Studies in Second Language Acquisition (2026), 1–25
doi:10.1017/S0272263126102009

METHODS FORUM

Automated measurement of complexity, accuracy, and fluency indices for Korean language data

Development and preliminary validation of K-text, a web-based natural language processing tool

Haerim Hwang

Department of English, The Chinese University of Hong Kong, Hong Kong
Email: haerimhwang@cuhk.edu.hk

(Received 16 January 2026 UTC; Revised 18 August 2026 UTC; Accepted 23 August 2026 UTC)

Abstract

This study introduces K-text, a website designed to measure Korean language proficiency from second language writing and speech data through 20 complexity, accuracy, and fluency indices using natural language processing. A key advantage of K-text is its ability to measure accuracy and fluency in addition to complexity. To test the 20 indices' preliminary validity, this study investigates whether they can predict overall criterion-based Korean proficiency from 1 to 6 (equivalent to A1 to C2) when they are used to analyze 45,855 second language learners' essays from the Korean Learner Corpus. The findings show that all of the individual K-text indices are significant predictors of overall Korean proficiency. Together, the 20 indices accounted for 44.50% of the variation observed in proficiency scores. These results suggest that K-text is a valuable resource for analyzing Korean language production data in relation to overall proficiency.

Keywords: accuracy; complexity; fluency; Korean; proficiency

Introduction

While natural language processing (NLP) has been expanding rapidly with far-reaching implications across multiple disciplines, its integration into (applied) linguistics remains challenging due to the programming expertise required for its use. Moreover, the majority of NLP research done in (applied) linguistics has centered on only a few languages, such as English, neglecting most others. Korean is one such underresearched language, even though it is gaining global popularity as a second language (L2). The 2025 Duolingo Language Report, based on one of the most widely used free learning applications, shows that Korean ranks sixth globally in terms of learner enrollment

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (http://creativecommons.org/licenses/by-nc-sa/4.0), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is used to distribute the re-used or adapted article and the original article is properly cited. The written permission of Cambridge University Press or the rights holder(s) must be obtained prior to any commercial use.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:08:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126102009

(Blanco, 2025). Further, Korean has numerous linguistic features that distinguish it from languages like English. Syntactically, it is a head-final language, with subject-object-verb word order, and it employs case markers to specify the grammatical case of arguments within a clause, as illustrated in (1).[1] Morphologically, it is an agglutinative language, characterized by a range of content morphemes (e.g., verb stems) and function morphemes (e.g., case markers), each conveying distinct meanings and functions (Sohn, 1999). These characteristics, along with the rising global interest in learning Korean, create a need for specific measures that can assess the proficiency or development of Korean as an L2.

(1) Sensayngnim-kkeyse haksayng-eykey pheyn-ul kenney
teacher-HON.NOM student-DAT pen-ACC hand.over
cwu-sy-ess-ta.
give-HON-PST-DECL
'The teacher handed over the pen to the student.'

Recent research has started to address this need (e.g., H. Hwang, 2024; H. Hwang & H. Kim, 2024) and to offer practical tools for researchers and educators within the complexity, accuracy, and fluency (CAF) framework (Housen, Kuiken, & Vedder, 2012). According to this framework, language proficiency comprises three key dimensions. The first is complexity, which encompasses two subdimensions: (a) global or system complexity, referring to "the degree of elaboration, size, breadth, width, or richness of the learner's L2 system or repertoire" (Housen et al., 2012, p. 25) and (b) local or structural complexity, which reflects either the transparency and multiplicity of form-meaning relationships or the number of discrete elements comprising a linguistic form. In this study, both types of complexity are considered at different linguistic levels. Specifically, morphological complexity is examined as a form of global/system complexity in order to capture morphological diversity, operationalized here as the breadth of a learner's L2 morphological system. This focus is particularly appropriate given the agglutinative nature of Korean and is therefore more suitable than lexical diversity. As a form of local/structural complexity, syntactic complexity in this study is operationalized as the number of specific syntactic units produced in the data; the syntactic indices of interest include per-sentence counts of morphemes, eojeols, coordinate phrases, relative clauses, adjunct clauses, verbs, nominal particle types, classifiers, tense markers, and honorific markers. The second is accuracy, which refers to "the extent to which an L2 learner's performance […] deviates from a norm" (Housen et al., 2012, p. 4). In the present study, accuracy is measured via the ratio of error-free sentences. The last is fluency, i.e., the ease, eloquence, and smoothness with which language is produced, reflecting the speaker's or writer's ability to communicate in a natural manner (Tavakoli & Skehan, 2005). This construct in the present study is defined in terms of the fluent and productive use of linguistic units at the morphological, eojeol,[2] and sentence levels.

[1]This study utilizes the following abbreviations for morpheme-by-morpheme glossing: ACC = accusative case particle; CAU = causative marker; CL = classifier; CONN = connective marker; DAT = dative case particle; DECL = declarative marker; HON = honorific marker; LOC = locative marker; NOM = nominative case particle; POL = politeness marker; PST = past tense marker; RL = relativizer; TOP = topic marker.

[2]The "eojeol" is a fundamental unit in written Korean, indicated by spacing, and provides both lexical and morphosyntactic information. For example, (1) illustrates a sentence containing five eojeols, each of which consists of one or more morphemes.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:08:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126102009

Building on Lu's (2011) syntactic complexity indices for English, H. Hwang and H. Kim (2024) proposed seven syntactic complexity indices for Korean, which were measurable via software they developed and called the Korean Syntactic Complexity Analyzer (KOSCA). Their analysis found that all seven indices, including language-independent measures (number of morphemes per sentence, number of coordinate phrases per sentence, number of relative clauses per sentence, number of adjunct clauses per sentence, number of verbs per sentence) and Korean-specific measures (number of eojeols per sentence, number of nominal particle types per sentence), contributed to predicting L2 Korean proficiency. In a follow-up study, H. Hwang (2024) focused on complexity in the morphological domain, developing another software, the Korean Morphological Richness Analyzer (KOMORA). Using this tool, she showed that the diversity of function morphemes and the diversity of content morphemes develop at different proficiency levels.

While existing research using NLP has shown the usefulness of certain indices in explaining L2 Korean proficiency, its focus has been limited to complexity alone, neglecting accuracy and fluency. Therefore, a more holistic approach is required, one that takes into account all these different components of proficiency. This approach would allow for a deeper understanding of L2 learners' production performance by reflecting their ability to produce sophisticated, accurate, and fluent language (Kuiken, Vedder, & Gilabert, 2010). Furthermore, the lack of user-friendly tools for analyzing Korean has made it difficult for researchers and educators to assess large amounts of Korean learner data. To address these issues in the present study, we develop K-text (https://k-text.com), a publicly available website that computes various CAF indices based on Korean language data entered by website users. We also test the preliminary validity of the indices available on K-text by assessing their ability to predict overall criterion-based proficiency, following one of the approaches adopted in previous studies (for an overview, see Kyle, Sung, Eguchi, & Zenker, 2024; e.g., Bulté & Roothooft, 2020; Treffers-Daller et al., 2018). To this end, we use a large dataset of L2 Korean essays (n = 45,855), which is at least twice the size of datasets employed in prior research in this area (e.g., H. Hwang & H. Kim, 2024).

Overall, this study makes both theoretical and practical contributions to the field of applied linguistics. First, the study introduces K-text, a valuable new resource for researchers, language educators, and learners, which will allow them to delve deeper into language use patterns. In addition, the study proposes and validates several new CAF indices, proving them effective in predicting Korean proficiency and thus useful in various research contexts, including language testing and education. In the long term, these indices could also contribute to the development of automated essay scoring systems for Korean.

K-text

K-text is a user-friendly and publicly available website that measures Korean language proficiency (see Figure 1). Over a period of ten months, its development process unfolded in three stages: first, creating Python scripts for analysis of Korean language data; second, testing the scripts using diverse text samples; and third, developing and releasing K-text using the Python web framework "Django" (https://www.djangoproject.com/) with the help of a web designer. The website features six sections dedicated to morphological complexity, syntactic complexity, accuracy, fluency, topic modeling, and a user-defined index. This study will focus on the sections related to morphological complexity, syntactic complexity, accuracy, and fluency. To analyze text, users can either enter Korean text in its original script (i.e., Hangul) directly into a text box on the website or upload up to

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:08:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126102009

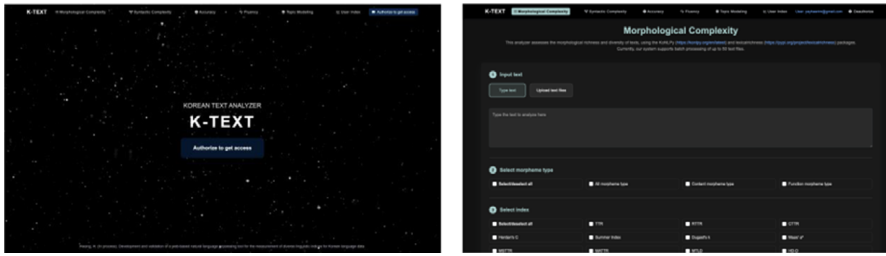


Figure 1. Graphic user interface of K-Text.

