Macrothink Institute™

International Research in Education
ISSN 2327-5499
2025, Vol. 13, No. 1

Exploring the Development of Lexical Bundles among L1 Arabic Learners of English

Hend Naif H. Almutairi

PhD Candidate, University of Wollongong, Australia

Faculty of the Arts, Social Sciences and Humanities

School of Humanities and Social Inquiry, Wollongong, Australia

ORCID iD: 0009-0007-4262-817X

Hind Shujaa Almutairi

School of Language and Linguistics, University of Melbourne, Australia

Parkville VIC 3052, Australia

Abdulaziz M Alodhialah

King Saud University, Saudi Arabia, Assistant Professor

College of Nursing, Department of Medical-Surgical Nursing

P.O. Box 22480, Riyadh 11495 Kingdom of Saudi Arabia

Received: January 1, 2025 Accepted: March 24, 2025 Published: March 28, 2025

doi: 10.5296/ire.v13i1.22756 URL: https://doi.org/10.5296/ire.v13i1.22756

Abstract

This study aims to explore the developmental aspects of lexical bundles in the learner data of L1 Arabic learners of English with different levels of English proficiency. It examines the structural distribution of lexical bundles to explain the connection between bundle use and language competence. The study employs a corpus-based methodology using data from two Arabic learners, Amina and Hassan, collected via an audio recording prompt. It analyzes the top 500 four-word lexical bundles from each learner's speech using LancsBox and a modified structural taxonomy adapted from Biber et al. (1999). The results show Amina, the more proficient learner, used more clause fragments and elaborated nouns both quantitatively and qualitatively. Hassan was less inclined to use objects with transitive verbs, indicating L1 interference. While Hassan used some structures more frequently, Amina demonstrated a

more competent use of structures qualitatively. The study concludes more competent L2 speakers make more sophisticated use of lexical bundles in terms of structure and function. It provides preliminary evidence of associations between bundle use and language proficiency, though more data is needed. The findings imply lexical bundle development may be used to track L2 proficiency growth longitudinally. A better understanding of bundle acquisition can inform teaching practices to help learners incorporate bundles into their developing interlanguage systems. More research is warranted on bundle development across various learner populations and proficiency levels.

Keywords: Arabic Learners, lexical bundles, L1, L2, English

1. Introduction

1.1 Problem Statement

Lexical bundles (LBs) are formulaic sequences of idiomatic, conventionalized expressions or phrases, also known as multiword lexical units (Cowie, 1992), phraseological units (Barros & Castro, 2017), formulaic sequences (Puimège & Peters, 2020), and formulaic expressions (Wang, 2019). For this study, lexical bundles refer to frequently occurring lexical combinations. Despite their recognized role in language development, literature on lexical units has largely overlooked their developmental role, particularly in second language (L2) learning (Shin, 2019). While computer-based methods of textual analysis have recently advanced research in this area, these studies often focus on cross-linguistic and cross-generic orientations rather than developmental aspects (e.g., Güngör & Uysal, 2020; Shin, 2019; Vo, 2019).

This study addresses this gap by exploring the developmental aspects of LBs in L2 learners, focusing on Arabic-speaking learners of English with varying proficiency levels. Using a corpus-based tool (Lancsbox), this research examines n-gram units of LBs to investigate their structural distribution and relationship to language competence. The findings aim to shed light on how interlanguage interference influences the acquisition of L2 LBs and enhance data analysis reliability.

1.2 Importance of the Problem

Understanding the role of LBs is crucial for advancing theories of lexical acquisition and improving language teaching methodologies. Although Meara (1997) highlights the absence of a comprehensive theory of lexical acquisition, this study seeks to contribute to filling this theoretical gap. By exploring how LBs develop and differ qualitatively and quantitatively across proficiency levels, the research offers practical insights for educators and theoretical implications for L2 acquisition models.

1.3 Relevant Scholarship

Early studies by Meara and associates (e.g., Meara, 1978, 1992; Wilks & Meara, 2002) emphasize the associative nature of the mental lexicon, supporting the connectionist

perspective that explains interlanguage as a behavioral outcome of linguistic regularities processed at lexical and syntactic levels (Waltz & Pollack, 1985). Ellis (1996) extends this by demonstrating the importance of sequencing in L2 acquisition, claiming that knowledge of word sequences aids in grammar abstraction and language production.

Longitudinal studies like Reppen (2009) and Crossley and Salsbury (2011) show that L2 exposure enhances the structural and functional use of LBs. However, contradictory findings (e.g., Huang, 2015) caution against equating frequency with accuracy. Recent corpus-driven studies (e.g., Yan, 2019; Chen, 2019) further highlight the nuanced relationship between proficiency and LB use, underscoring the need for more focused research on L1 influences and developmental patterns.

1.4 Hypotheses and Their Correspondence to Research Design

This study investigates the developmental patterns of LB acquisition among L1 Arabic learners of English, focusing on the following research questions:

1. What is the developmental pattern for the acquisition of lexical bundles among L1 Arabic learners of English?

2. How is language competence related to the structural distribution of lexical bundles among L1 Arabic learners of English?

3. How does interlanguage interference explain the developmental patterns for acquiring L2 lexical bundles among L1 Arabic learners of English?

To address these questions, the research employs a corpus-based tool (Lancsbox) for analyzing n-gram units. The study prioritizes both quantitative (frequency) and qualitative (structural and functional) aspects of LB usage, ensuring a robust design that aligns with its theoretical framework and objectives.

2. Method

A number of research studies in lexical bundles employed corpus-based methods to identify the frequency of occurrences of the formulaic combinations (see Gries, 2016). Granger (2002) believes that the use of corpus-based tools has two advantages: first, it makes learner corpus more manageable, and secondly, the raw learner corpus can be automatically annotated for machine-based analysis. The current study is also a corpus-based study that employs qualitative and quantitative techniques to identify and analyse the targeted combinations of lexical bundles. The current study employs a structural taxonomy of lexical bundles adapted from Biber et al. (1999) to quantitatively investigate the use of different structural types of lexical bundles in the spoken narrative data of two English learners with Arabic as their L1.

