The Role of Usage-Based Indices as Indicators of Syntactic Growth in Second Language Writing

Hyunwoo Kim[1] & Yangon Rah[2†]
[1]Yonsei University, [2]Republic of Korea Air Force Academy

ABSTRACT
Although syntactic complexity features both length-based syntactic complexity and syntactic sophistication, previous research has predominantly focused on length-based complexity. Against this background, this study aimed to test the validity of syntactic sophistication including usage-based indices in explaining L2 writing proficiency, in comparison to the validity of length-based complexity indices. In a series of multiple regression analyses including various types of indices as predictors of the proficiency levels reflected in 3,201 sample essays, we found that the model including usage-based predictors produced explanatory power comparable to that of the model including length-based indices. However, when we incorporated both usage- and length-based indices in a single regression model, we found that the integrated model had a stronger explanatory power than the length-based and usage-based models, respectively. These results suggest that the proposed usage-based indices can serve as reliable measures of L2 syntactic complexity and may supplement length-based indices.

Keywords: syntactic complexity, syntactic sophistication, L2 writing, usage-based approaches, multiple regression

1. Introduction

In the field of second language (L2) production, syntactic complexity has increasingly attracted researchers' attention as a core construct that measures the variation and sophistication of the syntactic structures used by L2 learners. Among various types of syntactic complexity metrics, length-based measures have been at the center of research on L2 acquisition, particularly in the domain of written production (e.g., Biber, Gray, & Poonpon, 2011; Hwang, Jung, & Kim, 2020; Lu, 2011; Norris & Ortega, 2009; Wolfe-Quintero, Inagaki, & Kim, 1998). Commonly adopted length-based measures include the mean number of words per clause, mean

† Corresponding author: yonah@mnd.go.kr

number of clauses per sentence, mean number of subordination or coordination per sentence, and mean number of complex nominals or verb phrases per clause (Lu, 2011, 2017).

Although a long line of research has identified some relationships between length-based indices and L2 written proficiency, there has been considerable debate about whether these measures can serve as reliable indicators of L2 syntactic growth (Biber et al., 2011; Kyle & Crossley, 2017). One major criticism concerns their lack of solid theoretical background. The use of length-based indices rests on an intuitive assumption that more proficient learners will produce longer sentences involving a greater number of syntactic units. Some researchers have questioned about the validity of this assumption, arguing that the length-based approach mostly taps into writing fluency rather than complexity, and thus, may fall short of fully assessing the degree of syntactic complexity and sophistication as a function of L2 syntactic development (e.g., Bardovi-Harlig, 1992; Biber et al., 2011; Norris & Ortega, 2009; Rimmer, 2006). For instance, the number of words per clause, one of the most widely adopted length-based measures, may be associated with a writer's vocabulary skill and writing fluency but hardly indicate how often the writer uses more difficult and sophisticated structures (Kyle & Crossley, 2017). To illustrate, sentence (1a) contains a greater number of words per clause than sentence (1b), and yet, it is difficult to argue that (1a) is structurally more complex and sophisticated than (1b).

(1) a. The man was reading the book at the café this afternoon.
    b. He showed me the book that he bought.

Against this backdrop, the current study proposes a set of usage-based indices as theory-driven measures that may supplement the traditional length-based indices. The usage-based indices adopted in this study are firmly grounded in the usage-based theories of language development (e.g., Bybee, 2008; Ellis, 2013; Fillmore & Kay, 1999; Goldberg, 1995, 2006). According to usage-based approaches, learners accrue language experience, allowing them to detect regular patterns underlying individual exemplars of language units and generalize them to abstract syntactic categories called constructions, which are defined as form–meaning pairings of language units such as words, phrases, and argument structure patterns (Goldberg, 1995, 2006). In the process of constructional development, language users expand their repositories of verbs and constructions, using more sophisticated verb and construction types

(Goldberg, 2006; Goldberg, Casenhiser, & Sethuraman, 2004; Ninio, 1999). In the domain of L2 production, ample evidence suggests that several usage-based indices contribute to accounting for L2 syntactic growth in written production by demonstrating that learners with higher proficiency tend to produce less frequent verbs (e.g., Ellis & Larsen-Freeman, 2009; Kim & Rah, 2016), less frequent constructions (e.g., Hwang & Kim, 2022), less frequent verb–construction combinations (e.g., Kyle & Crossley, 2017), more strongly associated verb–construction combinations (e.g., Kyle & Crossley, 2017; Mostafa & Crossley, 2020), and a greater variety of construction types (e.g., Hwang & Kim, 2022).

