Considerations and Challenges in Longitudinal Studies of Lexical Features in L2 Writing

Minkyung Kim

Nagoya University of Commerce and Business

Abstract

Exploring the longitudinal development of second language (L2) lexical use has been one of the important topics in L2 vocabulary research. One approach to examining longitudinal changes in L2 lexical use is to capture changes in lexical features as found in learner production, such as L2 writing, over time. To further facilitate this approach, the purpose of this article is to discuss considerations and challenges for conducting longitudinal studies on lexical features in L2 writing. The article first provides a summary of relevant previous studies, followed by the promise of longitudinal studies on lexical features in L2 writing. It then presents considerations and challenges in longitudinal studies of lexical features in L2 writing in terms of the data collection and analysis and the choice of lexical measures. More research on the longitudinal changes in lexical features in L2 learner production seems warranted. Ultimately, more longitudinal research in lexical features in L2 learner production will help us have a deeper understanding of L2 lexical development and design better vocabulary intervention in L2 classrooms.

Keywords: lexical features; lexical richness; longitudinal study; L2 writing

1 Introduction

In second language (i.e., language other than mother tongues; L2) vocabulary research, one of the important topics is to examine longitudinal developmental patterns of L2 vocabulary (lexical) use. One approach to exploring it is to track changes in lexical features as found in learner production, such as L2 writing, over time under the assumption that lexical use can be indicative of the underlying L2 vocabulary ability. An increasing body of research has examined developmental patterns of L2 lexical use in L2 writing (e.g., Gené-Gil, Juan-Garau, & Salazar-Noguera, 2015; Knoch et al., 2015; Mazgutova & Kormos, 2015; Polio & Shea, 2014; Verspoor et al., 2008; Zheng, 2016). The purpose of this article is to discuss considerations and challenges for conducting longitudinal studies of lexical features in L2 writing. In the following section, I first present previous studies that examined lexical features in L2 writing.

2 Previous studies on lexical features in L2 writing

While there are various conceptualisations of lexical features, this article focuses on Read's (2000) conceptualisation of lexical richness that includes

lexical density (i.e., proportion of content words [i.e., verbs, nouns, adjectives, and adverbs] as compared to function words [e.g., articles and prepositions]), lexical diversity (i.e., use of different types of words), lexical sophistication (i.e., use of advanced and difficult words), and lexical accuracy (i.e., errors in the use of words). Much cross-sectional research has investigated the relationships between lexical richness and L2 writing quality. While lexical density has shown little relation with L2 writing quality (Engber, 1995), lexical diversity, sophistication, and accuracy have shown close relations with L2 writing quality, such that higher-rated L2 texts tend to include more diverse words (e.g., Crossley & McNamara, 2012), more sophisticated words (e.g., Kim et al., 2018; Laufer & Nation, 1995), and fewer lexical errors (Crossley et al., 2019).

On the other hand, the longitudinal changes in lexical features in L2 writing remain more open. With respect to lexical density, a few L2 writing studies examined longitudinal changes (Zheng, 2016). Zheng examined lexical density over a year in Chinese upper-intermediate-level learners' English argumentative writing and reported no significant change in lexical density. Similarly, a few L2 writing studies explored longitudinal changes in lexical accuracy (Llach, 2011; Polio & Shea, 2014). Polio and Shea (2014) examined lexical errors produced in English as a second language (ESL) university-aged students' descriptive writing over a semester and reported no significant change in lexical errors. Llach (2011) explored lexical errors in the English letter writing of fourth grade Spanish EFL learners over 2 years and reported that the number of lexical errors, particularly for form-related errors, decreased.