2.1 Research Framework: Structural Taxonomy of Lexical Bundles

Biber et al.'s (1999) taxonomy for the structural configuration of the lexical bundles comprises three major categories of Noun-, Verb-, and Preposition-based bundles. The Noun- and

Verb-based lexical bundles are further divided into sub-categories. For the purpose of this study, certain changes were made to this original taxonomy proposed by Biber et al. (1999) to accommodate the learners' errors or incomplete structures. For example, the label of the sub-category of Noun-based+post-modifiers was changed to Other Noun-based LBs (see Figure 1). The reason for this change was to broaden the scope of this sub-category to include almost anything that carried noun-based lexical bundles, even if their post-modifiers are not used properly in the learner data. Secondly, Biber et. al.'s (1999) model did not include Clause-based bundles, but this category was added to take into account those 4-word lexical bundles that are either a complete clause fragment or an obj-less clause fragment. This category was added to Biber et al.'s (1999) taxonomy after the learner data revealed many complete and obj-less clause structures. It was believed that the frequency occurrences of the clause structures could provide the researcher with insights into the language development of the respondents. The modified model for the lexical bundles is presented below in Figure 1.

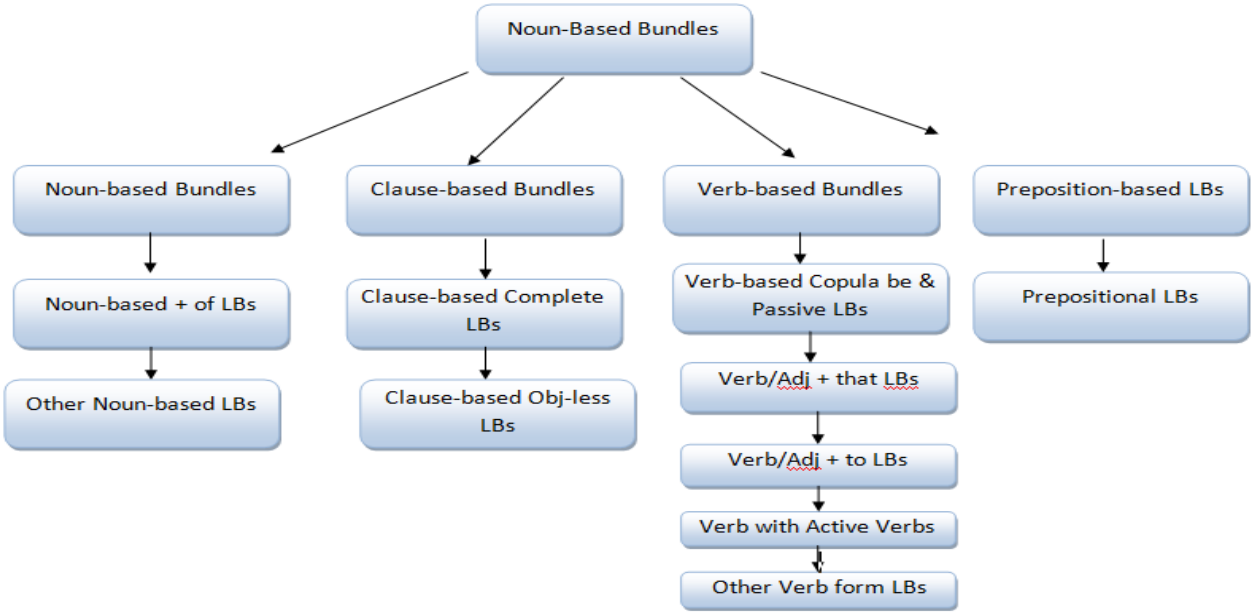


Figure 1. Adapted Structural Taxonomy for Lexical Bundles (Adapted from Biber et al., 1999)

2.2 Settings and Population

The learner corpus was collected from two Arabic natives, Hassan and Amina (pseudonyms), both living in Australia. Hassan and Amina are both Saudis by birth and have been living in Australia for two years and six months, respectively. Hassan is a 5-and-a-half-year-old Kindergarten boy and Amina is a 7-year-old girl. Hassan is a relatively less competent speaker of English than Amina, who has an intermediate level of competence in English. Both the target subjects attend school in Australia where they gain ample exposure to the target language. The relative difference in their English language competence could be explained by

the time they have been going to school. At home, they live in a multilingual setting where the parents quite often use Arabic with them but they prefer to watch English cartoon TV channels.

2.3 Tools used for Speech Production and Analysis

The following tools were used for data collection and analysis.

a) Smart Tab and Smartphone

The narrative speech data were collected by providing a speech prompt to the students in the form of a YouTube video that required the use of a smart tab to play the video. The speech data that were produced in response to the video was collected by using an audio recording application on the researcher's smartphone.

b) Story: Data Collection Prompt

As a prompt, the respondents were required to watch the story "Happy Prince" on the YouTube channel, T-Series Kids Hut. A relatively long video of 12:40 mins was used to collect data, as it was considered long enough for the respondents to speak for 4-5 minutes for data elicitation.

c) LancsBox Introduction

The LancsBox tool, developed by Lancaster University, was used to analyse the speech data for the presence of lexical bundles (http://corpora.lancs.ac.uk/lancsbox/). There are multiple text-analyses tasks that can be accomplished using LancsBox. It was used to a) measure the size of the corpus produced by Hassan and Amina, and b) identify the 4-word lexical bundles (4-grams) from the collected corpus. For some cases, the Key Word In Context (KWIC) tab was also used to look for the identified lexical bundles in the context to identify the correct grammatical category of the lexical bundles.

