

L2 Vocabulary Knowledge Across Modes of Production: A Pilot Study

Gavin Brooks

Kwansei Gakuin University, School of Policy Studies

Jennifer Jordan

Kwansei Gakuin University, School of Policy Studies

Reference Data

Brooks, G., & Jordan, J. (2021). L2 vocabulary knowledge and lexical diversity across three modes of academic production: A pilot study. In P. Clements, R. Derrah, & P. Ferguson (Eds.), *Communities of teachers & learners*. JALT. <https://doi.org/10.37546/JALTPCP2020-22>

The pilot study described in this paper builds on existing studies of lexical diversity to determine the extent to which learners' L2 vocabulary proficiency affects the lexical diversity of the texts they produce across three modes of production. To do this, the researcher examined the relationship between first-year students' ($n = 17$) L2 vocabulary knowledge, as measured by Webb, Sasao, and Ballance's (2017) updated Vocabulary Levels Test (uVLT), and the lexical diversity of their academic writing, academic presentations, and academic discussions. The results showed that even learners with higher scores on the uVLT tended not to use those low-frequency words in their academic writing and presentations. As a result, there was no statistically significant correlation between students' uVLT scores and the lexical diversity of their essays or presentations. However, the analysis did show a statistically significant correlation between vocabulary knowledge and the lexical diversity of their academic discussions.

本稿で検証する予備研究は、語彙の多様性に関する既存の研究を基に、学習者のL2語彙力が三形態にわたるテキスト産出の多様性にどの程度の影響を及ぼすかを明らかにするものである。そのために、本稿ではウェブ、笹尾、バランス(2017)の語彙レベルテスト更新版(uVLT)で測定した1年生($n = 17$)のL2語彙知識と、学術的なライティング、プレゼンテーション、および討論における語彙の多様性との関係を調査した。その結果、uVLTのスコアが高い学習者であっても、低頻出語を学術的ライティングやプレゼンテーションで使用しない傾向が見られた。調査結果では、学生のuVLTスコアと、ライティングやプレゼンテーションにおける語彙の多様性との間には、統計的に有意な相関は認められなかったが、語彙知識と学術的な討論における語彙の多様性との間には、統計的に有意な相関が示された。

While research has long acknowledged the importance of vocabulary and multi-word expressions (MWE) for university-level students (Laufer & Nation, 1995; Ruegg et al., 2011; Treffers-Daller et al., 2016), what is lacking is the detailing of how the effects of vocabulary knowledge differ over different modes of production and across different stages of learner development. To detail these kinds of changes, researchers need access to detailed linguistic corpora that allow for tracking changes in individual learners' vocabulary over different modes of production. To date, most large-scale learner corpora have tended to focus on specific aspects of language and language use, such as comparisons between L1 and L2 usage, as opposed to differences in different types of texts produced by individual learners (Granger & Bestgen, 2014; Lee et al., 2019; Verspoor et al., 2012). Previous research projects have either focused on spoken (Garner & Crossley, 2018; Qi & Ding, 2011) or written (Granger & Bestgen, 2014; Siyanova-Chanturia & Spina, 2019; Yoon, 2016) learner language but have not explored the difference between two modes of production for the same learners. The research described in this article builds on such earlier works' strengths by using the techniques outlined in these articles to explore language usage over three different modes of academic production: written, written for spoken (academic presentations), and spoken (small group discussions). The project described in this paper is small-scale and intended to act as a pilot project for a larger study that will look at language development in a large group of L2 English learners ($n = 1,000$) over a two-year period.