Despite their roles as potent indicators of L2 syntactic development in production, however, each of these usage-based indices has been tested across different studies involving different learner groups, making it difficult to precisely weigh their relative contributions to explaining L2 syntactic complexity. To address this issue, this study adopted several usage-based indices that have been used in separate studies and tested their explanatory accuracy of predicting L2 written proficiency in comparison with the predictive power of the traditional length-based syntactic complexity indices. Using a large-scale learner corpus consisting of 3,201 argumentative essays, we conducted a series of multiple regression analyses to determine which indices contribute more strongly to explaining the variance in L2 written proficiency. Before presenting our study, we provide an overview of previous research on L2 syntactic complexity.

2. Length-based Syntactic Complexity in L2 Written Production

Syntactic complexity is operationalized as a measure of the "variety and degree of sophistication of the syntactic structures" (Lu, 2017, p. 494). It is conceptually divided into two categories: (a) absolute complexity, which focuses on length-based features of syntactic units, and (b) relative complexity or sophistication, which measures the degree of difficulty of particular syntactic units (Bulté & Housen, 2012). Despite its dual characteristics, syntactic complexity has primarily been investigated in terms of absolute complexity, using length-based indices as shown in Table 1.

Table 1. Examples of length-based syntactic complexity indices

Category	Index code	Index name
Length of production	MLC	Mean length (= number of words) of clause
	MLS	Mean length (= number of words) of sentence
	MLT	Mean length (= number of words) of T-unit
Sentence complexity	C/S	Clauses per sentence
Subordination	C/T	Clauses per T-unit
	CT/T	Complex T-unit ratio (= Number of complex T-units divided by number of T-units)
	DC/C	Dependent clauses per clause
	DC/T	Dependent clauses per T-unit
Coordination	CP/C	Coordinate phrases per clause
	CP/T	Coordinate phrases per T-unit
	T/S	T-units per sentence
Particular structures	CN/C	Complex nominals per clause
	CN/T	Complex nominals per T-unit
	VP/T	Verb phrases per T-unit

Note: Adapted from Kyle & Crossley, 2017, p. 9, Table 5.

Numerous studies have reported these length-based indices' significant roles in distinguishing L2 proficiency levels (e.g., Lu, 2011; Ortega, 2003; Wolf-Quintero et al., 1998). For example, Lu (2011) investigated the extent to which the 14 indices as listed in Table 1 explained the variability in the proficiency level of L2 writing samples produced by Chinese-speaking adult learners of English. He found that many of these indices' values increased as an essay's proficiency level was higher. Specifically, complex nominals per clause (CN/C) and mean length of clauses (MLC) contributed most strongly to the discrimination of the proficiency level, followed by the contributions of complex nominals per T-unit[1] (CN/T), mean length of sentence (MLS), mean length of T-unit (MLT), coordinate phrases per clause (CP/C), and coordinate phrases per T-unit (CP/T).

The close relationship between length-based measures and L2 proficiency was also

1) A T-unit is defined as a combination of a main clause and any dependent clauses associated with it (Hunt, 1965).

reported among child L2 learners. Hwang et al. (2020) used the same length-based indices as Lu (2011) to analyze the written and spoken production of elementary-school students and found that different types of indices contributed to the students' English proficiency depending on the production mode. The mean length of sentence (MLS) was identified as the strongest contributor to proficiency variability in written production, whereas T-units per sentence (T/S) was the strongest contributor in spoken production. These results were taken to indicate that young learners with higher proficiency produced longer sentences with more words in written production and used more T-units per sentence in spoken production.

Despite the well-documented roles of length-based measures in explaining learner proficiency, however, some studies suggest that these indices alone may not accurately gauge L2 syntactic complexity (Bardovi-Harlig, 1992; Biber et al., 2011; Norris & Ortega, 2009; Rimmer, 2006). For example, Biber et al. (2011) analyzed 429 academic writing and 723 conversation texts using various types of syntactic complexity indices including length-based measures (e.g., mean length of T-unit, clauses per T-unit) and grammatical features (e.g., finite and nonfinite dependent clauses, dependent phrases). They found that length-based indices were appropriate for capturing the characteristics of conversation but not the characteristics of academic writing, which was more strongly associated with grammatical features rather than length-based measures. Based on these findings, Biber and his colleagues claimed that length-based measures may not be ideal for assessing L2 written production including academic writing. In a similar vein, Norris and Ortega (2009) cautioned against using length-based measures for assessing L2 writing complexity, stating that these indices are mainly associated with writing fluency but not with complexity. To address this concern, they proposed that for the purpose of capturing general aspects of L2 syntactic development, length-based measures should be supplemented by "more developmentally sensitive and interlanguage-based measures (p. 574)".