2.4 Data Collection

Data were collected from the respondents at their home. Their mother and the researcher's cousin were present during the recording to facilitate the data collection process. The respondents were totally relaxed and watched the video with great interest. The respondents could watch the story only once to minimise the priming effect of the video on their speech.

2.5 Data Analysis Procedure

After collecting the data, the collected corpus was transcribed verbatim into two Microsoft (MS) Word Document files, one for Amina and the other for Hassan (see Appendices A & B). These files were later converted into text files, as LancsBox can read and process only text files.

After uploading the data files into LancsBox, it was found that Amina's speech corpus had 158 types (number of different words) and 625 tokens (total words in the corpus), while Hassan had 179 types and 713 tokens in his speech corpus. Moreover, there were 546 *n*-gram types in Amina's data and 620 *n*-gram types in Hassan's data (only the top 500 were analysed for the study). Furthermore, by using LancsBox, lists for the 4-word lexical bundles were generated separately for the speech data produced by Amina and Hassan.

As *n*-grams refer to any type of lexical combinations with *n* numbers, for the 4-grams, LancsBox produced a long list of 4-word lexical bundles from the corpora. Many structures were meaningless combinations like "put aa immediately take" and "said ah a i". Learners with less competence in the target language are more likely to produce incomprehensible utterances in the learner corpus that become more meaningless when the corpus tool produces *n*-grams. Generally, these structures are not counted for the analysis and the same approach was also followed in this study. However, the approach in this study was slightly flexible in the inclusion of some grammatically incorrect or incomplete structures that provided an idea of the main grammatical phrase used in the bundle. This methodological decision was important to accommodate the limited size of the learner data that, as expected, had some wrong and less-comprehensible lexical units.

3. Results

There were 546 lexical bundles as shown in table 1 produced by LancsBox for Amina and 620 for Hassan. However, due to the limited scope of the study, the researcher analysed the top 500 lexical bundles for Amina and Hassan each (i.e., 1000 in total).

Table 1: Detailed List of Structural Types of Lexical Bundles

Lexical Bundle Category	Amina	Hassan
Noun-based Lexical Bundles	**Noun-based + of LBs:** 8 (e.g., *The end of the, End of the story*)	**Noun-based + of LBs:** 0
	Other Noun-based LBs: 25 (e.g., *Golden statue then he, The girl then the, The bird his eyes*)	**Other Noun-based LBs:** 14 (e.g., *The happy prince he, A story about a, Some oranges so he*)
	Total Noun-based LBs: 33	**Total Noun-based LBs:** 14
Clause-based Lexical Bundles	**Complete LBs:** 74 (e.g., *He was crying and, He talked to him, Then he waked up*)	**Complete LBs:** 26 (e.g., *Bird said no I, The mayor came and, She was very scared*)
	Transitive but Object-less LBs: 9 (e.g., *He take his eyes, Then he said I*)	**Transitive but Object-less LBs:** 6 (e.g., *The prince said oh, Then the prince said take*)
	Total Clause-based LBs: 83	**Total Clause-based LBs:** 32
Verb-based and Prepositional Lexical Bundles	**Copula Be & Passives:** 6 (e.g., *Is freezing then he, Was crying and and*)	**Copula Be & Passives:** 4 (e.g., *Am feeling much better, Was made out of (Passive)*)
	Verb/Adjective + That: 0	**Verb/Adjective + That:** 0
	Verb/Adjective + To: 3 (e.g., *He*	**Verb/Adjective + To:** 5 (e.g.,

Lexical Bundle Category	Amina	Hassan
	goed to kiss, Want to kiss you, She want to buy)	*Going to tell a, I have to go, To tell a story*)
	Other Verb Forms: 11 (e.g., *Talked to him then, Take his eyes then, See the prince and*)	**Other Verb Forms:** 13 (e.g., *Gave the feathers to, Sent to the writer, Go to any longer*)
	Prepositions: 8 (e.g., *Into the table then, About stone golden stone, Under his feet then*)	**Prepositions:** 12 (e.g., *In the bin then, To the writer he, Out of a blue (proverbial)*)
	Total Verb-based & Prepositional LBs: 28	**Total Verb-based & Prepositional LBs:** 34

The frequency occurrences of the lexical bundles were measured in accordance with the adapted structural taxonomy of lexical bundles. The distribution across Amina's and Hassan's data for the major structural categories of the lexical bundles is demonstrated in Figure 2.