Literature Review

Given the acknowledged differences between spoken and written vocabulary (Dang et al., 2017), the need to determine the differences in L2 English learners' written and spoken vocabulary usage is pressing. Researchers have previously shown that there is a relationship between the vocabulary profiles of learners and their language proficiency with both written (e.g., Daller & Phelan, 2013; Treffers-Daller et al., 2016) and spoken tasks (e.g., Clenton et al., 2020; de Jong et al., 2013). While these previous studies have

helped researchers to identify the specific types of vocabulary that learners need at different levels of proficiency for either written or spoken English, no single study has addressed the differences that exist between learners' written and spoken vocabulary knowledge (Jarvis, 2013a; Treffers-Daller et al., 2016). There is, therefore, a pressing need for studies that help researchers and educators to better understand the relationship between vocabulary knowledge and language proficiency across different modes of production. Without such knowledge, it will be difficult for researchers and teachers to provide learners with the specific language support they need in order to succeed at a specific type of task.

Lexical Diversity

One way of measuring learner proficiency is by looking at the lexical diversity of the texts that they produce. Lexical diversity is a measure of how many unique words are used in a text; in other words, it is the "proportion of words in a language sample that are not repetitions of words already encountered" (Jarvis, 2013b, p. 88). Lexical diversity is also a textual measure that has been increasingly discussed in the literature (Kyle et al., 2020; Zenker & Kyle, 2021) and shown to be an effective way to measure both the complexity and the quality of language use (Malvern & Richards, 2002; Treffers-Daller et al., 2016).

A wide variety of studies have looked at the relationship between lexical diversity and the written and spoken language proficiency of L2 learners (e.g., Crossley et al., 2010; Malvern & Richards, 2002). Engber (1995), for example, found a substantial correlation between the writing proficiency of 67 intermediate learners of English of different linguistic backgrounds and the lexical diversity of their academic writing. A more recent study of Turkish learners of German (Daller et al., 2003) found a significant correlation between learners' language proficiency and the lexical diversity of their oral production.

There is also a growing body of research that has shown a correlation between the lexical diversity of learner texts and the scores those texts were likely to receive from experienced raters (Gebriel & Plakans, 2014; Kyle & Crossley, 2014). For example, Yu (2009) found that the lexical diversity of the texts learners ($n=25$) produced for the written and spoken sections of the Michigan English Language Assessment Battery were positively correlated with the scores that they received on those texts. However, not all studies of lexical diversity have shown a correlation between L2 proficiency and the lexical diversity of the texts that those learners produce. In three different studies of the lexical diversity of Danish learners of English, Henriksen & Danelund (2015) found that

even proficient learners were reluctant to use low-frequency words in their academic composition, resulting in the compositions that they looked at in their study receiving low scores of lexical diversity across different levels of proficiency. In another study on the texts Chinese learners of English ($N=45$) produced for the Chinese National Matriculation English Test, Wang (2006) found no correlation between the lexical diversity of the texts and the proficiency of the learners. She also found no correlation between the lexical diversity of the texts and the scores that the participants received on those texts. The differences between the findings of these studies illustrate the need for further studies in this area of research.

Measuring Lexical Diversity

There are a number of different ways that researchers can measure lexical diversity. The traditional approach is to use the type-token ratio (TTR) (Malvern & Richards, 2002) which provides a ratio between the number of different words (type) and the total number of words (tokens) in a text. For instance, consider the following sentence "He studied English in Canada, so he gets good grades in his English classes." The sentence is 14 words long (14 tokens) but, because the words "he", "in", and "English" are repeated twice each, it only contains 11 different words (11 types). The TTR of this sentence would be 0.79. On the other hand, a sentence such as "I went for a walk.", where all the words in the text are unique, would have a TTR of 1. All texts can therefore be assigned a TTR between 0 and 1 and the closer the TTR of a text is to 1, the more lexically diverse that text is said to be.

A significant weakness with using TTR is that it is sensitive to variability in the length of the text; the longer a text is, the more likely it is to get a low TTR score (Zenker & Kyle, 2021). McCarthy & Jarvis (2010) attempted to address the issue of text length through an alternative measure of lexical diversity known as the Measure of Textual Lexical Diversity (MTLD). The MTLD is measured by first cutting the text into segments with a TTR of around 0.72. The MTLD can then be calculated by determining the average number of words in each of these segments (McCarthy & Jarvis, 2010). In the study described in this paper, the MTLD ranged from 35.52 for the least lexically diverse text to 73.05 for the most lexically diverse text. While there are still limitations with using MTLD, it allows researchers to better compare texts of different lengths and is generally thought to be one of the more accurate lexical diversity measures (Jarvis, 2013a).