3. Usage-based Approaches and Usage-based Indices in L2 Written Production

Recent studies have proposed a number of developmentally sensitive measures, such as syntactic sophistication indices, based on usage-based theories of language

development (e.g., Kyle & Crossley, 2017). Usage-based approaches assume that cumulative language experience enables learners to use a variety of sophisticated syntactic units (Langacker, 2008; Robinson & Ellis, 2008). The development of syntactic sophistication in language users is well illustrated by their use of verbs. For example, monolingual children in the early stages of language acquisition rely heavily on high-frequency verbs with prototypical and general senses, such as *go*, *get*, *take*, *give*, *put*, and *want*, which recur in various sentence patterns in the language use (Goldberg et al., 2004; Ninio, 1999; Tomasello, 2003). As children gain more experience with the ambient language, they expand their verb inventories to include diverse types of verbs including those with low frequency.

Usage-based approaches also assume that children's expansion of verbs takes place concurrently with the development of knowledge of argument structure constructions, which are defined as clause-level units that carry form and meaning in their own right (Goldberg, 1995, 2006). For example, sentences in (1) instantiate different argument structure types and denote different meanings despite involving the same verb *kick*.

(2) a. Transitive construction: *Paul kicked the ball.* (Meaning: X acts on Y)
b. Ditransitive construction: *Paul kicked me the ball.* (Meaning: X causes Y to receive Z)
c. Caused-motion construction: *Paul kicked the ball into the hole.* (Meaning: X causes Y to move Z)
d. Resultative construction: *Paul kicked the door open.* (Meaning: X causes Y to become Z)

Several studies have shown that monolingual children gradually expand their inventories of argument structure constructions and learn to generalize individual instances of sentences into abstract constructions (Ellis, Römer, & O'Donnell, 2016; Goldberg et al., 2004; Goldberg, Casenhiser, & White, 2007; Sethuraman, 2002). As a consequence, they show increased ability to produce less frequent and more diverse types of constructions and combine them with various types of verbs (Boyd & Goldberg, 2011; Goldberg, 1995).

Using insights from usage-based language learning, several studies have tested the validity of usage-based indices in explaining L2 written proficiency. For example, Kyle and Crossley (2017) adopted measures of verb frequency, construction frequency, frequency of verb–construction combination, and verb–construction association

strength and investigated their explanatory power in predicting L2 written proficiency scores. They found that L2 essays with higher scores tended to contain less frequent and more strongly associated verb–construction combinations. These findings led Kyle and Crossley to the conclusion that increased language experience allows learners to use less frequent verbs and constructions and combine them in a more native-like way (see Kim & Hwang, 2022, for similar findings among child L2 learners).

Hwang and Kim (2022) focused on the use of individual constructions and examined whether the diversity and frequency of construction types are associated with L2 writing proficiency. They calculated the proportions of 13 constructions listed in Table 2 in a native-speaker corpus (American National Corpus) and an L2 corpus (Yonsei English Learner Corpus) and found almost identical frequency distributions of the constructions across the L1 and L2 corpora. Crucially, their analysis of the L2 corpus showed that L2 written essays with higher proficiency contained more diverse types of constructions with lower frequency. For instance, essays with lower proficiency mostly contained high-frequency constructions, such as simple transitive, simple intransitive, and attributive constructions, whereas higher-proficiency essays contained diverse types of constructions, including low-frequency ones, such as passive, caused-motion, phrasal verb, and resultative constructions. Hwang and Kim interpreted these results as evidence for L2 learners' constructional development covarying with increased writing skills.

Table 2. Construction types investigated in Hwang and Kim (2022)

Construction	Example
Simple transitive	*The man kicked the ball.*
Simple intransitive	*I worked.*
Attributive	*She is a student.*
Passive	*It was folded.*
There-expletive	*There is a ball on the table.*
Caused motion	*She sent a letter to him.*
Intransitive motion	*I ran into the room.*
Ditransitive	*She sent him a message.*
Phrasal verb	*I put the hat on.*
Transitive resultative	*The girl made the can flat.*
Intransitive resultative	*The pond froze solid.*

The significant explanatory power of the usage-based indices attested in previous studies suggests that these measures may serve as reliable indices of L2 syntactic complexity in writing, potentially addressing some issues underlying the use of length-based measures. However, it is unclear which usage-based indices make stronger contributions to explaining L2 written proficiency and whether their explanatory power is as strong as or stronger than the traditional length-based measures. To address these questions, we built regression models that include usage-based and length-based indices as predictors and examined their relative contributions to explaining L2 written proficiency. We pose the following research questions:

Research question 1: Do length-based syntactic complexity indices reliably predict L2 written proficiency?