50 text files. Once the data is processed, the website presents the results in a tabular format, which can also be downloaded as a CSV file.

K-text is able to measure 20 indices, which are outlined in Table 1. Note that most of these indices use the sentence as the unit of analysis. This decision was made to account for unique features of Korean structure that complicate clause- or T-unit-based analysis. Because Korean is head-final, frequently omits subjects, and makes extensive use of connective endings to link clauses, identifying clause boundaries is often difficult. These factors render the sentence a more practical and consistent unit for analysis than the clause or T-unit. For the complexity and fluency indices, K-text makes use of the "Kkma" parser from the "KoNLPy" module to identify morphemes and their part-of-speech (POS) tags within each sentence (see http://kkma.snu.ac.kr/). This parser was selected over alternatives like Hannanum, Stanza, and UDpipe due to its excellent Python compatibility. It is also the parser that most comprehensively covers POS categories for Korean morphemes. Crucially, the parser is known for its high tagging accuracy, achieving 93.88% for morphemes in first language (L1) Korean data (Min, Lee, & Lee, 2022) and an F1 score of .91 for L2 Korean data (H. Hwang, 2024).

Table 1. Complexity, Accuracy, and Fluency Indices in K-Text

Construct	Subconstruct	Index
Complexity	Morphological complexity	Moving-average type-token ratio for all morphemes
		Moving-average type-token ratio for content morphemes
		Moving-average type-token ratio for function morphemes
		Measure of textual lexical diversity for all morphemes
		Measure of textual lexical diversity for content morphemes
		Measure of textual lexical diversity for function morphemes
	Syntactic complexity	Number of morphemes per sentence
		Number of eojeols per sentence
		Number of coordinate phrases per sentence
		Number of relative clauses per sentence
		Number of adjunct clauses per sentence
		Number of verbs per sentence
		Number of nominal particle types per sentence
		Number of classifiers per sentence
		Number of tense markers per sentence
		Number of honorific markers per sentence
Accuracy		Ratio of accurate sentences to total sentences
Fluency		Number of morphemes per text
		Number of eojeols per text
		Number of sentences per text

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:08:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126102009

To assess accuracy, K-text employs the open-source Korean Automatic Grammatical Error Annotation System (KAGAS; Yoon et al., 2023). Built on the Korean-specific KoBART Transformer model (Katsumata & Komachi, 2020), KAGAS is currently the most accurate system for automatically correcting grammatical errors in Korean. What makes the system particularly appropriate for our purposes is that it maintains high accuracy across different types of data: 93.93% on data from L1 Korean speakers and 87.34% on data from adult L2 learners (see Yoon et al., 2023).

Of the 20 indices, 13 were derived from prior research (H. Hwang, 2024; H. Hwang & H. Kim, 2024), and 7 were newly developed for K-text: 3 for syntactic complexity (number of classifiers per sentence, number of tense markers per sentence, number of honorific markers per sentence), 1 for accuracy (ratio of accurate sentences to total sentences), and 3 for fluency (number of morphemes per text, number of eojeols per text, number of sentences per text). The following sections describe each of the 20 indices and present the reliability of these indices.

Complexity

Morphological complexity

K-text assesses morphological complexity using six indices proposed by H. Hwang (2024), which are founded on two established lexical diversity measures: the moving-average type-token ratio (MATTR; Covington & McFall, 2010) and the measure of textual lexical diversity (MTLD; McCarthy & Jarvis, 2010). These measures are applied three times to each of three morphological categories: all morphemes, content morphemes, and function morphemes. Using a subset of the Korean Learner Corpus (National Institute of Korean Language, 2020) consisting of 863 narrative essays (13,551 sentences; 125,939 eojeols), H. Hwang's research showed that overall morphological complexity (MATTR and MTLD based on all morphemes) improved with L2 proficiency primarily at lower levels. However, further analyses revealed that while the MATTR and MTLD for function morphemes increased faster at lower proficiency levels, the MATTR and MTLD for content morphemes increased faster at higher proficiency levels. These results suggest that a comprehensive measurement of morphological complexity should account for the differences between content and function morphemes, treating them as distinct categories.

- Moving-average type-token ratio. K-text includes three MATTR-based indices: MATTR for all morphemes, MATTR for function morphemes, and MATTR for content morphemes. Unlike the traditional type-token ratio (TTR; Johnson, 1944), the MATTR is less susceptible to the influence of text length. TTR, which is calculated by dividing the frequency of word types by the frequency of word tokens, has been widely used to measure lexical diversity. A well-documented limitation of the traditional TTR is its tendency to decrease as text length increases, suggesting that it conflates lexical diversity with text length. To address this issue, McCarthy and Jarvis (2007) introduced the MATTR, which provides a more nuanced measure of lexical diversity. The MATTR calculation involves segmenting a text into overlapping windows of a specified length and computing the TTR for each window. The final MATTR score is the average of these individual TTR values. Following Zenker and Kyle (2021), K-text employs a window size set at 50 and computes the MATTR three times: once for all morphemes, once for content morphemes, and once for function morphemes. A higher MATTR score is indicative of a greater

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:08:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126102009

diversity of morphemes within a text, which in turn reflects the text's level of morphological complexity.
- Measure of textual lexical diversity. K-text also employs three MTLD-based indices: MTLD for all morphemes, MTLD for function morphemes, and MTLD for content morphemes. MTLD measures the average number of sequential segments in a text that satisfy a certain value of TTR. In K-text, the value is set to .720, in accordance with the method described by McCarthy and Jarvis (2010), and the MTLD is computed three times: once each for all morphemes, for content morphemes, and for function morphemes. Just like the MATTR indices, a higher MTLD value indicates a greater variety of morphemes within a text.

Syntactic complexity

For syntactic complexity, K-text incorporates 10 indices. Seven of these are syntactic measures derived from H. Hwang and H. Kim (2024). The researchers analyzed 20,223 written essays and 1,102 oral interview samples produced by L2 learners of Korean with diverse L1 backgrounds. They confirmed that all seven syntactic indices they proposed are valid measures of overall Korean proficiency; all showed an increase in value alongside the development of proficiency. A brief description of each index is provided below.

- Number of morphemes per sentence. This index quantifies the average number of morphemes per sentence by dividing the total morpheme count by the total sentence count in a text. The measurement counts morphemes of both types: (a) content morphemes, such as nouns (e.g., *chayk* "book") and verb stems (e.g., *ha-* "do"); and (b) function morphemes, such as demonstrative pronouns (e.g., *ce* "that"), particles (e.g., nominative case particle *-i/ka*), and sentence enders (e.g., declarative marker *-ta*). This index score is predicted to increase as learners demonstrate the ability to combine multiple morphemes in a sentence to convey more sophisticated meanings.
- Number of eojeols per sentence. This index computes the mean number of eojeols per sentence. Consistent with the number of morphemes per sentence, this index score is expected to increase as learners develop the capacity to combine multiple eojeols within a sentence, resulting in more elaborate expressions.
- Number of coordinate phrases per sentence. This index calculates the frequency of coordinate phrases marked by conjunctive particles, such as *-ko* "and" and *-[u]na* "but", and then expresses this value as a proportion of the total sentence count. A higher score on this index indicates that the learner has a greater capacity to add new information and combine diverse pieces of information.
- Number of relative clauses per sentence. This index specifically examines the use of relative clauses, which serve to add descriptive details or provide further information about a noun (person or thing) mentioned in the sentence. The index quantifies the use of relative clauses by measuring their frequency per sentence. It involves identifying specific relativizers, such as *-(u)n/nun*, *-l*, and *-ten*, within the text and then calculating their occurrence as a ratio of the total number of sentences. Prenominal adjectives were also included in this calculation, following the syntactic account that analyzes adjectives as reduced relative clauses (M-J. Kim, 2002; Y-K. Kim, 1997). A higher score on this index suggests that the learner can better elaborate on nouns using sophisticated structures for modification.
- Number of adjunct clauses per sentence. This index focuses on the use of adjunct clauses, a type of subordinate clause introduced by adverbializers like *because* and *if*. The index calculates the frequency of subordinating morphemes (e.g., *-e/aseo*

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:08:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126102009

"because", -(*u*)*myen* "if") by dividing their total count by the total number of sentences in the text. The higher the score on this index, the better the learner is at elaborating on a main idea and establishing logical connections between different concepts.