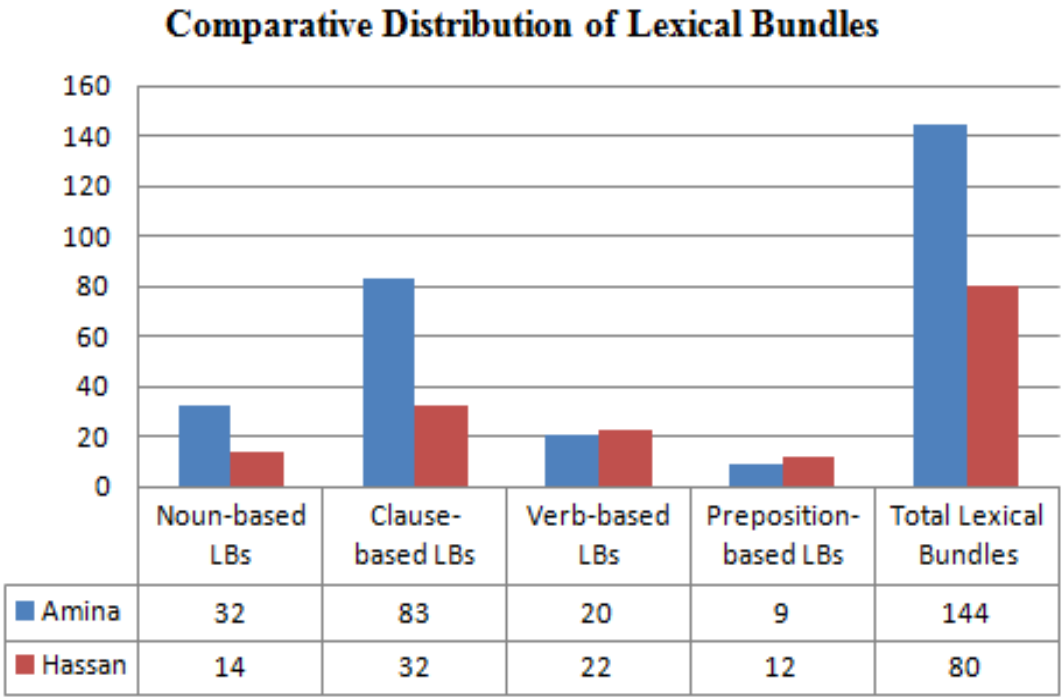


Figure 2. Comparative Distribution of Lexical Bundles

It can be observed from the above figure that, when compared with the other structural categories, the lexical bundle most commonly used by both the respondents is the Clause-based LB. However, the comparative results for the use of Noun-, Verb-, and Prepositional-based lexical bundles were mixed. It can be observed that Amina made significantly more use of Noun-based LBs compared to Hassan, and Hassan made marginally more use of Verb- and Prepositional-based lexical bundles compared to Amina. The total number of lexical bundles used by Amina (144) is significantly more than Hassan (80).

As Amina is more competent in English than Hassan, these data can also be interpreted in terms of the difference in the competence of the respondents. Importantly, Amina used more Clause-based lexical bundles, which suggests that she was making more use of subject + predicate constructions that carried the burden of the meaning. If this finding is compared with the use of Verb-based lexical bundles, it is noticeable that Hassan used a slightly greater (22) number of lexical bundles than Amina (20), which means that the use of Verb-based lexical bundles could not pick out subjects frequently, and only a few Verb-based lexical bundles could be realised in the form of clausal structures by Hassan, who is a less competent speaker. However, these conclusions need to be substantiated with the analysis of the sub-categories of the lexical bundles.

Table 1 and Figure 3 provide an overall distribution of the different structural types of lexical bundles for the sub-categories of Noun-, Verb-, Clause-, and Preposition-based lexical bundles.

Table 2. Distribution of Sub-categories of Lexical Bundles

Structural Types of Lexical Bundles (LBs)	**Amina**	**Hassan**
Noun-based Lexical Bundles	**Noun-based + of LBs:** 8 (*e.g., the end of the*)	**Noun-based + of LBs:** 0
	Other Noun-based LBs: 25 (*e.g., the girl see a*)	**Other Noun-based LBs:** 14 (*e.g., the happy prince he*)
	Total Noun-based LBs: 33	**Total Noun-based LBs:** 14
Clause-based Lexical Bundles	**Complete Clause-based LBs:** 74 (*e.g., he was crying and*)	**Complete Clause-based LBs:** 26 (*e.g., the bird went back*)
	Transitive but Object-less LBs: 9 (*e.g., he take his eyes*)	**Transitive but Object-less LBs:** 6 (*e.g., then the prince said*)
	Total Clause-based LBs: 83	**Total Clause-based LBs:** 32
Verb-based Lexical Bundles	**Copula Be & Passive LBs:** 6 (*e.g., is freezing then he*)	**Copula Be & Passive LBs:** 4 (*e.g., is feeling much better*)
	Verb/Adj + to LBs: 3 (*e.g., she want to buy*)	**Verb/Adj + to LBs:** 5 (*e.g., going to tell a*)

Structural Types of Lexical Bundles (LBs)	Amina	Hassan
	Verb/Adj + that LBs: 0	**Verb/Adj + that LBs:** 0
	Other Verb Forms: 11 (*e.g., talked to him then*)	**Other Verb Forms:** 13 (*e.g., gave the feathers to*)
	Total Verb-based LBs: 20	**Total Verb-based LBs:** 22
Prepositions	**Prepositions:** 9 (*e.g., into the table then*)	**Prepositions:** 12 (*e.g., in the bin then*)
	Total Prepositions: 9	**Total Prepositions:** 12

