4-30-2018

The Development of Collocations as Constructions in L2 Writing

James Garner

THE DEVELOPMENT OF COLLOCATIONS AS CONSTRUCTIONS IN L2 WRITING

by

JAMES GARNER

Under the Direction of Ute Römer (PhD)

ABSTRACT

Cross-sectional and longitudinal learner corpus studies utilizing phraseological, frequency, and association strength approaches to phraseological unit identification have shown how the use of phraseological units varies across proficiency levels and develops over time. However, these methods suffer from several limitations, such as a reliance on native speaker intuition, a limited focus on contiguous word sequences, and a neglect of part of speech information in association strength calculation. This study seeks to address these limitations by defining lexical collocations as constructions (henceforth "collconstructions") within the framework of Construction Grammar and investigating their cross-sectional variation and longitudinal development in two corpora of L2 writing. The cross-sectional corpus consisted of beginner and intermediate EFL learner texts assessed for overall writing proficiency, while the longitudinal corpus contained freewrites produced by ESL learners over the course of one year. Contiguous and non-contiguous adjective-noun, verb-noun, and adverb-adjective collconstruction tokens were extracted from each learner text in the two learner corpora. Each learner text was assessed for multiple constructional and collostructional indices of collconstruction production. Constructional indices included type frequencies, token frequencies,

and normalized entropy scores for each collconstruction category. Collostructional indices consisted of proportion scores for different categories of adjective-noun, adverb-adjective, and verb-noun collconstruction types and tokens based on covarying collexeme scores calculated using frequency information from an academic reference corpus. Variation across proficiency levels was evaluated both qualitatively and quantitatively. The qualitative analysis consisted of examining variation in the production of specific functional collconstruction subcategories from a Usage-based Second Language Acquisition perspective. The quantitative analysis consisted of the calculation of an ordinal logistic regression in order to determine whether any indices of collconstruction production were predictive of L2 writing quality. Longitudinal development at the group level was investigated through the use of linear mixed effects models. Development for individual learners was examined from a Dynamic Systems Theory perspective that focuses on the role of variability in language development as well as interconnected development for multiple indices of collconstruction production. This study has important implications for future research on L2 phraseology research and second language acquisition research as well as phraseology pedagogy.

THE DEVELOPMENT OF COLLOCATIONS AS CONSTRUCTIONS IN L2 WRITING

by

JAMES GARNER

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2018

THE DEVELOPMENT OF COLLOCATIONS AS CONSTRUCTIONS IN L2 WRITING

by

JAMES GARNER

Committee Chair:	Ute Römer
Committee:	Scott Crossley
	YouJin Kim
	Stefan Gries

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2018

DEDICATION

This dissertation is dedicated to three important woman in my life: my mom, who taught me the meaning of hard work, my wife Amanda, whose love, support, and patience was the driving force behind the completion of this project, and my daughter Amelia, who I cannot wait to meet.

ACKNOWLEDGEMENTS

This project represents the culmination of a massive amount of work and effort by an outstanding group of people.

First, I am deeply indebted to my advisor, Ute Römer. From our first meeting in 2014, she has been a tremendous mentor and colleague. She has given up countless hours sharing her expertise in corpus-driven research on phraseology and Construction Grammar, helping me navigate the publication process, and helping me understand what it means to be great researcher and great teacher. I owe so much of who I am as an applied corpus linguist and teacher to her guidance and mentorship.

I am also indebted to Scott Crossley, who has time and again patiently walked me through different types of statistical analyses and has consistently pushed me to be a better quantitative researcher. I also owe a debt of gratitude to YouJin Kim, whose SLA course sparked my interest in Second Language Acquisition in general and Dynamic Systems Theory specifically.

I also want to thank my team of essay raters, Jessica Lian, Jillian Siehlmann, Rurik Tywoniw, and Pamela Pearson. Without their dedication to reading and rating over one hundred learner texts over the span of a couple weeks, this project would not be possible.

I also owe a debt of gratitude to the faculty of the Department of Applied Linguistics and ESL, especially those who taught the PhD course I took. In particular, I want to thank Diane Belcher for her guidance in qualitative analysis and genre analysis as well the many conversations we've had around the break room water cooler. I also want to thank John Murphy for helping me become a better educator of future language teachers.

I want to thank my wife, Amanda Garner, for supporting me throughout the last three years of doctoral work and making me laugh when I needed it. I also want to thank my friends and classmates in the department, especially David Chiesa, YuKyoung Shin, MinKyung Kim, Cindy Berger, Stephen Skalicky, Meredith D'Arienzo, and Katia Monteiro. I have enjoyed our times discussing topics in Applied Linguistics, chatting about life outside of work, venting sessions, and moments of mutual procrastination. I hope for many more of these no matter where life takes us.

This project has also been supported by researchers outside of Georgia State. I am indebted to Stefan Gries, whose advice on normalized entropy and covarying collexeme analysis was incredibly beneficial to this project. I also want to thank Tom Salsbury for graciously allowing me to use his corpus. I also want to thank Marolijn Verspoor and Diane Larsen-Freeman for their advice on Dynamic System Theory research.

Lastly, I want to thank my mother, Kim Huey, who even when I was in the fourth grade taught me the value of good research and writing skills.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1   INTRODUCTION

Firth (1957) was among the first linguists to argue for a strong connection between the words in a language and the contexts in which they frequently occur when he wrote his now famous statement, "you shall know the word by the company it keeps." Other researchers have since expanded on this idea, arguing for the inseparability of lexis and grammar and for the centrality of phraseological units in language knowledge and production (Römer, 2009). Pawley and Syder (1983), for instance, suggested that native speakers routinely rely on multi-word sequences to produce idiomatic utterances. Sinclair (1991), using evidence from his corpus-based analysis of collocations, proposed that speakers typically rely on word combinations stored in their mental lexicon in order to create grammatically correct and idiomatic utterances, a concept he named the Idiom Principle. Still others working from a Cognitive Linguistics perspective, such as Fillmore (1988), Goldberg (1995) and Langacker (2000), have argued that language, rather than being the combination of distinct lexical and grammatical systems, is instead a unified system consisting of different types of constructions, ranging from morphemes and single words, via partially lexically filled phrases, to fully general phrasal patterns.

Over the last few decades, corpus research into native speech and writing has provided strong evidence of the phraseological nature of the English language. Although exact estimates vary, this research has shown that between 20% and 80% of all words in large corpora of native speech and writing occur in frequent word combinations (Altenberg, 1998; Biber, Johansson, Leech, Conrad, and Finegan, 1999; Erman and Warren, 2000). Additionally, corpus-based research has demonstrated that phraseological units serve important discourse functions in different speech communities (Hyland, 2008). Through the use of phraseological units common

to a speech community, writers and speakers can signal their membership within the community (Wray, 2002).

Psycholinguistic research has also shown that knowledge of phraseological units provides certain processing advantages to native (L1) and proficient non-native (L2) speakers. Studies have demonstrated that sentences containing phraseological units are read more quickly by L1 and advanced L2 speakers than sentences not containing them (Conklin and Schmitt, 2008; Tremblay, Derwing, Libben, & Westbury, 2011). They have also shown that phraseological units are more quickly and accurately recognized as grammatically correct than other less frequent grammatically correct word sequences as well as grammatically incorrect word sequences (Jiang & Nekrasova, 2007; Arnon & Snider, 2010). Studies focusing on collocation knowledge specifically have shown that words in collocations are more quickly and accurately recognized as words when occurring in collocations (Durrant & Doherty, 2010; Ellis, Frey, & Jalkanen, 2009). Lastly, research on phraseological units in language production has shown that high-frequency word combinations are produced more quickly and fluently than lower-frequency word combinations (Arnon & Cohen Priva, 2013; Janssen & Barber, 2012).

Given the ubiquity of phraseological units in speech and writing and the processing advantages that knowledge of such units can provide, there has been an increased interest in investigating the nature of this knowledge in L2 acquisition. Research utilizing corpora of learner speech and writing has shown that despite their frequency in usage, L2 learners often have difficulty acquiring a productive knowledge of English phraseological units (Ebeling & Hasselgård, 2015; Paquot & Granger, 2012). Studies comparing L1 and advanced L2 speakers and writers have demonstrated that L2 learners often produce a smaller range of phraseological units (Ädel & Erman, 2012; Juknevičienė, 2009; Chen & Baker, 2010; Paquot & Granger, 2012).

This research has also shown that a large proportion of units produced by advanced L2 learners are erroneous combinations (Juknevičienė, 2008; Nesselhauf, 2003). It has also shown that L2 writers often produce more phraseological units composed of high-frequency words, while their L1 counterparts produce more units composed of low-frequency words (Durrant & Schmitt, 2009). Comparing the use of phraseological units across proficiency levels, studies have shown that more proficient L2 writers produce a greater range of phraseological units and produce them more frequently than less proficient learners (Garner, Crossley, & Kyle, in press, under review; Hsu, 2007; Leńko-Szymańska, 2014; Ohlrogge, 2009; Vidakovic & Barker, 2010). Focusing on the longitudinal development of phraseological units in L2 writing, studies have shown that development is often slow, uneven, and highly dependent on the type of units examined (Bestgen & Granger, 2014; Chen, 2013; Crossley & Salsbury, 2011; Garner & Crossley, in press; Li & Schmitt, 2009, 2010).

These learner corpus studies of L2 phraseological unit production have relied on different approaches to the identification and analysis of phraseological units. Studies taking a more traditional approach towards phraseological unit identification have utilized native speaker judgments to identify units and judge whether or not they are well-formed. In contrast, frequency-based analyses, such as n-gram and lexical bundle approaches, have identified units on the basis of recurrence above certain frequency thresholds. Alternatively, frequency information has been used to calculate the strength of association between component words, with these association scores being used to identify units of interest. While studies utilizing each of these methods have revealed important information concerning L2 phraseological unit production and development, they also suffer from several limitations. Native speaker judgments run the risk of being too subjective, with different groups of native speakers varying in their

identification of phraseological units. Frequency analyses depend largely on frequency thresholds that are often arbitrarily set. These analyses also often ignore phraseological units that are non-contiguous. Studies that include association measures most often do not consider the role of part of speech information when calculating association strength scores. Additionally, all three methods share several limitations. These analyses often do not consider individual variation in development and limit their analysis to either a single type of phraseological unit or a single measure of phraseological unit knowledge. As a result, these studies provide only small glimpses into L2 productive phraseological knowledge rather than providing a more comprehensive picture of this knowledge and its development in L2 learner language.

This dissertation hopes to address some these limitations by defining lexical collocations as constructions ("collconstructions") within a Construction Grammar perspective and examines their cross-sectional and longitudinal development in L2 writing. This new construct of collconstructions brings together two approaches to phraseological unit analysis: collocations as lexical functions (Mel'čuk, 2007) and collostructional analysis (Stefanowitsch & Gries, 2003). The concept of collocations as lexical functions sees lexical collocations as word associations that fulfill specific lexical functions such as intensification (e.g. *infinite patience*) and support (e.g. *make an effort*). Collostructional analysis is an approach that measures the strength of association between lexemes and grammatical constructions using statistical association measures (Stefanowitsch & Gries, 2003). By combining these two approaches, the construct of collconstructions provides a more comprehensive view of word associations that takes into account both the functional characteristics of collocations and the strong associations lexemes may have with specific grammatical constructions.

In order to investigate cross-sectional collconstruction development, this dissertation examines variation in collconstruction production in a corpus of EFL student writing divided into high-beginner, low-intermediate, and high-intermediate proficiency levels. Variation in collconstruction production across proficiency levels was examined in two ways. First, a qualitative analysis of subcategories of collconstructions fulfilling specific lexical functions (e.g. [acquire] + *money* verb-noun collconstructions) was conducted through a comparison of type-token distributions for each subcategory across proficiency levels. This analysis was grounded in usage-based theories of second language acquisition that claim that construction acquisition occurs gradually as learners move from relying on fixed exemplars to more schematic patterns (Ellis & Wulff, 2014). The second analysis focused on cross-sectional variation in collconstruction production for each of the three structural collconstruction categories (i.e. verb-noun, adjective-noun, adverb-adjective). For this analysis, the learner texts were analyzed for several constructional and collostructional indices of adjective-noun, verb-noun, and adverb-adjective collconstruction production. Constructional indices included type and token frequencies and normalized entropy scores. Collostructional indices consisted of proportion scores for different categories of covarying collexeme scores for co-occurring words within the collconstructions. These indices were then submitted to an ordinal logistic regression in order to determine which, if any, indices of collconstruction production were predictive of language proficiency.

Longitudinal development in collconstruction production was examined using a corpus of freewrites produced by ESL learners over the course of one year. These texts were assessed for the same constructional and collostructional indices as the texts from the cross-sectional corpus. Development was assessed at both the group-level and the individual-level. At the group-level,

linear mixed-effects models were calculated in order to estimate the effects of time and language proficiency on multiple indices of collconstruction development. Longitudinal development for individual learners was investigated from a Dynamic Systems Theory (DST; Verspoor, 2015) perspective. DST and the related Complexity Theory (CT; Larsen-Freeman, 2006) view language as a dynamic and complex system that is composed of smaller subsystems self-organizing into one system that encompasses a speaker's linguistic capabilities. These theories see L2 development as a process in which a learner's interlanguage system emerges through the interaction of the multiple linguistic subsystems that comprise it and internal and external resources. This process is marked by alternating periods of variability and stability as learners gradually figure language out (van Geert & Verspoor, 2015). It is also marked by interactions between different types of linguistic knowledge in which these knowledge types either support each other's development or compete for the limited resources need for growth (van Geert, 1991).

These cross-sectional and longitudinal analyses were conducted in order to address the following research questions:

1. How does the use of collconstructions in L2 writing vary across proficiency levels?
2. How does the use of collconstructions in L2 writing develop over time?
3. How does the use of collconstructions in L2 writing develop in individual learners over time?

This dissertation is organized as follows. In Chapter 2, I discuss the literature regarding phraseology in learner corpus research. This discussion will begin with a short review of research highlighting the importance of productive phraseological knowledge for L2 learners. I will then discuss the different approaches to phraseological unit identification and analysis (e.g.

phraseological, n-gram, association strength) that have been utilized. This will be followed by a review of contrastive, cross-sectional, and longitudinal studies conducted on phraseological unit production in L2 speech and writing, focusing on the major findings of each study. I will end the chapter by discussing some of the limitations of the different existing approaches to phraseological unit analysis that have been applied to studies of learner phraseological unit production.

In Chapter 3, I present the working definition of lexical collocations as a type of construction within a Construction Grammar perspective that will be used in this dissertation. The first part of this chapter will discuss the major tenets of Construction Grammar, including the key features of linguistic constructions. The second section will review the literature concerning L2 constructional knowledge and its development. Next, the characteristics of lexical collocations and how they meet some of the criteria for constructional status will be presented. The chapter will end with a brief discussion of how this definition of lexical collocations can improve upon previous approaches to L2 phraseology research.

In Chapter 4, I present the data and methods utilized in order to address each of the research questions in the study. I introduce the cross-sectional learner corpus, the longitudinal learner corpus, and the L1 reference corpus used in this dissertation. I then discuss how collconstructions were extracted from each of the three corpora and how the multiple constructional and collostructional indices were calculated. I end the chapter with a discussion of the statistical and graphical methods for assessing development across proficiency levels and over time used in the subsequent chapters.

In Chapter 5, I address research question 1 by investigating the production of collconstructions in the writing of high-beginner, low-intermediate, and high-intermediate L1

Korean EFL learners. The chapter starts with a qualitative analysis of specific functional collconstruction categories and their use across proficiency levels. This is followed by a quantitative analysis that examines whether learners from each of the three proficiency levels are significantly different from each other in terms of their collconstruction and if these differences are predictive of human judgments of writing proficiency. The chapter ends with a summary of the results and how they relate to previous research.

In Chapter 6, I address research question 2 by examining the development of productive collconstruction knowledge in a corpus of freewrites produced by a group of ESL learners over the course of one year. Following the selection of a subset of freewrites, longitudinal growth for a selection of constructional and collostructional indices is assessed through the use of linear mixed-effects modeling. The chapter ends with a summary of the results and how they relate to previous research.

In Chapter 7, I address research question 3 by investigating collconstruction development for individual learners in the longitudinal learner corpus from a DST perspective. The chapter starts by discussing some of the key tenets of DST approaches to L2 development and the findings of previous DST studies on L2 development. This is followed by a more in-depth presentation of the graphical methods used in the chapter. The results for the selected learners for each structural collconstruction category are then presented. These results, and their relationship to previous longitudinal L2 phraseology and DST research, are discussed in the final section of the chapter.

In Chapter 8, I conclude the dissertation by reviewing the overall results concerning collconstruction development across proficiency levels and over time. This review is followed by a discussion of the implications the findings in this dissertation have for the study of L2

phraseology development and for phraseology instruction in the L2 writing classroom. I also discuss the limitations of the dissertation in this chapter.

2   PHRASEOLOGY IN LEARNER CORPUS RESEARCH

Acquiring a productive knowledge of English phraseology is an essential component of gaining accuracy and fluency in a second language (L2). Corpus research has demonstrated that phraseological units such as collocations, idioms, and n-grams are ubiquitous in language and form the basis of a large proportion of utterances (Altenberg, 1998; Biber, Johansson, Leech, Conrad, & Finegan, 1999; Erman & Warren, 2000; Römer, 2009). Corpus-based research has also highlighted the importance of appropriate use of phraseological units for indicating membership in discourse communities (Durrant, 2015; Hyland, 2008; Wray, 2002). Additionally, psycholinguistic research has shown that knowledge of phraseological units provides certain processing advantages in receptive and productive language tasks (Arnon & Cohen Priva, 2013; Arnon & Snider, 2010; Conklin & Schmitt, 2008; Durrant & Doherty, 2010; Ellis, Frey, & Jalkanen, 2009; Janssen & Barber, 2012; Jiang & Nekrasova, 2007; Siyanova-Chanturia & Martinez, 2015; Tremblay, Derwing, Libben, & Westbury, 2011).

Although productive phraseological unit knowledge is important, learner corpus research has shown that L2 learners often have difficulty acquiring it (Ebeling & Hasselgård, 2015; Paquot & Granger, 2012). Contrastive studies have revealed that learners tend to rely on a smaller variety of phraseological units than their native-speaking counterparts (Ädel & Erman, 2012; Juknevičienė, 2009; Chen & Baker, 2010; Paquot & Granger, 2012), with a sizeable amount of them being either erroneous combinations (Juknevičienė, 2008; Nesselhauf, 2003) or inappropriate for the communicative situation (Ädel & Erman, 2012; Chen & Baker, 2010; De Cock, 1998; Juknevičienė, 2009). Research has also shown that L2 writers tend to rely on phraseological units composed of high-frequency words, while their L1 counterparts tend to produce more units composed of low-frequency words (Durrant & Schmitt, 2009). Cross-

sectional studies have shown that both the range and frequency of phraseological units tend to increase with increasing language proficiency (Garner, Crossley, & Kyle, under review, in press; Hsu, 2007; Huang, 2015; Leńko-Szymańska, 2014; O'Donnell, Römer, & Ellis, 2013; Ohlrogge, 2009; Vidakovic & Barker, 2010). At the same time, longitudinal studies have shown that development of productive phraseological knowledge can be slow, uneven, and highly dependent on the type of units examined (Bestgen & Granger, 2014; Chen, 2013; Crossley & Salsbury, 2011; Garner & Crossley, in press; Li & Schmitt, 2009, 2010).