- Number of verbs per sentence. Recognizing the crucial role of verbs in conveying core sentence information and argument structure (Hovav et al., 2010; Levin, 1993), this index measures verb usage. It counts all types of verb stems in Korean: typical verbs (e.g., *ha-* "do"), adjectival verbs (e.g., *yeyppu-* "beautiful"), auxiliary verbs (e.g., *cwu-* "give"), and copulars (e.g., *-i-* "be"). The index is then calculated by dividing the total verb count by the total number of sentences in the given text. An increased score on this index indicates that the learner demonstrates the ability to pack more actions and states into a single sentence.
- Number of nominal particle types per sentence. This index evaluates the use of nominal particles per sentence, which is vital for understanding argument structure in Korean. Due to Korean's flexible word order, particles provide key morphosyntactic and semantic cues that clarify the grammatical relationships between nominal elements in a sentence (M. Lee, 2019). The index is computed by dividing the count of distinct nominal particle types by the total number of sentences in the given text. The computation process incorporates all types of nominal particles, including the nominative case particle *-i/ka*; the accusative case particle *-ul/lul*; the genitive case particle *-uy*; the complement particle *-i/ka*; the adverbial particles, including the locative particle *-eyse* and -(*u*)*lo* and the comparative particle *-pota*; the quotation particle *-lako*; and the auxiliary particles, including the topic marker *-un/nun* and the focus marker *-to*. A higher score on this index suggests that the learner has a greater command of the morphosyntactic and semantic information associated with nouns.

Complementing the indices reviewed above, this study proposes three new syntactic complexity indices, which are expected to provide a more comprehensive understanding of L2 Korean production and L2 Korean development.

- Number of classifiers per sentence. This index is computed by dividing the count of classifiers by the total number of sentences in the given text. Classifiers, often referred to as numeral quantifiers, are a required element in noun phrases that include a numeral in Korean. These classifiers categorize the associated noun based on its characteristics, such as shape, material, type, or function (Gao & Malt, 2009). Modern Korean employs 93 classifiers (S. Hwang, Kwon, & Yoon, 2010); for example, *kwon* is used for books (e.g., *chayk han-kwen* book one-CL), while *tay* applies to machines, instruments, and vehicles (e.g., *cha han-tay* car one-CL). Because classifiers must grammatically agree with their associated nouns, their use involves syntactic rather than purely lexical knowledge, making them particularly challenging for L2 learners. In fact, E. Lee and Lardiere (2016) found that even fairly high-proficiency L2 learners of Korean did not reach target-like performance in judging the grammaticality of sentences with classifiers; however, the learners showed evidence of improvement with increasing proficiency. Therefore, we hypothesize that more advanced learners will use a greater number of classifiers in each sentence.
- Number of tense markers per sentence. This index is calculated by dividing the count of tense markers by the total number of sentences in the given text. Conceptualizing and using tense in an L2 can be challenging, especially for learners of an agglutinative language like Korean, where tense is marked by an explicit morpheme that appears after the verb stem and before the sentence ender. Hahn (2005), in her analysis of written narratives by L2 learners of Korean, observed a gradual increase in the use of

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:08:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126102009

past tense markers as learners progressed in proficiency. Based on this finding, we predict that learners with higher proficiency will demonstrate the ability to use a greater number of tense markers in each sentence.

- Number of honorific markers per sentence. This index is computed by dividing the count of the honorific marker -(*u*)*si*- by the total number of sentences in the given text.[3] This unique Korean morpheme type is added to verbs to indicate honorification of the subject. Specifically, when the subject refers to an honorable person, such as grandfather or teacher, the verb should syntactically agree with the subject through the use of the honorific marker -(*u*)*si*-. L2 learners of Korean often struggle with the acquisition and processing of honorific markers. For example, Mueller and Jiang (2013) found that, unlike L1 speakers of Korean, L2 learners did not show sensitivity to honorific markers in a word-by-word self-paced reading comprehension task. However, there is evidence that some advanced learners do succeed in mastering the honorific system (J.-B. Hwang et al., 2008). Thus, we hypothesize that more advanced learners will be able to use a greater number of the honorific marker -(*u*)*si*- in each sentence.

Accuracy

K-text follows the convention of measuring accuracy with a single, global metric, in this case, the ratio of accurate sentences to total sentences (e.g., Larsen-Freeman, 2006; Vercellotti, 2017). Our use of only one index was primarily motivated by the fact that this study constitutes an initial attempt to automate accuracy assessment in Korean; for this reason, we adopted the most traditional global measure as a starting point. A further practical consideration was that our accuracy measurement depended on the sentence-based grammatical error correction system (i.e., KAGAS; Yoon et al., 2023), which did not facilitate the ready computation of other accuracy indices, such as error-free morphemes or error-free eojeols. KAGAS is designed to generate grammatically correct sentences after identifying a range of errors related to word insertion and deletion, spacing, word order, spelling, punctuation, and shortening, as well as errors involving verbs, adjectives, nouns, particles, enders, modifiers, and conjugations. By processing each sentence through KAGAS, K-text coded sentences containing any of the specified errors as 0, while coding error-free sentences as 1. The accuracy index is finally calculated as the sum of these codes divided by the total number of sentences in the text.

Fluency

For assessing fluency, K-text's indices are based on production length, following Barrot and Agdeppa (2021), Choi and Sung (2020), and Wolfe-Quintero, Inagaki, and Kim (1998). Faerch, Haarstrup, and Phillipson (1984) and Hedge (2000) maintain that fluency is closely linked to the ability to effectively make use of linguistic resources for successful communication. When learners can seamlessly combine linguistic units, they achieve a state of communicative fluency. From this viewpoint, each linguistic unit represents a distinct reference, idea, or message, and the number of units employed by a learner can be taken as an indicator of their fluency level (Barrot & Agdeppa, 2021; see also Foster, Tonkyn, & Wigglesworth, 2000; Fraser, 2014). It is

[3]It should be noted that the Korean language features other honorific markers, including the honorific nominative case marker -*kkeyseo*. This study's analysis only considered the most productive honorific marker -*(u)si*-.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:08:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126102009

important to note that length-based indices have sometimes been treated as measures of complexity. Yet, longer sentences or texts do not necessarily demonstrate greater syntactic complexity, which means that such indices fail to represent the nature of complexity (e.g., Bulté, Housen, & Pallotti, 2024). To measure the construct of fluency, we propose and test the three indices below.

- Number of morphemes per text. This index is computed by counting the total number of morphemes in a given text. It aims to provide information about a Korean learner's fluency in using morphemes.
- Number of eojeols per text. This index is calculated by counting the total number of eojeols in a given text. Its purpose is to provide insights into a Korean learner's fluency at the eojeol level.
- Number of sentences per text. This index is a traditional, large-grained measure of the overall length of a text. For the calculation, K-text tokenizes a text into individual sentences and then counts the total number of sentences.

Reliability analysis

To analyze the reliability of the K-text indices, we partially adopted an approach from prior research (Mizumoto, 2025; Polio & Yoon, 2018). For this analysis, two independent human coders with graduate training in applied linguistics and expertise in Korean grammar manually coded most syntactic complexity, accuracy, and fluency indices; for MATTR and MTLD, the coders instead counted the token and type frequencies of all morphemes, content morphemes, and function morphemes, since these measures depend primarily on those counts. Before the coding process started, the author had individual meetings with the human coders to explain each index in detail using the information presented in the previous sections, without providing details of how K-text automatically calculates each index. There was no separate written coding manual.

The data were then coded by human coders across three rounds. For this purpose, a total of 54 essays were selected from the written section of the Korean Learners' Corpus (National Institute of Korean Language, 2025), with the number determined on the basis of practical feasibility. The sampling procedure was carried out using the RAND() function in Excel so as to secure both random selection and coverage of different genres and proficiency levels, with the top three essays selected from the randomized list for each genre and proficiency level. As a result, the subset of data for reliability analysis included three essays from each of three genres (argumentative, descriptive, narrative) and from each of six proficiency levels (see Section "Method"). In the first round, two coders independently coded 11 sentences from one randomly chosen sample. After examining and comparing the final output, the author identified areas of disagreement, shared these with the two coders, and provided further guidance so that agreement could be achieved. For example, the first coder was told that the ratio of accurate sentences to total sentences should be determined according to descriptive, rather than prescriptive, grammar, and the second coder was instructed to include all relative clauses, as some had not been counted. In the second round, two coders independently coded 263 sentences from 17 essays. In the third round, Coder 1 coded the remaining 248 sentences from 18 essays, while Coder 2 coded a separate set of 264 sentences from another 18 essays.