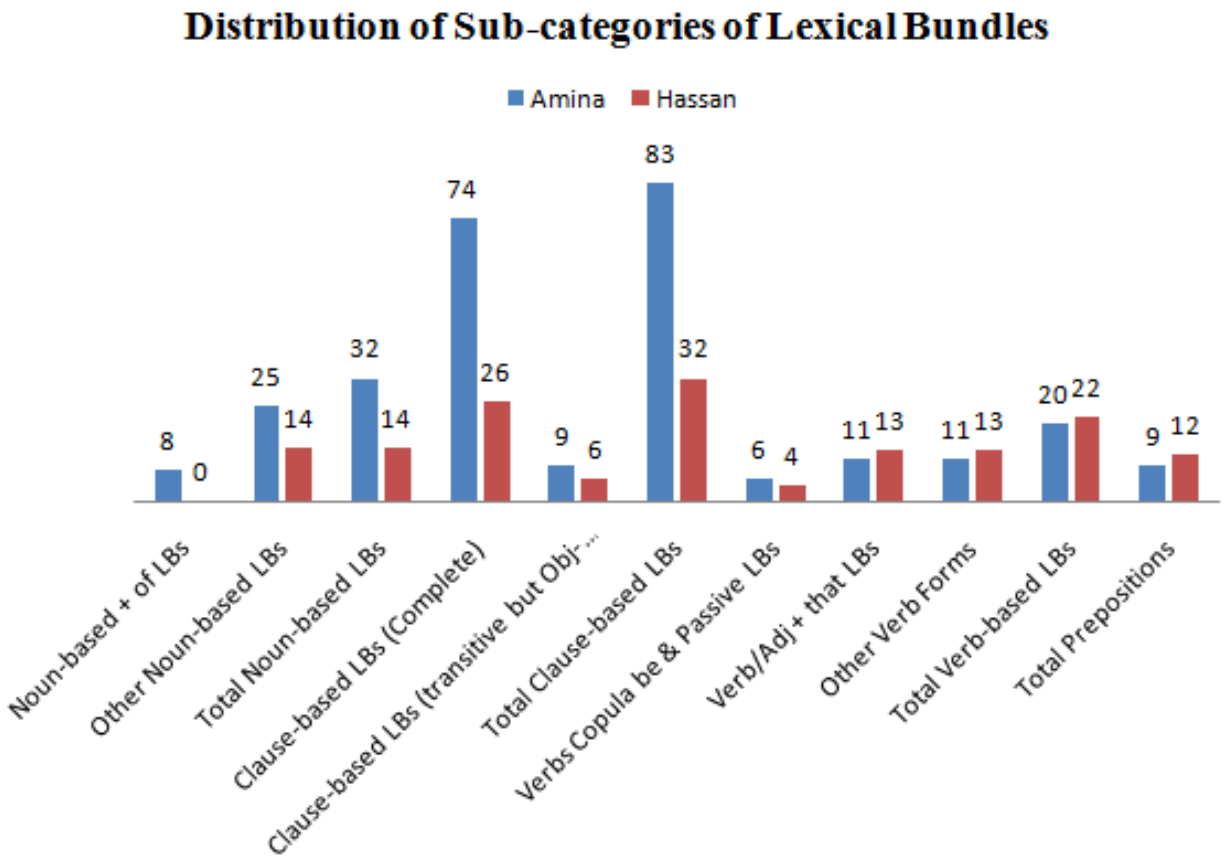


Figure 3. Distribution of Sub-categories of Lexical Bundles

The data show that for Noun-based lexical bundles, Amina was more likely to use Noun-based+of and Other Noun-based lexical bundles than Hassan. Although there is a small amount of data, it is quite likely that Hassan has not acquired the Noun-based+of structure yet as no such structure was present in Hassan's corpus data (see transcription in Appendix B). For Other Noun-based lexical bundles, Hassan employed a reasonable number (14) of lexical

bundles, but still less than that of Amina (25). Furthermore, for this sub-category, most of the nouns used by Hassan were also used by Amina (see the detailed list of LBs in Appendix C). Furthermore, Amina not only used a higher number of Noun-based lexical bundles, but also employed more adjectives to modify nouns. Interestingly, for the sub-category of Other Noun-based lexical bundles, Amina used the word "golden" seven times to modify nouns. No doubt, the use of adjectives with nouns reflects that Amina is more competent in using complex word (noun) combinations than Hassan.

For the last structural category of lexical bundles, Hassan made use of more prepositional lexical bundles than Amina. Nonetheless, for 22 prepositional structures, Hassan uttered only four different types of prepositions and Amina presented six different types of prepositions for 20 prepositional structures. This finding suggests that frequency occurrences of a structure should not be treated as the only marker of competence, and a more qualitative analysis is imperative to have better insights into the competent use of a structure by the speaker.

4. Discussion

The current study identified quantitative and qualitative differences in the use of lexical bundles by two Arabic native speakers with different levels of competence in English. The primary limitation of the study is the limited data, which meant there were certain structural sub-categories of lexical bundles that were not represented in the data. Overall, the current study highlights that more competent speakers make more use of clause fragments and elaborated nouns, both in quantity and quality. There was also evidence for language interference from Arabic to English, as Hassan was less inclined to use objects for transitive verbs. Perhaps a future study involving a larger amount of data collected from a larger number of speakers with different levels of competence would reveal possible associations between the use of clause fragments and L2 competence.

References

Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes, 31*(2), 81–92. https://doi.org/10.1016/j.esp.2011.08.004

Bal, B. (2010). Analysis of four-word lexical bundles in published research articles written by Turkish scholars (Unpublished MA thesis). USA: Georgia State University.

Barros, E. H., & Castro, M. B. (2017). Analysing phraseological units in legal translation: Evaluation of translation errors for the English-Spanish language pair. In *Phraseology in legal and institutional settings* (pp. 41–60). Routledge.

Bychkovska, T., & Lee, J. J. (2017). At the same time: Lexical bundles in L1 and L2 university student argumentative writing. *Journal of English for Academic Purposes, 30*, 38–52. https://doi.org/10.1016/j.jeap.2017.10.008

Chen, A. C. H. (2019). Assessing phraseological development in word sequences of variable lengths in second language texts using directional association measures. *Language Learning, 69*(2), 440–477. https://doi.org/10.1111/lang.12334

Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology, 14*(2), 30–49.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*(4), 397–423. https://doi.org/10.1016/j.esp.2003.12.001

Crossley, S., & Salsbury, T. L. (2011). The development of lexical bundle accuracy and production in English second language speakers. *International Review of Applied Linguistics in Language Teaching, 49*(1), 1–26. https://doi.org/10.1515/iral.2011.001

Crossley, S., Salsbury, T., & McNamara, D. (2009). Measuring L2 lexical growth using hypernymic relationships. *Language Learning, 59*(2), 307–334. https://doi.org/10.1111/j.1467-9922.2009.00508.x

Daller, 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

Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition, 18*(1), 91–126. https://doi.org/10.1017/S0272263100014698

Ellis, R. (1995). Modified oral input and the acquisition of word meanings. *Applied Linguistics, 16*, 409–435. https://doi.org/10.1093/applin/16.4.409

Farghal, M., & Obiedat, H. (1995). Collocations: A neglected variable in EFL. *International Review of Applied Linguistics in Language Teaching, 33*(4), 315–331. https://doi.org/10.1515/iral.1995.33.4.315

Granger, S. (2002). A bird's-eye view of learner corpus research. *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching, 6*, 3–33.

Gries, S. T. (2016). *Quantitative corpus linguistics with R: A practical introduction*. Taylor & Francis.