This chapter is divided into four sections. First, the importance of productive phraseological knowledge for second language acquisition (SLA) will be discussed, with special reference to corpus-based and psycholinguistic research on English phraseology. Second, current approaches to the operationalization of phraseological units in learner corpus research will be presented. This will be followed by a review of the research into phraseological unit production in L2 speech and writing, including contrastive, cross-sectional, and longitudinal studies. The chapter will end with a critical discussion of the limitations of the current approaches in learner corpus-based L2 phraseology research.

2.1   The Importance of Phraseology for Second Language Acquisition

Over the past three decades, there has been a considerable increase in interest in the phraseological nature of English. Pawley and Syder (1983) were two of the earliest scholars to suggest that native speakers rely on conventionalized multi-word sequences to construct idiomatic utterances. They define this capacity of native-like selection as,

> the ability of a native speaker routinely to convey his [or her] meaning by an expression
>
> that is not only grammatical but also nativelike; … he [or she] selects a sentence that is

natural and idiomatic from among the range of grammatically correct paraphrases, many

of which are non-nativelike or highly marked usages. (p. 191)

Similar to the idea of native-like section is the Idiom Principle formulated by John Sinclair.

Based on evidence from corpus-based analysis of collocations, Sinclair claimed that a language

user "has available to him or her a large number of semi-preconstructed phrases that constitute

single choices, even though they might appear to be analyzable into segments" (Sinclair, 1991, p.

110). According to this principle, language users employ combinations of words that have taken

on their own meaning or function in order to construct grammatically correct and idiomatic

utterances. For Sinclair, word combinations such as *of course*, *set eyes on,* and *in some cases* are

not constructed item by item in utterances, but rather represent single choices made by language

users.

Large-scale corpus-based research has provided support for the concepts of nativelike

selection and the Idiom Principle, showing how a large proportion of native utterances are

composed of phraseological units. Altenberg (1998) examined the phraseological nature of

spoken English using the London-Lund Corpus. His analysis, which relied on identifying

phraseological units on the basis of the frequency of word combinations, showed that around

80% of words in the corpus formed part of a recurrent word combination. Biber, Johansson,

Leech, Conrad, and Finegan (1999), in their analysis of the Longman Spoken and Written

English Corpus, provided much smaller estimates, most likely due to their use of a narrower

definition of phraseological units. Focusing on three- and four-word lexical bundles (recurrent

word combinations occurring above certain range and frequency thresholds), they found that

about 30% of the running words in conversation and about 21% of the running words in

academic prose occur in lexical bundles. Erman and Warren (2000) estimated that between 52%

and 58% of words occurring in native speech and writing occur in phraseological units of two words or longer. Regardless of the exact estimates, all three studies have provided evidence for the importance of acquiring both a receptive and productive knowledge of English phraseology.

Corpus-based research has also highlighted the importance of acquiring productive phraseological knowledge for L2 speakers and writers in academic contexts. This research has shown that different academic disciplines utilize different types of phraseological units in order to shape the discourse of the discipline. Hyland (2008) found considerable differences in the types of units employed across four academic disciplines (Business Studies, Applied Linguistics, Electronic Engineering, Biology). Theses, dissertations, and research articles in Business Studies and Applied Linguistics were found to exhibit a greater concentration of text-oriented units in order to contextualize and evaluate information and express authorial stance. The texts from Electronic Engineering and Biology, on the other hand, contained more research-oriented units as well as units for guiding reader interpretations of data. Durrant (2015) found that phraseological units employed in undergraduate course papers receiving top marks from course instructors also differed depending on the discipline. His findings showed that Humanities and Social Sciences texts contained more units that express stance, refer to abstract constructs, highlight contrast, and set claims in context. In contrast, Science and Technology texts contained more units that describe physical characteristics of objects, signal causative relationships, and that refer to data in figures and tables. For L2 writers, then, it is important that they acquire a productive knowledge of the preferred phraseological units of their chosen discipline. In doing so, they may be better able to signal their membership in their chosen community and, as a result, their ability to take part in its discourse (Wray, 2002).

Lastly, psycholinguistic research has shown that knowledge of phraseological units can provide certain processing advantages when comprehending and producing fluent speech. Studies utilizing self-paced reading tasks have shown that both L1 and advanced L2 speakers read sentences containing phraseological units significantly more quickly than they read sentences not containing them (Conklin and Schmitt, 2008; Tremblay, et al., 2011). Results from grammaticality judgment tasks have shown that L1 and advanced L2 speakers more quickly and accurately recognize phraseological units as grammatically correct and possible than both grammatically correct word sequences and ungrammatical sequences (Jiang & Nekrasova, 2007; Arnon & Snider, 2010). Similarly, findings from lexical decision tasks have shown that words are more quickly and accurately recognized as words when encountered in a collocation (Durrant & Doherty, 2010; Ellis, Frey, & Jalkanen, 2009). Although the processing of phraseological units in language production remains relatively underrepresented in the literature, several studies have shown that L1 speakers are sensitive to phrasal frequency when producing language. Janssen and Barber (2012), for instance, found that L1 Spanish speakers and L1 French speakers orally produced more frequent noun-adjective, noun-noun (Spanish group only), and determiner-noun-adjective (French group only) phrases in their native language more quickly than less frequent ones. In addition, Arnon and Cohen Priva (2013), using elicited and spontaneous speech data from L1 speakers, found that higher frequency phraseological units experienced a greater degree of phonetic reduction than low-frequency word sequences. This phonetic reduction was also found to not be affected by constituent word frequency or verb-phrase constituency, indicating that it was phrasal frequency that affected how quickly L1 speakers produced phraseological units.

2.2   **Current Approaches to Phraseological Unit Analysis**

Corpus research into L2 phraseological unit production has typically proceeded from one of two approaches to phraseology. The first approach, known as either the traditional or phraseological approach, attempts to identify phraseological units according to top-down linguistic criteria (Granger & Paquot, 2008). One of the most commonly used linguistic criteria for identifying phraseological units is that of non-compositionality. Word sequences are considered non-compositional, and hence idiomatic, if the meaning of the whole sequence cannot be derived from the meanings of the individual words. Thus, a sequence such as *blow a fuse* would be considered an idiom, while *blow a trumpet* would be considered a free combination and outside the realm of phraseology. Between these two extremes lie other categories that differ in degree of compositionality. For example, restricted collocations, such as *make a decision*, are comprised of one freely chosen lexeme and another lexeme with a figurative meaning that is selected based on the first (Mel'čuk, 1998). In regards to learner corpus research, studies adopting a phraseological approach usually identify idioms or restricted collocations either on the basis of native speaker judgments or reference dictionary information. For example, Nesselhauf (2003) used advanced learner dictionaries, large-scale corpus data (BNC), and native speaker judgments to investigate the use of verb-object-noun restricted collocations in the writing of advanced L1 German learners of English.

The second major approach to identifying phraseological units is the distributional approach. Also known as the frequency-based approach, this method is more inductive in that phraseological units are identified on the basis of statistics derived from corpus data (Granger & Paquot, 2008). Within the distributional approach, there are three main types of analysis that differ in terms of how frequency data is used. The first type of analysis is n-gram analysis, which

uses raw frequency to identify recurring contiguous sequences of two or more words (e.g. *on the other hand, in addition to*). Studies taking this approach often investigate the number of different n-gram types in a learner corpus (O'Donnell et al., 2013) or examine L2 learners' use of frequent n-grams found in larger reference corpora of native speech or writing (Kyle & Crossley, 2015). Similarly, lexical bundles analysis uses raw frequency to identify frequently recurring word combinations. It differs from n-gram analysis, however, in that n-gram analysis focuses on all recurring contiguous word sequences (although a minimum occurrence frequency may be specified), while lexical bundle analysis limits focus to only those sequences that occur above certain range and frequency thresholds (Altenberg, 1998; Biber, Conrad, & Cortes, 2004). Lexical bundle research also tends to focus on four-word sequences, although some studies have also investigated three-word and five-word lexical bundle production. For instance, in their investigation of lexical bundles in university teaching and textbooks, Biber, Conrad, and Cortes (2004) identified recurrent four-word sequences as lexical bundles only if they occurred at least forty times per million words and appeared in at least five texts in their corpus. By setting frequency and range cut-offs for lexical bundle status, researchers attempt to limit their analysis to contiguous word sequences that are more likely to be important in spoken and written discourse (Altenberg, 1998). The use of a range threshold also guards against idiosyncratic uses by individual writers or speakers (Biber, Conrad, & Cortes, 2004).

The third type of analysis in the distributional approach to phraseology is association measure analysis. This type of analysis classifies word combinations as collocations if the constituent words occur together within a certain span more often than is expected by chance. In order to determine this, researchers employ statistical association measures such as mutual information (MI) and t-score (Evert, 2009). These measures differ in that MI tends to highlight

collocations consisting of low-frequency words while t-score tends to highlight collocations consisting of high-frequency words. In addition to calculating association strengths for collocating words over a certain span, association measures can be calculated for bigrams (2-word n-grams). Most learner corpus studies that use association measure analysis calculate MI and t-scores for collocations in a learner text using frequency data from large reference corpora. Collocations are then classified into bands of MI and t-scores ranging from non-collocation to high association strength. Type and token frequencies for each band are then calculated for each learner text. Type and token frequencies are also calculated for "Below Threshold" word combinations that occur too infrequently (less than five occurrences) in the reference corpus for association score calculation. Alternatively, rather than using MI and t-score bands, average MI and t-scores and "Below Threshold" proportions can be calculated for each text. In either case, these measures are averaged over all learner texts in the learner corpus or subcorpus, allowing researchers to investigate individual variation (Bestgen & Granger, 2014; Durrant & Schmitt, 2009; Granger & Bestgen, 2014; Paquot, 2017).

2.3   Phraseological Unit Production in L2 Speech and Writing

2.3.1   Contrastive Studies of L2 Phraseological Unit Production

The majority of learner corpus studies of L2 phraseological unit production have tended to be contrastive in that they examine differences between L1 and advanced L2 speakers and writers. Regardless of approach, findings from these studies have highlighted the issues learners have in employing phraseological units. From a phraseological perspective, studies have shown that advanced L2 writers often misuse and underuse restricted collocations. Nesselhauf (2003), for example, found that, among all erroneous verb-object-noun combinations employed by advanced L1 German writers, one fifth were collocation errors. The L1 Lithuanian writers in

Juknevičienė (2008) produced half as many collocations in their English argumentative writing as their L1 counterparts. The collocations they did produce were often those more typical of speech and generally acquired early (e.g. *take place, take care*). The most significant cause of erroneous collocations in these studies was L1 transfer, with both groups of learners relying on this strategy to compensate for their limited collocation knowledge.

Frequency-based contrastive studies have also revealed significant differences in phraseological unit production between native and nonnative speakers and writers. For instance, studies investigating lexical bundle production have shown that L2 writers and speakers tend to rely on a smaller set of lexical bundles and produce them less frequently than L1 speakers and writers. Juknevičienė (2009), in her analysis of four-word lexical bundle use by advanced L1 Lithuanian writers, found that these writers employed significantly fewer lexical bundle types and used them less frequently than native speaking writers. Similarly, Chen and Baker (2010) found that L1 Chinese writers enrolled in British universities produced a smaller range of lexical bundles less frequently than their native-speaking counterparts and published academic writers. Ädel and Erman (2012) compared the writing by L1 Swedish and native British English speaking university students of English linguistics, focusing on the use of four-word lexical bundles. They found that the Swedish writers employed fewer than half the number of lexical bundle types than the native writers. Focusing on L2 speech, De Cock (2004) found that, overall, the L1 French speakers in her study underused two- to six-word bundles compared to native speakers.

In addition, frequency-based contrastive research has also provided evidence that L2 writers and speakers have difficulties employing context appropriate bundles. Juknevičienė (2009) found that advanced L1 Lithuanian writers overused bundles characteristic of spoken discourse, including verb phrase bundles and personal stance bundles. Chen and Baker (2010)

and Ädel and Erman (2012) found that the advanced L2 writers in their studies underused hedging lexical bundles. Ädel and Erman (2012) also found that L1 Swedish writers underused bundles containing unattended "this" and existential "there" while overusing anticipatory-it bundles containing informal words such as "easy" and "hard". Regarding L2 speech, De Cock (2004) found that her L1 French speakers underused bundles that indicate vagueness (e.g. *sort of, kind of*), and overused emphatic bundles (e.g. *of course*). Taking an association measures perspective and comparing L1 and L2 writers' production of adjective-noun and noun-noun bigrams, Durrant and Schmitt (2009) found that L2 writers significantly overused bigrams consisting of high-frequency words (as indicated by t-score), while underusing those comprised of low-frequency words (as indicated by MI score). Subsequent type-token analyses further revealed that these L2 writers relied on a small range of both types of bigrams.

2.3.2   Cross-sectional Studies of L2 Phraseological Unit Production

Researchers have also begun to investigate the development of L2 phraseological unit production across proficiency levels using cross-sectional learner corpora. In these studies, proficiency is determined based on either human or computer judgments of writing quality or on time spent studying English. Studies taking a phraseological approach have demonstrated that, overall, more advanced learners produce a greater range of restricted collocations more frequently than less proficient learners. Hsu (2007) examined correlations between essays scores produced by an automated scoring system and restricted collocation type and token frequencies in essays written by L1 Chinese university students. Significant positive correlations were found between both restricted collocation frequency measures and holistic essay scores. Comparing type and token frequencies, it was found that the former correlated most strongly with essay scores, indicating that variety of restricted collocations may be a better indicator of writing

scores than quantity. Comparing different structural collocation types, Hsu found that verb-noun and adjective-noun restricted collocations were most strongly correlated with essay scores. Ohlrogge (2009) investigated the connection between multiple types of formulaic sequences and essay scores for intermediate EFL learners. His analysis focused on eight categories of formulaic sequences: restricted collocations, idioms, phrasal verbs, personal stance markers, transitions, language copied from the prompt, generic rhetoric, and irrelevant biographical information. Strong positive correlations were found between idiom and restricted collocation frequencies and essay scores as well as between stance markers and essay scores. In contrast, a significant negative correlation was found for copied text and essay scores. Laufer and Waldman (2011) examined verb-noun restricted collocation production in essays written by L1 Hebrew learners of English across basic, intermediate, and advanced proficiency levels. In this study, proficiency was not determined by essay scores, but rather time spent studying English. In addition to calculating frequencies, restricted collocations were classified as either well-formed or incorrect based on dictionary information. Results showed that more advanced learners produced more restricted collocations than learners in the other two levels; however, this difference was only statistically significant for the comparison between basic and advanced learners. Furthermore, results showed that advanced learners produced as many erroneous combinations as the beginning and intermediate learners.

Studies employing n-gram analysis have also shown that, with increasing proficiency, L2 writers and speakers use phraseological units more frequently. O'Donnell, Römer, and Ellis (2013) explored phraseological unit production in L1 and L2 academic writing across different levels of writing expertise, focusing on undergraduate, graduate, and published academic writing. Their exploration of phraseological unit production involved the use of n-gram and

association score (specifically MI) measures. Focal n-grams ranged in length from three words to five words. The n-gram analysis showed a significant effect of expertise, with the least proficient L1 and L2 writers producing fewer frequency-defined n-grams types than either graduate or expert writers. No significant differences, however, were found between the graduate and expert writers in terms of their n-gram use. Kyle and Crossley (2015) analyzed written texts produced by ESL writers and rated for holistic lexical proficiency by expert raters. They also analyzed transcripts from the spoken section of the TOEFL exam that were rated for holistic speaking proficiency. Using the Tool for the Automatic Analysis of Lexical Sophistication (TAALES), bigram (two-word sequence) and trigram (three-word sequence) production in the texts and transcripts was compared to bigram and trigram production in the spoken and written subsections of the British National Corpus (BNC Consortium, 2007). Their results indicated that more lexically proficient L2 texts contained more high-frequency trigrams from the written BNC and a higher proportion of frequent bigrams in the spoken BNC. Their analysis of the spoken transcripts also revealed that higher speaking proficiency scores were predicted by the production of more high-frequency trigrams found in the written BNC.

Lexical bundle research has also provided evidence of increasing phraseological unit range and frequency with increasing language proficiency. Vidakovic and Baker (2010) examined four-word lexical bundle production in the essays written for the Cambridge Skills for Life Examination across five proficiency levels corresponding to A1 (beginner) to C1 (advanced) on the Common European Framework of Reference for Languages (CEFR). Frequency results showed that both the variety and overall frequency of bundles were higher at more advanced proficiency levels. Leńko-Szymańska (2014) investigated three-word lexical bundle production in essays composed by learners in six EFL contexts and across three proficiency levels (low,

medium, high) based on school year. Type and token frequencies as well as lexical bundle proportions (i.e. proportion of running words accounted for by the lexical bundles) were calculated for each learner text. Results showed that, across all L1 backgrounds, learners produced a greater variety of lexical bundles as time spent learning English increased. Similar increases for lexical bundle token frequency, however, were only observed for three groups of learners (Spanish, Polish, and Austrian). Lexical bundle proportions were also found to increase as time spent learning English increased; nevertheless, they remained far below that of native texts. Huang (2015) examined the production of three- to five- word lexical bundles in timed argumentative essays written by Chinese EFL learners enrolled in their junior and senior years of English major study at multiple Chinese universities. Similar to Leńko-Szymańska (2014), Huang's results showed that the more experienced senior students produced a greater range of lexical bundles than their less experienced counterparts. Results also showed that the senior learners produced lexical bundles more frequently than the junior students.

Learner corpus studies examining lexical bundle production have also focused on the different types of lexical bundles produced by learners at multiple proficiency levels. In their study, Vidakovic and Barker (2010) found that stance bundles gradually increase across all proficiency levels. Discourse bundles, on the other hand, were found to only increase in frequency at the high-intermediate and advanced levels. Chen and Baker (2016) investigated the structural and functional characteristics of four-word lexical bundles in essays written by intermediate and advanced L1 Chinese writers. Essays written by L1 Chinese learners were evaluated by trained raters and divided into three subcorpora representing the B1 (low-intermediate), B2 (high-intermediate), and C1 (advanced) CEFR levels. Their results showed that the advanced writers were better able to employ bundles characteristic of academic prose. This

included a greater use of noun phrase and prepositional phrase based, framing, and impersonal stance bundles. B1 learners, on the other hand, tended to rely on bundles more characteristic of speech, preferring to employ verb phrase based, quantifying, and personal stance bundles. B2 writers formed a sort of middle ground, using bundles characteristic of both speech and writing fairly evenly.

Conversely, some lexical bundle analyses of L2 production have shown that more proficient L2 writers produce fewer lexical bundles than less proficient learners. For instance, Staples, Egbert, Biber, and McClair (2013) divided responses to the TOEFL iBT writing section into three proficiency levels (low, intermediate, high) based on percentile ranks for the mean test score. The results of their lexical bundle analysis showed that the high scoring responses contained significantly fewer four-word lexical bundles than responses in the low proficiency group. Appel and Wood (2016) divided written responses to a college entrance language exam into two proficiency levels: low and high. Texts in the low proficiency group received scores lower than half of the total possible points, while the high proficiency texts received scores higher than half. The authors found that the low proficiency texts contained more four-word lexical bundles. This overuse of lexical bundles was strongest for stance and discourse-organizing bundles. However, the results from both studies may have been affected by the low-scoring learners relying on bundles taken directly from the writing prompts. In fact, Staples et al. (2013) found that the learners in the low proficiency group used significantly more prompt-provided bundles than learners in the other two groups.