Lexical Diversity in Learner Corpora

One area of research where lexical diversity is of particular importance is the field of corpus linguistics. Previous studies have been able to use corpora to help identify what vocabulary and expressions learners need to know in specific contexts, such as in oral presentations (Wood & Nambe, 2013) and academic writing (Nattinger & DeCarrico, 1994). However, these studies have primarily focused on L2 vocabulary usage in comparison to measures taken from texts produced by L1 speakers. This is problematic because it has been shown that L2 speaker usage of vocabulary and MWE differs significantly from L1 speakers' usage, even at the highest proficiency levels (Hou et al., 2016; Sinclair, 2004). This is because the native academic writer's or presenter's vast lexicon is nearly unattainable for most L2 learners. This is why learner corpora are so important; such corpora can help to establish the types of lexical knowledge required by learners of different proficiency levels (Treffers-Daller et al., 2016).

One problem with existing learner corpora, however, is the lack of corpora that span multiple modes of production. Despite the clear differences in the vocabulary needed for written and spoken texts (Dang et al., 2017), previous studies have tended to focus on written corpora, and there are few corpora that are composed of both spoken and written texts.

Research Questions

The present study has two research questions:

- RQ1. To what extent does vocabulary knowledge affect learners' ability to produce lexically diverse texts across the three modes of production (written, written for spoken, and spoken)?
- RQ2. Is there a relationship between the lexical diversity of learners' texts over these three different modes of production?

Methodology

Participants and Setting

The participants in this study were 17 first-year university students enrolled in a private university in Japan. All of the participants were required to take four academic English courses a week as part of their degree. These courses included a listening class, an academic presentation class, an academic writing class, and a seminar class where students were required to research and discuss academic topics in a small group setting. Each class met once a week for the 14-week semester. All of the written and spoken

content was produced as part of the students' classwork. Students were also asked to take the updated Vocabulary Levels Test during class time in week 12. All participants gave informed consent to participate in the study, and the project was cleared with the university's institutional review board prior to commencement.

Data Collection

The learner-produced texts used in this study were collected from the work that students did as part of their English for Academic Purposes (EAP) classes. During the semester students were required to write four academic essays ranging in length from three to five paragraphs long, give three academic presentations, and participate in four group discussions about an academic topic they had studied and discussed in class. For this study, the first of each of these assignments was analyzed. The essays were four paragraphs long on the topic, "The features of a successful store." The presentations were two to three minutes long on the topic, "A good place to visit." The seminar discussions were done in groups of three or four; groups had eight to ten minutes, depending on the size of the group, to discuss the topic "Success and failure at university." While the students were aware that the seminar discussions, presentations, and essays they produced for their classes could be used as part of the research project, they were not told which of the assignments would be analyzed for this project.

Because it was part of their regular course work, the participants were able to write their essays and the script for their presentations at home. While having the participants write these texts at home means that they could have accessed additional tools while writing, this was done for a number of reasons. Firstly, this approach to collecting student essays mirrors previous studies (e.g., Higginbotham & Reid, 2019), which have argued that this type of essay is closer to the type that students at a university level will be asked to write as part of their course than one done under exam conditions. As these essays were a large part of their course grade, it is our opinion that the participants approached this exercise as they would any other essay that they would have been asked to write for class. This means that the lexical diversity of these individual essays would be reflective of the lexical diversity of their academic English writing at large. Finally, one of the purposes of the larger study is to track if and how the use of lexical items is transferred between this type of deliberate written English to more spontaneous speech produced in classroom settings. The need to track this transfer means that it was important for the authors to analyze the types of written texts that students are likely to produce during their time in the program. However, we are aware of the differences between the conditions in which these different types of texts were produced and realize

that writing and study styles were probably much more influential in determining the types of vocabulary participants used in their written texts than in their discussions.

