A theoretical and empirical evaluation. *Language Testing, 24*(4), 459-488.

MacWhinney, B. (2000). The CHILDES project: Tools for analyzing talk: Volume I: Transcription format and programs, volume II: The database.

Meara, P. (2005). LLAMA language aptitude tests*: The manual. Swansea: Lognostics*.

Park, S. (2021). Predicting L2 Speaking Proficiency Using Syntactic Complexity Measures: A Corpus-Based Study. 3L: Language, Linguistics, Literature®, 27(4), 101-113.

Park, S. (2022). Syntactic complexity in a learner written corpus and L2 speaking quality: Suggestions for distinguishing L2 speaking proficiency. Journal of Language and Linguistic Studies, 18(1), 361-371.

Park, S. K. (2013). Lexical Analysis of Korean University Students' Narrative and Argumentative Essays. *English Teaching*, 68(3).

Read, J. (2000). *Assessing vocabulary*. Oxford: Oxford University Press.

Richards, B. J., & Malvern, D. D. (1997). *Quantifying lexical diversity in the study of language development*. University of Reading, Faculty of Education and Community Studies.

Schnur, E., & Rubio, F. (2021). Lexical complexity, writing proficiency and task effects in Spanish Dual Language Immersion. *Language Learning & Technology, 25*(1), 53-72.

Sio, J., & Ismail, R. (2019). Binary logistic regression analysis of instructional leadership factors affecting English language literacy in primary schools. *3L: Language, Linguistics, Literature®, 25*(2).

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied linguistics, 30*(4), 510-532.

Strömqvist, S., Johansson, V., Kriz, S., Ragnarsdottir, H., Aisenman, R., & Ravid, D. (2002). Toward a cross-linguistic comparison of lexical quanta in speech and writing. *Written Language & Literacy, 5*(1), 45-67.

Ure, J. (1971). Lexical density: A computational technique and some findings. *Talking about text,* 27-48.

Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998*). Second language development in writing: Measures of fluency, accuracy, & complexity* (No. 17). University of Hawaii Press.

Yu, G. (2010). Lexical diversity in writing and speaking task performances. *Applied linguistics, 31*(2), 236-259.

Youssef, A. (2019). *Syntactic Complexity and Lexical Diversity in English Conference Abstracts*, Vol. 8, Issue No.2, April 2019, pp.33-70.