On the other hand, changes in lexical diversity and sophistication over time have been increasingly examined in the L2 writing literature. Previous studies have found that lexical diversity in L2 writing tends to increase when investigation durations were more than 1 year (e.g., Gené-Gil et al., 2015; Verspoor et al., 2008; Zheng, 2016). For example, Gené-Gil et al. (2015) examined the English narrative writing of Spanish-Catalan secondary EFL learners over 3 years and reported an increasing pattern of lexical diversity over time. In a similar vein, developing patterns in lexical sophistication over time have been reported, such that L2 writers tend to use more advanced words over time (e.g., Mazgutova & Kormos, 2015; Storch & Tapper, 2009; Zheng, 2016). For example, Mazgutova and Kormos (2015) found that after 1-month-long intensive English course, undergraduate ESL students' argumentative essays included more advanced content words (i.e., less frequently used content words). However, some other L2 writing studies have reported no changes in lexical diversity or sophistication over time when university-aged L2 learners' writing was analysed for a shorter span of time (e.g., one semester; Bulté & Housen, 2014; Yoon & Polio, 2017) and when university-aged L2 learners did not participate in language-focused programmes during data collection periods (e.g., Knoch et al., 2015). These findings indicate that to observe L2 writers' developmental patterns in lexical diversity and sophistication, researchers may need to investigate L2 writing for a longer span of time (at least 1 year) or involve L2-focused instruction.

In sum, while previous longitudinal L2 writing studies on lexical features have provided important information about the developmental patterns of L2

lexical use, it still seems that we do not know much of L2 lexical development, and more research is needed.

3 The promise of longitudinal studies on lexical features in L2 writing

Considering that more longitudinal studies are needed, to help researchers think about their potential research on longitudinal changes in lexical features, in this section, I present four main research domains.

First, more research can describe longitudinal changes in lexical features over time. While previous studies have examined longitudinal changes in lexical features, more research is needed to better understand developmental patterns of L2 lexical use. Researchers can focus on various levels of L2 proficiency, learners from various first language (L1) backgrounds, and learners from different school levels. In addition, when describing longitudinal changes, researchers should understand the importance of examining longitudinal development at not only group levels through group statistics but also individual levels through individual case studies (Lowie & Verspoor, 2019).

Second, research can explore longitudinal interrelationships among linguistic features. Changes in lexical features do not occur in isolation. Instead, they likely take place in relation to changes in other linguistic features, such as syntactic and cohesive features. For example, when writers make greater use of more information-dense, advanced words in a clause, the length of that clause may become shorter not to hold too much information in it. Studies from a Dynamic Systems Theory perspective (Larsen-Freeman, 2006; Verspoor et al., 2008) have focused on developmental relationships of linguistic features, suggesting four types of longitudinal relationships among linguistic features: support (i.e., jointly developing), competition (i.e., developing in an opposite direction), precursor (growth of one linguistic feature precedes that of another), and asymmetry (i.e., relations between linguistic features varying over time; Verspoor & van Dijk, 2011). Such relationships have not yet examined thoroughly, and more research is needed to shed light on the complex nature of longitudinal interrelationships among linguistic features.

Third, research can also examine longitudinal interrelationships between lexical features and L2 writing quality. Over time, it is ideal for L2 learners to develop both their L2 vocabulary ability and their L2 writing ability. In addition, L2 writing quality (indicative of L2 writing ability) tends to be influenced by linguistic features, including lexical features (indicative of L2 vocabulary ability), found in L2 writing samples. Thus, linking lexical features and L2 writing quality over time may be promising to answer a question of what lexical features predict L2 writing quality longitudinally. If some lexical features consistently predict L2 writing quality over time, they can be the focus of vocabulary intervention in L2 writing classrooms.

Lastly, after identifying longitudinal developmental patterns in lexical features, the next important step that researchers can take is to explain what factors lead to such developmental patterns. Researchers can examine learner-related

factors, such as cognitive strategies and motivation, and contextual factors, such as exposure to L2s and instruction. Essentially, explaining longitudinal changes in lexical features should be an important goal to figure out how to design and implement intervention that considers learner and contextual factors.

4 Considerations and challenges in longitudinal studies on lexical features in L2 writing

To conduct longitudinal studies on lexical features in L2 writing, there are many basic questions about the research design, which also impose some challenges. These questions should be answered prior to collecting and analysing longitudinal data. The answers to these answers will help guide the research design and impact the degree to which the results of the study can generalise to other settings and populations. In this section, I present a set of questions along with considerations and challenges associated with key research design decisions that need to be made. Of course, each research design decision will be guided by the research questions of the study.

Who is the target learner population? The learners' age ranges and proficiency levels will depend on the research questions (e.g., adolescent or adult learners, beginning-level or advanced learners). The target learner population should be decided first to create writing tasks appropriate to that learner population.