Research question 2: Do usage-based syntactic sophistication indices reliably predict L2 written proficiency?

Research question 3: Which types of indices more strongly predict L2 written proficiency?

4. Method

4.1. Corpus data

The production data analyzed in this study include 3,201 written essays taken from the Yonsei English Learner Corpus (YELC; Rhee & Jung, 2012). The YELC writing samples consist of descriptive and argumentative essays produced by undergraduate students at Yonsei University and were aligned at 11 proficiency levels from A1 (lowest) to C2 (highest), which were measured with reference to the Common European Framework of Reference (Council of Europe, 2011). Among them, we focused on argumentative essays to control for genre and narrowed down proficiency groups to six levels (A1+, A2, B1, B1+, B2, B2+) that contained a sufficient number of samples for statistical analyses. Table 3 presents a summary of the data used in the current study.

Table 3. Overview of the YELC written data

Proficiency level (sample size)	Number of words		Number of constructions	
	Mean	SD	Mean	SD
A1+ (185)	119.1	79.4	19.7	13.1
A2 (684)	194.8	76.9	31.6	13.0
B1 (1,165)	234.7	63.0	37.2	11.1
B1+ (705)	271.2	40.9	42.4	8.5
B2 (378)	285.7	39.8	43.9	8.1
B2+ (81)	299.6	36.7	45.0	8.0

4.2. Variable selection

We employed 14 length-based syntactic complexity indices (see Table 1) and five usage-based indices (verb frequency, construction frequency, frequency of verb–construction combinations, verb–construction association strength, diversity of construction type) as the main variables for analysis.

The length-based indices and four usage-based indices (verb frequency, construction frequency, frequency of verb–construction combinations, verb–construction association strength) were measured using the Tool for the Automatic Analysis of Syntactic Sophistication and Complexity (TAASSC) version 1.3.8[2] (Kyle, 2016). For the calculation of frequency measures for verbs, constructions, verb–construction combinations, and verb–construction association strength, we referred to frequency information extracted from four subsections (academic, fiction, magazine, newspaper) of the Corpus of Contemporary American English (COCA; Davies, 2010), a native-speaker reference corpus used in TAASSC. For the values of verb–construction association strength, we used delta P, which was calculated as the difference between the probability that a verb co-occurs with a certain construction and the probability that the verb occurs with other constructions (Kyle & Crossley, 2017). Each frequency value was log-transformed for normal distribution.

To obtain constructional diversity scores for our samples, we used the Constructional Diversity Analyzer (CDA) version 1.0[3] (Hwang & Kim, 2022). This

2) https://www.linguisticanalysistools.org/taassc.html

3) https://haerimhwang.github.io/tools/English-constructional-diversity-analyzer

tool automatically calculates the number of types and tokens of 13 constructions (see Table 2 for the list). Using this information, we obtained a construction diversity score for each sample by dividing the number of construction types by the number of construction tokens and converting the resulting score into a log-transformed value.

4.3. Data analysis

To explore the extent to which the selected length-based syntactic complexity and usage-based indices explain the variability of L2 written proficiency, we performed stepwise multiple regression analyses that included various categories of indices as predictors of the proficiency level. Initially, we generated a regression model to assess the predictive power of the length-based syntactic complexity indices. We then compared this model to another model that included the usage-based indices as predictors. Finally, we created an integrated model that incorporated both length-based and usage-based indices to determine which indices contribute more strongly to the proficiency level.

Before conducting regression analyses, we checked for collinearity among the indices to ensure that our regression models include unique, uncorrelated predictors (Tabachnick & Fidell, 2013). Tests for collinearity were conducted using correlation coefficients and variance inflation factors (VIFs). Thresholds for acceptable limits for inclusion as predictors were set as a correlation coefficient of less than .7 and a VIF value of less than 4 (Kyle, 2016; Kyle & Crossley, 2017). Among the length-based syntactic complexity indices, 8 failed to meet these conditions and were excluded, leaving 6 variables including MLC, T/S, CT/T, CP/C, CN/T, and CN/C. The usage-based indices were not strongly correlated with one another (all rs < .7), and none had a VIF larger than 4. We thus included all the usage-based indices in our models.