Webb et. al's (2017) Updated Vocabulary Levels Test (uVLT) was used to measure the students' vocabulary knowledge. The uVLT tests the learner's receptive knowledge of the first five 1000-word frequency bands. Each word band consists of 30 items with items presented to the test taker in a table format (see Figure 1). Test takers have to match a simple definition of the word (given in the rows of the table) with the correct word (given in the columns). Each table in the test consists of three definitions and six words, and each 1000-word band consists of 10 tables.

Figure 1
An Example of a Question on the uVLT With the Answers Given

| | eye | father | year | van | voice | night |
|-----------------------------|-----|--------|------|-----|-------|-------|
| body part that sees | ✓ | | | | | |
| parent who is a man | | ✓ | | | | |
| part of the day with no sun | | | | | | ✓ |

Note: Adapted from "The Updated Vocabulary Levels Test" by S. Webb, Y. Sasao, and O. Balance, 2017, *International Journal of Applied Linguistics*, 168(1), p. 61.

The uVLT was chosen over the Vocabulary Levels Test (Schmitt et al., 2001) because it was developed using frequency bands taken from a more modern corpus, the BNC/COCA (Nation, 2017). Despite being a test of receptive knowledge, uVLT was chosen because existing tests of productive vocabulary require more time than receptive tests to both administer and assess. Given that the main study could potentially include up to 1,000 participants over three years, it was not thought to be feasible to administer this type of assessment to such a large number of participants. Furthermore, previous studies have shown that while learners tend to do better on tests of receptive vocabulary, there is a clear relationship between productive and receptive vocabulary knowledge (e.g., Fan, 2000; Webb, 2008).

Transcribing and Processing the Data

The presentations and seminar discussions were recorded as part of the class. As the presentations were recorded in video format, the audio track was first separated from the

video and saved as a new file. Any identifying information was then removed from both sets of recordings using the application Fission (<https://rogueamoeba.com/fission/>). After this information had been removed, the recordings were sent to Rev (<https://rev.com>) to be transcribed. The subsequent transcriptions were then cleaned prior to analysis. First, the transcription was checked for accuracy and places where the person doing the transcription had been unable to transcribe the audio properly were corrected. Proper nouns, Japanese words, disfluencies, and multiple occurrences of a single word where students used this repetition as a place holding technique were then removed from the transcriptions. This was done to ensure that these words and the disfluencies, which are part of everyday conversational English, would not affect the participants' lexical diversity scores. The resulting files were saved as plain text documents. Each document was given a name that included a unique identifier for the student, two letters to describe the type of text ('ap' for academic presentations and 'ad' for academic discussions), followed by the assignment number (for this study, all documents were coded as 'assignment_1').

The essays were also cleaned prior to analysis. The title, the student's name, and their identification numbers were removed from each essay. Spelling mistakes were manually changed to ensure that these were not erroneously counted as an additional word type, potentially increasing the lexical diversity scores of the text. To do this, the following principles were applied: If it was clear what word the student had intended to write words with one missing letter, one wrong additional letter, or one incorrectly placed letter were corrected. For example, "peopple", "peple", and "pepole" would all have been corrected to "people." Grammatical mistakes were not corrected. The cleaned files were then saved as plain text documents with the same naming convention as the transcriptions with the letters 'aw' used to designate the text type as academic writing.

Data Analysis

After the transcription was completed, the text files were uploaded to TextInspector (<https://textinspector.com>), an online tool for analyzing textual diversity. The MTLTD was calculated for each of the files and that, along with the length of the text, was recorded into an Excel file. The length and MTLTD for each of the presentations were then entered into Excel.