Association measure analyses have also demonstrated significant differences between L2 writers at different proficiency levels. In the previously discussed study, O'Donnell et al. (2013) found that the graduate-level L2 writers produced significantly more high MI score three- to

five-word n-grams than the undergraduate writers. Granger and Bestgen (2014) investigated the production of noun-noun, adjective-noun, and adverb-adjective bigrams by intermediate and advanced French, German, and Spanish EFL writers. Results indicated that the intermediate learners' texts contained a smaller proportion of low-frequency bigrams (as measured by MI score) than the texts written by advanced learners. Conversely, the intermediate learners employed a larger proportion of high-frequency bigrams (as measured by t-score). Comparing results for different structural bigram types, the strongest effects were found for modifier-noun, adjective-noun, and all types combined categories.

In a second study, Bestgen and Granger (2014) analyzed the bigram production by university-age ESL learners using association measure scores. Essays collected from learners over the course of a semester were evaluated by expert raters for overall proficiency and vocabulary and language use. Significant positive correlations were found between mean MI type and token scores and both categories of essay quality. Correlations were stronger for MI type scores than MI token scores, indicating that the variety of strongly associated bigrams may have a stronger effect on essay quality than repetition. Negative correlations were also found between proportions of absent bigrams and essay scores, indicating a possible effect of ungrammatical or unidiomatic bigrams on judgments of essay quality.

Paquot (2017) examined the use of adjectival (adjective-noun), adverbial (adverb-adjective, adverb-verb, adverb-adverb), and direct object (verb-noun) collocations in academic texts written by French EFL learners. The texts were judged by trained raters and placed into proficiency levels ranging from level B2 to C2 on the CEFR. Her results showed statistically significant increases in mean MI score for all three types of combinations across the three proficiency levels. However, significant differences were not observed for all pairwise

comparisons. For instance, while B2 and C2 differed significantly from each other for all three collocation types, B2 and C1 only differed significantly in adverbial mean MI score, while C1 and C2 only differed in direct object mean MI score.

In contrast to several of the previously reviewed association measure studies that focused on intermediate and advanced learners, Garner, Crossley, and Kyle (in press) examined differences in association strength for bigrams and trigrams across high-beginner, low-intermediate, and high-intermediate L1 Korean EFL writing. Additionally, they utilized multiple indices of association strength (e.g. MI, t-score, ΔP, and collexeme score) based on the spoken and academic subsections of the Corpus of Contemporary American English (COCA; Davies, 2009) in order to determine which, if any, were most predictive of human judgments of writing proficiency. Their results demonstrated that the intermediate writers produced more strongly associated bigrams common to academic writing and more strongly associated trigrams common to spoken English than the beginner learners. Comparing the different association measure indices, results showed that ΔP outperformed all the other measures in terms of its predictive power.

In another study, Garner, Crossley, and Kyle (under review) combined raw frequency and association strength analyses in order to determine what, if any, indices of bigram and trigram production are most predictive of writing quality. Their corpus consisted of placement exam essays written by L1 Korean EFL learners rated for overall writing proficiency from A2 to C1 on the CEFR. Their indices included measures of frequency, proportion (i.e. a measure that reports the proportion of n-grams in a text that occur frequently in a reference corpus), and association strength that were based on the academic and spoken subsections of COCA. Their results showed that a combination of proportion and association strength indices explained almost one

fifth of the variance in human judgments of writing proficiency. This finding suggested that writers judged as more proficient produced more bigrams and trigrams common to academic writing and more strongly associated bigrams and trigrams than writers who were judged as less proficient.

2.3.3 Longitudinal Studies of L2 Phraseological Unit Production

While not as common as contrastive or cross-sectional studies, learner corpus researchers have also begun to produce studies that examine phraseological unit development longitudinally. While cross-sectional studies have demonstrated that more advanced learners produce a greater range of phraseological units and produce them more frequently, longitudinal analyses have revealed more mixed results. These studies have shown that developmental trajectories are often nonlinear and vary according to how phraseological units are operationalized. From a phraseological perspective, Li and Schmitt (2009) investigated the development of formulaic language production in texts written by one highly advanced L1 Chinese learner of English. The learner's written data consisted of course papers and a dissertation, all of which were written over the course of ten months. The type of formulaic sequence examined in this study was lexical phrases, which the authors defined as frequently occurring conventionalized form/function composites that have idiomatic meaning (e.g. *a summary can be drawn from, it seems that, there is no consensus*). Results showed that the learner developed a greater range and frequency of lexical phrases in her writing and was better able to employ them appropriately over time. However, this development was non-linear, with several peaks and valleys in range, frequency, and appropriateness occurring at different points over the 10-month period.

Chen (2013) analyzed the development of phrasal verb production by university-level L1 Chinese learners. Subjects wrote timed essays twice a year over the course of three years. Results

showed that there was no link between phrasal verb production and time studying English. Both type and token phrasal verb frequencies fluctuated over time, decreasing from the first year to the second and increasing from the second to the third. Qi and Ding (2011) investigated changes in formulaic sequence use by EFL Chinese speakers in their first and fourth years of university study. Formulaic sequences in transcripts of these learners' monologues were identified manually and checked for formulaicity using a dictionary and native speaker judges. Results showed that over the course of three years, the learners had increased their knowledge of formulaic sequences. They produced a wider range of sequences in their fourth year interview than in their first year one and used some formulaic sequences more accurately in their fourth year, although a few errors still remained.

Frequency-based analyses utilizing both n-gram measures and association strength measures have also demonstrated mixed patterns of development in learner phraseological unit production. Li and Schmitt (2010) combined both types of analyses in their investigation of adjective-noun bigram development for a group of highly advanced L1 Chinese ESL learners. Similar to Li and Schmitt (2009), the data consisted of course papers and dissertations written by the learners over the course of one year. Group results indicated that the subjects used a less diverse range of adjective-noun bigrams and used them less frequently over time. Individual results showed that one subject experienced a general decline across all indices, while another subject increased across all indices. Results for the third subject fluctuated, eventually ending in almost the exact same position she was in in the beginning. The last student showed mixed results, decreasing the range of bigrams produced, but increasing her use of strongly associated bigrams. In the previously discussed study of bigrams in ESL texts, Bestgen and Granger (2014) examined the development of bigram production by ESL learners over the course of one

semester of study. While their cross-sectional analysis indicated a strong positive relationship between high MI score bigrams and essay quality scores, there was no significant change in this measure over time. Average t-scores were the only indices to show any significant change during the semester, decreasing from the first essay to the last. However, the effect sizes for these measures were found to be quite small, indicating that the statistically significant changes in average t-score were still not that strong.

In addition to focusing on L2 writing, several studies have looked at the longitudinal development of phraseological unit production in L2 speech. Crossley and Salsbury (2011) focused on the development of bigram accuracy in the spoken output of six ESL learners over the course of a year. Correlations between bigram frequencies in the L2 interviews and a spoken L1 reference corpus indicated that, while individual differences existed, learners became more accurate in their bigram use. Specifically, four learners showed increasing bigram accuracy, while two learners showed static patterns of bigram production. Significant correlations were also found between TOEFL scores and time learning English, indicating that improvements in spoken bigram accuracy coincided with increases in language proficiency. Garner and Crossley (in press) examined development in phraseological unit production using multiple indices of bigram and trigram use (frequency, association strength, proportion). Their data consisted of transcripts of naturalistic conversations between L1 and L2 conversation partners occurring once a month for four months. The L2 speakers ranged in proficiency from high-beginner to advanced. Using an approach to longitudinal data analysis known as Latent Curve Modeling (LCM; Bollen & Curran, 2006), the authors found that L2 speakers produced a greater proportion of targetlike bigrams and trigrams over time, with the less proficient speakers showing the greatest amount of growth in bigram proportions. Their results also showed that L2

speakers whose conversation partner was another L2 speaker experienced greater growth in their use of frequent bigrams from a spoken reference corpus.

2.4   Limitations of Current Approaches

Each of the above-reviewed approaches to the study of L2 phraseological unit production has provided valuable information concerning its development across proficiency levels and time. However, each approach also suffers from limitations that negatively affect the strength of the results of studies utilizing them. Concerning phraseological studies, a reliance on native speaker judgments can cause several issues. First, some phraseological units may not be identified as units relevant to the analysis because they may not be immediately salient to native speakers, possibly causing a large amount of phraseological units to be overlooked. For instance, word combinations such as *of course*, *on the other hand,* and *as a result* have a clear meaning or discourse function that is immediately recognizable to native speakers. For other highly frequent combinations such as *in the middle of, a number of,* and *the nature of*, their meaning or discourse function may not be immediately discernable, potentially leading them to be overlooked. Second, the subjective nature of native speaker judgments of phraseological unit correctness or nativelikeness makes the use of such judgments problematic. Some units may be considered unidiomatic by one group of native speakers but idiomatic and correct by another group. One native speaker may claim that an unfamiliar collocation is infrequent when in fact it is frequent in language.

N-gram and lexical bundle analyses avoid these problems to an extent because they allow researchers to identify and analyze a wide variety of phraseological units that may not be easily identified by native speakers (De Cock, 2004). However, they also have their own drawbacks. In regards to n-gram analysis, some word combinations occur frequently in speech or writing not

because they are phraseological units, but rather because their component words are highly frequent. The bigram *at the* is a highly frequent word combination, but it would not be recognized as a phraseological unit because it lacks its own meaning or discourse function. Lexical bundles analysis addresses this limitation to an extent by imposing minimum range and frequency thresholds on word sequences for lexical bundle status and by focusing on longer sequences (e.g. four- and five-word sequences). However, a dependency on frequency thresholds for phraseological unit status runs the risk of ignoring a lot of possibly interesting units. For example, in Vidakovic and Barker (2010), a four-word sequence had to occur at least 40 times per million words to be considered a lexical bundle and included in subsequent analyses. In setting such a high threshold, the authors might have excluded phraseological units that still occurred frequently, just not *that* frequently, possibly impacting their findings. For instance, discourse organizing bundles, which were found to be absent at lower proficiency levels, could have occurred repeatedly at those levels, but not frequently enough to meet the minimum threshold for lexical bundle status.

A second critique of n-gram and lexical bundle analyses is that they often treat learner corpora as whole texts rather than a collection of individual texts. In many n-gram and lexical bundle studies, type and token frequencies as well as type-token ratios are calculated for an entire corpus or subcorpus of learner texts. This approach is warranted in that it allows for the identification of phraseological units frequently produced by large groups of learners rather than a single learner. However, this approach makes any investigation of individual variation in phraseological unit production impossible (O'Donnell et al., 2013). By calculating type and token frequencies and type-token ratios for an entire subcorpus rather than for individual texts, researchers miss the opportunity to investigate how learners at the same proficiency level may

differ from one another and how individual variation in phraseological unit production may change across proficiency levels. It also disallows the use of statistical procedures that could provide more detailed information about significant differences in phraseological unit production across proficiency levels. O'Donnell et al. (2013) showed how this issue can be resolved by using a sampling procedure that produces subsamples within a corpus. Nevertheless, their subsamples still contained texts from multiple writers, masking the individual differences of those writers.

Lastly, association measure analyses have several limitations related to the most widely used measures themselves as well as the approach as a whole. First, as previously stated, t-score tends to favor word combinations composed of high-frequency words. It is also heavily dependent on corpus size, making t-score comparisons difficult across corpora that differ in length (Gablasova, Brezina, & McEnery, 2017). Another issue with the use of t-score as an association measure is that t-score is based on an assumption of normally distributed data, an assumption very rarely met with corpus data (Evert, 2009; Stefanowitsch & Gries, 2003). Regarding MI score, as previously noted, this association measure tends to highlight word combinations composed of low frequency words. However, among combinations of low frequency words, it tends to highlight those that occur less frequently. This means that two collocations, both comprised of low frequency words that occur almost exclusively in these collocations, may receive different MI scores simply because one is less frequent than the other (Gablasova et al., 2017). To use an example from Gablasova et al. (2017), the word combinations *ceteris paribus* and *jampa ngodrup* in the BNC would receive different MI scores, even though the constituent words of each occur exclusively in those combinations. The difference in MI score is, therefore, solely a result of *ceteris paribus* occurring more frequently

than *jampa ngodrup*. Thus, MI may not be a reliable measure of association between low-frequency words. A solution that has been proposed to address this issue is the use of MI^2, a variation of MI that counteracts the bias towards low-frequency combinations (Evert, 2009). However, this association measure has, so far, not been used in L2 phraseology research (Gablasova et al., 2017).

In addition to their individual limitations, both MI and t-score are limited in that they do not account for the directionality of word combinations. Similar to Gries' (2013) example with the bigram *of course*, most speakers would recognize that the bigrams *result of* and *of result* are two separate bigrams with different strengths of association. In the former bigram, *result* is much more likely to prime the use of *of*, while the same could not be said for *of* priming the use of *result*. However, because MI and t-scores are bidirectional measures of association, this difference would remain unaccounted for, resulting in both bigrams receiving the same association strength score. A possible solution for this issue is the use of ΔP, a measure from associative learning research that accounts for the directionality of word combination (Ellis, 2006). ΔP measures the probability of one word in a collocation given the occurrence of the other word in the bigram (e.g. the probability of *of* given the occurrence of *result*). Gries (2013) has shown how this measure can measure the directional associations of bigrams identified by more traditional bidirectional measures. In addition, two of the above reviewed studies (Garner, Crossley, & Kyle, in press, under review) showed how this measure outperforms other association measures in predicting human judgments of L2 writing proficiency.

Arguably the most significant drawback to association measure analyses as a whole is their treatment of part of speech information in association strength calculation. Some studies, such as Bestgen and Granger (2014), calculate average MI and t-scores for all bigrams in a

learner text regardless of their structural (e.g. part-of-speech, grammatical) characteristics. As a result, information regarding the effects of grammar on the development of collocation association strength is overlooked. Other studies, such as Durrant and Schmitt (2009) and Granger and Bestgen (2014) acknowledge the influence of grammar on bigram and collocation production by analyzing different structural types of bigrams (e.g. adjective-noun, adverb-adjective). Paquot (2017) also acknowledges the influence of grammar by analyzing direct object (verb-noun), adjectival (adjective-noun), and adverbial (adverb-adjective, adverb-adverb, adverb-verb) collocations separately. However, part of speech was not taken into account in these studies when MI and t-scores were calculated. These measures calculate association strength using expected frequencies based on the chance of two words co-occurring at random given the total number of words in the corpus. All words in a corpus are thus given the same chance to co-occur. This is in spite of the fact that a word's grammatical characteristics would probably influence their chances of co-occurring with certain words. For example, adjectives, because they modify referents in nominal expressions, are much more likely to occur before nouns than before verbs. Because MI and t-score do not account for this, they may not accurately reflect the true nature of syntagmatic word associations in the speaker's mental lexicon.

Furthermore, Hoey (2005), in his theory of lexical priming, claims that words are not only primed to co-occur with certain words, but also may be primed to occur in certain grammatical structures, a phenomenon known as lexical colligation. For instance, some adjectives may be positively primed to occur before nouns in an attributive position (e.g. *cold coffee*) and negatively primed to occur in a predicative position (e.g. *The coffee is cold.*), making it much more likely to occur in attributive adjective-noun collocations. Alternatively, some adjectives may have the opposite priming, making them less likely to occur in attributive

adjective-noun collocations. Biber et al. (1999), for example, demonstrated that adjectives such as *industrial, local,* and *social* show a very strong preference for attributive position, while over 95% of the occurrences of *grateful, impossible,* and *responsible* are in a predicative position. Collostructional analyses have also tended to support this notion of colligational priming. These analyses have shown that words can be strongly associated with certain grammatical constructions (e.g. *give* and the ditransitive construction [*John gave Mary a book*]) and can be negatively associated with other semantically similar constructions (e.g. *give* and the prepositional dative [*John gave a book to Mary*]) (Stefanowitsch & Gries, 2009). This research has also shown that words can be strongly associated with each other within constructions (e.g. *ask* and *question* in the ditransitive [*John asked Mary a question*]). Thus, in order to accurately reflect the true nature of syntagmatic word associations in the mental lexicon of L2 speakers, it is important that future research take into account the influence of syntactic information on association strength, something that so far research involving MI and t-score has failed to do.

2.5   Summary

This chapter has reviewed the current state of learner corpus phraseology research. This review included an overview of the different approaches researchers have taken towards identifying and analyzing phraseological units in learner corpora and some of the major findings from studies utilizing these approaches. It has shown that, while acquiring a strong productive phraseological knowledge is important for L2 learners, these learners often fail to attain nativelike proficiency in the use of a wide range of phraseological units. This research has shown that, even at advanced proficiency levels, learners produce a smaller range of phraseological units compared to their native speaking counterparts. It has also been shown that advanced L2 learners have difficulty in using phraseological units that are appropriate for the given

communicative context. For instance, advanced L2 writers have been shown to overuse units characteristic of spoken discourse and underuse academic units in their academic writing. L2 speakers have also been found to overuse emphatic units while underusing vagueness units. Cross-sectional research has shown that, as learners become more proficient in their overall language use, they produce a wider range of phraseological units and, in some cases, produce them more frequently. They also tend to produce units that are more frequent in native speech and writing and more strongly associated collocations comprised of low-frequency words. Longitudinal research, however, has provided mixed results, showing that phraseological development for individual learners often occurs nonlinearly and varies according to the type of analysis (e.g. phraseological, n-gram, lexical bundle, association strength).

The chapter ended with a brief discussion of the limitations of current approaches in L2 phraseology research. Phraseological research, by using native-speaker judgments in phraseological unit identification and analysis, runs the risk of overlooking frequently occurring units that are not immediately salient. N-gram analyses may misidentify word sequences as phraseological units simply because they are highly frequent due to the high frequency of the component words. Lexical bundle analyses, on the other hand, risk overlooking important phraseological units that may not occur above the researcher's arbitrarily chosen frequency or range thresholds. Lastly, association measure analysis utilizing t-score and MI scores are limited in that these measures tend to favor certain items over others. They also fail to consider the importance of directionality in word associations. Most importantly, though, these studies fail to consider the syntactic characteristics of component words in a phraseological unit, thus providing an imperfect measure of the syntagmatic word associations in an L2 learner's mental lexicon.

 In the next chapter, I will present a definition of lexical collocations, a type of phraseological unit, as a construction within the framework of Construction Grammar. In doing so, I hope to address some of the limitations of current approaches to L2 phraseological unit analysis.

3   COLLOCATIONS AS CONSTRUCTIONS

In order to address some of the previously discussed limitations of current methods in L2

phraseological unit research, this chapter proposes a new approach to the definition and analysis

of collocations. Specifically, it argues for a definition of collocations as a type of linguistic

construction within the framework of Construction Grammar (Goldberg, 2006). Although the

current dissertation predominantly focuses on lexical collocations as defined by Granger and

Paquot (2008), it is believed that this argument can serve as the basis for the inclusion of other

types of phraseological units (e.g. grammatical collocations, phrase-frames, complex

prepositions, etc.) in Construction Grammar research.

This chapter is divided into three sections. The first section will present the theoretical

basis of Construction Grammar. The second section will discuss the status of constructions in L2

learner language and a review of some of the empirical research examining L2 construction

acquisition from a usage-based SLA perspective. The third section will outline the defining

characteristics of lexical collocations and show how lexical collocations can be defined as

constructions. This section will also show how this definition can improve upon current practices

in L2 collocation analysis.