Inter-rater reliability was then assessed using a two-way mixed-effects model for all indices calculated for each essay from the coding of individual sentences. As shown in Table 2, these analyses showed that intraclass correlation coefficients (ICCs) between the two human coders ranged from .784 to .998 across all indices for 18 essays measured

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:08:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126102009

Table 2. Reliability of Indices/Measures for K-Text

Index/Measure	Reliability between Two Human Coders	Reliability between K-text and Human Coders
Token frequency of all morphemes	.985	.997
Token frequency of content morphemes	.971	.985
Token frequency of function morphemes	.976	.987
Type frequency of all morphemes	.983	.922
Type frequency of content morphemes	.974	.914
Type frequency of function morphemes	.976	.907
Number of morphemes per sentence	.984	.921
Number of eojeols per sentence	.997	.930
Number of coordinate phrases per sentence	.805	.851
Number of relative clauses per sentence	.978	.953
Number of adjunct clauses per sentence	.868	.772
Number of verbs per sentence	.947	.898
Number of nominal particle types per sentence	.974	.825
Number of classifiers per sentence	.784	.857
Number of tense markers per sentence	.982	.987
Number of honorific markers per sentence	.911	.938
Ratio of accurate sentences to total sentences	.979	.768
Number of morphemes per text	.985	.995
Number of eojeols per text	.998	1.000
Number of sentences per text	.998	.964

in the first two rounds of coding, indicating substantial agreement (i.e., above .61) and, in most cases, excellent agreement (i.e., above .81; see Fleiss & Cohen, 1973). For the reliability analysis of the K-text indices, ICCs were calculated by comparing them with the average measures of the two human coders for 18 essays from the first two rounds of coding and with the measures of either Coder 1 or Coder 2 for 36 essays from the final round of coding. The overall reliability of the K-text indices fell within the substantial-to-excellent range (range of ICCs: .768–1.000).

The present study

This study aims to validate 20 complexity, accuracy, and fluency indices measurable through K-text. Specifically, it examines whether these indices can predict overall criterion-based Korean language proficiency based on L2 learners' essays. The following research questions (RQs) guide the present study.

RQ1: Do the individual K-text indices for complexity, accuracy, and fluency significantly predict L2 proficiency?

RQ2: What is the overall contribution of the complete set of K-text indices in explaining variance in L2 proficiency?

Method

Data

The preliminary validity of K-text's indices was assessed using the most recently updated version of the Korean Learners' Corpus (National Institute of Korean Language, 2025). This corpus comprises both written and spoken data from learners of Korean, representing a wide range of L1s, including (in order of representation in the corpus): Mandarin Chinese, Vietnamese, Japanese, Bengali, English, Cantonese, Thai,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:08:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126102009

Russian, Mongolian, Indonesian, Spanish, French, Arabic, Swedish, and others. The corpus provides information about each contributor's levels on the Test of Proficiency in Korean (TOPIK), which is a standardized assessment that measures the overall listening, reading, and writing proficiency of L2 learners of Korean (https://www.topik.go.kr/). The levels of proficiency span from beginner at level 1, equivalent to the A1 level of the Common European Framework of Reference (CEFR; Council of Europe, 2020), to highly proficient at level 6, equivalent to the CEFR C2 level.

The analysis focused on written texts within the corpus, which were composed in approximately 60 minutes. These texts include various genres, such as argumentative essays, autobiography, descriptive essays, letters, narratives, and travelogues, and they cover diverse topics, like future plans, success, environmental conservation, economic development, shopping, water pollution, and leisure activities. Out of 53,070 written texts, we extracted 50,865 essays of argumentation, description, and narration, as these genres are the most common in the corpus. The corpus background information sheet suggests that each text was given a unique sample ID and that each participant contributed one text. Of these, 238 essays were excluded because the writers' proficiency information was unavailable or unclear. We further excluded 4,712 essays that contained fewer than 50 morphemes, as this is the minimum amount needed to calculate the MATTR. Finally, 60 more essays were removed from the dataset due to a probable error in which the writers' L1 was listed as Korean. The final dataset included 45,855 essays from 45,855 learners, which had a total of 705,132 sentences (M = 15.38; SD = 7.08).

An a priori power analysis (Soper, 2025) indicated that the model for a preliminary analysis for RQ2—with more variables than that for RQ1—required a sample size of 162 to reach the target power level of 0.8. The size of the dataset was thus considered adequate to detect the effects to be tested. Table 3 provides a summary of the dataset, detailing information for the three genres and six proficiency levels.

Analysis

The complete R code, as well as the dataset, used for our analyses is accessible at https://osf.io/e28qy. For statistical analyses, we partially adopted the approach of H. Hwang and H. Kim (2024) and Kyle and Crossley (2017). As a preliminary step, we converted proficiency levels into continuous scores from 1 to 6 in line with the practice of treating

Table 3. Summary of the Data Used in the Present Study

Genre	Argumentative		Descriptive		Narrative	
Proficiency	Number of Samples	Mean Number of Sentences	Number of Samples	Mean Number of Sentences	Number of Samples	Mean Number of Sentences
1	431	17.56 (SD = 7.62)	2311	19.40 (SD = 8.03)	5323	16.99 (SD = 8.30)
2	737	14.39 (SD = 7.34)	3860	15.82 (SD = 5.58)	6366	14.60 (SD = 7.09)
3	1722	15.51 (SD = 6.34)	2494	14.50 (SD = 5.66)	5527	14.74 (SD = 6.90)
4	4034	13.87 (SD = 4.85)	1255	15.19 (SD = 5.74)	2624	16.16 (SD = 9.99)
5	4370	14.39 (SD = 4.38)	858	14.44 (SD = 5.17)	761	19.52 (SD = 12.49)
6	2382	13.52 (SD = 4.61)	270	13.41 (SD = 5.26)	530	18.91 (SD = 12.03)

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:08:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126102009

ordinal variables with more than four levels as continuous (e.g., Robitzsch, 2020). In addition, we identified and removed outliers for each index based on a threshold of ±3 standard deviations (e.g., Kisselev, Klimov, & Kopotev, 2022), which led to the removal of 11.80% of the written data in total. To normalize the data, we rescaled the indices using a combination of log-transformation and Laplace smoothing, specifically applying the formula log(index score + 1). The resulting normalized values were then rounded to three decimal places for precision. Following this data preparation, we proceeded to examine the data in relation to our two RQs, as detailed in the subsequent sections.

Mixed-effects linear regression: Contribution of each K-text index to predicting proficiency

We addressed RQ1 by running 20 separate mixed-effects linear regression models in R (R Core Team, 2025) using the "lme4" package (Bates et al., 2015). Each model included one of the indices as a fixed effect, genre and learners' L1 as random effects on the intercept, and proficiency scores as the dependent variable. Because 20 models were constructed on the same dataset, a Bonferroni correction was applied, and statistical significance was interpreted at $p = .0025$ (.05/20).

Stepwise linear regression: Contribution of the complete set of K-text indices to predicting proficiency

For RQ2, we conducted a forward stepwise linear regression analysis to determine the extent to which the K-text indices, as a group, explain learner proficiency, with all complexity, accuracy, and fluency indices considered as predictors. In this analysis, it was essential that the final stepwise model include only unique explanatory variables. To reduce the chance of incorrectly identifying significant results from certain indices (Type I error), we initially fitted a mixed-effects regression model with proficiency as the dependent variable, including all indices simultaneously as fixed effects to control for one another, and genre and L1 as random intercepts. Indices that were not significant at the $p < .05$ level were subsequently removed from the stepwise model. To further address multicollinearity issues in the final model, we iteratively removed the indices with the highest Variance Inflation Factors (VIFs) until all remaining indices had VIFs of less than 4, a commonly used threshold above which predictors are considered correlated with one another (e.g., Fox, 1991).

After refining the index set, we conducted a stepwise regression analysis based on the Akaike information criterion (AIC) method (Akaike, 1974), utilizing the "olsrr" package in R (R Core Team, 2025). This analysis begins with the strongest predictor and refines the model by adding variables one at a time. The variable that most reduces the AIC is added at each step. The process stops when no remaining variable can improve the model's predictive power. To confirm the robustness of the stepwise regression model, a 10-fold cross-validation was performed via R's "caret" package. The dataset was randomly divided into 10 approximately equal folds. For each iteration, nine folds were used as the training set and the remaining fold was used as the validation set. After completing all ten iterations, the performance estimates were averaged across the ten folds.