The multiple regression analyses were conducted using the *ols_step_both_p* function in R version 4.2.1 (R Core Team, 2022). Final model outcomes were obtained by starting with an empty model with no predictor and then adding each predictor to the model in a stepwise manner until the final model contained no significant predictor to be added.

5. Results

5.1. Length-based syntactic complexity indices

We first report outcomes from the model including the length-based syntactic complexity indices as predictors. Descriptive statistics on the index values are presented in Table 4.

Table 4. Mean scores (standard deviations) of the length-based syntactic complexity indices

Index	A1+	A2	B1	B1+	B2	B2+
MLC	7.38 (1.50)	7.81 (1.24)	8.18 (1.17)	8.38 (1.14)	8.70 (1.24)	8.98 (1.24)
T/S	1.05 (0.18)	1.07 (0.14)	1.07 (0.10)	1.09 (0.10)	1.10 (0.11)	1.11 (0.11)
CT/T	0.35 (0.24)	0.41 (0.18)	0.45 (0.16)	0.47 (0.16)	0.52 (0.14)	0.54 (0.15)
CP/C	0.11 (0.12)	0.13 (0.10)	0.14 (0.10)	0.14 (0.09)	0.15 (0.09)	0.15 (0.08)
CN/T	1.75 (2.77)	1.59 (0.57)	1.71 (0.53)	1.81 (0.51)	1.99 (0.54)	2.03 (0.55)
CN/C	0.97 (0.40)	1.03 (0.29)	1.08 (0.27)	1.11 (0.27)	1.16 (0.28)	1.17 (0.29)

Note: MLC = mean length of clause, T/S = T-units per sentence, CT/T = complex T-unit ratio, CP/C = coordinate phrases per clause, CN/T = complex nominals per T-unit, CN/C = complex nominals per clause.

As shown in Table 5, the final regression model included five significant predictors: MLC, CT/T, T/S, CN/T, and CP/C. These indices jointly explained 15.8% ($R^2 = 0.158$, $p < .001$) of the variance of the proficiency level. Among them, MLC had the strongest contribution to the model, as indicated by its standardized beta, suggesting that the increase rate in the number of words per clause with rising proficiency was greater than that of the other indices.

Table 5. Summary of regression model including length-based predictors

Entry	Predictors	*β*	*SE*	*B*	*t*	*p*
1	MLC	0.317	0.018	0.346	17.804	< .001
2	CN/T	1.889	0.122	0.284	15.463	< .001
3	T/S	0.931	0.164	0.093	5.681	< .001
4	CN/T	−0.114	0.026	−0.084	−4.410	< .001
5	CP/C	−0.454	0.213	−0.039	−2.134	.033

Note: *β* = unstandardized beta, *SE* = standard error, *B* = standardized beta.

5.2. Usage-based indices

Descriptive statistics for the usage-based indices' values are presented in Table 6, followed by model outcomes in Table 7. The final model included two significant predictors: construction diversity and verb frequency, which jointly accounted for 13.6% ($R^2 = 0.136$, $p < .001$) of the variance of the proficiency. These results suggest that essays with higher proficiency tended to include more diverse types of constructions and less frequent verbs. The standardized beta was greater for construction diversity than verb frequency, indicating its stronger contribution to distinguishing the proficiency levels.

Table 6. Mean scores (standard deviations) of the usage-based indices

Index	A1+	A2	B1	B1+	B2	B2+
VF	5.39 (0.67)	5.43 (0.26)	5.40 (0.22)	5.37 (0.21)	5.33 (0.22)	5.28 (0.20)
CF	4.76 (0.67)	4.76 (0.29)	4.74 (0.25)	4.72 (0.23)	4.70 (0.23)	4.64 (0.23)
VCF	3.45 (0.72)	3.47 (0.40)	3.44 (0.34)	3.40 (0.31)	3.37 (0.31)	3.32 (0.26)
VCA	0.023 (0.024)	0.025 (0.015)	0.025 (0.013)	0.025 (0.012)	0.025 (0.011)	0.025 (0.011)
CD	0.68 (0.11)	0.75 (0.08)	0.78 (0.07)	0.80 (0.07)	0.81 (0.06)	0.83 (0.07)

Note: VF = verb frequency, CF = construction frequency, VCF = verb–construction frequency, VCA = verb–construction association strength, CD = construction diversity.