How long is the research duration? The research duration will depend on the research questions of the study (e.g., whether interested in weekly, monthly, or yearly changing trends). If the duration is too short, it may be difficult to detect changes. On the other hand, longer research designs in duration may lead to higher attrition rates.

How frequently do learners write? The rates for how frequently learners produce writing samples will also depend on the research questions of the study. If researchers focus on learner variation over time, they can collect samples frequently (e.g., weekly), which may enable them to zoom in on various changing patterns including gradual changes, sudden spurs, and stagnation (Lowie & Verspoor, 2015). On the other hand, if researchers focus on learner development, they can collect samples less frequently (e.g., bi-monthly) but with a longer research design in duration because L2 development does not tend to take place in a short term. To track both learner variation and learner development, longer research durations with high frequency samples may be ideal but may not be always feasible.

What are writing tasks? It is important to consider the appropriateness of the writing tasks to the research questions. Choosing writing tasks must be meticulous because the choice of writing tasks (e.g., genre types, topics) influence learners' lexical use (Crossley, 2020). For longitudinal data collection, if researchers use the same writing prompt for each data collection occasion, it may result in the cofounding effect of the repeated practice on learner performance. That is, learners may perform better over time not because of their development, but because of their repeated practice. On the other hand, if researchers use different writing prompts across data collection occasions, they should make several key decisions.

First, the genre of the writing tasks must be consistent throughout the research duration because different genres (e.g., narratives vs. persuasion) will lead to different lexical use. Second, it is ideal to have at least two different prompts per data collection occasion to counterbalance and reduce the potential prompt effects. Lastly, it is important to be cautious when interpreting the results and comparing writing samples produced in response to different prompts (Read, 2000).

Is there instruction during the data collection period? The amount, intensity, and quality of L2 instruction during the data collection will influence the rates of changes in lexical features in L2 writing. Information about instruction needs to be collected, and instructional factors need to be considered when interpreting the results.

What is the basic unit of lexical analysis? For data analysis, basic units of lexical analysis can be tokens, lemmas (i.e., base forms of words, such as *confirm* for *confirmed* and *confirms*), and word families (i.e., word forms that share a common meaning, such as *confirm* for both inflected forms [e.g., *confirmed*] and derived forms [e.g., *confirmation*]). Generally, a widely used basic unit of lexical analysis is lemma, so that inflected forms and their base forms are counted as instances of the same lemma (Read, 2000).

Are function words included for the analysis? When measuring L2 vocabulary knowledge in receptive forms, in most cases, the focus of analysis is on content words. However, when examining lexical use in L2 writing, many previous studies included function words for analysis. Researchers need to decide whether to include function words for their analysis.

Are words related to prompts included for the analysis? An important consideration in analysing lexical features in written texts is whether to include words related to prompts. If learners use words that appear in the prompt in their texts, those words may not represent leaners' vocabulary ability, but rather reflect the prompt wording. Thus, while some previous studies did not control for prompt-based words, it may be advisable to remove content words that appear in the prompt prior to analysis.

5 Considerations and challenges in choosing measures of lexical features

In addition to the questions provided above, an important, but challenging, task is to choose measures for lexical features. Various measures have been used in assessing lexical features. In this section, I present brief explanations and considerations for the use of measures of the four main lexical features: lexical density, diversity, sophistication, and accuracy.

Lexical density is calculated by the proportion of content words to function words in the text. Content words tend to have specific semantic content, belonging to open classes of words (i.e., new members can be added), whereas FWs tend to have a non-conceptual meaning, belonging to closed classes of words (i.e., new members are rarely added; Corver & Van Riemsdijk 2001). Different approaches to defining function words have been suggested. For example, Halliday (1985) suggested that phrasal verbs which consist of one verb and one or more prepositions

(e.g., *pick up*) as one lexical item. However, generally, function words are considered as single lexical items.