Güngör, F., & Uysal, H. H. (2020). Lexical bundle use and crosslinguistic influence in academic texts. *Lingua, 242*, 102859. https://doi.org/10.1016/j.lingua.2020.102859

Hakuta, K. (1974). Prefabricated patterns and the emergence of structure in second language acquisition. *Language Learning, 24*(2), 287–297. https://doi.org/10.1111/j.1467-1770.1974.tb00509.x

Hickey, T. (1993). Identifying formulas in first language acquisition. *Journal of Child Language, 20*, 27–41. https://doi.org/10.1017/S0305000900009122

Huang, K. (2015). More does not mean better: Frequency and accuracy analysis of lexical bundles in Chinese EFL learners' essay writing. *System, 53*, 13–23. https://doi.org/10.1016/j.system.2015.06.011

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*(1), 4–21. https://doi.org/10.1016/j.esp.2007.06.001

Hyland, K., & Jiang, F. K. (2018). Academic lexical bundles: How are they changing? *International Journal of Corpus Linguistics, 23*(4), 383–407. https://doi.org/10.1075/ijcl.17080.hyl

Lu, X., & Deng, J. (2019). With the rapid development: A contrastive analysis of lexical bundles in dissertation abstracts by Chinese and L1 English writers. *Journal of English for Academic Purposes, 38*, 49–61. https://doi.org/10.1016/j.jeap.2019.01.003

Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics, 33*(3), 299–320. https://doi.org/10.1093/applin/ams010

Nesi, H., & Gardner, S. (2012). *Genres across the disciplines: Student writing in higher education*. Cambridge University Press. https://doi.org/10.1017/CBO9781139050060

Pan, F., & Liu, M. (2019). Lexical bundles in Chinese doctoral students' dissertation writing. *System, 87*, 102–152. https://doi.org/10.1016/j.system.2019.102152

Paquot, M. (2010). *Academic vocabulary in learner writing: From extraction to analysis*. Bloomsbury Publishing.

Peters, A., & Menn, L. (1993). False starts and filler syllables: Ways to learn grammatical morphemes. *Language, 69*(4), 742–777. https://doi.org/10.1353/lan.1993.0074

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 1–22). John Benjamins Publishing. https://doi.org/10.1075/lllt.9

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics, 31*(4), 487–512. https://doi.org/10.1093/applin/amp058

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.

Staples, S., & Reppen, R. (2016). Understanding first-year L2 writing: A lexico-grammatical analysis across L1s and genres. *Journal of Second Language Writing, 32*, 17–35. https://doi.org/10.1016/j.jslw.2016.02.002

Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes, 12*(3), 214–225. https://doi.org/10.1016/j.jeap.2013.05.002

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press. https://doi.org/10.1017/CBO9780511519772

Wood, D. (2006). Uses and functions of formulaic sequences in second language speech: An exploration of the foundations of fluency. *Canadian Modern Language Review, 63*(1), 13–33. https://doi.org/10.3138/cmlr.63.1.13

Wood, D. (2015). *Fundamentals of formulaic language: An introduction*. Bloomsbury Publishing.

Xu, J., & Ding, Y. (2014). Lexical bundles in business English: Variations across spoken and written registers. *English for Specific Purposes, 34*, 83–93. https://doi.org/10.1016/j.esp.2013.11.002

Yoon, C. (2019). Measuring and tracking the development of collocational competence: Word frequency versus mutual information. *System, 80*, 176–187. https://doi.org/10.1016/j.system.2018.11.011

Zhang, X. (2017). Writing for peer-reviewed journals: Challenges and strategies for non-English-speaking scholars. *International Journal of English for Academic Purposes, 1*(1), 24–39.

Acknowledgments

I would also like to express my appreciation to my family and friends for their encouragement, patience, and understanding during this process. Your faith in me has been a constant source of motivation. Special thanks go to the participants who contributed their time and data to this study, as well as to the faculty members and colleagues who provided valuable insights and feedback. Your input enriched the study and expanded its scope.

Funding

Not Applicable.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Macrothink Institute.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

Open access

Copyrights

Appendix A

Amina's Data

Hey,

Aaaa. Aaaa, I was, I have story of Guinea toward about stone, golden stone, statue..... and he all have feathers.

All of his place and and his eyes are blue and statue men he has something on his soul soul and and all of the body all of the body came then aaaa aaa then happy and they came then then then they did they all of them happy. And the somebody was drop something then he see the prince and then he be happy then then aa some girl and her mom wants ice cream and his mom says no and the girl see aa aaaa strong golden stone and she said always be like him. Then the bird then the bird says to his mommy said can I go somewhere and his mommy said no and recently his mommy says no and he be angry of her mommy....

then he flied away then he see a golden stone, golden statue then he said I will rest under his feet then the came the night then then he sleeped and then he was crying and and aaaa and there is waterdrop and and he sleep again and then he say up and then he see he was crying and the the golden prince statue, golden statue then he tooked him about a baby have have a fever and he is crying and crying then then he said take the golden aaa his sword and he said no and he said ok.then he go and put it into the table then then he gone to the baby stommy and then he flup and then the mommy waked up and she see the girl then the then the the the then the aaaa back to the golden statue then she said why you crying and he said he talked to him then he said a woman old woman the old man he always waiting and he stepped here to me when he is called then he take his eyes then put it in the table he get a stick then he put it in the fire then then then he go back then he waked up then he see the eyes and he is aaaaa he take crying he take his eyes and he cried again and he take his because eyes and he cried again and he take his {…}eyes because some girl she want to buy and his dad say if you don't come with money I will be angry and then he takes his another eyes then then he he left the store the statue golden statue then then the…. bird go back then then he goes back because he doesn't have a eyes.