Table 7. Summary of regression model including usage-based predictors

Entry	Predictors	β	*SE*	*B*	*t*	*p*
1	CD	5.012	0.234	0.353	21.444	< .001
2	VF	−0.491	0.069	−0.117	−7.108	< .001

Note: β = unstandardized beta, SE = standard error, B = standardized beta.

The variance explained by the usage-based model (13.6%) was lower than that explained by the length-based model (15.8%). Despite their numerical difference, however, the explanatory power of the two models was not statistically different. A Fisher *r*-to-*z* transformation (Dunn & Clark, 1969) showed no significant difference in the coefficients between the two models (z = 1.31, p = .190). These results suggest that the model derived from the usage-based indices had predictive power comparable to the length-based syntactic complexity model.

5.3 Integrated model including all indices

In order to examine the relative weighting of each index, we included both length-based and usage-based indices in a single regression model. Table 8 presents the model outcome.

Table 8. Summary of regression model including length-based predictors

Entry	Predictors	Category	β	*SE*	*B*	*t*	*p*
1	CD	Usage-based	3.929	0.227	0.296	17.315	< .001
2	MLC	Length-based	0.267	0.017	0.291	15.249	< .001
3	CT/T	Length-based	1.646	0.120	0.248	13.748	< .001
4	T/S	Length-based	0.772	0.157	0.077	4.926	< .001
5	CN/T	Length-based	−0.144	0.027	−0.107	−5.332	< .001
6	VF	Usage-based	−0.438	0.093	−0.104	−4.716	< .001
7	VCF	Usage-based	0.140	0.064	0.045	2.199	.028
8	CP/C	Length-based	−0.364	0.203	−0.031	−1.793	.073

Note: β = unstandardized beta, *SE* = standard error, *B* = standardized beta.

The integrated model explained 23.5% (R^2 = 0.235, p < .001) of the variance in the proficiency level. Among the indices, the model identified construction diversity as the strongest predictor (standardized beta = 0.296), suggesting that the increase rate of construction diversity as a function of rising proficiency was the greatest among the variables included in the model. The next strongest predictors were four length-based indices, including MLC, CT/T, T/S, and CN/T, followed by verb frequency (VF) and verb–construction frequency (VCF). Finally, CP/C was included as a predictor in the model, but its contribution was only marginally significant (p = .073).

6. Discussion

The objective of this study was to compare the explanatory value of usage-based indices and length-based measures in accounting for the variance of L2 learners' writing proficiency. To address issues with length-based indices, we proposed several usage-based indices within the framework of usage-based theories of language development. Our regression analyses revealed that the predictive power of the usage-based model was as strong as that of the length-based model. When combining both types of indices in a single model, we found that construction diversity, a usage-based index, had the strongest contribution to the variance of the proficiency level in our samples.

Our findings that the proposed usage-based indices can serve as reliable measures of L2 syntactic complexity that may supplement the traditional length-based indices. The regression model that included the length-based indices as predictors significantly explained the variance of the proficiency level, with mean length of clause contributing most strongly to the model. These results are consistent with previous findings that the mean number of words per clause or sentence serves as a strong indicator of L2 syntactic growth in writing (e.g., Hwang et al., 2020; Kyle & Crossley, 2017; Lu, 2011). However, as noted in the introduction, researchers have constantly raised criticism against using length-based indices due to their lack of a solid linguistic basis (e.g., Biber et al., 2011; Rimmer, 2006). In contrast, the usage-based indices adopted in this study appeal to usage-based theories, providing a theoretical ground for using them as measures for assessing L2 syntactic development. As a result, we found that the usage-based indices captured the variance of L2 proficiency in writing as accurately as the length-based indices.

Among the usage-based indices included in the model, construction diversity was the strongest contributor. This finding aligns with previous research showing that L2 learners exhibit an increased ability to comprehend and produce more diverse construction types as they progress toward higher proficiency (e.g., Ellis & Larsen-Freeman, 2009; Gries & Wulff, 2005; Hwang & Kim, 2022; Kim & Rah, 2016). Our findings provide evidence that the use of constructions is a significant indicator of L2 syntactic development in writing. The strong relationship between construction diversity and L2 writing proficiency found in our study is illustrated in Table 9, which presents two samples from the A1+ group and the B2 group, respectively. The essay from the A1+ group (ID: #623) contained only simple transitive and attributive constructions, indicating limited constructional diversity. In contrast, the essay from the B2 group (ID: #2524) included not only simple transitive and attributive constructions but also more complex and less frequent constructions, such as caused-motion (e.g., *driving wrong-going-students into the right road*), passive (e.g., *dislike being confined at school*), and phrasal verb constructions (e.g., *to cut it out*). As demonstrated in these samples, essays with higher proficiency were more likely to contain both high-frequency and low-frequency constructions, reflecting the writers' improved skills to express their ideas using various types of constructions.