Lexical diversity indicates how much the text includes unique words. It is traditionally related to a type-token ratio (TTR) which calculates the number of unique words divided by the total number of words in the text. However, TTR has a limitation that it is influenced by text length. Beyond TTR, various measures not influenced by text length have been suggested including the measure of textual lexical diversity (MTLD; the mean length of sequential word strings in a text that maintain a given TTR value, such as .720; McCarthy & Jarvis, 2010) and hyper-geometric distribution diversity (HD-D; the probability of encountering any of its type in a random sub-sample drawn from the text; McCarthy & Jarvis, 2010). When measuring lexical diversity, a caveat is that it is difficult to be convinced that lexical diversity purely assesses writers' vocabulary ability because there is a possibility that lower levels of lexical diversity may be due to writers' purposeful repetition of the same words to build cohesion by lexical repetition.

Lexical sophistication is considered as the use of advanced words in a text (Read, 2000). Advanced words were originally based on word frequency (i.e., how frequently a word is used) with less frequently used words being more advanced. Recently, using various computational indices, beyond word frequency, the operationalisation of lexical sophistication has been expanded, such that advanced words include academic words, words whose meanings are more specific, words that are more abstract and less familiar, and words acquired at a later age (Kim et al., 2018). However, the measure that may best represent lexical sophistication has not been agreed upon and may differ depending on L2 proficiency and writing tasks. Thus, researchers should carefully select lexical sophistication features for their analysis. In addition, given the richness of lexical sophistication measures and to avoid redundancy, if researchers use various lexical sophistication measures, they should consider construct distinctiveness, so that each measure targets different sub-constructs of lexical sophistication.

To measure lexical accuracy, various classification schemes of lexical errors have been suggested (e.g., Engber, 1995; Llach, 2011). For example, Engber (1995) distinguished errors resulting from lexical choice (i.e., meaning-related lexical errors) from those from lexical forms. One of the challenges in examining lexical accuracy lies in the subjective nature of lexical error detection, and thus, it is difficult to establish inter-rater reliability. Indeed, when examining various types of errors (e.g., preposition errors, verb phrase errors) in L2 writing, the inter-rater reliability of lexical errors was the lowest (0.54; Polio & Shea, 2014). In addition, beyond the number of lexical errors, types of lexical errors also need to be considered because while the number of lexical errors may not change over time, the specific types of lexical errors may change as L2 proficiency increases. For example, lower proficiency learners tend to produce L1-influenced errors, while higher proficiency learners tend to produce more target-language-oriented errors (Llach, 2011).

Lexical features presented above can be coded and calculated manually, but it will be labour-intensive and time-consuming. With the advent of natural language processing (NLP) tools, automated indices of lexical features can be

calculated quickly, flexibly, and reliably. I present some freely accessible automated tools for lexical features. The Coh-Metrix calculates lexical diversity and sophistication indices (Graesser et al., 2004). The Lexical Complexity Analyzer measures lexical density, variation, and sophistication (Lu, 2012). The Tool for the Automatic Analysis of Lexical Diversity (TAALED; Kyle et al., 2021) and the Tool for the Automatic Analysis of Lexical Sophistication (TAALES, Kyle & Crossley, 2015) provide measures of lexical diversity and lexical sophistication, respectively. The Grammar and Mechanics Error Tool (GAMET; Crossley et al., 2019) provide several measures related to lexical errors.

6 Conclusion

In this article, I provided a summary of previous studies on longitudinal studies on lexical features in L2 writing, and considerations and challenges associated with capturing longitudinal changes in lexical features in L2 writing. To disentangle the unpredictability and nonlinearity of L2 lexical development and to clearly distinguish *development* (progressing towards more target-like behaviour), which we aim at understanding, from mere *change* (becoming different over time), it seems clear that more research on longitudinal changes in lexical features in L2 learner production is warranted. Ultimately, more longitudinal research in lexical features found in L2 learner production will help us have a deeper understanding of L2 lexical development and implement better vocabulary intervention in L2 classrooms.