The bird his eyes and he turned away of him and then he talked to him, then he take one left to give it to everone and golden {…}.to him he give it to everyone. then he don't have ..... and the ice is everywhere and and the bird is freezing then he said I want to kiss you then he goed to kiss him then he then he then he died. The bird died. Then the statue his heart is broken then the mayor said get my stone and then the break the aaaaa golden stone,… then… then he go back and then he take his heart and throwed in a rubbish bin then then then the bird inside the rubbish bin. God and… angel take the bird and the heart then the god he said I fix the{…} bird and heart {…}. and finished.

That's the end of the story.

Appendix B

Hassan's Data

Hay

I am going to tell a story about a [ey] happy prince. One day there was a happy prince. He was made out of leaves and [and] his eyes was made out of a blue stones and his sword [was a] has a red ruby. [he] everybody likes story. They gave him a worth. [one] Then..[aaa] then the bay querry had [ah] had fallen his flower. He so.. so… the happy prince he said oh I wish I was like him. {…} [one] [aaa]There was a little girl with her mother. Her mother doesn't [ah] buy ice-cream for her. She said. " [ah a] I wish I was like this happy prince".[amm]

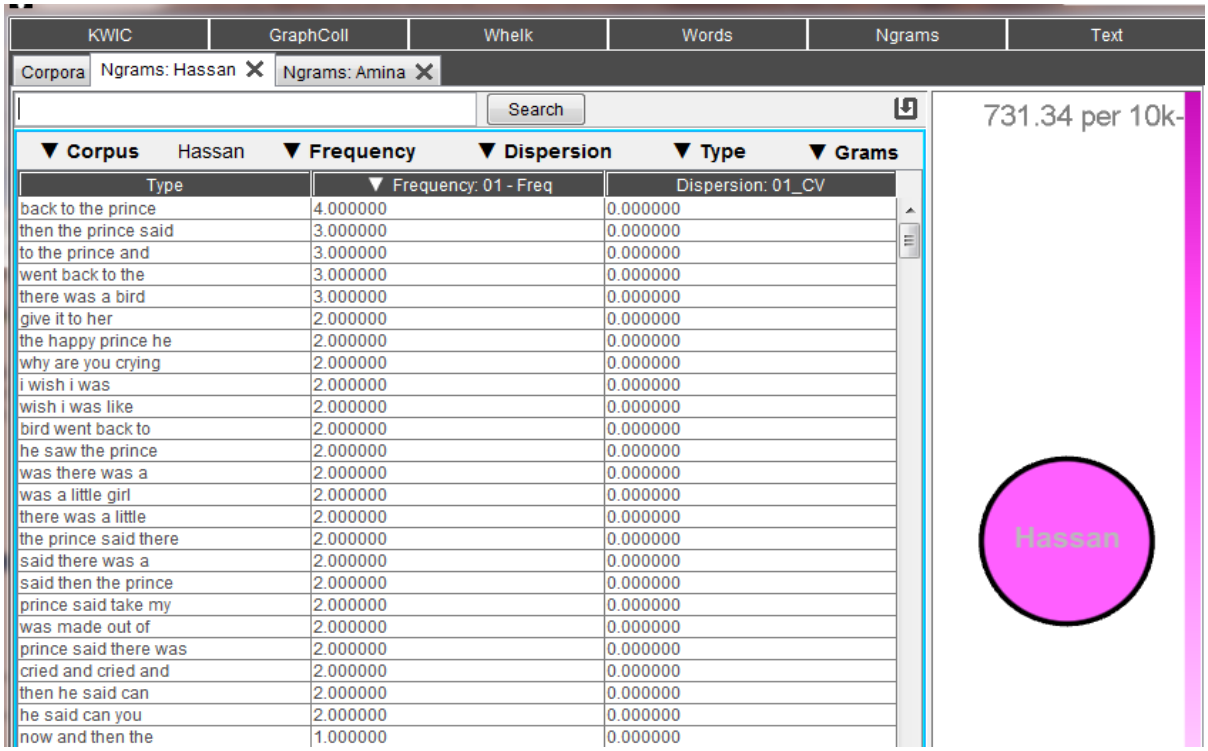
There was a little. [was] There was a bird. {…} There was a bird. {…} [Oh] There was. [oh] there was {…} hmmm. {…} there was. There was a bird under. The bird is under the happy prince. He [he] talked him. "Why are you crying?". He said, "I am [I am] crying because I am feeling someone a girl was stitching her red fingers and she has blood and [aaa] her baby has fever. He cried he wants some oranges. So he cried and cried and cried and his little and his mother has only water and then [aaa] the [aaaa] happy prince said, "Take my ruby to her, please". So he went to the. He gave the ruby to the women. Then He went back to the prince. Then the little baby said, "I am feeling much better". Than The Mumm was really happy.

Soo soo the bird [the bird] flew back to the prince and said [amm] and said and when he come back he saw the prince crying. He said, [hmm] "why are you crying?". [aaa] The bird said. Then the prince said, "there was a man. He was a write up. He doesn't have a feel and he was cold. So he said can you see come only one night and so the the girl said, "okay only one night". Take from my eyes and go give it to the writer he sent to the writer and found him waiting because he was very cold and tired and he doesn't have fit so the bird gave him his the princes eyes then he go to some joysticks put on the fire and so so the hmmm so he went back he came then the writer write up. So this eye and then bird went back to the prince and then it was a night time so he saw the prince sad again crying and then [ah] the bird said what happen to you. The prince said {…} there was a {…} there was a little girl put a fire for her stuff then [then] she was very scared for her father and then the prince said take my other eye and give it to her . One day he give it to her she went back skipping and then and {…} then [aa] the bird went back to the prince and so the [aa] prince happy. Then he said [can you take] I will stay with you because I can be your eyes. Then he gave the feathers to all the poor people.