3.1   Construction Grammar

Constructions are form-meaning pairings that become entrenched in the minds of fluent

language users through their experiences with language (Goldberg, 2006). They exist at all levels

of granularity and range from simple to complex. They can be as small as morphemes (*un-, pre-,
-able*) and single words (*apple, car, person*) and as large as multi-word idiomatic expressions

(*Too many cooks spoil the broth*). More complex constructions include partially filled idiomatic

expressions (*the* <adjective>, *the better*)*,* phrasal constructions ([P NP]), and abstract syntactic

frames (ditransitive construction: Subj V Obj_1 Obj_2). Even low-frequency or unusual patterns, such as prepositional phrases with bare count nouns (*in prison, for work, on vacation*) can be considered constructions (Goldberg, 2013). In general, for a linguistic pattern to be a construction, some aspect of its form or function cannot be predicted based on its component parts. For the preposition-bare count noun construction for example, its function of conveying the involvement of stereotypical activities associated with the noun is not predicted by the preposition or the noun. A pattern can also achieve constructional status if it occurs frequently enough to become represented in a fluent language user's linguistic system (Goldberg, 2006). According to constructionist approaches, utterances are built through the combination of constructions at all levels of size and complexity. Thus, the sentence *Archer poured Lana a drink* is composed of the constructions *Archer, Lana, pour, a, drink*, the past tense *–ed* ending, three noun phrase constructions, a verb phrase construction, and a benefactive ditransitive construction.

Goldberg (1995) claims that a fluent language user's inventory of constructions, known as their constructicon, exists as an interlinked network with formally and semantically similar constructions connected through three main types of inheritance relations: instance links, polysemy links, and subpart links. The first two types of inheritance relations are hierarchical, with higher level, more general and abstract constructions being connected to more specific constructions at lower levels. Instance links connect abstract constructions with special cases of the construction. For example, the preposition-bare count noun construction discussed earlier is a special case of the general prepositional phrase construction, inheriting its word order from the prepositional phrase. It differs from the more general construction in that it specifies an unmodified noun rather than a noun phrase as its daughter. In terms of function, it differs in that

it refers to the stereotypical activity related to the noun rather than identifying location. Phrasal verbs are another example of constructions connected through inheritance links. According to Gilquin (2015), phrasal verbs exist in a three-level hierarchy. At the highest level is the superconstruction phrasal verb [PV]. At the second level are three different structural instantiations of the [PV] construction: [V Prt], [V Prt OBJ], and [V OBJ Prt]. All three inherit the general characteristics of phrasal verbs, but differ in how they are realized. The lowest level of the hierarchy consists of partially lexically filled constructions with the verb and particle specified and the object allowed to vary. These lower level constructions inherit their structural features from constructions at the second level, but differ in their verb and particle constituents.

The second type of inheritance relation is the polysemy link, which connects a particular sense of a construction to any extensions of this sense within the same construction (Goldberg, 1995). For instance, the ditransitive construction has the core sense of "agent causes patient to receive theme". Extensions of this core sense include instantiations of the ditransitive that express intentionality ("agent intends to cause patient to receive theme"), enablement ("agent permits patient to receive theme"), and resistance ("agent causes patient to not receive theme").

Lastly, subpart links are non-hierarchical in that they connect constructions at the same level of abstraction that share formal or semantic characteristics (Hilpert, 2014). For instance, the caused motion construction (*Kirk pushed Spock into the room*) and the intransitive motion construction (*Spock walked into the room*) share the same semantic characteristic of a person or object moving towards a goal. They also share the structural characteristics of movement being expressed by the verb and goal being expressed by an oblique.

Experimental research has provided evidence for claims that a native speaker's linguistic system is organized as a collection of interconnected constructions. Specifically, research has

shown that when determining sentence meaning, native speakers are more likely to rely on information provided by the construction than the verb. In Bencini & Goldberg (2000), participants were presented with a sentence sorting task in which they were asked to sort sixteen sentences into four different piles based on overall meaning. The sixteen sentences in the task were created using four different verbs and four different verb-argument constructions (VACs), with each construction being paired with each verb. Results showed that just under half of the participants sorted exclusively according to constructions, while no participants sorted exclusively based on verbs. For the participants that produced mixed sorts, it was found that the sorts more closely resembled constructional sorts than verb-based sorts. Research has also shown that native speakers' knowledge of VACs is sensitive to their usage patterns. Ellis, O'Donnell, and Römer (2014) presented a large group of native English speakers with target VAC frames (e.g. s/*he* ___ *across the…, it* ___ *under the…*). In the first experiment, participants were asked to produce the first word that came to mind. In the second, they were asked to produce as many verbs as they could think of in a minute. Results from both experiments revealed that how frequently a verb type was produced in the VACs was determined by the frequency of verbs in the VACs in a large reference corpus of native language production, the contingency between verb types and the VAC, and semantic prototypicality for the VAC.

Concerning language acquisition, Construction Grammar perspectives view this process as an inherently usage-based, with linguistic knowledge being built up over time and through experience with the language (Ellis, 2006a). In doing so, it rejects the Generative Grammar concept of abstract grammars built through the use of language-specific cognitive mechanisms (Tribushina & Gillis, 2017). Rather, children learn their first language using general cognitive mechanisms that are applied to all types of learning (Ellis, 2006a). One of these mechanisms is

the ability to notice frequencies in the language they encounter and generalize from these language experiences. According to Ellis (2002, 2006a), language learners are intuitive statisticians with language learning being driven by frequency and contingency between form and function. As learners repeatedly encounter the same construction in the input, this construction comes to be represented in their linguistic systems. Exposure to different variants of the construction can also help learners generalize an abstract construction that encompasses all the variants. While raw frequency appears to be important for construction acquisition, form-function contingency for constructions plays an equally, if not more important, role. The more reliably the form of a construction matches with its communicative function (i.e. meaning), the more likely it is that the construction will be acquired (Ellis, 2006a).

Other factors besides frequency and contingency can also influence the acquisition of constructions by language learners. These factors include the saliency, conceptual complexity, and functional load of constructions. Saliency refers to how strong a stimulus (in this case, a construction) is in the input (Ellis, 2006a), with more salient constructions being more likely to be noticed and acquired by children. For example, in the sentence *John walked to the store yesterday*, the time adverbial *yesterday* is a more salient marker of past time than the *–ed* ending. Thus, *yesterday* will most likely be noticed, and hence acquired, earlier than the past tense morpheme. Conceptual complexity has also been demonstrated to influence first language acquisition. For instance, Tribushinina (2013) found that more complex spatial adjectives (e.g. *dik* "thick/fat") in Dutch are acquired later by Dutch children. Lastly, functional load refers to how frequently a language element is used to make meaningful distinctions in language. Research has shown that consonant phonemes with greater functional load (i.e. used more often

to make meaning distinctions) are acquired earlier by English-speaking children (Stokes & Surendran, 2005).

3.2 Second Language Constructions

Research into second language constructional knowledge has also provided evidence that knowledge of an L2 includes knowledge of L2 constructions. Gries and Wulff (2005) replicated the study by Bencini and Goldberg (2000) with German EFL learners. These learners were presented with sixteen sentences composed of four different verbs crossed with four different constructions and were asked to sort them based on overall meaning. Similar to the native speakers in Bencini and Goldberg's study, the German learners tended to sort the sentences according to the construction that was used to create the sentence. In a follow-up study, Gries and Wulff (2009) again tested the constructional knowledge of German EFL learners. Participants were shown a sentence consisting of either a *to* construction (e.g. *to cross a road*) or an *–ing* construction (e.g. *talking to his mother*). These constructions were shown immediately following verbs that were distinctive for either construction (as determined by a distinctive collexeme analysis). After reading each sentence, participants were instructed to rate its acceptability as a sentence. These sentences also served as primes for target sentence fragments that required students to provide either a *to* or an *–ing* construction after the verb. The results of both the acceptability ratings and sentence completions showed strong evidence for learners' constructional knowledge. Participants were more likely to rate a construction as acceptable if it accompanied a verb that preferred the construction than a verb that did not. Similarly, when the verb in the target sentence fragment was distinctive for a construction, the sentence was more likely to be completed with that construction. Ellis, O'Donnell, and Römer (2014b) used another free association task to investigate the constructional knowledge of advanced Czech, German,

and Spanish EFL learners and L1 speakers of English. Similar to their study involving only L1

participants (Ellis, O'Donnell, & Römer, 2014a), participants were shown a series of VAC

frames and asked to fill in the blank with the first word that came to mind. Results showed that,

for all four participant groups, verb generation in the constructions was predicted by

contingency, frequency, and semantic prototypicality of verbs in VACs in usage.

Additionally, research has investigated the nature of L2 constructional knowledge, how it

differs from L1 knowledge, and its development across proficiency levels and over time. In

Wulff and Gries (2011), the authors present two case studies of constructions and their

associations in L1 and L2 writing using distinctive collexeme analysis. The first case study

focused on the production of the ditransitive (e.g. *Ben gave Leslie a waffle*) and the prepositional

dative (e.g. *Ben gave a waffle to Leslie*) constructions. Their results indicated that advanced L1

German and L1 Dutch learners recognized differences in verb preferences for the two

constructions, forming stronger generalizations in the case of the prepositional dative compared

to L1 writers. In the second case study, it was found that, although learners seemed to be aware

of verb preferences for infinitival (e.g. *Vida began to chase the squirrels*) and gerundial (e.g.

Vida began chasing the squirrels) complementation constructions, they still made some non-

native like verb choices. In a follow-up study, Martinez-Garcia and Wulff (2012) compared the

production data for L1 German and native speakers in Wulff and Gries (2011) to production data

for L1 Spanish learners. Results showed that Spanish learners were aware of the semantic

preferences for the gerundial and infinitival constructions, but differed from the L1 German and

native speakers in terms of their preferred verbs. In comparison, the L1 German learners were

more attuned to the specific verbs that occupy the target constructions. Taken together, both

studies demonstrate that inaccuracy in construction use by L2 speakers is not due to incorrect verb choices, but rather unidiomatic or nonnative-like choices.

Findings from studies in Usage-based Second Language Acquisition (UBSLA) have demonstrated that constructions emerge in a learner's interlanguage through a process of increasing schematicity driven by affordances in the immediate communicative context. In a series of studies, Eskildsen (2009, 2012, 2015) analyzed the development of negation constructions, WH- and yes/no-question constructions, and *can*-constructions for two L1 Mexican Spanish learners in a corpus of classroom interactions. His analyses showed that, for all four constructions, the learners' inventories became more schematic, dynamic, and targetlike over several years of study. In most cases, this development was driven by the use of one fixed exemplar (e.g. *I don't know*) that became more schematic with continued use in different situational contexts. His findings also indicated that these changes were brought about by changes in the learner's communicative context and interactive needs. Mellow (2006), in a study of relative clause development in the interlanguage of one young L1 Spanish learner, found that functional needs drove the emergence of some relative clauses types. For instance, the learner's need to clearly distinguish between two referents in a story led her to start producing relative clauses to modify subjects. Myles, Hooper, and Mitchell (1998) found that, over the course of two years, L2 French learners were able to break down three first person pronoun-verb formulaic sequences (*j'aime, j'adore, j'habite*) and use them more schematically. Analysis of spoken transcripts revealed that this breaking down of sequences occurred as the learners moved from classroom tasks about themselves to discussions about third parties.

In two studies, Ellis and Ferreira-Junior (2009a, 2009b) demonstrated that the nature of constructional input that learners receive also influences their learning of L2 constructions. The

data for both studies consisted of transcripts of spoken conversations occurring every four to six weeks over 30 months between native-speaking interviewers and seven ESL learners living in England. In the first study, the authors examined the influence of verb frequency, verb distribution, and semantic prototypicality on the acquisition of three VAC types (movement VL [verb locative], caused motion VOL [verb object locative], double-object ditransitive VOO). Results indicated that, in both L1 and L2 speech, verb distributions for each VAC were Zipfian. A significant correlation was also found between the rank order of verb types in the VACs in the L1 and L2 data sets. The most frequent verbs in each VAC were also found to occur first in the L2 data, serving as the pathbreaking verb. Development in L2 VAC production occurred as the learners began to employ a more diverse range of verbs in the VACs. Results also indicated that verbs serving as pathbreaking verbs for each VAC were prototypical in meaning for the constructions as well as general in their semantics. In the second study, Ellis & Ferreira-Junior (2009b) extended their analysis to focus on other slots in the VACs (e.g. subject, object). In addition, the authors included distinctive collexeme analysis and contingency analysis for verbs in the VAC. Results for the verb analysis showed that the learners' VAC acquisition was generally driven by the most frequent, prototypical, and generic verb type in the VACs. It was also seen that the acquisition of verbs in the VACs was strongly predicted by collexeme strength and contingency between the verb and the VAC. Results were similar for the other slots in the VACs. The distribution of items in each slot were Zipfian, with the most frequent items in the NS data being the first acquired and most frequently used by the NNS. It was also found that the most frequent items in each slot were prototypical of the slot's function.

Taken together, this research has demonstrated how L2 learners' constructional knowledge emerges through their continued interaction with language. Early on in their language

development, learners rely on a small set of fixed constructional exemplars that are highly frequent in language and are general in their semantics. Over time, they begin to break down these fixed constructions and use them more schematically with a wider range of lexemes occupying different slots within the constructions. This movement towards greater constructional schematicity is driven by a combination of greater amounts of input and more complex communicative demands. With greater amounts of input, learners encounter a greater range of items occupying slots within the construction and build categories of related items they can use in each slot. More complex communicative demands push learners to use their constructional inventory in new ways to express a wider range of meanings.

3.3   Lexical Collocations as Constructions

Lexical collocations are, according to the classification scheme in Granger and Paquot (2008), a subtype of referential phrasemes. In this scheme, lexical collocations are defined as preferred relations between lexemes that co-occur frequently in specific syntactic patterns, such as adjective-noun (*heavy rain*), verb-object noun (*reach an agreement*), and adverb-adjective (*statistically significant*). These preferred relations between lexemes are usage-determined, meaning that their representation in the minds of fluent language users is based on mechanisms such as frequency and contingency. The more two lexemes co-occur in a syntactic pattern and the more contingent the occurrence of one lexeme is on the occurrence of the other, the more likely they are to be represented as lexical collocations (Ellis, 2002). In terms of semantics, both lexemes contribute to the overall meaning of the collocation; however, they differ in how much they contribute. According to Granger and Paquot (2008), the base of the collocation is selected first by the language user for its independent meaning. It is the main referent of the collocation. The second element, known as the collocator, serves to enhance the meaning of the base or focus

attention to a characteristic of it. It is selected by the first lexeme and is semantically dependent on it. In the collocation *black coffee,* the base is *coffee* and the collocator is *black*. While *coffee* refers to its usual meaning of the hot drink made from roasted coffee beans, the collocator *black* does not refer to its normal meaning but instead indicates the absence of milk products.

As referential phrasemes, lexical collocations are used to convey content messages by referring to objects, facts, and other phenomena. According to Mel'čuk (1998, 2007), they do so by expressing lexical functions. Lexical functions refer to general abstract meanings that can be lexically expressed in a variety of ways depending on the lexical unit to which the meaning applies. Concerning lexical collocations, these items often express nominal, adjectival, adverbial, or verbal lexical functions. For instance, adjective-noun collocations can, among others, express the functions of **Magn** or "intensification" (*infinite patience, close shave*), **Ver**, which corresponds to genuineness or the meaning of "as it should be" (*genuine surprise*, *precise instrument*), or **Bon,** which corresponds to positive evaluation (*tempting proposal*, *valuable aid*). In regards to verb-noun collocations, two of the most common functions fulfilled by these collocations include **Oper** ("support"), in which the verb supports the meaning of the noun (e.g. *make an effort, take measures*), and **Real** ("fulfillment"), in which the verb expresses the meaning "do what one is supposed to do with the noun" (e.g. *watch television, drink champagne*).

Based on this definition, it is clear that lexical collocations meet the basic requirements laid out by Goldberg (2006, 2013) for being considered constructions. They can be considered form-function pairs because lexical collocations of all structural types (e.g. adjective-noun, verb-object noun) fulfill various lexical functions, such as intensification, support, and fulfillment. While exact estimates vary, research has shown that collocations and other phraseological units

are pervasive in language (Römer, 2009). As such, they most likely occur frequently enough to become represented in the mental lexicons of fluent language users. They are also unpredictable in form and function. For instance, although it is theoretically possible for all intensifying adverbs to occur with all adjectives, research has shown that this is not the case (Kennedy, 2003). Lexical collocations also exist in inheritance networks in which they share syntactic and semantic links with more general and abstract phrases. For example, adjective-noun collocations exist within the general attributive adjective construction (e.g. *the large dog, a green car*), which, according to Hilpert (2014), is an instance of the more general head-complement construction (e.g. Adj-N, Adv-Adj). They share an instance link with the general attributive adjective construction in that they inherit their word order and semantic characteristics from the general attributive adjective construction, but differ from it in that they are not fully productive. While it is possible for *bottomless* to be used to intensify *patience*, *infinite* is the preferred collocator. In addition to sharing links with the abstract attributive adjective construction, adjective-noun collocations also share a subpart link with adverb-adjective collocations due to the fact that they can fulfill similar functions (e.g. intensifying the meaning of the base).

Defining lexical collocations as constructions can provide several benefits for L2 phraseological unit analysis using learner corpus data. First, this definition allows for a more comprehensive approach to analyzing collocation production that combines the analysis of collocations as lexical functions (Mel'čuk, 1998, 2007) and co-varying collexeme analysis (Stefanowitsch & Gries, 2005). Co-varying collexeme analysis is a type of collostructional analysis that assesses the association strength between the collocating words (e.g. *infinite* and *patience*) within a specific grammatical construction (e.g. the attribute adjective construction). By combining co-varying collexeme analysis with an analysis of the lexical functions of

collocations, this new approach to analyzing L2 collocation production is comprehensive in that it accounts for both the functional and lexicogrammatical aspects of collocations. For instance, studies investigating collocations within a specific grammatical construction (e.g. collocations in the attributive adjective-noun construction) could examine the production of collocations across different functional subcategories in the grammatical construction. In doing so, these studies can better account for the influence of lexical functions on collocation use. Alternatively, studies focusing on collocation production within a specific functional collocation category (e.g. support verb-noun collocations) could, in addition to analyzing their frequency or diversity, examine the association strength for collocations within the category using co-varying collexeme analysis. The use of co-varying collexeme analysis would allow these studies to account for the influence of associations between the collocating lexemes and the grammatical construction they co-occur in.

This new definition of lexical collocations as constructions also has the benefit of connecting collocation research with UBSLA. As previously noted, findings from UBSLA research have shown that a learner's inventory of L2 constructions emerges as learners move from relying on fixed exemplars that are general in their semantics to using constructions more schematically with a wider range of lexemes (Ellis & Ferreira-Junior, 2009a, 2009b; Eskildsen, 2009, 2012, 2015; Mellow, 2006; Myles et al., 1998). In defining lexical collocations as constructions, researchers could examine the possibility that collocations also develop along this trajectory. It may be the case that learners initially express specific lexical functions for a word with only one or two collocates that are semantically general and can be used with a wide variety of words. With more exposure to language, learners may develop their ability to express this

lexical function with a wider range of collocates that are less general in their semantics and more strongly associated with the target lexeme.