Results

Contribution of each K-text index to predicting proficiency

This section reports the results of a series of linear mixed-effects regression analyses conducted to examine whether individual K-text indices of complexity, accuracy, and fluency significantly predict L2 proficiency (RQ1). As detailed in Table 4, the results indicate that all 20 indices are significant predictors of learner proficiency scores under the Bonferroni-corrected significance threshold (all ps < .001). Visual representations

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:08:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126102009

Table 4. Linear Mixed-Effects Model Output for Each K-Text Index

Category	Index	*b*	95% CI	*SE*	*t*	*p*	R^2_M	R^2_C
Complexity	Moving-average type-token ratio for all morphemes	14.99	[14.58, 15.40]	.21	71.55	< .001	.07	.47
	Moving-average type-token ratio for content morphemes	3.82	[3.56, 4.08]	.13	28.43	< .001	.01	.47
	Moving-average type-token ratio for function morphemes	11.25	[10.99, 11.52]	.14	83.20	< .001	.10	.49
	Measure of textual lexical diversity for all morphemes	1.44	[1.40, 1.48]	.02	69.71	< .001	.07	.47
	Measure of textual lexical diversity for content morphemes	.35	[0.33, 0.38]	.01	29.68	< .001	.01	.47
	Measure of textual lexical diversity for function morphemes	1.85	[1.81, 1.89]	.02	87.15	< .001	.11	.49
	Number of morphemes per sentence	2.33	[2.29, 2.37]	.02	117.24	< .001	.22	.50
	Number of eojeols per sentence	2.29	[2.25, 2.33]	.02	113.90	< .001	.20	.50
	Number of coordinate phrases per sentence	1.76	[1.71, 1.82]	.03	62.39	< .001	.06	.46
	Number of relative clauses per sentence	2.08	[2.05, 2.12]	.02	107.40	< .001	.19	.48
	Number of adjunct clauses per sentence	2.03	[1.99, 2.09]	.02	82.15	< .001	.10	.48
	Number of verbs per sentence	2.29	[2.24, 2.33]	.02	100.32	< .001	.15	.50
	Number of nominal particle types per sentence	2.55	[2.49, 2.62]	.03	76.15	< .001	.10	.46
	Number of classifiers per sentence	.85	[0.71, 0.98]	.07	12.34	< .001	.00	.47
	Number of tense markers per sentence	.59	[0.53, 0.64]	.03	20.38	< .001	.01	.48
	Number of honorific markers per sentence	3.32	[2.66, 3.98]	.34	9.85	< .001	.00	.47
Accuracy	Ratio of accurate sentences to total sentences	−1.37	[−1.45, −1.29]	.04	−32.30	< .001	.01	.47
Fluency	Number of morphemes per text	1.41	[1.37, 1.44]	.02	78.90	< .001	.10	.48
	Number of eojeols per text	1.42	[1.39, 1.45]	.02	81.38	< .001	.10	.48
	Number of sentences per text	−0.29	[−0.33, −0.25]	.02	−15.10	< .001	.00	.46

Note: Model formula: lmer(Proficiency ~ Index + (1 | genre) + (1 | L1)); *b* = unstandardized coefficient, *SE* = standard error, R^2_M = marginal R^2, R^2_C = conditional R^2.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:08:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126102009

of these relationships in Figure 2 show that as learners became more proficient, the majority of the indices also increased. However, the ratio of accurate sentences to total sentences and the number of sentences per text show an inverse relationship, decreasing with higher proficiency.

In sum, the findings provide preliminary evidence that all K-text indices are valid for predicting the overall criterion-based L2 proficiency of Korean learners. However, the relationship between these indices and proficiency is not consistent in our research context, as seen in the case of the ratio of accurate sentences to total sentences and the number of sentences per text, which tend to decrease as proficiency increases.

Contribution of the complete set of K-text indices to predicting proficiency

This section presents the results of a stepwise regression analysis conducted to evaluate the overall contribution of the complete set of K-text indices in explaining variance in L2 proficiency (RQ2). Based on our preliminary analyses, 11 indices were eliminated prior to the stepwise linear regression to control for Type I errors and multicollinearity. First, the mixed-effects regression model incorporating all 20 indices showed that the number of eojeols per sentence (p = .210), number of nominal particle types per sentence (p = .693), number of classifiers per sentence (p = .238), MATTR for content morphemes (p = .927), MTLD for all morphemes (p = .372), MTLD for content morphemes (p = .056), and number of eojeols (p = .178) did not have significant effects on proficiency. Further, the following indices showed VIF values higher than 4: number of morphemes per text (VIF = 2080.45), number of verbs per sentence (VIF = 7.60), number of morphemes per sentence (VIF = 6.29), and MATTR for function

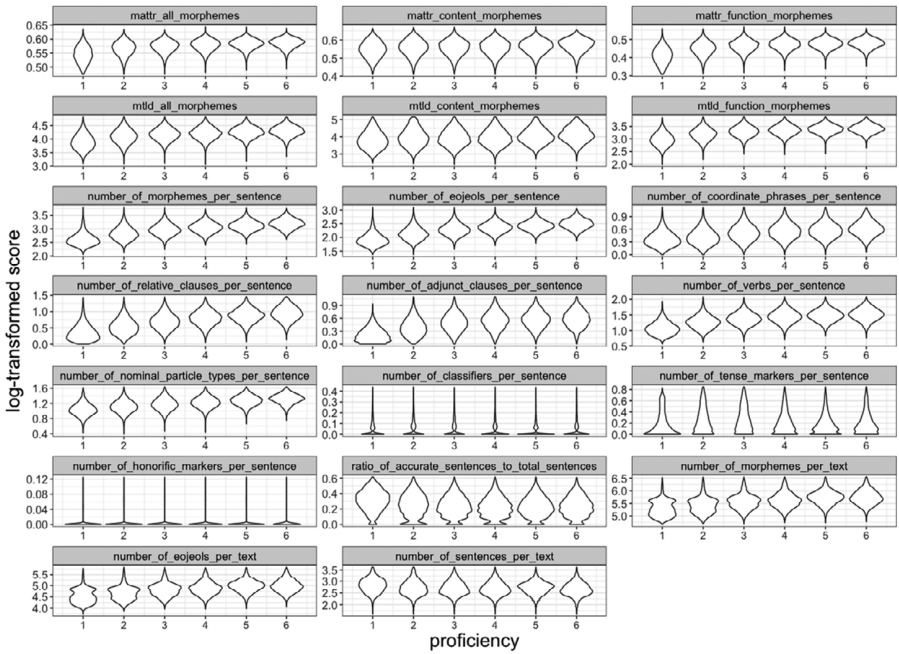


Figure 2. Relationship between proficiency and log-transformed scores of K-text indices.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:08:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126102009

morphemes (VIF = 4.78). The process of removing these indices ensured that all of the remaining nine indices had VIFs lower than 4.

The results of the stepwise regression, presented in Table 5, showed that the final model was statistically significant: $F(9, 40{,}438) = 3602.07$, $p < .001$. This model explained 44.50% of the variance in proficiency scores ($R^2 = .445$). The model's robustness was confirmed using a 10-fold cross-validation, which revealed that the model accounted for 44.50% of the variance in proficiency scores ($R^2 = .445$).

With all nine indices identified as significant predictors, the number of relative clauses per sentence contributed the most to the final model, as evidenced by the largest decrease in AIC compared to the initial model (i.e., 147127.29 – 129638.44 = 17488.85). The remaining indices, ranked by their predictive power, were MTLD for function morphemes (AIC = 129638.44 – 126431.08 = 3207.36), number of adjunct clauses per sentence (AIC = 126431.08 – 125379.42 = 1051.66), number of sentences per text (125379.42 – 124376.41 = 1003.01), ratio of accurate sentences to total sentences (AIC = 124376.41 – 123833.52 = 542.89), and so forth.

To summarize, our stepwise linear regression analyses, using a substantial learner dataset, demonstrated that indices related to complexity, accuracy, and fluency all significantly predicted proficiency. These findings provide preliminary support for the validity of the K-text indices in explaining the overall criterion-based L2 proficiency of Korean learners.

Discussion

Theoretical implications

Our mixed-effects linear regression analyses of K-text's indices using a large number of essays demonstrated that each index assessing specific dimensions of complexity, accuracy, and fluency significantly contributed to predicting overall criterion-based L2 Korean proficiency. First, for complexity, all 10 syntactic indices were found to be significant predictors of overall L2 Korean proficiency. Three of these 10 indices were newly introduced in this study, and these indicated that higher-proficiency learners used more classifiers, tense markers, and honorific markers per sentence, suggesting their greater command of the grammatical resources needed to classify nouns with numerals and to express tense and politeness through verb morphology. In addition, all six morphological indices were identified as significant predictors of L2 Korean proficiency. Regardless of the specific measure of morphological diversity (e.g., MATTR, MTLD), higher-proficiency learners exhibited greater diversity in using morphemes overall, while showing more varied use of both content and function morphemes.

As for fluency, the number of morphemes and the number of eojeols per text were significant predictors of L2 Korean proficiency. Higher-proficiency learners demonstrated better mastery of morphemes and eojeols, as evidenced by their ability to incorporate greater numbers of these units into their texts, thereby enhancing the overall smoothness of their writing.