Kendall's tau correlation was used to determine the relationship between the lexical diversity of the different types of learner produced texts and the students' uVLT scores. The small size of the data set meant that it was not possible to calculate the *p*-value

using the normal formulas, so bootstrapping (Field et al., 2012) was used. Bootstrapping gets around the problems caused by a small sample size by estimating the properties of the sampling distribution from the sample data (Bruce, 2015). The correlational coefficient was calculated using this method, and the standard deviation of the sampling distribution of the bootstrapped samples was used to estimate the standard error for the correlational coefficients (Wright et al., 2011). This was done in R (<https://www.R-project.org/>) by using the bootstrapping function to resample each of the data sets 2,000 times to determine the correlational coefficients and the *p* values of each data set.

Results

An analysis of the uVLT (see Table 1) showed that the students' vocabulary knowledge was consistent with intermediate Japanese L2 English learners (Brooks, 2020). The majority of the participants were able to master the first two bands of the uVLT, showing that they were proficient with the higher-frequency word bands. However, very few of the participants were able to master the lower-frequency bands, the 3000 to 5000 bands.

Table 1
Number of Students Who Achieved Mastery for Each Frequency Bands

| | 1000-word Band | 2000-word Band | 3000-word Band | 4000-word Band | 5000-word Band |
|-------------------------|----------------|----------------|----------------|----------------|----------------|
| Number (<i>n</i> = 17) | 17 | 17 | 5 | 2 | 1 |
| Percentage | 100% | 100% | 29% | 12% | 6% |

Note: Master is considered to be a score of 86% or higher on that frequency band.

The descriptive statistics of the three different corpora were then examined. The corpus derived from the written compositions was the largest and included 6,604 words with a mean length of 388 words for each of the compositions. The next largest corpus was the academic presentations, which was 3,419 words long, and the mean length of the transcripts of the presentations was 201 words. The average presentation time was 2:15 minutes. The corpus compiled from the seminar discussions was almost as long as the academic presentation corpus. This corpus was 3,369 words long with an average length of 198 words for each discussion transcript. Each discussion consisted of three students talking for the 10-minute discussion.

The vocabulary profiles of the three different modes of production were very similar (see Table 2) with most of the words for each of the texts coming from the first two 2000-word bands of the BNC/COCA (Nation, 2017). Of the three different text types, academic presentations had the highest percentage of low-frequency words.

Table 2
Vocabulary Profiles of the Different Texts

| Text Type | 1000-word Band | 2000-word Band | 1000- + 2000-word Bands | 3000-word Band | 4000-word Band | 5000-word Band | Off List |
|---------------|----------------|----------------|-------------------------|----------------|----------------|----------------|----------|
| Written | 79.5% | 15.7% | 95.2% | 3.4% | 0.3% | 0.5% | 0.6% |
| Presentations | 89.7% | 0.1% | 89.8% | 0.1% | 0.4% | 9.6% | 0.2% |
| Discussions | 89.4% | 5.7% | 95.1% | 0.4% | 0.1% | 3.8% | 0.6% |

An analysis of the correlation between the different text types and the students' vocabulary knowledge (see Table 3) showed that the academic discussions were the only text type where there was a significant correlation between lexical diversity and MTLT ($\tau = .37, p = .039$).

Table 3
The Correlation Between uVLT and the MTLT of the Different Text Types

| | Written | Presentation | Discussion |
|------------------|---------|--------------|------------|
| Kendall's Tau | 0.18 | -0.07 | 0.37* |
| <i>p</i> -values | 0.406 | 0.512 | 0.039 |

Note: * = correlation significant at $p < .05$; ** correlation significant at $p < .01$; *** correlation significant at $p < .001$

An analysis of the MTLT of the different text types (see Table 4) showed that the only text types that were significantly correlated were the written and presentation texts ($\tau = .41, p = .014$).