Table 9. Construction diversity information of two sample essays

ID (level)	Sample sentence	Construction diversity score
#623 (A1+)	I don't think it. Physical punishment is good. Because this is straight pass. Student often isn't followed teacher but also teacher isn't student of change. However Physical punishment is easy finish. Student remember physical punishment that don't forget everyday.	0.48
#2524 (B2)	Some people thinks that physical punishment is just hurting and giving stress to students and violating student's right. But I think there should be some physical punishment to lead students, and I also think it's better for students themselves. First, I think it's needed for driving wrong-going-students into the right road. Students, especially teenaged ones, tend to dislike being confined at school. So they seek for fun outside of school. It's hard to make students who are addicted to these kinds of stuffs to cut it out.	0.95

Another important finding of this study was that the model integrating both length-based and usage-based indices as predictors was superior to each of the individual models in terms of the amount of variance of proficiency explained by

these models. The integrated model explained 23.5% of the variance, while the length-based and usage-based models explained 15.8% and 13.6% of the variance, respectively. This result suggests that using both length-based and usage-based indices can provide a more accurate assessment of L2 writing proficiency than focusing on either index category. Therefore, researchers examining L2 syntactic complexity in writing may consider using both types of indices within a single model. The advantage of integrating length-based and usage-based indices may also extend to the educational domain. Writing teachers often face challenges in assessing a large number of student essays under limited time constraints. This burden can be alleviated with the help of the resources used in the current study. The natural language processing tools used for calculating the length-based and usage-based measures (i.e., TAASSC and CDA) are publicly available online and provide user-friendly interfaces. Teachers can thus run these computer programs to obtain preliminary information on their students' writing and use this information as reference for grading.

Alongside these implications, we also acknowledge some limitations that may call upon further enhancement. First, the current study focused solely on written argumentative essays, and the findings obtained from our analyses have limited generalizability. To determine whether our findings apply to other types of production, further research is necessary to test the validity of length-based and usage-based indices in a variety of genres such as narratives, interviews, and book reports. Second, the sample data used in this study only consisted of Korean-speaking learners of English. Therefore, it remains unclear whether our findings are specific to this group of learners or can be applied to other learner groups with different L1 backgrounds. Future work should address this issue by including a broad range of learner groups with diverse L1 backgrounds.

Notwithstanding these limitations, the current study offers a promising framework for investigating L2 written production. The integration of length-based and usage-based indices may allow for a detailed understanding of the syntactic complexity of L2 writing, which is characterized by both absolute, length-based complexity and sophistication or relative complexity (Bulté & Housen, 2012). Such an integrated approach may provide new opportunities for expanding the scope of research on syntactic complexity, which has predominantly focused on length-based measures, to include theory-driven usage-based indices. We hope that our findings will provide useful resources for researchers and educators to assess L2 syntactic complexity in writing.

References

Bardovi-Harlig, K. (1992). A second look at T-unit analysis: Reconsidering the sentence. *TESOL Quarterly, 26*, 390-395.

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, *45*(1), 5-35.

Boyd, J. K., & Goldberg, A. E. (2011). Learning what not to say: The role of statistical preemption and categorization in a-adjective production. *Language, 87*, 55-83.

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken & F. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21-46). Amsterdam: John Benjamins.

Bybee, J. (2008). Usage-based grammar and second language acquisition. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 216-236). New York: Routledge.

Council of Europe (2011). *Common European framework of reference for languages: Learning, teaching, assessment*. Council of Europe.

Davies, M. (2010). The corpus of contemporary American English as the first reliable monitor corpus of English. *Literary and Linguistic Computing, 25*, 447-464.

Dunn, O. J., & Clark, V. (1969). Correlation coefficients measured on the same individuals. *Journal of the American Statistical Association, 64*, 366-377.

Ellis, N. C. (2013). Construction grammar and second language acquisition. In G. Trousdale & T. Hoffmann (Eds.), *Oxford handbook of construction grammar* (pp. 365-378). Oxford: Oxford University Press.