Somewhat counterintuitively, the sentence-level fluency index and the single accuracy index revealed inverse relationships with overall proficiency, with higher-proficiency learners exhibiting lower scores on these measures. This pattern may be understood with reference to the presumed developmental sequence whereby complexity emerges before accuracy and fluency. In particular, learners at higher proficiency levels may prioritize the production of more complex language, and such

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:08:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126102009

Table 5. Stepwise Multiple Regression Model Output

Step	Effects	AIC	R^2	*b*	95% CI	*SE*	*t*	*p*
0	Base model (intercept)	147127.29	.000	−4.93	[−5.21, −4.64]	.14	−34.11	< .001
1	Number of relative clauses per sentence	129638.44	.351	2.02	[1.97, 2.07]	.03	80.52	< .001
2	MTLD function morphemes	126431.08	.401	.96	[0.90, 1.03]	.03	28.25	< .001
3	Number of adjunct clauses per sentence	125379.42	.416	1.05	[0.99, 1.11]	.03	33.62	< .001
4	Number of sentences per text	124376.41	.430	.70	[0.66, 0.74]	.02	35.85	< .001
5	Ratio of accurate sentences to total sentences	123833.52	.438	−.97	[−1.05, −0.90]	.04	−24.17	< .001
6	Number of coordinate phrases per sentence	123467.21	.443	.58	[0.52, 0.65]	.03	17.92	< .001
7	Number of tense markers per sentence	123396.55	.444	−.22	[−0.28, −0.17]	.03	−8.21	< .001
8	MATTR all morphemes	123357.27	.445	2.04	[1.43, 2.66]	.31	6.50	< .001
9	Number of honorific markers per sentence	123332.58	.445	−1.68	[−2.32, −1.04]	.33	−5.17	< .001

Note: AIC = Akaike information criterion; *b* = unstandardized coefficient; *CI* = confidence interval; *SE* = standard error.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:08:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126102009

development in complexity may temporarily outpace gains in accuracy and fluency. As they experiment with more complex grammar, including adjunct and relative clauses, their production may become more intricate and longer but also more error-prone and less sentence-dense. From this perspective, the observed trade-off may reflect not regression, but a developmental focus on greater linguistic complexity at higher proficiency levels.

Furthermore, the nuanced and dynamic relationship between accuracy and proficiency observed in this study aligns, at least in part, with findings from previous research, which suggest that accuracy does not necessarily improve in tandem with proficiency. For instance, Brown and Iwasaki's (2013) research showed that L2 Korean learners' accuracy with the postpositions *-eyse* and *-ey* decreased before increasing in both spoken and written production, forming a U-shaped developmental pattern. Similarly, Yadollahi, Moloodi, and Rahimi's (2024) study on L2 English writing showed complex patterns in the development of accuracy. They found a U-shaped trend in punctuation accuracy; however, article and number agreement accuracy showed a different pattern, with significant improvement from beginner to intermediate levels, but little change from intermediate to advanced levels, indicating a more stable pattern at higher levels of proficiency. These findings underscore the importance of developing and testing more fine-grained accuracy indices that target specific linguistic features, such as case particles, to provide a more detailed understanding of L2 learners' accuracy in the Korean language.

The final stepwise regression model, incorporating the complete set of CAF indices, explained 44.50% of the variance in overall proficiency. Importantly, the model retained at least one index from each dimension—complexity (e.g., number of relative clauses per sentence), accuracy (e.g., ratio of accurate sentences to total sentences), and fluency (e.g., number of sentences per text)—highlighting the unique contribution of each to overall proficiency. These results align with the CAF framework, which posits that "L2 proficiency is not a unitary construct but […] is multicomponential in nature, and […] its principal components can be fruitfully captured by the notions of complexity, fluency and accuracy" (Housen et al., 2012, p. 3). By empirically testing the preliminary validity of indices that capture these components, the current study contributes to addressing the CAF framework's core question as to how L2 proficiency can be adequately measured.

Notably, the stepwise regression process resulted in a final model containing nine indices. Five of these were syntactic complexity indices, such as the number of relative clauses per sentence (rank 1), number of adjunct clauses per sentence (rank 3), number of coordinate phrases per sentence (rank 6), number of tense markers per sentence (rank 7), and number of honorific markers per sentence (rank 9). The strong predictive power of the first three indices, which all relate to connecting different linguistic units for elaboration, indicates that higher-proficiency learners are better able to organize and expand their ideas (e.g., Jiang, Bi, & Liu, 2019; H. Hwang & H. Kim, 2024), using a wider range of syntactic tools, like relativizers, subordinators, and coordinators. The last two syntactic complexity indices pertain to the use of verb morphemes, which are difficult to master in agglutinative languages. This result indicates that higher usage of verbal morphemes is closely tied to higher proficiency, which is in line with the findings of Hahn (2005) and J.-B. Hwang et al. (2008).

To further explore the usage of verbal morphemes, we conducted a qualitative analysis of descriptive essays on the topic of an admired person. This analysis showed that more advanced learners consistently used correct past tense markers, whereas less

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:08:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126102009

Table 6. Excerpts from Descriptive Essays by Two L2 Korean Learners at High and Low Proficiency Levels

Participant Code	Excerpt from Essay
19100 (Proficiency level: 6)	*Yengkwuk* *cen* *chongli* *eynsuniitun-i* *yengkwuk* UK prior prime ministry Anthony Eden-NOM UK *kwicok* *kamwun-eyse* *thayen-ass-ko* *yengkwuk-eyse* nobility family-LOC born-PST-CONN UK-LOC *meylithu* *kyoyuk-ul* *pat-ass-ta.* elite education-ACC receive-PST-DECL "Former British Prime Minister Anthony Eden was born into an aristocratic family in the UK and received an elite education there." *Ely-ess-ul* *ttay-pwuthe* *oykwuk-e-eytayhan* young-PST-RL time-from foreign-language-about *kwansim-i* *manh-ass-ko* *oykyokwan-i* *toy-ko* interest-NOM a lot-PST-and diplomat-NOM become-CONN *siph-e-hay-se* *20tay* *ttay* *Oxford* *tayhak-ey* want-CONN-do-because 20s time Oxford university-LOC *ka-se* *oykyohak-ul* *cenkong-hay-ss-ta.* go-CONN diplomacy-ACC major-do-PST-DECL "From a young age, he had a strong interest in foreign languages and aspired to become a diplomat, so in his twenties, he entered Oxford University and majored in diplomacy."
29442 (Proficiency level: 2)	*halwu* *congil* *pyengwen-eyse* *ku* *chinkwu-lul* *cal* day all day hospital-LOC that friend-ACC well *chayng-ky-e-se* *machi* *apeci* *kath-a-yo.* take.care-CAU-CONN-because like father like-DECL-POL "(Tami) takes [intended: took] such good care of that friend at the hospital all day—she is [intended: was] like her father." *tami* *sensayngnim* *tekpwuney* *chinkwu-ka* *mwusa-hal* Tami teacher because friend-NOM stay safe-do *swu* *iss-ta.* can be-DECL "Because of Ms. Tami, my friend is [intended: was] able to stay safe."

advanced learners often omitted the past tense marker where it was warranted, as illustrated in Table 6.

Besides the accuracy index (ratio of accurate sentences to total sentences, rank 5) and the fluency index (number of sentences per text, rank 4), the stepwise regression model revealed that two morphological complexity indices (MTLD for function morphemes, rank 2; MATTR for all morphemes, rank 8) were significant predictors of L2 Korean proficiency. These findings emphasize the unique characteristics of Korean, where meaning is densely packed through the combination of content and function morphemes. Relying solely on English-based measures, like the subordination ratio, risks overlooking this key aspect of L2 Korean development. The current study therefore

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:08:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126102009

argues that incorporating language-specific indices is crucial for a valid assessment of L2 proficiency and recommends this approach in future research across diverse languages.

Practical implications

This study demonstrates the soundness and utility of K-text's indices for research, assessment, and pedagogical purposes. While previous work (e.g., H. Hwang & H. Kim, 2024; H. Hwang, 2024) offered downloadable software for analyzing some of these indices, software is susceptible to issues related to operating system upgrades and computer architecture differences. To address these limitations, K-text was developed as a website platform that functions efficiently across different computer environments. This makes K-text a valuable and accessible resource for the automated computation of linguistic indices, removing the burden of manual calculation from users (e.g., Y. Kim, Nam, & Lee, 2016; Seo, 2009).

In terms of language assessment, the K-text indices enable a nuanced evaluation of L2 Korean proficiency, encompassing the distinct aspects of complexity, accuracy, and fluency. By examining a range of indices, each targeting a specific component of proficiency, assessors can develop a detailed profile of a learner's language abilities, which may then be used to estimate their overall proficiency. In the long term, the theoretically grounded indices proposed in this study could contribute to the development and improvement of automated essay scoring (AES) systems. Unlike many current AES systems that rely on nonlinguistic features, such as character count, our approach, based on interpretable linguistic features, should help create more transparent and explainable AI systems for essay assessment and proficiency prediction.

Pedagogically, the K-text indices may offer a useful tool for teachers, allowing them to identify specific areas for improvement in their students' language production. By analyzing weaknesses in particular dimensions, whether complexity, accuracy, or fluency, or in certain subcomponents, like the use of function morphemes, teachers can develop more targeted interventions and refine their teaching materials. In this way, K-text's use could contribute to a higher quality of education that is tailored to students' individual needs and developmental levels, thus promoting more effective teaching and learning.

While not central to this study, tools for topic modeling and user-defined indices are also provided by K-text. In particular, the user-defined index feature allows users without programming skills to create custom indices by selecting part-of-speech (POS) categories from a dropdown menu. For example, users can analyze nominal and accusative case particles separately from all case particles to gain a deeper understanding of how their usage varies by type. This user-friendly application of NLP techniques has the potential to advance theoretical and methodological innovation, benefiting the academic community interested in Korean language production.