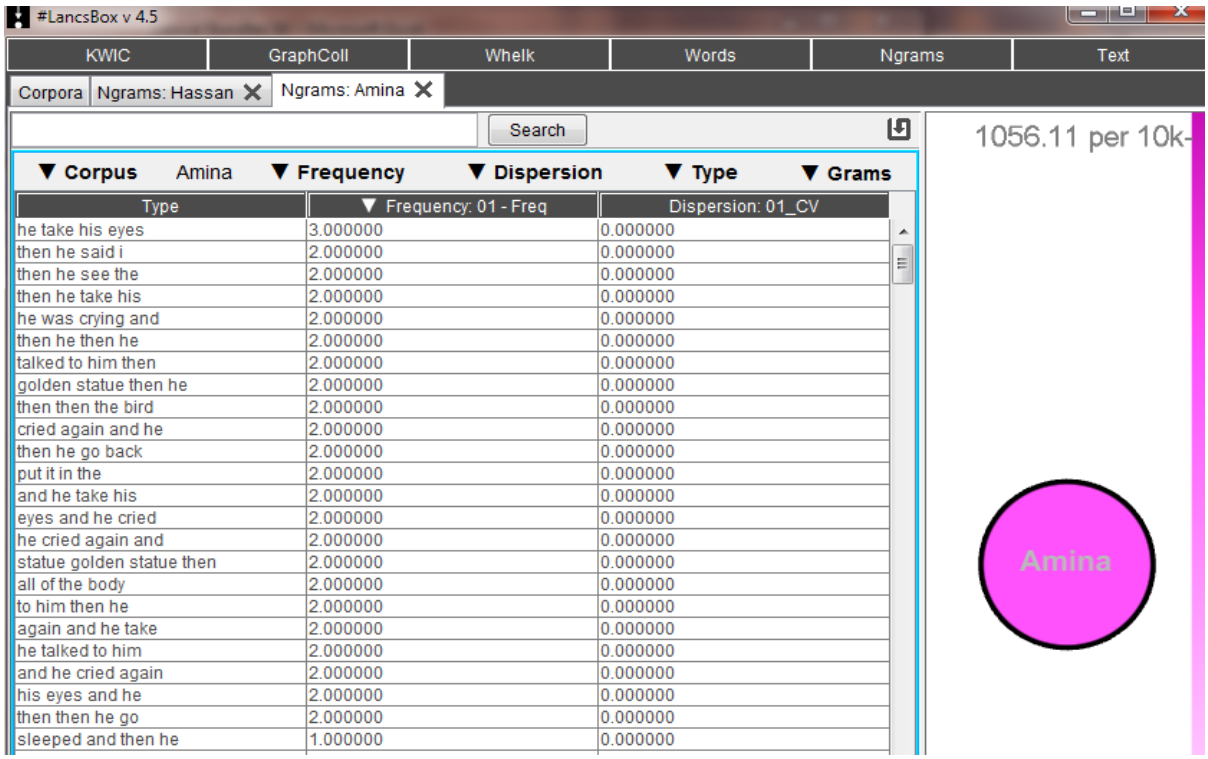
Then, everybody was happy. [Ah] then it was very winter and it's snowy. And the little bird said I can't go to any longer. [ah] he said then the prince said, "you can go back to the Egypt now" and {…} then the bird said, "No, I have to go back to put and then he said can I kiss you and so the prince said oh yes you can. And then he kissed him then he [he faa] died. The little bird died then the princes' back was banged. Then the [the] mayor came and said what an ugly prince. {…} Can you put [aa] immediately. Take it down. Can you put my picture? Put my statue instead so they chocked this statue you in the bin in the bin then [oh] in the bin. The end.

Appendix C

Screenshots from LancsBox



Corpus: Hassan	Frequency	Dispersion	Type	Grams
Type	Frequency: 01 - Freq	Dispersion: 01_CV		
back to the prince	4.000000	0.000000		
then the prince said	3.000000	0.000000		
to the prince and	3.000000	0.000000		
went back to the	3.000000	0.000000		
there was a bird	3.000000	0.000000		
give it to her	2.000000	0.000000		
the happy prince he	2.000000	0.000000		
why are you crying	2.000000	0.000000		
i wish i was	2.000000	0.000000		
wish i was like	2.000000	0.000000		
bird went back to	2.000000	0.000000		
he saw the prince	2.000000	0.000000		
was there was a	2.000000	0.000000		
was a little girl	2.000000	0.000000		
there was a little	2.000000	0.000000		
the prince said there	2.000000	0.000000		
said there was a	2.000000	0.000000		
said then the prince	2.000000	0.000000		
prince said take my	2.000000	0.000000		
was made out of	2.000000	0.000000		
prince said there was	2.000000	0.000000		
cried and cried and	2.000000	0.000000		
then he said can	2.000000	0.000000		
he said can you	2.000000	0.000000		
now and then the	1.000000	0.000000		



Corpus: Amina	Frequency	Dispersion	Type	Grams
Type	Frequency: 01 - Freq	Dispersion: 01_CV		
he take his eyes	3.000000	0.000000		
then he said i	2.000000	0.000000		
then he see the	2.000000	0.000000		
then he take his	2.000000	0.000000		
he was crying and	2.000000	0.000000		
then he then he	2.000000	0.000000		
talked to him then	2.000000	0.000000		
golden statue then he	2.000000	0.000000		
then then the bird	2.000000	0.000000		
cried again and he	2.000000	0.000000		
then he go back	2.000000	0.000000		
put it in the	2.000000	0.000000		
and he take his	2.000000	0.000000		
eyes and he cried	2.000000	0.000000		
he cried again and	2.000000	0.000000		
statue golden statue then	2.000000	0.000000		
all of the body	2.000000	0.000000		
to him then he	2.000000	0.000000		
again and he take	2.000000	0.000000		
he talked to him	2.000000	0.000000		
and he cried again	2.000000	0.000000		
his eyes and he	2.000000	0.000000		
then then he go	2.000000	0.000000		
sleeped and then he	1.000000	0.000000		