In addition, defining and analyzing lexical collocations as constructions can address some of the limitations of current approaches to phraseology (discussed in Section 2.4), such as an overreliance on native speaker intuition and lack of syntactic information in association strength calculation. First, collocations would no longer be identified based on native speaker intuitions or dictionary information; instead, they would be automatically extracted along with all other instances of their constructions. Their collocational status would be determined not by subjective judgments, but rather on the basis of their co-occurrence information. While this is similar to n-gram and association measure analyses, it also improves on these approaches. In regards to n-gram analysis, co-varying collexeme analyses do not necessarily have frequency thresholds for designation as an important item. Rather, all collexemes can be included in the analysis. In comparison to most association measure approaches, co-varying collexeme analysis of collocations would include syntactic information in the determination of association strength. A co-varying collexeme analysis of adjective-noun collocations would focus on the co-occurrence of adjectives and nouns in the [ADJ N] construction irrespective of their frequencies outside of the construction. As previously alluded to, this would remove a lot of the noise in statistical analysis and provide a truer measure of association between words in the mental lexicon. Additionally, this approach would more closely capture how language users, both native and nonnative, go about processing and producing lexical collocations.

3.4   Summary

This chapter has presented a working definition of lexical collocations as constructions (henceforth "collconstructions") within a Construction Grammar framework. It has done so by

showing how lexical collocations, a type of frequently occurring referential phraseme, meet all the criteria for being considered constructions, that is form-function mappings that exist within inheritance networks in the minds of fluent language users. It has also shown how, by defining lexical collocations as constructions, researchers can account for the structural and functional characteristics of collocations when analyzing how learners acquire productive collocational knowledge. This not only helps overcome some of the previously mentioned limitations of current approaches to L2 phraseological unit analysis, but also provides a more comprehensive analysis of L2 collocation acquisition. Specifically, by including grammatical information in the analysis of association strength between collocates, researchers can better investigate the interaction between grammar and collocational knowledge. In addition, the inclusion of functional subcategories allows for the analysis of how the functional characteristics of collconstructions influence their use. Lastly, viewing lexical collocations as constructions can ground future learner corpus studies of phraseological unit production in the framework of UBSLA, helping provide a more theoretically sound background for investigating development.

4   DATA AND METHODS

Phraseological, n-gram, lexical bundle, and association strength analyses of L2 writing have demonstrated how productive phraseological knowledge develops both across proficiency levels and over time. However, these approaches towards phraseological unit analysis often fail to consider how the grammatical characteristics of the component words influence their association. The current dissertation aims to address this limitation by investigating the cross-sectional variation in collconstruction production and the longitudinal development of collconstruction production in L2 writing. Focal collconstructions included verb-noun, adjective-noun, and adverb-adjective combinations. Contiguous and non-contiguous combinations were extracted from the learner texts and used to calculate multiple constructional and collostructional indices. Constructional indices included the calculation of type and token frequencies and normalized entropy for each of the target collconstructions, while collostructional indices consisted of proportion scores for different categories of collconstructions based on covarying collexeme scores. Cross-sectional data was collected from a corpus of essays written by beginner and intermediate L1 Korean EFL learners that were assessed for overall language proficiency. Longitudinal data was collected from a corpus of freewrites by six ESL learners (their L1s include Korean, Arabic, Spanish, and Japanese) over the course of one year of intensive English study. Variation across proficiency levels was assessed qualitatively by examining the production of different functional collconstruction subcategories and quantitatively using an ordinal logistic regression model that determined what indices of collconstruction production for the three structural categories were predictive of language proficiency. Longitudinal development for the six learners was assessed using linear mixed-effects models that allow for the estimation of growth over time. Longitudinal development for individual learners was

investigated through graphical methods that allow for the analysis of individual developmental trends, intraindividual variation in development, and interconnected development between indices of collconstruction production. As indicated at the end of Chapter 1, the research questions that drive this dissertation are:

1. How does the use of collconstructions in L2 writing vary across proficiency levels?

2. How does the use of collconstructions in L2 writing develop over time?

3. How does the use of collconstructions in L2 writing develop in individual learners over time?

4.1   Learner Corpora

4.1.1   ICNALE

The cross-sectional study in this dissertation makes use of a sample of written texts from the International Corpus Network of Asian Learners of English (ICNALE, Ishikawa, 2013). ICNALE comprises 10,000 spoken and written productions by L1 and L2 speakers of English. Of these, 5,200 are written texts produced by 2,600 university-age EFL learners from 10 Asian countries. Each learner produced two texts in which they were asked to agree or disagree with the following statements, "It is important for college students to have a part-time job" and "Smoking should be completely banned at all the restaurants in the country". These texts were written on a computer at home or in class. Learners were given 40 minutes to complete each essay and were instructed to not use a dictionary or other resources (Ishikawa, 2013). In addition to the written texts, learners were asked to complete questionnaires designed to collect basic demographic information (age, sex, country, university major, year of study), scores on English proficiency tests (e.g. TOEIC, TOEFL, IELTS, TEPS), information regarding their integrative and instrumental motivations for learning English, and information regarding their experiences

learning English (Ishikawa, 2013). Learners were also required to take a vocabulary size test

(VST) that included 50 items from the 1,000 to 5,000 word levels of the VST monolingual

Version (Nation & Beglar, 2007). Based on either self-reported proficiency test scores or VST

scores that had been converted to TOEIC scores, learners in ICNALE were placed into four

proficiency levels (A2, B1_1, B1_2, B2) corresponding to the Common European Framework of

Reference for Languages (CEFR; Council of Europe, 2001). The sample of ICNALE texts

utilized in the current study comprises 300 texts (totaling 69,592 words) written by L1 Korean

EFL learners responding to the first prompt (i.e., "It is important for college students to have a

part-time job"). This sample was chosen because this L1 group is the most balanced across

proficiency levels, with no level accounting for more than 30% of total texts. These texts were

also on average longer than those responding to the other prompt. More information regarding

this corpus, including the number of texts, total word count, and words per text in each

proficiency level, is provided in Table 4.1.

Table 4.1 Overview of L1 Korean subsection of ICNALE

	A2	B1_1	B1_2	B2
Texts	75	61	88	76
Words	16,804	13,634	20,490	18,664
Words per text	224.05	223.51	232.84	245.58

While the texts in ICNALE are separated according to proficiency level, these ratings are

problematic because they are based on either test scores reported by the participants or on a

converted VST score. As such, they may not accurately reflect the true writing proficiency of the

participants. For this reason, it was decided that the texts would be rated for overall writing

proficiency by independent raters and re-classified into three proficiency levels according to the

CEFR. The essay rating procedure occurred in three stages, the first two of which served as

norming. In the first stage, the researcher met with four fellow PhD students who had

volunteered to serve as raters. Each of these volunteer raters indicated that they had at least one year experience of either teaching or assessing ESL writing. During this meeting, the researcher introduced the raters to the CEFR-based writing rubric to be used in the study (Council of Europe, 2003). This rubric asks raters to assign a holistic score to each text. This rubric has previously been used in Chen and Baker (2014) and is shown in Appendix A. After introducing the rubric, the researcher and the volunteer raters rated five sample texts from the L1 Chinese and Japanese sections of ICNALE. These texts were written in response to the same prompt as the texts in the dataset used in this study. Two of these texts were from the A2 proficiency level while there was one text each from the B1_1, B1_2, and B2 proficiency levels in ICNALE. For the first two texts, the researcher and volunteer raters read the text, discussed it, and assigned a CEFR score together. For the next three texts, the raters read the text and assigned a score individually. This was followed by a brief discussion of each text in which the researcher and raters discussed the reasoning behind their ratings. The second stage of the rating procedures consisted of two rounds of independent rating in which the researcher and raters rated 20 texts (total of 40 texts over the two rounds) from the L1 Chinese and Japanese subsections of ICNALE. These texts were written in response to the same prompt as the essays in the study dataset and were evenly distributed across the four proficiency levels in ICNALE. The purpose of this stage of the rating procedures was to give the raters more practice using the rubric to rate ICNALE texts.

Following the two norming stages of the rating procedures, the raters were each sent a set of 120 randomly assigned essays from the study dataset. They were instructed to assign each text a score from high beginner (A2) to upper-intermediate (B2) on the CEFR. The researcher also rated 120 essays from the study dataset. The texts were assigned to the raters in a manner that

ensured each essay received a rating from two raters. The researcher and raters rated the essays independently and did not discuss the essays with each other at any time during the rating process.

During the rating process, several issues with some of the learner texts were discovered. First, it was discovered that some of the texts were written with the assistance of Google Translate, even though the subjects were instructed not to use any outside electronic resources. These texts contained clear signs that Google Translate had been used, such as garbled sentences (e.g. "But once we get real downside were backed guests will come.") and words that Google was unable to translate. For instance, one of the Korean words for part-time job, 알바, was translated in the learner texts to "Alba". Second, two essays were found to be plagiarized from online sources. Lastly, one of the essays in the dataset was written in response to an entirely different prompt not included in ICNALE (the benefits of studying in groups versus studying alone). In total, 24 essays in the dataset were found to fall into these categories. These essays were removed from the dataset, resulting in a final dataset of 276 essays, totaling 62,271 words.

Once all essay scores were collected from the raters, statistical analyses were conducted in order to determine inter-rater reliability. Overall, the two main raters assigned the same rating for 242/276 essays, resulting in an inter-rater reliability score of 0.798. The 35 essays that were assigned different scores by the two main raters were sent to a third rater, who assigned a third score. Based on scores assigned through this rating procedure, essays in the dataset were divided into three separate subcorpora representing CEFR A2, B1, and B2 proficiency levels. The number of texts, total words, and words per text in each of the subcorpora are presented in Table 4.2.

Table 4.2 Overview of final cross-sectional learner corpus dataset (from ICNALE-Korean)

	A2	B1	B2	Total
Texts	106	121	49	276
Words	22,810	27,274	12,247	62,271
Words per text	215.19	225.40	246.35	225.62

4.1.2   Salsbury Written Corpus

The longitudinal analyses in the current study make use of the Salsbury written corpus (Salsbury, 2000). This corpus consists of 337 untimed and unstructured freewrites (63,700 words) written by six ESL learners enrolled in an intensive English program at a university in the United States. Freewrites were collected from learners every one to two weeks over the course of one year, with an average time between first and last text of 49.33 weeks. Participants chose to write on a wide variety of topics over the course of the year with topics including descriptions of daily life and discussions of controversial issues. In terms of L1 backgrounds, three participants were L1 Arabic speakers, one was an L1 Spanish speaker, another was L1 Korean, and one was an L1 Japanese speaker. In addition to freewrites, participants completed institutional TOEFL exams at six points over the course of the year. Table 4.3 provides more detailed information regarding this corpus, including the learner's L1s, genders, number of texts written, weeks between first and last text, total number of words written, and average number of words per text.

Table 4.3 Overview of longitudinal learner corpus data (Salsbury Corpus)

Name	L1	Gender	Number of Texts	Weeks Between First and Last Text	Total Number of Words	Words per Text
Faisal	Arabic	Male	39	49	6,305	161.67
Kamal	Arabic	Male	26	50	4,389	168.81
Jalil	Arabic	Female	43	47	10,400	241.86
Marta	Spanish	Female	53	50	11,574	218.38
EunHui	Korean	Female	89	50	13,072	146.88
Takako	Japanese	Female	87	50	17,960	206.44
Total			**337**		**63,700**	**189.02**

4.2   Reference Corpus

The reference corpus used in the current dissertation is the offline version of the academic subsection of the Corpus of Contemporary American English (COCA; Davies, 2009). This corpus will be referred to as COCA Academic throughout this dissertation. COCA Academic contains around 88 million words from academic journals published from 1990 to 2012. COCA Academic was chosen as the reference corpus for two main reasons. First, compared to the other subcorpora of COCA, the linguistic features within COCA Academic are more likely to approximate the types of features the ICNALE learners were attempting to employ in their writing tasks. Similarly, the ESL learners in the Salsbury Corpus were most likely working towards improving their academic writing abilities over a year of study in a university intensive English program, making an academic English corpus a suitable reference. The second reason for using COCA Academic as the reference corpus in this dissertation is that previous research has shown that bigram measures derived from COCA Academic are predictive of human judgments of writing quality, even at lower levels of writing proficiency (Garner, Crossley, & Kyle, forthcoming; Garner, Crossley, & Kyle, under review). This is in comparison to similar bigram measures derived from the other subsections of COCA, such as COCA Spoken.

4.3   Collconstruction Extraction

Collconstruction tokens were extracted from ICNALE-Korean and the Salsbury Written Corpus in a similar fashion. First, all texts were tagged for part-of-speech (POS) information using TagAnt (Anthony, 2015). TagAnt is a freeware POS tagger that uses TreeTagger (Schmid, 1994) to assign POS information to lexical items in a text. TagAnt was chosen as the tagger in this study due to its ease of use and because previous research has shown it to be relatively accurate with written L2 German and French data (Reznicek & Zinsmeister, 2013; Thouësny,

2011; Zinsmeister & Breckle, 2010). TreeTagger was also used in Yoon's (2016) study of verb-noun collocations in L1 and L2 writing. Using POS tags and regular expression searches in AntConc (Anthony, 2014), concordance lines for contiguous and non-contiguous adjective-noun, adverb-adjective, and verb-noun collconstructions were extracted from the learner texts. For adjective-noun collconstructions, a regular expression that extracts sequences of adjectives followed by zero to two words, followed by a noun was used (to extract collconstructions such as *good vibes, good musical vibes, good classical music vibes*). In the case of verb-noun collconstructions, a regular expression that extracts sequences of verbs followed by zero to four words, followed by a noun was used (to extract collconstructions such as *consider danger, consider the danger, consider the biggest danger, consider the very real danger, consider the biggest and scariest danger*). For adverb-adjective collconstructions, a regular expression that extracts adverbs immediately followed by an adjective was used (to extract collconstructions such as *extremely dangerous*). These spans were chosen based on pilot searches conducted on texts from ICNALE. The search spans listed above were found to be most efficient in finding most of the possible syntactic configurations of collconstructions occurring in learner writing.

Once concordance lines for all collconstruction tokens were retrieved, results were filtered manually in order to ensure that extracted sequences were true adjective-noun, verb-noun, and adverb-adjective collconstructions. For all three collconstruction categories, instances in which the tagger incorrectly tagged a word and instances in which the words co-occurred across phrasal or clausal boundaries (e.g. *ran and then the athletes*) were removed. For adjective-noun, instances in which the two words were separated by a preposition (e.g. *ordinary for some people*) were also removed. For verb-noun collconstructions, results were also filtered in order to ensure that the verb-noun combinations occurred only in verb-direct object constructions. This

included removing instances in which they occurred in passive (e.g. *were caused by problems*), phrasal verb (e.g. *write about a story*), and verb-indirect object (e.g. *give their children money*) constructions.

This extraction process resulted in three separate lists of collconstruction tokens (one for each collconstruction category) for all texts included in ICNALE-Korean and all texts by the same learner in the Salsbury Written Corpus. The ICNALE-Korean lists were then split according to the proficiency level and individual learner, resulting in tables containing collconstructions produced at each proficiency level and collconstructions produced in each individual text. A sample results table for collconstruction tokens in an individual ICNALE text produced by a B2 level Korean learner is shown in Table 4.4. For the Salsbury Written Corpus, collconstruction token lists for each learner were split according to the week that the learner produced each collconstruction token. These lists look similar to the one shown for the ICNALE text in Table 4.4.

Table 4.4 Collconstructions in a learner text from ICNALE (level B2)

Adjective-Noun	Adverb-Adjective	Verb-Noun
certain concept	very helpful	earn money
economic standard		make money
korean situation		experience society
part-time job		experience society
		have reason
		support statement
		have concept
		spend money
		build concept
		give chance
		handle problem
		shape standard
		experience job

4.4   Constructional Analysis

The studies in this dissertation investigate the development of collconstruction production from constructional and collostructional perspectives. The constructional analysis involved the use of construction type and token frequency and constructional diversity measures for adjective-noun, verb-noun, and adverb-adjective collconstructions. The type and token frequency measures consisted of normalized (per 100 words) type and token frequencies for all three categories in each learner text. Constructional diversity was assessed by calculating normalized entropy scores for each collconstruction category in a learner text. Normalized entropy measures the uncertainty of a probability distribution, in this case the distribution of collconstruction types among the collconstruction tokens in a text, and ranges from 0 to 1 (Kumar, Kumar, & Kapur, 1986). Values closer to 1 indicate more even distributions in which all collconstruction types occur equally, while values closer to 0 indicate a more uneven distribution in which one or more collconstruction types are used very frequently. Normalized entropy was chosen as the diversity measure in this dissertation because it is more sensitive to Zipfian frequency distributions than other more commonly used diversity measures such as type-token ratio and root type-token ratio (Gries & Ellis, 2015). Research has also shown that lower normalized entropy scores indicate a higher degree of semantic cohesion for a set of words in a construction (Ellis & O'Donnell, 2014).

4.5   Collostructional Analysis

The collostructional analysis involved the use of collexeme analysis. Collexeme analysis refers to a range of approaches developed by Stefanowitsch and Gries (2003) that attempt to quantitatively analyze associations between constructions and words. The specific type of collexeme analysis adopted in this dissertation is covarying collexeme analysis, which measures

the association between two words occurring in different slots within the same construction (Stefanowitsch & Gries, 2005). Covarying collexeme analysis is carried out by comparing the actual co-occurrence frequency of two words in a construction against their expected co-occurrence frequency using a 2-by-2 distribution table. An example table for verb-noun collconstructions is shown in Table 4.5. This table contains actual co-occurrence frequencies of two words in the construction, the frequency of other words in the first slot co-occurring with the target word in the second slot, the frequency of the target word in the first slot occurring before other words in the second slot, and the frequency of the construction when both target words are absent. These frequencies are then submitted to statistical tests, traditionally a Fisher-Yates exact test and log-transformation, to yield values that indicate the strength of association between two words within a grammatical construction.

Table 4.5 Covarying collexeme analysis table for verb-noun collconstructions

	N (Noun in VN CollCon)	ON (Other Nouns in VN CollCon)
V (Verb in VN CollCon)	Freq (V + N)	Freq (V + ON)
OV (Other Verbs in VN CollCon)	Freq (OV + N)	Freq (OV + ON)

The current study differs from previous collexeme analyses in two ways. First, instead of calculating collexeme scores for learner construction production by submitting word and construction frequencies in L2 data to a Fischer-Yates exact test (Deshors, 2016; Gilquin, 2015; Martinez-Garcia and Wulff, 2012; Wulff and Gries, 2011), covarying collexeme scores were assigned to collconstruction tokens based on frequency information from COCA Academic. The necessary frequency data for covarying collexeme analysis was extracted from the offline version of COCA Academic in a similar fashion to the method employed for collconstruction extraction from the learner texts. The corpus was first part-of-speech tagged using TagAnt (Anthony, 2015). Using POS tags and span searches in WordSmith Tools (Version 6, Scott,

2012), collconstruction frequencies were extracted from COCA Academic by first extracting frequencies for collconstruction tokens with the first lexeme in the construction, then extracting frequencies for collconstruction tokens with the second lexeme in the construction. For instance, frequency data for the verb-noun collconstruction *earn money* was extracted by first searching for instances of the lexeme *earn* (i.e., *earn, earns, earning,* and *earned*) occurring within five words to the left of all nouns. This was followed by searches for all instances of *money* (including the plural form *monies*) occurring within five words to the right of all verbs. Subsequent searches were done in order to account for instances in which the lexemes were separated by prepositions, conjunctions, or punctuation marks. This extra step ensured that frequency data extracted from COCA Academic truly represented collconstruction frequencies and not simple textual co-occurrence. Lastly, the overall frequency of verb-noun, adjective-noun, and adverb-noun collconstructions were extracted from COCA Academic by searching for all instances of verbs and nouns, adjectives and nouns, and adverbs and adjectives combinations within their set spans. Subsequent searches for verb-noun and adjective-noun combinations were conducted in order to filter out instances in which these word categories were separated by prepositions, conjunctions, and punctuation. This process resulted in frequency tables for each collconstruction token similar to the one for the collconstruction token *earn money* in Table 4.6.