Table 4
The Correlation Between the MTL D of the Different Text Types

| | Written | Presentation | Discussion |
|---------------|---------|--------------|------------|
| Written | - | | |
| Presentations | .41* | - | |
| Discussion | .21 | .17 | - |

Note: * = correlation significant at $p < .05$; ** correlation significant at $p < .01$; *** correlation significant at $p < .001$

Discussion

In response to the first research question, similar to early studies (e.g., Henriksen & Danelund, 2015; Wang, 2006) no correlational relationship was found between the participants' written (essays) and written for spoken (academic presentations) texts and their vocabulary knowledge. However, there was a statistically significant relationship between their academic discussions and their vocabulary knowledge. This also mirrors earlier studies (e.g., Yu, 2009) which have shown that there is a stronger relationship between vocabulary knowledge and lexical diversity in learners' spoken as compared with their written production.

In response to research question two, there was a weak but significant relationship between the participants' written (essays) and written for spoken (academic presentations) texts. This may have been due to the fact that the participants had additional resources available to them while writing their presentations and essays. The participants who put more time into writing their essays were also more likely to do the same when writing their presentations, meaning that these participants were probably more likely than the other participants to use a dictionary and take the time to use more difficult words when constructing these texts. On the other hand, students were forced to rely on vocabulary that they already had knowledge of when doing their academic discussions. As a result, the participants' vocabulary knowledge, as opposed to their study style, was probably more likely to have an effect on the lexical diversity of their academic discussions.

Conclusion

The results of this study are promising as they show that there may indeed be differences in how learners use vocabulary depending on the mode of production. It also

showed that, despite the fact that both academic presentations and academic discussions are given as spoken texts, academic presentations have more in common with academic writing than they do with discussions. In this study, even students with lower levels of vocabulary proficiency were able to use additional resources to improve the diversity of the vocabulary in their essays and presentations. As such, teachers in presentation classes may be wise to focus more on teaching students how to choose and use appropriate vocabulary, than in teaching them a lot of new words.

One weakness with this study is that it was a small-scale study with only 17 participants who produced only one sample of writing or speaking for each of the text types. Consequently, the sample size was probably not large enough to allow a weak relationship between textual lexical diversity and vocabulary knowledge to show up as being significant. However, as this study was intended primarily as a pilot study, it is hoped that some of the issues related to the small sample size will be answered in the larger study. While small in size, this research project did demonstrate that the vocabulary used by L2 learners for different text types is different and, in doing so, highlighted the need for more multimodal corpora. Because most vocabulary lists in use at the moment were developed using written texts produced by L1 English speakers, the development of corpora like the ones described in this paper will allow researchers to develop tools designed specifically to support L2 learners of English based on the type of texts they are being asked to produce. These tools are important because they can be used to support learners in the classroom and are something that the researchers involved in this project hope they can begin to produce in the future.

Bio Data

Gavin Brooks received his Master of Applied Linguistics from the University of New England with a focus on the stages of language acquisition. He has taught English in Japan, Indonesia, Colombia, and Ecuador. For the last 15 years, he has been teaching in Japan and currently works as an Associate Lecturer of English at Kwansai Gakuin's School of Policy Studies. <gavinbrooks@gmail.com>

Jen Jordan has been teaching in Japan for 10 years; at KGU's School of Policy Studies for 8 of them. Before Japan, she taught for 11 years in South Korea, mostly in Daegu, but also in Daejeon. She is currently working on a PhD from Cardiff University in Wales. Her research interests include writing pedagogy, learner corpus linguistics, and language program management and development. <jordan.jen@gmail.com>