Limitations and future directions

One limitation of the present study is the reliance on TOPIK-based proficiency scores as the dependent variable in the regression models conducted. These scores represent an aggregated measure of multiple language skills, including reading, writing, and listening, whereas the current study is primarily concerned with proficiency constructs

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:08:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126102009

in writing. As such, there is a mismatch between the construct being predicted and the construct the study seeks to validate. This means that predicting general TOPIK level from writing features cannot be taken as direct validation of writing proficiency indices. To address this issue, future work should employ independent assessments of writing proficiency that closely match the focus of the analysis.

Also, the texts in the dataset differed in genre and topic across proficiency levels. Specifically, lower proficiency learners produced relatively more narrative essays, whereas higher-proficiency learners produced more argumentative essays. Such an imbalance in genre and topic distribution may have affected the findings of the present study, and so they should be interpreted with caution.

pt?>A further limitation concerns the treatment of outliers. While outliers were removed in accordance with common practice in the field to ensure the validity of the statistical analyses, the exclusion of 11.80% of the data may have influenced the distributional properties of the dataset across the six proficiency levels. Further research should examine the extent to which different outlier-treatment methods influence the results.

Regarding the NLP-based analysis, POS tagging can be sensitive to spacing, morphological, and orthographic errors in learner Korean, particularly at lower proficiency levels. These deviations may compromise tagging accuracy and, as a result, affect the computation of certain K-text indices. To partially address this issue, we investigated discrepancies between K-text and the human coders by proficiency level, drawing on the available reliability data. Our preliminary analysis showed that most indices were reliable across the proficiency levels considered, with values above .61, suggesting substantial agreement (Fleiss & Cohen, 1973; see Figure A at https://osf.io/e28qy). However, several indices showed relatively lower agreement for certain proficiency levels: number of eojeols per sentence for level 5 (ICC = .564), number of coordinate phrases per sentence for level 2 (ICC = .476), number of adjunct clauses per sentence for levels 3 (ICC = .545) and 4 (ICC = .541), number of nominal particle types per sentence for levels 4 (ICC = .591) and 5 (ICC = .500), and ratio of accurate sentences to total sentences for level 6 (ICC = .451). These results suggest that the effects of these indices in predicting overall proficiency may partly reflect NLP-related errors rather than genuine developmental patterns. Nevertheless, all the indices at issue fall within the moderate reliability range of .41 to .61 (Fleiss & Cohen, 1973). Also, their relatively low reliability may be attributable to the small number of essays examined because an ICC analysis was based on only nine essays per proficiency level—a sample size not recommended for reliability analysis. In comparison, when a greater number of essays was analyzed (n = 18) across the three genres (argumentative, descriptive, and narrative), all indices showed at least substantial reliability, with values above .61 (see Figure B at https://osf.io/e28qy). Accordingly, this issue warrants further investigation in future research.

From a linguistic standpoint, the scope of K-text could be expanded to include other levels of analysis, such as discourse. Currently, its indices primarily focus on morphology and syntax when assessing complexity and fluency. Furthermore, some indices are rather coarse-grained as noted by reviewers. Regarding the number of relative clauses per sentence, several possible patterns are identifiable, such as (a) traditional relative clauses (e.g., (2)), (b) fact-S type clauses (e.g., (3)), (c) headless relative clauses (e.g., (4)), and (d) coreferent-opaque clauses (e.g., (5)), where the embedded predicate may take either a verbal (e.g., (2a)) or adjectival from (e.g., (2b)), with (e.g., (2c)) or without a modifier (e.g., (2b)) (Sohn, 1999). For

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:08:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126102009

accuracy, K-text only measures the overall proportion of error-free sentences. The coarse-grained nature of this single metric may be associated with the observed trade-off between complexity and accuracy. Incorporating alternative accuracy indices, such as morpheme-based or eojeol-based measures, could yield a more comprehensive understanding of how accuracy evolves alongside complexity, as these are likely to be less influenced by variations in sentence length and complexity. In addition, reliability analyses comparing K-text with the human coders showed relatively lower agreement for the ratio of accurate sentences to total sentences, especially at level 6 (see above). Moving forward, we plan to develop more fine-grained indices that capture complexity, accuracy, and fluency at diverse linguistic levels while building a deep learning-based system specifically trained to judge accuracy in learner data. This approach will allow us to capture a more detailed picture of L2 Korean production and development such that different types of relative clauses may follow different developmental paths across proficiency levels, and morphological accuracy may emerge before sentential accuracy.

(2) a. nay-ka manna-n yeca
I-NOM meet-RL woman
"the woman whom I met"

(adapted from Sohn, 1999, p. 310, (79a))

b. yeypp-un yeca
beautiful-RL woman
"the woman who is beautiful"

c. acwu yeypp-un yeca
very beautiful-RL woman
"the woman who is very beautiful"

(3) nay-ka yeca-lul manna-n sasil
I-NOM woman-ACC meet-RL fact
"the fact that I met a woman"

(adapted from Sohn, 1999, p. 312, (83a))

(4) Minho-nun thayngkhu-ka ka-nun kes-ul po-ass-ta.
Minho-TOP tank-NOM go-RL fact-ACC see-PST-DECL
"Minho saw a tank going."

(adapted from Sohn, 1999, p. 313, (85a))

(5) na-nun kapcaki koki kwup-nun naymsay-lul math-ass-ta.
I-TOP suddenly meat bake-RL smell-ACC smell-PST-DECL
"I suddenly smelled the smell of (somebody) broiling meat."

(adapted from Sohn, 1999, p. 314, p. 86)

Finally, future research would benefit from taking a longitudinal approach to the study of Korean proficiency development. Such a design would allow researchers to

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:08:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126102009

track changes in learners' complexity, accuracy, and fluency over time and to gain deeper insight into the developmental trajectories of these constructs within the same learners.

Conclusion

The current study introduces K-text, a user-friendly website designed to measure a range of complexity, accuracy, and fluency indices from user-inputted data. Through a large-scale learner corpus analysis, the study offers preliminary evidence supporting the validity of the tool's indices, demonstrating their individual and combined contributions to predicting overall criterion-based proficiency. This finding underscores the utility of K-text and its associated indices. We expect that both the tool and the insights derived from this study will inform future research, enhance language assessment practices, and support pedagogical improvement.

Data availability statement. The complete R code, as well as the dataset, used for our analyses is accessible at https://osf.io/e28qy.

Acknowledgments. The data used in this study were drawn from the Korean Learner Corpus. I am grateful to the National Institute of Korean Language for making the data available. My thanks also go to Joonhee Kim and Yena Lee, who worked as coders for the reliability analysis. Generative AI tools (e.g., Gemini 3 Pro, GPT-5.1) were used between May 1, 2025, and August 18, 2026, via the https://chat.chatbot.app/ and https://www.genspark.ai/ platforms to review grammar and proofread selected draft sections that were originally written by the author.

Funding. This research was supported by the 2025 Korean Studies Grant Program of the Academy of Korean Studies (AKS-2025-R-018).

Declaration of Competing interests. The author declares none.

Ethics approval. The research procedure was approved by the Survey and Behavioural Research Ethics Committee at the Chinese University of Hong Kong (Reference No.: SBRE-24-0530).

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723. https://doi.org/10.1109/TAC.1974.1100705

Barrot, J. S., & Agdeppa, J. Y. (2021). Complexity, accuracy, and fluency as indices of college-level L2 writers' proficiency. *Assessing Writing*, *47*, 100510. https://doi.org/10.1016/j.asw.2020.100510

Bates, D., Machler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-8. http://CRAN.R-project.org/package=lme4

Blanco, C. (2025). *2025 Duolingo language report*. https://blog.duolingo.com/2025-duolingo-language-report/

Brown, L., & Iwasaki, N. (2013). Cross-linguistic influence in the L2 acquisition of Korean case particles by Japanese-speaking and English-speaking learners: L1-L2 proximity and learner perceptions. *Electronic Journal of Foreign Language Teaching*, *10*, 176–195.

Bulté, B., Housen, A., & Pallotti, G. (2024). Complexity and difficulty in second language acquisition: A theoretical and methodological overview. *Language Learning*. Advance online publication. https://doi.org/10.1111/lang.12669

Bulté, B., & Roothooft, H. (2020). Investigating the interrelationship between rated L2 proficiency and linguistic complexity in L2 speech. *System*, *91*, Article 102246. https://doi.org/10.1016/j.system.2020.102246

Choi, J., & Sung, M. C. (2020). Utterance-based measurement of L2 fluency in speaking interactions: A constructionist approach. *English Teaching*, *75*, 105–126.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:08:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126102009

Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Council of Europe Publishing. https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4

Covington, M. A., & McFall, J. D. (2010). Cutting the Gordian knot: The moving-average type-token ratio (MATTR). *Journal of Quantitative Linguistics*, *17*, 94–100. https://doi.org/10.1080/09296171003643098

Faerch, C., Haarstrup, K., & Phillipson, R. (1984). *Learner language and language learning*. Multilingual Matters.

Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted Kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, *33*, 613–619. https://doi.org/10.1177/001316447303300309

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, *21*, 354–375. https://doi.org/10.1093/applin/21.3.354

Fox, J. (1991). *Regression diagnostics*. Sage.

Fraser, S. (2014). Assessing fluency: A framework for spoken and written output. In T. Muller, J. Adamson, P. Brown, & S. Herder (Eds.), *Exploring EFL fluency in Asia* (pp. 178–195). Palgrave Macmillan.

Gao, M. Y., & Malt, B. C. (2009). Mental representation and cognitive consequences of Chinese individual classifiers. *Language and Cognitive Processes*, *24*, 1124–1179. https://doi.org/10.1080/01690960802018323

Hahn, H. R. (2005). Past tense acquisition by less successful learners in the Korean EFL classroom setting. *English Teaching*, *60*, 109–132.

Hedge, T. (2000). *Teaching and learning in the language classroom*. Oxford University Press.

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

Hovav, M. R., Doron, E., & Sichel, I. (2010). *Lexical semantics, syntax, and event structure* (No. 27). Oxford University Press.

Hwang, J.-B., Ahn, H.-D., Han, J.-I., Lim, I., Hong, J.-J., & Sim, C.-Y. (2008). A study of foreign learners' acquisition of some combinations of Korean prefinal endings. *Language Sciences*, *15*, 183–199.

Hwang, H. (2024). Development of morphological diversity in second language Korean: An NLP analysis using the Korean Morphological Richness Analyzer 1.0. *System*, *121*, 103260. https://doi.org/10.1016/j.system.2024.103260

Hwang, H., & Kim, H. (2024). Korean Syntactic Complexity Analyzer (KOSCA): An NLP tool for the analysis of syntactic complexity in second language production. *Language Testing*. Advance online publication. https://doi.org/10.1177/02655322231222596

Hwang, S., Kwon, H.-C., & Yoon, A. (2010). KorLexClas 1.5: A lexical semantic network for Korean numeral classifiers. *Journal of KIISE: Software and Applications*, *37*, 60–73.

Jiang, J., Bi, P., & Liu, H. (2019). Syntactic complexity development in the writings of EFL learners: Insights from a dependency syntactically annotated corpus. *Journal of Second Language Writing*, *46*, 100666. https://doi.org/10.1016/j.jslw.2019.100666

Johnson, W. (1944). Studies in language behavior 1: A program of research. *Psychological Monographs*, *56*, 1–15.

Katsumata, S., & Komachi, M. (2020). Stronger baselines for grammatical error correction using a pretrained encoder-decoder model. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing* (pp. 827–832). Association for Computational Linguistics.

Kim, Y., Nam, J., & Lee, S.-Y. (2016). Correlation of proficiency with complexity, accuracy, and fluency in spoken and written production: Evidence from L2 Korean. *Journal of the National Council of Less Commonly Taught Languages*, *19*, 147–181. https://doaj.org/article/5b33f84d53354f8d8ac35353aa2c107e

Kim, Y.-K. (1997). Agreement phrases in DP. *UCL Working Papers in Linguistics*, *9*, 1–24.

Kim, M.-J. (2002). Does Korean have adjectives? *MIT Working Papers in Linguistics*, *43*, 71–89.

Kisselev, O., Klimov, A., & Kopotev, M. (2022). Syntactic complexity measures as linguistic correlates of proficiency level in learner Russian. In A. Leńko-Szymańska & S. Götz (Eds.), *Complexity, accuracy and fluency in learner corpus research* (pp. 51–80). John Benjamins. https://doi.org/10.1075/scl.104.03kis

Kuiken, F., Vedder, I., & Gilabert, R. (2010). Communicative adequacy and linguistic complexity in L2 writing. In I. Bartning, M. Martin, & I. Vedder (Eds.), *Communicative proficiency and linguistic development: Intersections between SLA and language testing research* (pp. 81–100). European Second Language Association.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:08:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126102009

Kyle, K., & Crossley, S. A. (2017). Assessing syntactic sophistication in L2 writing: A usage-based approach. *Language Testing*, *34*, 513–535. https://doi.org/10.1177/0265532217712554

Kyle, K., Sung, H., Eguchi, M., & Zenker, F. (2024). Evaluating evidence for the reliability and validity of lexical diversity indices in L2 oral task responses. *Studies in Second Language Acquisition*, *46*(1), 278–299. https://doi.org/10.1017/S0272263123000402

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*, *27*, 590–619. https://doi.org/10.1093/applin/aml029

Lee, E., & Lardiere, D. (2016). L2 acquisition of number marking in Korean and Indonesian: A feature-based approach. In D. Stringer *et al.* (Eds.), *Proceedings of the 13th Generative Approaches to Second Language Acquisition Conference (GASLA 2015)* (pp. 113–123). Cascadilla Proceedings Project.

Lee, M. (2019). Effects of case-marking on the anticipatory processing of Korean sentences. *Journal of Cognitive Science*, *20*, 339–364.

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. University of Chicago Press.

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, *45*, 36–62. https://doi.org/10.5054/tq.2011.240859

McCarthy, P. M., & Jarvis, S. (2007). Vocd: A theoretical and empirical evaluation. *Language Testing*, *24*, 459–488. https://doi.org/10.1177/0265532207080767

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, *42*, 381–392. https://doi.org/10.3758/BRM.42.2.381

Min, M., Lee, J. J., & Lee, K. (2022). Detecting illegal online gambling (IOG) services in the mobile environment. *Security and Communication Networks*, Article ID: 3286623, 1–12. https://doi.org/10.1155/2022/3286623

Mizumoto, A. (2025). Automated analysis of common errors in L2 learner production: Prototype web application development. *Studies in Second Language Acquisition*, *47*, 867–884. https://doi.org/10.1017/S0272263125100934

Mueller, J., & Jiang, N. (2013). The acquisition of the Korean honorific affix (*u*)*si* by advanced L2 learners. *The Modern Language Journal*, *97*, 318–339. https://doi.org/10.1111/j.1540-4781.2013.12005.x

National Institute of Korean Language. (2020). *Korean Learners' Corpus*. https://kcorpus.korean.go.kr/

National Institute of Korean Language. (2025). *Korean Learners' Corpus*. https://kcorpus.korean.go.kr/

Polio, C., & Yoon, H. J. (2018). The reliability and validity of automated tools for examining variation in syntactic complexity across genres. *International Journal of Applied Linguistics*, *28*, 165–188. https://doi.org/10.1111/ijal.12200

R Core Team. (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Robitzsch, A. (2020). Why ordinal variables can (almost) always be treated as continuous variables: Clarifying assumptions of robust continuous and ordinal factor analysis estimation methods. *Frontiers in Education*, *5*, 589965. https://doi.org/10.3389/feduc.2020.589965

Seo, S. J. (2009). *Study on the interlanguage development of Korean language learners by syntactic proficiency assessment: Based on analyzing syntactic features of writings* [Master's thesis, Yonsei University]. Yonsei University Library. https://library.yonsei.ac.kr/search/detail/CATTOT000000732514

Sohn, H. M. (1999). *The Korean language*. Cambridge University Press.

Soper, D. S. (2025). *A priori sample size calculator for hierarchical multiple regression [Software]*. https://www.danielsoper.com/statcalc

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239–273). John Benjamins Publishing Company.

Treffers-Daller, J., Parslow, P., & Williams, S. (2018). Back to basics: How measures of lexical diversity can help discriminate between CEFR levels. *Applied Linguistics*, *39*, 302–327. https://doi.org/10.1093/applin/amw009

Vercellotti, M. L. (2017). The development of complexity, accuracy, and fluency in second language performance: A longitudinal study. *Applied Linguistics*, *38*, 90–111. https://doi.org/10.1093/applin/amv002

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:08:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126102009

Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency, accuracy and complexity*. University of Hawaii Press.

Yadollahi, S., Moloodi, A., & Rahimi, M. (2024). Analyzing the development of a longitudinal error-tagged English learner corpus: Exploring accuracy development in Iranian EFL learners' writing. *Journal of Modern Research in English Language Studies*, *12*, 147–171.

Yoon, S., Park, S., Kim, G., Cho, J., Park, K., Kim, G., Seo, M., & Oh, A. (2023). Towards standardizing Korean grammatical error correction: Datasets and annotation. In A. Rogers, J. Boyd-Graber, N. Okazaki (Eds.), Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 6713–6742). Association for Computational Linguistics.

Zenker, F., & Kyle, K. (2021). Investigating minimum text lengths for lexical diversity indices. *Assessing Writing*, *47*, Article 100505. https://doi.org/10.1016/j.asw.2020.100505

Cite this article: Hwang, H. (2026). Automated measurement of complexity, accuracy, and fluency indices for Korean language data: Development and preliminary validation of K-text, a web-based natural language processing tool. *Studies in Second Language Acquisition,* 1–25. https://doi.org/10.1017/S0272263126102009

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:08:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126102009