Table 4.6 Covarying collexeme analysis table for the verb-noun collconstruction earn money

	money	Other Nouns	Row Totals
earn	58	2,540	2,598
Other Verbs	7,638	2,574,819	2,582,457
Column Totals	7,696	2,577,359	2,585,055

The second deviation from previous L2 collexeme research in this dissertation is the choice of statistical test used to measure association. Instead of calculating covarying collexeme scores using the negative log of the Fisher-Yates exact test, the current dissertation used log odds

ratio as the covarying collexeme score. Log odds ratio is preferred over the Fisher-Yates test because it is less sensitive to overall collconstruction frequency (Gries, personal communication, 2017).

Similar to previous association strength analyses of L2 phraseological unit production (Durrant & Schmitt, 2009; Granger & Bestgen, 2014), collconstruction types and tokens in the learner texts were divided into four categories based on their collostructional scores. The first category ("N/A") consisted of collconstructions in the learner texts that were found to be absent in COCA Academic. The second category consisted of collconstructions that had a negative log odds ratio, indicating that the two lexemes in the collconstructions exhibit a negative association with, or repulsion from, each other. Collconstructions were placed into the third category, labeled "non-collexemes", if they exhibited a positive, yet non-significant, association. Collconstructions were placed into the "collexemes" category if their collexeme score was above the threshold value for significance. A log odds ratio score of 2.5 was set as the threshold value for collexeme status given that it roughly corresponds to an MI score of 3, the standard threshold value for collocation status in the corpus linguistics literature (Hunston, 2002). Table 4.7 provides a breakdown of each category. Proportion scores for each category were calculated for each learner text. This approach to analyzing collexeme analysis was adopted over taking average collexeme scores for each learner text because of the low type and token frequencies for adjective-noun, adverb-adjective, and verb-noun collconstructions in each learner text. For instance, a text containing two strongly associated collconstructions and one negative collconstruction might have an average score that corresponds to the non-collexeme category. When compared to a text that contains three non-collexeme collconstructions, it may appear that these two texts are more similar in the use of collconstructions than they actually are. The

categorization of collconstructions based on association strength was done for collconstruction types as well as tokens in order to determine whether change in construction production is marked by changes in the repetition of strongly associated collexemes or changes in the variety of strongly associated collexemes. In total, each learner text was assessed for 33 indices of collconstruction production: normalized type and token frequencies, normalized entropy, type and token proportions scores for N/A, negative collexeme, non-collexeme, and collexeme collconstructions (11 indices times 3 collconstruction structural categories). In the case of ICNALE-Korean, these indices were also calculated for all adjective-noun, verb-noun, and adverb-adjective collconstructions in each subcorpus. An overview of these indices is shown in Table 4.8.

Table 4.7 Collconstruction categories based on collexeme score

Category	Criteria	Example
N/A	Not found in reference corpus	*correlate experience*
Negative Collexeme	Log odds ratio < 0	*common student*
Non-Collexeme	Log odds ratio between 0 and 2.5	*extremely high*
Collexeme	Log odds ratio above 2.5	*ride bicycle*

Table 4.8 Indices of collconstruction production

Index	CollConstruction	Label
Token Frequency	Adjective-Noun	JN_Tok
Type Frequency	Adjective-Noun	JN_Typ
Normalized Entropy	Adjective-Noun	JN_Ent
N/A Proportion Token	Adjective-Noun	JN_NA_Tok
N/A Proportion Type	Adjective-Noun	JN_NA_Typ
Negative Proportion Token	Adjective-Noun	JN_NG_Tok
Negative Proportion Type	Adjective-Noun	JN_NG_Typ
Non-Collexeme Proportion Token	Adjective-Noun	JN_NC_Tok
Non-Collexeme Proportion Type	Adjective-Noun	JN_NC_Typ
Collexeme Proportion Token	Adjective-Noun	JN_CO_Tok
Collexeme Proportion Type	Adjective-Noun	JN_CO_Typ
Token Frequency	Adverb-Adjective	RJ_Tok
Type Frequency	Adverb-Adjective	RJ_Typ
Normalized Entropy	Adverb-Adjective	RJ_Ent
N/A Proportion Token	Adverb-Adjective	RJ_NA_Tok
N/A Proportion Type	Adverb-Adjective	RJ_NA_Typ
Negative Proportion Token	Adverb-Adjective	RJ_NG_Tok
Negative Proportion Type	Adverb-Adjective	RJ_NG_Typ
Non-Collexeme Proportion Token	Adverb-Adjective	RJ_NC_Tok
Non-Collexeme Proportion Type	Adverb-Adjective	RJ_NC_Typ
Collexeme Proportion Token	Adverb-Adjective	RJ_CO_Tok
Collexeme Proportion Type	Adverb-Adjective	RJ_CO_Typ
Token Frequency	Verb-Noun	VN_Tok
Type Frequency	Verb-Noun	VN_Typ
Normalized Entropy	Verb-Noun	VN_Ent
N/A Proportion Token	Verb-Noun	VN_NA_Tok
N/A Proportion Type	Verb-Noun	VN_NA_Typ
Negative Proportion Token	Verb-Noun	VN_NG_Tok
Negative Proportion Type	Verb-Noun	VN_NG_Typ
Non-Collexeme Proportion Token	Verb-Noun	VN_NC_Tok
Non-Collexeme Proportion Type	Verb-Noun	VN_NC_Typ
Collexeme Proportion Token	Verb-Noun	VN_CO_Tok
Collexeme Proportion Type	Verb-Noun	VN_CO_Typ

4.6 Addressing Research Question 1

In order to examine variation in collconstruction production across proficiency levels (RQ 1), this study employed both qualitative and quantitative analyses of differences across proficiency levels. The focus of the qualitative analysis was on variation across proficiency levels in the use of certain collconstructional subcategories that fulfilled specific lexical

functions (Mel'čuk, 2007). In order to identify these functional subcategories, type-token lists for each structural collconstruction category in each subcorpus were created. The lists were then searched for collconstruction types that occurred particularly frequently as well as for the most frequent lexical functions being expressed by those types. Once a most frequent functional collconstruction subcategory for each structural category was identified, separate type-token lists were created for these subcategories in each subcorpus. These lists were then examined in order to identify differences in the use of these subcategories across the three proficiency levels represented in ICNALE-Korean.

The quantitative analysis aimed at addressing Research Question 1 consisted of the creation of an ordinal logistic regression model with the purpose of finding out which, if any, indices related to collconstruction production are predictive of human judgments of L2 writing proficiency. Before being entered into the logistic regression, all of the indices were standardized in order to avoid issues that can occur when variables in a regression model are on different scales. Indices were then checked for multi-collinearity in order to ensure that each index is measuring a distinct construct and to avoid overfitting the model. If any two variables were found to be correlated at $r \geq .700$ (Tabachnick & Fidell, 2001), the variable with the strongest correlation with proficiency was kept. All remaining variables were entered into a backward stepwise ordinal logistic regression that determined the model that best fit the data using the Akaike Information Criterion (AIC; Akaike, 1973). The *polr* function in the R package *MASS* (Ripley, 2017) was used to calculate the ordinal logistic regression model, while the *stepAIC* function in *MASS* was used for model selection.

4.7   Addressing Research Question 2

In order to investigate the longitudinal development of productive collconstruction

knowledge (RQ 2), this study employed linear mixed-effects (LME) models. The R package

lmerTest (Kuznetsova, Brockhoff, & Christensen, 2015) was used to construct LME models and

derive p-values for individual fixed effects. The *r.squaredGLMM* function from the *MuMIn*

package (Nakagawa, & Schielzeth, 2013) was also used in order to obtain effect size measures.

This function computes an R^2 value for marginal R^2 for the variance explained by the fixed

factors as well as a conditional R^2 for the variance explained by both fixed and random factors.

Before being entered into mixed effects models, all indices were checked for normality and the

absence of multi-collinearity. Any indices found to be non-normally distributed or multi-

collinear with another index that had a higher correlation with time were removed from further

analysis. From the remaining indices, one constructional (either token frequency, type frequency,

or normalized entropy) and two collostructional indices for adjective-noun, verb-noun, and

adverb-adjective collconstructions were selected, resulting in a total of 9 different models. These

indices were selected based on their correlations with time. "Time" (in weeks) and "proficiency"

(measured by institutional TOEFL score) were entered into the models as fixed effects. The

variable "subjects" was entered into the models as a random effect, with a random slope for time

added to "subjects". Estimates, standard errors, t-values, and p-values for the fixed effects and

variance and standard deviations for the random effect were estimated using restricted maximum

likelihood.

4.8   Addressing Research Question 3

In order to investigate collconstruction development for individual learners in the

Salsbury written corpus (RQ 3), this study employed methods common to Dynamic Systems

Theory (DST; Verspoor, 2015) research on L2 development that highlight the roles of intraindividual variability and interconnected linguistic knowledge in L2 development (Larsen-Freeman, 2006; Verspoor, de Bot, & Lowie, 2011). First, development for individual learners was assessed by plotting moving min-max graphs for each collconstruction category for each learner over time (van Dijk, Verspoor, & Lowie, 2011). Min-max graphs calculate and plot minimum and maximum scores for subsets of data using moving windows (time frames that move one data collection point at a time). In this study, a moving window of three observations plotted the minimum and maximum scores for the first three observations in the dataset. The next moving window plotted minimum and maximum scores for the second to fourth data points, with this process continuing until the final data point. Once complete, the min-max graphs allow for a visual inspection of the dynamic development in observed scores over time and how intraindividual variability in scores relates to long-term development.

Rather than create min-max graphs for all indices for all learners, which would have resulted in 198 graphs, a sample of indices was plotted for one learner per structural collconstruction category. The indices consisted of one frequency (either token or type) index, normalized entropy, and two collostructional indices. This resulted in a total of 12 min-max graphs being plotted. The selection of indices and individual learners will be discussed in Chapter 7. This analysis resulted in separate graphs for each of the four chosen indices in a structural collconstruction category for one learner that contain a developmental trend line, minimum and maximum score lines, and a polynomial trend line.

In order to examine the nature of interconnected development for the multiple indices of collconstruction development, moving correlation coefficients between selected frequency, normalized entropy, and collostructional proportion over time were plotted. Similar to moving

min-max graphs, moving correlation coefficients measure the association for two variables over a set window of data collection points, in the case of this dissertation three data points. Positive correlations indicate a supportive relationship, while negative correlations indicate a competitive one. By moving the window one point at a time over the entire dataset, longitudinal changes in these relationships can be observed (Verspoor & van Dijk, 2011). In order to examine interconnected development for different types of collconstruction knowledge (e.g. frequency, diversity, association strength), moving correlation analysis was carried out within structural collconstruction categories (e.g. verb-noun token frequency to non-collexeme type proportion score). Moving correlation analysis was also conducted for the same indices of collconstruction knowledge across collconstruction categories (e.g. verb-noun token frequency to adjective-noun token frequency) in order to investigate interconnected development for the same type of collconstruction knowledge in different structural collconstruction categories.

5 COLLCONSTRUCTION DEVELOPMENT ACROSS PROFICIENCY LEVELS

In this chapter, I examine the development of productive collconstruction knowledge across high-beginner, low-intermediate, and high-intermediate proficiency levels. For this analysis, texts from ICNALE-Korean (Ishikawa 2013), a corpus of texts written by EFL Korean learners, were assessed for overall writing proficiency and, based on these ratings, placed into three subcorpora corresponding to A2, B1, and B2 on the CEFR. Learner texts were also analyzed for several indices of adjective-noun, verb-noun, and adverb-adjective collconstruction development, including (1) token and type frequencies, (2) normalized entropy, and (3) proportion scores for different categories (N/A, negative, non-collexeme, collexeme) of collexeme strength (see Chapter 4 for an in-depth discussion of these indices). Qualitative analysis investigated variation in the production of multiple functional subcategories across the three proficiency levels. This was followed by a quantitative analysis that examined variation for structural categories through the creation of ordinal logistic regression models that determined which, if any, indices of collconstruction production can be used to predict human judgments of writing proficiency. These analyses were conducted in order to address research question 1: How does the use of collconstructions in L2 writing vary across proficiency levels?

5.1 Qualitative Analysis of Collconstruction Production for Functional Subcategories

In this section, I examine development for select functional subcategories of adjective-noun, verb-noun, and adverb-adjective collconstructions produced by the learners in ICNALE-Korean. Specifically, I examine type and token frequencies, normalized entropy, and covarying collexeme scores for functional subcategories and their tokens in order to identify possible changes in collconstruction production across proficiency levels. Although token and type frequencies for individual functional collconstruction categories are too small for the calculation

of inferential statistics, this section hopes to reveal developmental trends that warrant further analysis in future research.

5.1.1 Adjective-Noun Collconstructions

As a first step in investigating adjective-noun collconstruction production in ICNALE-Korean, I created lists of the most frequent adjective-noun collconstruction types, along with their normalized frequency and collexeme category, at each proficiency level (Table 5.1). For inclusion in this list, each collconstruction had to occur at last .04 times per 100 words, which is roughly equivalent to 1 occurrence per 10% of the texts in each subcorpus.

Table 5.1 Adjective-noun collconstructions in learner subcorpora occurring > .04 times per 100 words

A2			B1			B2		
Collconstruction	Frq	Cat	Collconstruction	Frq	Cat	Collconstruction	Frq	Cat
part-time student	.149	CO	social experience	.077	NG	social experience	.098	NG
social life	.114	NC	social life	.070	NC	social skill	.057	NC
good experience	.070	NG	good experience	.065	NG	social life	.057	NC
part-time work	.061	NC	social skill	.059	NC	future job	.057	NC
social experience	.061	NG	important thing	.051	NC	real world	.049	CO
			various people	.051	NG	korean student	.049	NC
			real world	.044	CO	negative effect	.041	CO
			future job	.040	NC	valuable experience	.041	NC
						good grade	.041	NC

Note. NA, NG, NC, and CO stand for N/A, negative, non-collexeme, and collexeme collconstruction categories.

Looking at the lists of adjective-noun collconstructions, there exist several similarities between the three learner groups. First, adjective-noun collconstructions appear to not occur very frequently at three proficiency levels. The types in the A2 list occur a total of 0.456 times per 100 words. In the B1 list, the combined frequency is 0.458 tokens per 100 words, while for the B2 list it is 0.490 tokens per 100 words. Second, several collconstructions occur frequently in at least two of the subcorpora. For instance, *good experience* occurs frequently in the A2 and B1 subcorpora, while *real world, social skill,* and *future job* are shared between the B1 and B2

subcorpora. In addition, *social life* and *social experience* are shared across all three learner subcorpora. The high frequency of these collconstruction types across the three subcorpora are most likely the result of the prompt, which asked students to discuss the benefits or drawbacks of students having a part-time job. Second, the combined frequencies for these most frequent collconstructions are similar across the three subcorpora. The five collconstructions in the A2 list have a combined frequency of .456 tokens per 100 words, while the eight in the B1 list have a combined frequency of .458 tokens per 100 words and the nine in the B2 list have a combined frequency of .490 tokens per 100 words.

However, the data in Table 5.1 also reveal several differences between the three groups. The first difference concerns the distribution of collexeme categories in each list. In the A2 subcorpus, the most frequent adjective-noun collconstructions are split between negatively and positively associated collconstructions, while one (*part-time student*) is significantly associated. The B1 list is also fairly evenly split, with three negatively associated, four positively associated, and one significantly associated collconstructions. At the B2 level, though, this balance is shifted, as only one of the nine collconstructions (*social experience*) is negatively associated. The other eight are either positively or significantly associated.

In order to investigate cross-sectional variation for functional adjective-noun collconstruction subcategories, I chose to focus on the subcategory [positive] + *experience*. This subcategory was chosen for two reasons. First, the lexeme *experience* was the most frequently occurring noun in adjective-noun collconstructions in ICNALE-Korean. Second, the function of positively evaluating *experience*, which is equivalent to Mel'čuk's (2007) lexical function of **Bon** ("positive evaluation"), was the most frequently occurring lexical function expressed with this noun. Evidence for this can be found in the lists in Table 5.1, with the collconstruction *good*

experience occurring among the most frequent collconstructions in the A2 and B1 corpora. In order to examine change across proficiency levels for this functional subcategory, I extracted all adjective-noun collconstructions containing the noun *experience* with positive evaluation in each subcorpus along with their frequencies per 100 words. I then calculated normalized entropy scores for each group to determine how evenly different positive evaluation adjectives were distributed in this functional subcategory. The results of this analysis (provided in Table 5.2) suggest that learners are becoming less reliant on the general positive adjective *good* for evaluating *experience*. Not only is the percentage of positive *experience* collconstruction tokens with *good* decreasing across proficiency levels, but also the entropy scores for these collconstructions are increasing. This provides some evidence that, at higher proficiency levels, learners are better able to positively evaluate *experience* with a wider range of less semantically general adjectives. It is also worth noting that the most frequently occurring adjective besides *good* in the B1 and B2 subcorpora is *valuable*. In fact, in the B2 subcorpus, *valuable experience* is the most frequently occurring token in this subcategory. Given the fact that this combination has the highest covarying collexeme score of any combination in this functional subcategory (1.849), this provides some evidence that more advanced learners are more targetlike in their production of this functional subcategory. From a Construction Grammar perspective, this finding, along with that for normalized entropy, suggest that, at higher proficiency levels, learners are less reliant on general purpose adjectives, are able use adjective-noun collconstructions with a wider variety of semantically similar lexemes, and are more targetlike in their choices.

Table 5.2 Positive adjective-noun collconstructions with experience

	good + experience		*[positive] + experience*		
	% Tok	Types	% Tok	most frequent (freq/100 words)	Entropy
A2	51.61%	9	48.39%	*great* (.017)	.715
B1	39.13%	12	60.87%	*valuable* (.029)	.772
B2	15.38%	6	84.62%	*valuable* (.041)	.890

5.1.2 Verb-Noun Collconstructions

Table 5.3 includes lists of the most frequent verb-noun collconstruction types, along with their normalized frequency and collconstruction category at each proficiency level. Similar to the lists for adjective-noun collconstructions, a collconstruction had to occur at least .04 times per 100 words to be included in a list.

Table 5.3 Verb-noun collconstructions in learner subcorpora occurring > .04 times per 100 words

A2			B1			B2		
Collconstruction	Frq	Cat	Collconstruction	Frq	Cat	Collconstruction	Frq	Cat
make money	.215	NC	earn money	.352	NC	earn money	.286	NC
earn money	.210	NC	get job	.183	NC	spend time	.122	NC
do job	.162	NA	make money	.176	NC	make money	.122	NC
get job	.136	NC	do job	.158	NA	do job	.106	NA
meet people	.066	NG	get money	.110	NC	get job	.090	NC
find job	.053	NC	meet people	.088	NG	meet people	.073	NG
have experience	.053	NG	need money	.084	NC	have experience	.065	NG
spend money	.048	NC	have experience	.084	NG	have time	.065	NG
need money	.048	NC	spend money	.077	NC	get experience	.065	NG
save money	.044	CO	have chance	.059	NG	find job	.057	NC
have money	.044	NG	get experience	.051	NG	afford college	.049	NC
			experience society	.044	NA	save money	.041	CO
			spend time	.044	NC	spend money	.041	NC
			know value	.044	NG	learn skill	.041	NC
			help student	.040	NC	get grade	.041	NC
			have money	.040	NG	have effect	.041	NC
						enjoy life	.041	NG
						use money	.041	NG
						experience thing	.041	NG

Note. NA, NG, NC, and CO stand for N/A, negative, non-collexeme, and collexeme collconstruction categories.