References

- Brooks, G. (2020). Vocabulary usage in second language presentations. *JALT Postconference Publication*, 2019(1), 288–295. <https://doi.org/10.37546/jaltpcp2019-34>
- Bruce, P. C. (2015). *Introductory statistics and analytics*. John Wiley & Sons.
- Clenton, J., de Jong, N. H., Clingwall, D., & Fraser, S. (2020). Investigating the extent to which vocabulary knowledge and skills can predict aspects of fluency for a small group of pre-intermediate Japanese L1 users of English (L2). In J. Clenton & P. Booth (Eds.), *Vocabulary and the four skills* (pp. 126–145). Routledge. <https://doi.org/10.4324/9780429285400-15>
- Crossley, S. A., Salsbury, T., McNamara, D. S., & Jarvis, S. (2010). Predicting lexical proficiency in language learner texts using computational indices. *Language Testing*, 28(4), 561–580. <https://doi.org/10.1177/0265532210378031>
- Daller, M. H., & Phelan, D. (2013). Predicting international student study success. *Applied Linguistics Review*, 4(1), 173–193. <https://doi.org/10.1515/applirev-2013-0008>
- Daller, M. H., van Hout, R., & Treffers-Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics*, 24(2), 197–222. <https://doi.org/10.1093/applin/24.2.197>
- Dang, T. N. Y., Coxhead, A., & Webb, S. (2017). The academic spoken word list. *Language Learning*, 67(4), 959–997. <https://doi.org/10.1111/lang.12253>
- de Jong, N. H., Groenhout, R., Schoonen, R., & Hulstijn, J. H. (2013). Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics*, 36(2), 223–243. <https://doi.org/10.1017/s0142716413000210>
- Engber, C. A. (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](https://doi.org/10.1016/1060-3743(95)90004-7)
- Fan, M. (2000). How big is the gap and how to narrow it? An investigation into the active and passive vocabulary knowledge of L2 learners. *RELC Journal*, 31(2), 105–119. <https://doi.org/10.1177/003368820003100205>
- Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. SAGE Publications.
- Garner, J., & Crossley, S. (2018). A latent curve model approach to studying L2 N-gram development. *The Modern Language Journal*, 102(3), 494–511. <https://doi.org/10.1111/modl.12494>
- Gebril, A., & Plakans, L. (2014). Assembling validity evidence for assessing academic writing: Rater reactions to integrated tasks. *Assessing Writing*, 21, 56–73. <https://doi.org/10.1016/j.asw.2014.03.002>
- Granger, S., & Bestgen, Y. (2014). The use of collocations by intermediate vs. advanced non-native writers: A bigram-based study. *International Review of Applied Linguistics in Language Teaching*, 52(3), 229–252. <https://doi.org/10.1515/iral-2014-0011>
- Henriksen, B., & Danelund, L. (2015). Studies of Danish L2 learners' vocabulary knowledge and the lexical richness of their written production in English. In P. Pietilä, K. Doró, & P. Renata (Eds.), *Lexical issues in L2 writing* (pp. 29–55). Cambridge Scholars Publishing.
- Higginbotham, G., & Reid, J. (2019). The lexical sophistication of second language learners' academic essays. *Journal of English for Academic Purposes*, 37, 127–140. <https://doi.org/10.1016/j.jeap.2018.12.002>
- Hou, J., Loerts, H., & Verspoor, M. H. (2016). Chunk use and development in advanced Chinese L2 learners of English. *Language Teaching Research*, 22(2), 148–168. <https://doi.org/10.1177/1362168816662290>
- Jarvis, S. (2013a). Defining and measuring lexical diversity. In S. Jarvis & M. Daller (Eds.), *Vocabulary Knowledge: Human ratings and automated measures* (pp. 13–44). John Benjamins Publishing. <https://doi.org/10.1075/sibil.47.03ch1>
- Jarvis, S. (2013b). Capturing the diversity in lexical diversity. *Language Learning*, 63, 87–106. <https://doi.org/10.1111/j.1467-9922.2012.00739.x>
- Kyle, K., & Crossley, S. A. (2014). 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. (2020). Assessing the validity of lexical diversity indices using direct judgements. *Language Assessment Quarterly*, 1–17. <https://doi.org/10.1080/15434303.2020.1844205>
- 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>
- Lee, J. J., Bychkovska, T., & Maxwell, J. D. (2019). Breaking the rules? A corpus-based comparison of informal features in L1 and L2 undergraduate student writing. *System*, 80, 143–153. <https://doi.org/10.1016/j.system.2018.11.010>
- Malvern, D., & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, 19(1), 85–104. <https://doi.org/10.1191/0265532202lt2210a>
- McCarthy, P. M., & Jarvis, S. (2010). MTL, 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>
- Nation, P. (2017). *The BNC/COCA Level 6 word family lists (Version 1.0.0)* [Data set]. <http://www.victoria.ac.nz/lals/staff/paul-nation.aspx>
- Nattinger, J. R., & DeCarrico, J. S. (1994). *Lexical phrases and language teaching*. Oxford University Press.
- Qi, Y., & Ding, Y. (2011). Use of formulaic sequences in monologues of Chinese EFL learners. *System*, 39(2), 164–174. <https://doi.org/10.1016/j.system.2011.02.003>