Ellis, N. C., & Larsen-Freeman, D. (2009). Constructing a second language: Analyses and computational simulations of the emergence of linguistic constructions from usage. *Language Learning, 59*, 90-125.

Ellis, N. C., Römer, U., & O'Donnell, M. B. (2016). *Language usage, acquisition, and processing: Cognitive and corpus investigations of construction grammar*. Malden, MA: Wiley-Blackwell.

Fillmore, C. J., & Kay, P. (1999). Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language, 64*, 501-538.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. London: University of Chicago Press.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: University Press.

Goldberg, A. E., Casenhiser, D. M., & Sethuraman, N. (2004). Learning generalizations. *Cognitive Linguistics, 15*, 289-316.

Goldberg, A. E., Casenhiser, D., & White, T. (2007). Constructions as categories of

language. *New Ideas in Psychology, 25*, 70-86.

Gries, S. T., & Wulff, S. (2005). Do foreign language learners also have constructions? *Annual Review of Cognitive Linguistics, 3*, 182-200.

Hunt, K. W. (1965). *Grammatical structures written at three grade levels* (Research Report No. 3). Champaign, IL: National Council of Teachers of English.

Hwang, H., Jung, H., & Kim, H. (2020). Effects of written versus spoken production modalities on syntactic complexity measures in beginning-level child EFL learners. *The Modern Language Journal, 104*, 267-283.

Hwang, H., & Kim, H. (2022). Automatic analysis of constructional diversity as a predictor of EFL students' writing proficiency. *Applied Linguistics*. Advance online publication, https://doi.org/10.1093/applin/amac046

Kim, H., & Hwang, H. (2022). Assessing verb-construction integration in young learners of English as a foreign language: Analyses of written and spoken production. *Language Learning, 72*, 497-533.

Kim, H., & Rah, Y. (2016). Effects of verb semantics and proficiency in second language use of constructional knowledge. *The Modern Language Journal, 100*, 716-731.

Kyle, K. (2016). *Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication* (Unpublished doctoral dissertation). Georgia State University.

Kyle, K., & Crossley, S. (2017). Assessing syntactic sophistication in L2 writing: A usage-based approach. *Language Testing, 34*, 513-535.

Langacker, R. (2008). The relevance of cognitive grammar for language pedagogy. In S. De Knop, & T. De Rycker (Eds.), A*pplications of cognitive linguistics 9: Cognitive approaches to pedagogical grammar* (pp. 7-35). Berlin: Mouton de Gruyter.

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, *45*(1), 36-62.

Lu, X. (2017). Automated measurement of syntactic complexity in corpus-based L2 writing research and implications for writing assessment. *Language Testing, 34*, 493-511.

Mostafa, T., & Crossley, S. A. (2020). Verb argument construction complexity indices and L2 writing quality: Effects of writing tasks and prompts. *Journal of Second Language Writing, 49*, Article 100730.

Ninio, A. (1999). Pathbreaking verbs in syntactic development and the question of prototypical transitivity. *Journal of Child Language, 26*, 619-653.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in SLA: The case of complexity. *Applied Linguistics, 30*, 555-578.

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics, 24*, 492-518.

R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Rhee, S. C., & Jung, C. K. (2012). Yonsei English Learner Corpus (YELC). Paper

presented at the *First Yonsei English Corpus Symposium*, March.
Rimmer, W. (2006). Measuring grammatical complexity: The Gordian knot. *Language Testing, 23*, 497-519.
Robinson, P., & Ellis, N. (2008). An introduction to cognitive linguistics, second language acquisition, and language instruction. In P. Robinson & N. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 3-24). New York: Routledge.
Sethuraman, N. (2002). *The acquisition of verbs and argument structure constructions* (Unpublished doctoral dissertation). University of California.
Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Harlow, UK: Pearson Education.
Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second language development in writing: Measures of fluency, accuracy, & complexity*. University of Hawaii Press: Honolulu.

Hyunwoo Kim
Assistant Professor
Department of English Language and Literature
Yonsei University
50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea
E-mail: hyunwoo2@yonsei.ac.kr

Yangon Rah
Professor
Department of Language and Literature
Republic of Korea Air Force Academy
635, Danjae-ro, Namil-myeon, Sangdang-gu, Cheongju 28187, Korea
E-mail: yonah@mnd.go.kr

Received: February 27, 2023
Revised version received: March 29, 2023
Accepted: April 4, 2023