Three immediate observations can be made by looking at the results in Table 5.3. First, similar to adjective-noun collconstructions, these collconstruction types do not occur very frequently in the learner texts, with the highest combined token frequency still well below 2 tokens per 100 words. Second, there is an increase in the number of verb-noun collconstruction types above the frequency threshold. In addition, the combined frequency for the B1 list is higher (1.635 tokens/100 words) than the A2 list (1.078 tokes/100 words) and the B2 list (1.429 tokens/100 words). Third, the three learner groups produce some of the same verb-noun collconstructions frequently. Seven collconstructions (*do job, earn money, get job, have experience, make money, meet people, spend money*) appear in all three lists. The collconstructions *need money* and *have money* are shared between A2 and B1, while *find job* and *save money* are shared between A2 and B2 and *get experience* and *spend time* are shared between B1 and B2. Similar to the shared adjective-noun collconstructions across proficiency levels, these shared verb-noun collconstructions are most likely due to the topic the students were writing about (i.e. part-time jobs). In addition, the most frequent verb-noun collconstructions in the three subcorpora are also similar in how they are distributed across the four collexeme categories. The largest category in each list is non-collexeme, followed by negative collexeme, then N/A, and last collexeme.

In order to investigate changes for individual functional verb-noun collconstruction subcategories, I examined the use of verb-noun collconstructions with the noun *money*. Besides *job*, this noun was the most frequently occurring noun in verb-noun collconstructions. Specifically, I focused on *money* verb-noun collconstructions with the lexical function of expressing 'acquisition' (e.g. *earn money, get money, make money*), which was the most frequent

function used with *money*. Lists of [acquire] + *money* collconstructions and their normalized

frequencies are presented in Table 5.4.

Looking first at the combined token frequencies for each list, it can be seen that the B1

learners made the greatest use of this group of collconstructions, producing them .678 times per

100 words, with B2 learners producing them the least at .433 tokens per 100 words. This

decrease in token frequency may be a result of the less advanced learners focusing most of their

texts on how part-time jobs can help students earn more money, while the more advanced

learners wrote about other reasons for having a part-time job in addition to earning money. For

instance, in one A2 text, the learner used their entire text to write about how having a part-time

job could help them afford things like computers and school tuition. In contrast, one writer of a

B2 text wrote about how part-time jobs could help students earn money, make new friendships,

and develop practical skills.

Focusing on the number of types in each list in Table 5.4, it can be seen that there is a

decrease in the number of [acquire] + *money* collconstruction types from A2 to B2. High-

beginner learners produced the greatest range of [acquire] + *money* collconstructions, including

the greatest number of collconstructions produced only once in the subcorpus (*accept money,*

borrow money, find money, raise money, take money). In contrast, the B2 learners used the

smallest range of types, with only two of them, the positively associated and highly frequent

earn money and *make money*, occurring more than once. Additionally, *earn money*, which is the

most strongly associated collconstruction occurring in this functional category (collexeme score

= 2.04), increases in terms of its proportion of all collconstruction tokens in the category. It

accounts for 41.38% of all [acquire] + *money* tokens in the A2 subcorpus, 51.89% in the B1

subcorpus, and 66.04% in the B2 subcorpus.

Table 5.4 [acquire] + money *collconstructions in each of the learner subcorpora (normalized frequencies)*

A2		B1		B2	
Collconstruction	Frq	Collconstruction	Frq	Collconstruction	Frq
make money	.215	earn money	.352	earn money	.286
earn money	.210	make money	.176	make money	.122
get money	.035	get money	.110	receive money	.008
receive money	.018	receive money	.022	get money	.008
collect money	.009	gain money	.011	acquire money	.008
accept money	.004	borrow money	.004		
borrow money	.004	take money	.004		
find money	.004				
raise money	.004				
take money	.004				
Total	**.509**	**Total**	**.627**	**Total**	**.433**

From a Construction Grammar perspective, these results suggest a developmental trend for verb-noun collconstructions in which more advanced learners use certain functional verb-noun collconstruction categories in a more targetlike fashion and with greater semantic coherence. At the A2 level, learners are able to produce highly frequent and strongly associated verb-noun combinations such as *make money*, *earn money,* and *get money*. However, they still produce a range of verbs with *money* that are less frequent, less strongly associated, and less semantically coherent. At the B1 level, the range of verbs decreases as learners begin to rely more on a set of highly frequent and positively associated verb-noun collconstructions while reducing the number of less frequent and weakly associated combinations. At the B2 proficiency level, the range of combinations decreases even further as the learners come to predominantly rely on a single most strongly associated collconstruction type (*earn money*).

5.1.3 Adverb-Adjective Collconstructions

Table 5.5 presents the most frequent adverb-adjective collconstruction types, along with their normalized frequency and collconstruction category at each proficiency level. It is immediately clear from the lists in Table 5.5 that adverb-adjective collconstructions occurred

very infrequently in ICNALE-Korean. Only two types occurred above .040 times per 100 words in the A2 and B2 subcorpora, while only three types occurred above the same threshold in the B1 subcorpus. It is also clear that all three learner groups were similar in their use of adverb-adjective collconstructions. All seven types in the three lists have an intensifying function, with *very* occurring in five of the seven types. Additionally, the adjective *important* occurred in five of the seven types.

Table 5.5 Adverb-adjective collconstructions in learner subcorpora occurring > .04 times per 100 words

A2			B1			B2		
Collconstruction	Frq	Cat	Collconstruction	Frq	Cat	Collconstruction	Frq	Cat
very important	.096	NC	very important	.084	NC	very important	.057	NC
very hard	.061	NC	more important	.047	NC	most important	.049	NC
			very expensive	.040	NC			

Given the comparatively frequent use of an intensifying adverb with the adjective *important,* I chose to focus on this functional subcategory. The type list for this subcategory is shown in Table 5.6. It can be seen in the table that the distribution of adverbs in this subcategory becomes more even as proficiency increases. In the A2 subcorpus, *very important* occurs the most frequently and accounts for over 65% of the total number of [intensification] + *important* collconstruction tokens. While it remains the most frequent collconstruction type, its proportion of total [intensification] + *important* collconstruction tokens decreases to 43.39% in the B1 subcorpus to 38.89% in the B2 subcorpus. *More important* and *most important*, on the other hand, increase in proportion from A2 (14.71% and 17.65%) to B2 (16.67% and 33.33%). Similar to the findings for [positive] + *experience* adjective-noun collconstructions, these findings suggest that more proficient learners are better able to use adverb-adjective collconstructions with a wider range of semantically similar lexemes.

Table 5.6 [intensification] + important *adverb-adjective collconstructions in each subcorpus*

A2		B1		B2	
Collconstruction	Frq	Collconstruction	Frq	Collconstruction	Frq
very important	.096	very important	.084	very important	.057
most important	.026	more important	.048	most important	.049
more important	.022	most important	.026	more important	.024
so important	.004	really important	.011	much important	.008
		so important	.011	really important	.008
		less important	.004		
		much important	.004		
		quite important	.004		
		truly important	.004		

5.2 Quantitative Analysis of Collconstruction Production across Proficiency Levels

In this section, I quantitatively examine variation in adjective-noun, verb-noun, and adverb-adjective collconstruction production across proficiency levels. The focus in this section is on all collconstructions within each structural category (e.g. adjective-noun) regardless of lexical function. In order to examine variation, I first calculate means and standard deviations for each of the 11 indices of collconstruction production per collconstruction category for the learner texts in each subcorpus. I then calculate an ordinal logistic regression for the purposes of determining which, if any, collconstruction production indices are predictive of human judgments of writing proficiency. This is followed by a more in-depth analysis that provides more information concerning the predictive indices.

5.2.1 Descriptive Statistics

The means and standard deviations for all adjective-noun, verb-noun, and adverb-adjective constructional and collostructional indices in the three learner subcorpora are presented in Tables 5.7 to 5.9. Looking at the tables, it is clear that the target collconstructions did not occur very frequently in learner texts. Adjective-noun collconstructions occurred less than 3.5 times per 100 words on average in learner texts across proficiency levels, while adverb-adjective collconstructions on average never occurred above 1 time per 100 words in the learner texts.

Verb-noun collconstructions occurred the most frequently, but still only achieved a maximum average frequency of 5.41 times per 100 words. Although these low frequencies are not entirely surprising given bigram and collocation frequencies reported in previous L2 phraseology research (Granger & Bestgen, 2014; Siyanova-Chanturia, 2015), they do require a level of caution when interpreting the findings from any statistical analyses. These low frequencies also highlight one of the pitfalls of examining collocation production in shorter texts produced by beginner and intermediate L2 writers.

Table 5.7 Means and standard deviations for all adjective-noun collconstruction indices

	A2		B1		B2	
	Mean	SD	Mean	SD	Mean	SD
JN_Tok	3.06	1.60	2.46	1.27	3.05	1.70
JN_Typ	2.82	1.49	2.28	1.20	2.72	1.56
JN_Ent	.980	.098	.968	.130	.989	.020
JN_NA_Tok	22.24	22.25	20.37	24.81	16.20	16.12
JN_NG_Tok	17.12	21.27	18.26	23.32	16.55	19.01
JN_NC_Tok	41.59	26.94	44.99	29.56	46.14	27.67
JN_CO_Tok	19.05	20.88	14.79	20.54	17.00	19.57
JN_NA_Typ	23.06	22.36	21.03	25.14	16.84	16.23
JN_NG_Typ	17.28	20.83	18.20	23.03	15.76	17.79
JN_NC_Typ	41.10	27.58	44.79	28.94	46.22	27.13
JN_CO_Typ	18.54	20.31	14.39	19.77	17.12	19.36

Note. See Table 4.8 in Chapter 4 for an explanation of these indices

Table 5.8 Means and standard deviations for all verb-noun collconstruction indices

	A2		B1		B2	
	Mean	SD	Mean	SD	Mean	SD
VN_Tok	4.56	1.69	5.41	1.80	4.69	1.54
VN_Typ	4.05	1.34	4.57	1.43	4.09	1.26
VN_Ent	.981	.025	.981	.024	.980	.022
VN_NA_Tok	25.00	16.37	23.06	14.79	21.50	13.80
VN_NG_Tok	33.75	18.37	35.88	17.28	37.54	16.40
VN_NC_Tok	36.89	18.28	37.61	16.17	36.34	15.36
VN_CO_Tok	4.36	7.70	3.46	5.96	4.63	6.52
VN_NA_Typ	24.88	15.65	24.01	14.91	22.49	13.99
VN_NG_Typ	35.03	18.51	37.44	16.88	38.12	16.60
VN_NC_Typ	35.33	16.89	34.79	14.87	34.36	14.54
VN_CO_Typ	4.76	8.40	3.76	6.43	5.03	7.01

Note. See Table 4.8 in Chapter 4 for an explanation of these indices

Table 5.9 Means and standard deviations for all adverb-adjective collconstruction indices

	A2		B1		B2	
	Mean	SD	Mean	SD	Mean	SD
RJ_Tok	0.97	0.80	0.77	0.69	0.56	0.60
RJ_Typ	0.94	0.77	0.71	0.63	0.56	0.60
RJ_Ent	0.74	0.44	0.74	0.44	0.79	0.41
RJ_NA_Tok	12.47	24.30	6.51	19.72	9.86	28.44
RJ_NG_Tok	8.09	20.96	8.68	21.20	8.45	22.59
RJ_NC_Tok	57.23	40.35	58.50	43.77	51.06	47.11
RJ_CO_Tok	4.26	16.47	0.69	4.36	4.08	19.99
RJ_NA_Typ	12.51	24.33	6.39	19.55	9.86	28.44
RJ_NG_Typ	8.22	21.13	8.66	21.08	8.45	22.59
RJ_NC_Typ	57.04	40.34	58.57	43.73	51.06	47.11
RJ_CO_Typ	4.29	16.50	0.75	4.77	4.082	19.99

Note. See Table 4.8 in Chapter 4 for an explanation of these indices

5.2.2 Ordinal Logistic Regression

In order to examine whether or not differences in adjective-noun, verb-noun, and adverb-adjective collconstruction production across proficiency levels are significant and predictive of human judgments of writing proficiency, an ordinal logistic regression was conducted. Before conducting the regression, all indices were standardized and checked for multi-collinearity. Unsurprisingly, the type and token versions of each type of index (e.g. frequency, normalized entropy, N/A proportion) were found to be correlated above $r \geq .700$. In order to determine which of the multi-collinear indices to remove, pairwise correlations were calculated between each of the indices and proficiency level (coded as "1" for A2, "2" for B1, and "3" for B2). The type or token version of each index that was found to have the lowest correlation with proficiency level was then removed. This resulted in the removal of 15 indices, leaving 18 total indices (shown in Table 5.10) to be entered into a backward stepwise ordinal logistic regression model with proficiency level as the ordered dependent variable.

Table 5.10 Indices entered into ordinal logistic regression

Adjective-noun	Verb-noun	Adverb-adjective
JN_Typ	VN_Tok	RJ_Typ
JN_Ent	VN_Ent	RJ_Ent
JN_NA_Typ	VN_NA_Tok	RJ_NA_Typ
JN_NG_Typ	VN_NG_Tok	RJ_NG_Tok
JN_NC_Typ	VN_NC_Typ	RJ_NC_Tok
JN_CO_Tok	VN_CO_Typ	RJ_CO_Typ

The final ordinal logistic regression model yielded a significant model, χ^2 (2) = 19.39, $p <$ .001. However, this model did not perform well in predicting human judgments of writing proficiency. The McFadden's R^2 of .078 indicates that this model accounted for only 7.8% of the variance in human essay ratings. This model correctly classified texts according to proficiency level in 48.91% of cases, with the reported Kappa = 0.119 indicating only slight agreement between actual and predicted proficiency levels for the texts. The confusion matrix showing how the logistic regression performed in classifying the learner texts is shown in Table 5.11, while the classification accuracy, precision, and recall indices for each proficiency level are presented in Table 5.12.

Table 5.11 Confusion matrix for ordinal logistic regression for all three groups

	A2	B1	B2
A2	**45**	32	10
B1	61	**89**	38
B2	0	0	**1**

Table 5.12 Accuracy, precision, and recall statistics for ordinal logistic regression

	Precision	Recall
A2	.517	.425
B1	.473	.736
B2	1.000	.020

The final model included two indices of collconstruction production: verb-noun normalized entropy and adverb-adjective type frequency. The coefficients, standard errors, odds ratios, and odds ratio confidence intervals for these two indices are shown in Table 5.13. Both of these indices were found to have a significant effect on human judgments of writing proficiency.

For verb-noun normalized entropy, the model indicates that a decrease of .024 (i.e. the standard deviation for this index across the entire dataset) in normalized entropy for verb-noun collconstructions in a learner text was associated with a 25.9% increase in the odds of that text being judged as more proficient. The odds ratio for adverb-adjective type frequency indicated that a decrease of 0.694 types (i.e. the standard deviation for this index across the entire dataset) per 100 words in a learner text was associated with a 35.6% increase in the odds of that text being judged as more proficient.

Table 5.13 Coefficients, standard errors, and confidence Intervals for ordinal logistic regression

| | | | | 95% CI for Odds Ratio | | |
	B	*S.E.*	*p*	Lower	Odds Ratio	Upper
VN_Ent	-0.299	0.116	.010	0.589	0.741	0.930
RJ_Typ	-0.441	0.174	< .001	0.505	0.644	0.813

Note. See Table 4.8 in Chapter 4 for an explanation of these indices

5.2.3   Interim Discussion

In this section of the chapter, I investigated cross-sectional variation in the production of all adjective-noun, verb-noun, and adverb-adjective collconstructions regardless of lexical function. The findings from the ordinal logistic regression model indicated that texts containing fewer adverb-adjective collconstruction types and having lower normalized entropy scores were more likely to be judged as more proficient. Focusing first on adverb-adjective type frequencies, these results may be partially due to the fact that, compared to the A2 subcorpus, a greater proportion of learners in the B1 and B2 subcorpora produced no adverb-adjective collconstructions in their texts. At the B1 level, 31 out of 121 (25.61%) learners produced no adverb-adjective collconstructions. 19 out of 49 B2 learners (26.53%) produced no adverb-adjective collconstructions. This is compared to only 19 out of 106 (17.92%) learners in the A2 subcorpus producing no adverb-adjective collconstructions.