References

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing, 26*, 42–65. https://doi.org/10.1016/j.jslw.2014.09.005

Corver, N., & van Riemsdijk, H. (Eds.). (2001). *Semi-lexical categories.* Mouton de Gruyter. https://doi.org/10.1515/9783110874006

Crossley, S. A. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research, 11*(3), 415–443. https://doi.org/10.17239/jowr-2020.11.03.01

Crossley, S. A., Bradfield, F., & Bustamante, A. (2019). Using human judgments to examine the validity of automated grammar, syntax, and mechanical errors in writing. *Journal of Writing Research, 11*(2), 251–270. https://doi.org/10.17239/jowr-2019.11.02.01

Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading, 35*, 115–135. https://doi.org/10.1111/j.1467-9817.2010.01449.x

Engber, C. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing, 4*(2), 139–155. https://doi.org/10.1016/1060-3743(95)90004-7

Gené-Gil, M., Juan-Garau, M., & Salazar-Noguera, J. (2015). Development of EFL writing over three years in secondary education: CLIL and

non-CLIL settings. *The Language Learning Journal, 43*(3), 286–303. https://doi.org/10.1080/09571736.2015.1053278

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, and Computers, 36*, 193–202. https://doi.org/10.3758/BF03195564

Halliday, M. A. K. (1985). *Spoken and written language*. Deakin University.

Kim, M., Crossley, S. A., & Kyle, K. (2018). Lexical sophistication as a multidimensional phenomenon: Relations to second language lexical proficiency, development, and writing quality. *The Modern Language Journal, 102*(1), 120–141. https://doi.org/10.1111/modl.12447

Knoch, U., Rouhshad, A., Oon, S. P., & Storch, N. (2015). What happens to ESL students' writing after three years of study at an English medium university? *Journal of Second Language Writing, 28*, 39–52. https://doi.org/10.1016/j.jslw.2015.02.005

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly, 49*(4), 757–786. https://doi.org/10.1002/tesq.194

Kyle, K., Crossley, S. A., & Jarvis, S. (2021). Assessing the validity of lexical diversity using direct judgements. *Language Assessment Quarterly, 18*(2), 154–170. https://doi.org/10.1080/15434303.2020.1844205

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

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics, 16*(3), 307–322. https://doi.org/10.1093/applin/16.3.307

Llach, A. (2011). *Lexical errors and accuracy in foreign language writing*. Multilingual Matters.

Lowie, W., & Verspoor, M. (2015). Variability and variation in second language acquisition orders: A dynamic reevaluation. *Language Learning, 65*(1), 63088. https://doi.org/10.1111/lang.12093

Lowie, W. M., & Verspoor, M. H. (2019). Individual differences and the ergodicity problem. *Language Learning, 69*, 184–206. https://doi.org/10.1111/lang.12324

Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal, 96*(2), 190–208. https://doi.org/10.1111/j.1540-4781.2011.01232_1.x

Mazgutova, D., & Kormos, J. (2015). Syntactic and lexical development in an intensive English for Academic Purposes programme. *Journal of Second Language Writing, 29,* 3–15. https://doi.org/10.1016/j.jslw.2015.06.004

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*(2), 381–392. https://doi.org/10.3758/BRM.42.2.381

Polio, C., & Shea, M. C. (2014). An investigation into current measures of linguistic accuracy in second language writing research. *Journal of Second Language Writing, 26*, 10–27. https://doi.org/10.1016/j.jslw.2014.09.003

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

Storch, N., & Tapper, J. (2009). The impact of an EAP course on postgraduate writing. *Journal of English for Academic Purposes, 8*(3), 207–223. https://doi.org/10.1016/j.jeap.2009.03.001

Verspoor, M., Lowie, W., & van Dijk, M. (2008). Variability in second language development from a Dynamic Systems perspective. *The Modern Language Journal, 92*(2), 214–232. https://doi.org/10.1111/j.1540-4781.2008.00715.x

Verspoor, M., & van Dijk, M. (2011). Visualizing interactions between variables. In M. Verspoor, K. de Bot, & W. Lowie (Eds.), *A dynamic approach to second language development: methods and techniques* (pp. 85–98). John Benjamins Publishing Company.

Yoon, H.-J., & Polio, C. (2017). The linguistic development of students of English as a second language in two written genres. *TESOL Quarterly, 51*, 275–301. https://doi.org/10.1002/tesq.296

Zheng, Y. (2016). The complex, dynamic development of L2 lexical use: A longitudinal study on Chinese learners of English. *System, 56*, 40–53. https://doi.org/10.1016/j.system.2015.11.007