Brooks & Jordan: *L2 Vocabulary Knowledge Across Modes of Production: A Pilot Study*

- Ruegg, R., Fritz, E., & Holland, J. (2011). Rater sensitivity to qualities of lexis in writing. *TESOL Quarterly*, 45(1), 63–80. <https://doi.org/10.5054/tq.2011.240860>
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18(1), 55–88. <https://doi.org/10.1191/026553201668475857>
- Sinclair, J. (2004). New evidence, new priorities, new attitudes. In J. Sinclair (Ed.), *How to use corpora in language teaching: Studies in corpus linguistics* (pp. 271–299). John Benjamins Publishing Company. <https://doi.org/10.1075/scl.12.20sin>
- Siyanova-Chanturia, A., & Spina, S. (2019). Multi-word expressions in second language writing: A large-scale longitudinal learner corpus study. *Language Learning*, 70(2), 420–463. <https://doi.org/10.1111/lang.12383>
- Treffers-Daller, J., Parslow, P., & Williams, S. (2016). Back to basics: How measures of lexical diversity can help discriminate between CEFR levels. *Applied Linguistics*, 39(3), 302–327. <https://doi.org/10.1093/applin/amw009>
- Verspoor, M. H., de Bot, K., & Lowie, W. (2012). *A dynamic approach to second language development: Methods and techniques*. Wiley. <https://doi.org/10.1075/llt.29>
- Wang, X. (2006). The relationship between lexical diversity and EFL writing proficiency. *University of Sydney Papers*, 9, 65–88. https://faculty.edfac.usyd.edu.au/projects/usp_in_tesol/pdf/volume09/Article03.pdf
- Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition*, 30(1), 79–95. <https://doi.org/10.1017/s0272263108080042>
- Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test. *International Journal of Applied Linguistics*, 168(1), 33–69. <https://doi.org/10.1075/itl.168.1.02web>
- Wood, K., & Nambe, D. (2013). Focused instruction of formulaic language: Use and awareness in a Japanese university class. *Proceedings of the Third Asian Conference on Language Learning*, 3, 213–212. http://25qt511nswfi49iayd31ch80-wpengine.netdna-ssl.com/wp-content/uploads/papers/acll2013/ACLL2013_0113.pdf
- Wright, D. B., London, K., & Field, A. P. (2011). Using bootstrap estimation and the plug-in principle for clinical psychology data. *Journal of Experimental Psychopathology*, 2(2), 252–270. <https://doi.org/10.5127/jep.013611>
- Yoon, H.-J. (2016). Association strength of verb-noun combinations in experienced NS and less experienced NNS writing: Longitudinal and cross-sectional findings. *Journal of Second Language Writing*, 34, 42–57. <https://doi.org/10.1016/j.jslw.2016.11.001>
- Yu, G. (2009). Lexical diversity in writing and speaking task performances. *Applied Linguistics*, 31(2), 236–259. <https://doi.org/10.1093/applin/amp024>
- Zenker, F., & Kyle, K. (2021). Investigating minimum text lengths for lexical diversity indices. *Assessing Writing*, 47. <https://doi.org/10.1016/j.asw.2020.100505>