The findings on verb-noun normalized entropy scores may be partially explained by the fact that, compared to the B1 and B2 subcorpora, a greater proportion of learners in the A2 subcorpus used each verb-noun collconstruction type only once in their text, resulting in a perfect normalized entropy score (1.000). As can be seen in Figure 5.1, the A2 subcorpus had the largest percentage of texts in which each verb-noun collconstruction only occurred once. The B1 subcorpus had the lowest percentage, while the percentage for the B2 subcorpus fell between those for the other two. In contrast, Figure 5.2 shows that the proportion of texts with a normalized entropy score below .950 increased from A2 to B2. Taken together, these findings suggest that less proficient learners tend to produce each verb-noun collconstruction only once per text, while their more proficient counterparts are better able to repeat some collconstructions more frequently.
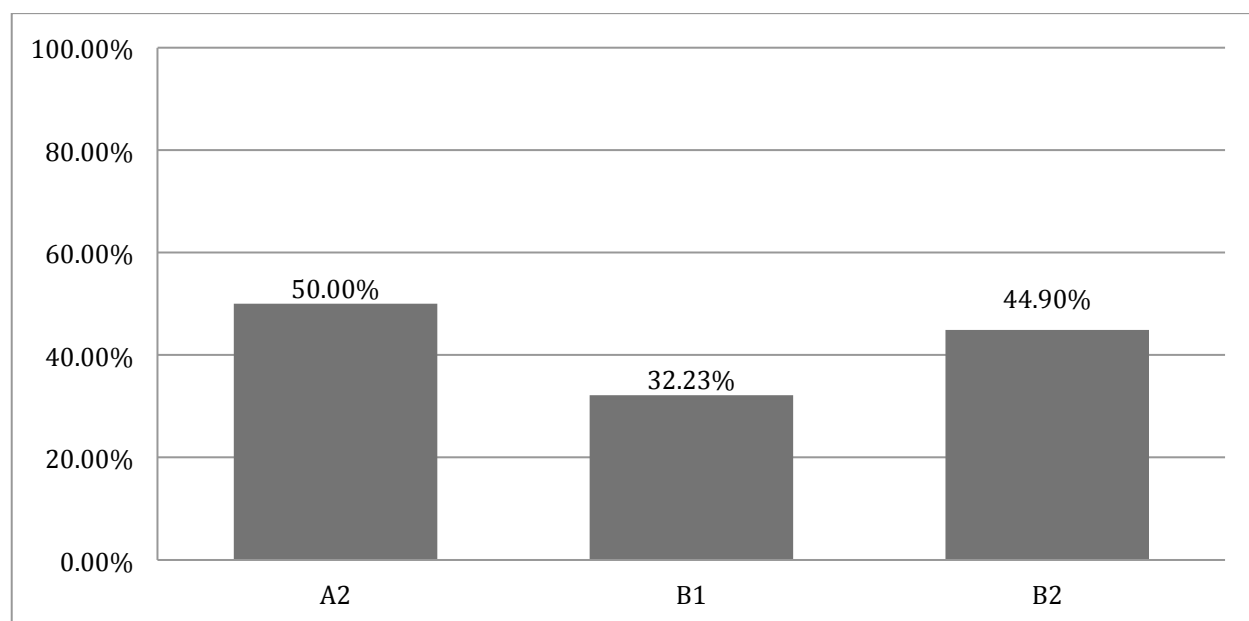
Figure 5.1 Percentage of texts in each subcorpus with a H_{norm} = 1.000
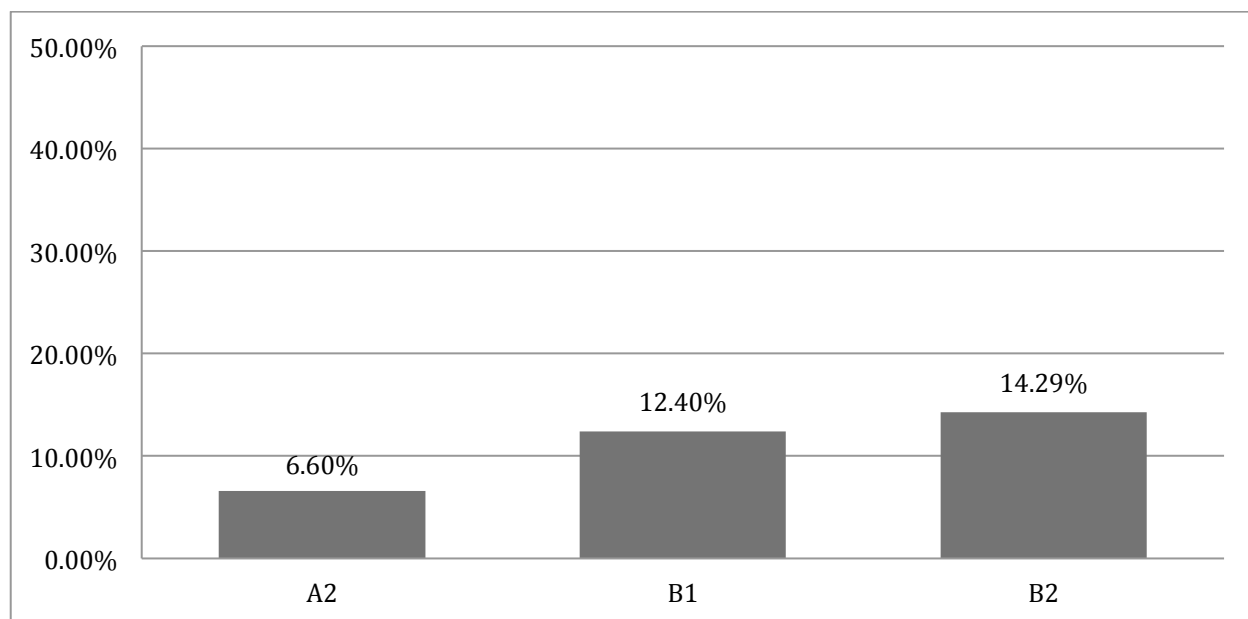
Figure 5.2 Percentage of texts with H$_{norm}$ = .950

5.3   Summary and Discussion

This chapter explored how productive adjective-noun, verb-noun, and adverb-adjective collconstruction knowledge varies across proficiency levels in a corpus of essays written by Korean EFL learners responding to the same prompt. It did so by examining cross-sectional variation in the use of different functional collconstruction subcategories as well as variation in the use of all collconstructions in each structural category (e.g. adjective-noun). Regarding the former, results suggest that, across proficiency levels, learners may have differed in how they used certain functional groups of collconstructions. It was found that, for some collconstructions, the more proficient learners may have been less reliant on general purpose adjectives, verbs, or adverbs for expressing lexical functions. For instance, low- and high-intermediate learners in the study produced a wider range of adjectives in the collconstruction [positive] + *experience* than high-beginner learners, who instead mostly relied on the general purpose adjective *good*. In addition, the more advanced learners produced more strongly associated [positive] + *experience* collconstructions (e.g. *valuable experience, meaningful experience*) more frequently. These

findings support those from previous research that has examined learner language from a Usage-based Second Language Acquisition (UBSLA) perspective. Similar to the results for [positive] + *experience*, UBSLA research has shown that, as proficiency increases, learners' production of constructions becomes more schematic and targetlike as they move from relying on fixed constructional exemplars that are more general in their semantics to using a wider range of lexemes within the constructions (Ellis & Ferreira-Junior, 2009a; Eskildsen, 2009, 2012, 2015, Mellow, 2006; Myles et al., 1998).

The findings also suggest an opposite trend in change across proficiency levels for some collconstructions, with learners' lexical inventories possibly being more restricted, semantically coherent, and targetlike at higher proficiency levels. Specifically, it was found that, compared to lower proficiency learners, more proficient learners relied on a narrower range of verbs to express the lexical function [acquire] + *money*. In fact, the high-intermediate learners overwhelmingly relied on only two verbs, the positively associated, highly frequent, and idiomatic *earn money* and *make money*. Additionally, the more proficient learners were also found to produce a significantly more restricted range of verb-noun collconstructions compared to their less proficient counterparts. These findings support those from Römer and Garner (under review) on construction development in spoken learner English. The authors found that, compared to their low-intermediate counterparts, advanced L2 speakers produce a more restricted range of targetlike verbs in verb-argument constructions (VACs).

Regarding variation across proficiency levels for the three structural collconstruction categories, it was found that high-beginner, low-intermediate, and high-intermediate learners are overall very similar in their collconstruction production. Results from the ordinal logistic regression demonstrated that only verb-noun normalized entropy and adverb-adjective

collconstruction type frequency were predictive of human judgments of writing proficiency. The coefficients and odds ratios for each of these indices indicated that the more proficient writers in the corpus were more likely to produce fewer adverb-adjective collconstructions and be more able to repeat some verb-noun collconstructions in their writing. None of the collexeme proportion score indices for any of the structural collconstruction categories were found to be predictive of human judgments of writing quality. The finding that no collexeme proportion score indices were found to be predictive of writing proficient seems to go against the findings in Garner, Crossley, and Kyle (in press). In their study, several bigram association score indices were found to be predictive of human judgments of writing quality in a different corpus of texts by high-beginner, low-intermediate, and high-intermediate L1 Korean writers. These contradictory findings suggest that, between beginner and intermediate stages of L2 writing, development in the use of strongly associated phraseological units may be occurring for other types of phraseological units, such as grammatical collocations (e.g. *interested in, aware of, agree with*). In combination with previous research that has shown significant differences in collocation or bigram production in intermediate and advanced writing (Hsu, 2007; Granger & Bestgen, 2014, Paquot, 2017), the findings for the collostructional indices in this chapter suggest that some types of productive collocational knowledge may develop at more advanced stages of writing proficiency. At beginner and intermediate proficiency levels, L2 writers may be focused on developing their ability to use adjective-noun, verb-noun, and adverb-adjectives collocations frequently regardless of the strength of association between constituent words. Through more written and spoken interaction in English, however, learners may begin noticing collocational patterns in use and integrate them into their own speech and writing at more advanced proficiency levels.

Taken together, these results also highlight the importance of taking an approach towards collconstruction analysis that focuses on categories of collconstructions based on the lexical functions that they fulfill. While the broader analysis of all collconstructions in learner language showed very little variation in collconstruction production across proficiency levels, the more fine grained analysis of individual functional groups of collconstructions did provide some evidence that high-beginner, low-intermediate, and high-intermediate learners differ in their collconstruction production. Future research should therefore not only include analysis of all collconstructions within a specific structural category (e.g. verb-noun, adjective-noun), but should also include a more specific analysis that focuses on different functional categories of collconstructions (e.g. verb-noun fulfillment collconstructions, adjective-noun intensification collconstructions) and the highly frequent exemplars of those categories. Through such an analysis, researchers may be better able to examine the multi-faceted nature of phraseological development in L2 writing across proficiency levels.

There are several limitations that need to be noted concerning the cross-sectional corpus used in this analysis that limit the generalizability of these results. First, this corpus only contained writing from EFL L1 Korean learners and did not include learners from any other L1 background. Second, this corpus was also small and contained texts that were relatively short, with most texts being under 300 words. These features limited overall collconstruction token and type frequencies in the learner texts. This may have led to higher degrees of variability in the proportion scores. It also limited the amount of analysis on functional collconstruction subcategories that was possible. Lastly, all of the texts in the cross-sectional corpus were written on the same topic of part-time jobs. Even though prompt-based collconstruction types (e.g. "have

job" and "part-time job") were removed, this single prompt may have narrowed the range of collconstructions types the learners produced.

In conclusion, this chapter has shown how productive collconstruction knowledge varies between different groups of learners at different levels of writing proficiency, providing some insight into how development of this knowledge occurs cross-sectionally. However, it is also essential to investigate how collconstructional knowledge develops for individual learners longitudinally over an extended period of English study. The next two chapters focus on this form of development for a small group of ESL learners studying English over the course of one year.

6   COLLCONSTRUCTION DEVELOPMENT OVER TIME: MIXED-EFFECTS MODELS

In the previous chapter, I investigated the development of adjective-noun, verb-noun, and adverb-adjective collconstructions across high-beginner (A2), low-intermediate (B1), and high-intermediate (B2) proficiency levels. This analysis showed that more proficient learners produce a smaller range of adverb-adjective collconstructions, repeat verb-noun collconstruction types more frequently, and are more targetlike in their use of some functional collconstruction categories (e.g. the [positive] + *experience* collconstruction). In the present and the following chapters, I shift focus from differences in collconstruction production across proficiency levels to examining how learners develop their productive collconstruction knowledge over time. I do so through the analysis of parts of the Salsbury written corpus (Salsbury, 2000), a longitudinal learner corpus of untimed and unstructured freewrites produced by six ESL learners enrolled in an intensive English program over the course of one year (more details about this corpus can be found in section 4.1.2). These texts were analyzed for several indices of adjective-noun, verb-noun, and adverb-adjective collconstruction development, including (1) token and type frequencies, (2) normalized entropy, and (3) proportion scores for different categories of collexeme strength. For this chapter, linear mixed effects models were calculated for a selection of these indices in order to determine their development over the course of the study period. In addition, TOEFL scores for each of the learners were included in the models in order to examine the effect that overall language proficiency had on collconstruction production. These analyses were conducted in order to address research question 2: How does the use of collconstructions in L2 writing develop over time?

6.1   Methods

6.1.1 Data

For this analysis, a subcorpus of texts from the Salsbury corpus was created. This subcorpus consists of ten texts from each of the six learners written every 3 weeks from Weeks 3 to 15 and every 5 to 7 weeks from Weeks 21 to 50. Each text consists of all the freewrites produced by the learner during a particular week. For instance, all five of EunHui's freewrites in Week 9 were included in her Week 9 text. A summary of this data, including the number of words for each text and total word counts for each learner, is provided in Table 6.1. This subcorpus is the same as the one used by Kyle (2016) in his study of longitudinal syntactic complexity development. For more information about each of these learners, such as their L1, see section 4.1.2. Looking at the table, it is immediately clear to see that there is a lot of intraindividual variation in terms of how much the learners wrote each week. For instance, total weekly word counts for Takako ranged from a low of 201 to a high of 1,927 words. The table also shows a lot of interindividual variation, with some learners producing longer texts than other learners. However, these differences should not affect the results of the current analyses as the calculated indices are based on normalized frequencies and proportion scores.

In addition to completing freewrites, the learners in the Salsbury corpus took institutional TOEFL exams every two months. Each text was therefore assigned a proficiency score based on the learner's score on the most recent TOEFL exam. So, for example, a learner's text written in Week 9 was assigned that learner's score on the first TOEFL exam that was administered in mid-October. A linear mixed effects model for proficiency scores was constructed with a fixed effect of time (in weeks) with learners being entered into the model as a random effect and a random slope for time added to learners. The results of this mixed effects model showed a significant

effect of time (estimate = 2.402, SE = 0.316, $t = 7.603$, $p < .001$). The two R^2s for this model are

$R^2_{marginal} = .389$ and $R^2_{conditional} = .844$. This finding indicates that, over the course of one year, the

learners in this corpus made significant gains in language proficiency as measured by the

institutional TOEFL exam.

Table 6.1 Number of words collected per participant in the Salsbury corpus

Week	EunHui	Faisal	Jalil	Kamal	Marta	Takako
3	379	339	294	102	279	242
6	120	252	179	184	180	302
9	552	244	501	125	458	467
12	443	676	517	88	244	450
15	153	718	904	235	278	436
21	392	381	489	144	676	1,117
26	241	513	343	215	440	201
34	225	270	480	271	228	228
43	246	126	477	318	377	1,927
50	226	1007	450	313	231	244
Total	**2,977**	**4,526**	**4,634**	**1,995**	**3,391**	**5,614**

Note. See section Chapter 4.1.2 for more information about these learners

6.1.2   Variable Selection

Once all 33 indices of adjective-noun, verb-noun, and adverb-adjective collconstruction

production were calculated for all texts in the subcorpus (see Table 4.8 for a complete list of

indices), they were checked for normality. Although linear mixed effects models are relatively

robust against violations against the assumption of normality (Winter, 2013), non-normal

distributions can still negatively effect the results of the model. In order to check for normality,

skewness and kurtosis scores for each of the indices were calculated, with any indices scoring

greater than 3 for either skewness or kurtosis removed. This led to the removal of 9 indices.

Correlations were then conducted between the remaining indices in order to check for

multicollinearity. If any two indices were correlated above $r = .700$, only the measure with the

highest correlation with time was retained. This led to the removal of 10 indices. From the

remaining 14 indices, three indices from each structural collconstruction category (i.e. adjective-

noun, verb-noun, adverb-adjective) were chosen for use in the linear mixed effects models.

These indices included one constructional index (either token frequency, type frequency, or

normalized entropy) and either N/A and non-collexeme type proportion scores or N/A and non-

collexeme token proportion scores. For the constructional indices, the one showing the strongest

correlation with time was chosen. N/A and non-collexeme proportion scores were chosen over

negative and collexeme proportion scores because these two indices consistently showed strong

correlations with time across all three structural collconstruction categories. The 9 indices used

in the linear mixed effects models in this chapter are presented in Table 6.2 along with their

correlations with time. An alpha level of 0.0055 with a Bonferroni correction for multiple

comparisons was used for all fixed effects in the models.

Table 6.2 Collconstructional indices used in linear mixed-effects models and their correlations with time

Adjective-Noun Collconstructions		Verb-Noun Collconstructions		Adverb-Adjective Collconstructions	
Index	r	Index	r	Index	r
JN_Tok	.204	VN_Tok	.037	RJ_Ent	-.058
JN_NA_Tok	-.236	VN_NA_Typ	-.267	RJ_NA_Tok	.083
JN_NC_Tok	.458	VN_NC_Typ	.219	RJ_NC_Tok	-.137

6.2   Results

6.2.1   Adjective-Noun Collconstructions

For adjective-noun collconstructions, the token frequency, N/A token proportion scores,

and non-collexeme token proportion scores were entered into three separate mixed effects

models with time and proficiency entered in as fixed effects and random intercepts and slopes

calculated for the random effect of learners. The coefficients, standard errors, t-scores, and p-

values for the fixed effects in each of the three models are shown in Tables 6.3 to 6.5. The model

for adjective-noun collconstruction token frequency revealed no significant main effect of time.

It also showed no main effect of proficiency. Similar non-significant findings for time and

proficiency were also found for N/A token proportion and non-collexeme token proportion.

Furthermore, the $R^2_{marginal}$ values for each model (shown in Table 6.6) indicate that the fixed

factors of time and proficiency explain between 3.91% and 20.50% of the variance in the

learners' adjective-noun collconstruction over the course of one year. Overall, these findings

suggest that the six learners showed little measurable change in their production of adjective-

noun collconstructions over the course of one year and that changes in proficiency had no

significant effect on the learners' adjective-noun collconstruction production.

Table 6.3 Results of linear mixed effects model for adjective-noun collconstruction token
frequency

Fixed Effects	coefficient	SE	t	p
Intercept	2.194			
Time	0.015	0.020	0.758	.468
Proficiency	0.001	0.004	0.205	.839
Random Effect	Name	sd		
Learner	Intercept	1.239		
	Slope	0.001		

Table 6.4 Results of linear mixed effects model for adjective-noun collconstruction N/A token
proportion

Fixed Effects	coefficient	SE	t	p
Intercept	70.607			
Time	-0.106	0.278	-0.380	.709
Proficiency	-0.103	0.088	-1.166	.255
Random Effect	Name	sd		
Learner	Intercept	12.570		
	Slope	0.145		

Table 6.5 Results of linear mixed effects model for adjective-noun collconstruction non-
collexeme token proportion

Fixed Effects	coefficient	SE	t	p
Intercept	4.817			
Time	0.699	0.319	2.189	.046
Proficiency	0.044	0.095	0.466	.647
Random Effect	Name	sd		
Learner	Intercept	14.216		
	Slope	0.317		

Table 6.6 Marginal and Conditional R^2s for the three adjective-noun collconstruction linear mixed effects models

	$R^2_{marginal}$	$R^2_{conditional}$
JN_Tok	.039	.308
JN_NA_Tok	.087	.301
JN_NC_Tok	.205	.388

6.2.2 Verb-Noun Collconstructions

Verb-noun collconstruction token frequency, N/A type proportion scores, and non-collexeme type proportion scores were entered into three separate mixed effects models with time and proficiency entered in as fixed effects and random intercepts and slopes calculated for the random effect of learners. The coefficients, standard errors, t-scores, and p-values for the fixed effects in each of the four models are shown in Tables 6.7 to 6.9. Similar to the results for adjective-noun token frequency, the model for verb-noun token frequency found no significant main effects for time or proficiency. Additionally, no significant main effects for time or proficiency were found for N/A type proportion and non-collexeme type proportion. The values for $R^2_{marginal}$ (Table 6.10) show that the fixed effects of time and proficiency only accounted for between 3.00% and 9.08% of the variance in the three verb-noun collconstruction indices. Taken together, these findings indicate that there was little measurable development in the learners' use of verb-noun collconstructions over time and that changes in their proficiency as measured by their TOEFL test scores also had no significant effect on their verb-noun collconstruction production.

Table 6.7 Results of linear mixed effects model for verb-noun collconstruction token frequency

Fixed Effects	*coefficient*	*SE*	*t*	*p*
Intercept	6.338			
Time	0.019	0.018	1.072	.288
Proficiency	-0.006	0.005	-1.320	.192
Random Effect	Name	*sd*		
Learner	Intercept	0.002		
	Slope	0.001		

Table 6.8 Results of linear mixed effects model for verb-noun collconstruction N/A type proportion

Fixed Effects	*coefficient*	*SE*	*t*	*p*
Intercept	19.461			
Time	-0.493	0.207	2.380	.021
Proficiency	0.066	0.056	1.179	.243
Random Effect	Name	*sd*		
Learner	Intercept	0.000		
	Slope	0.001		

Table 6.9 Results of linear mixed effects model for verb-noun collconstruction non-collexeme type proportion

Fixed Effects	*coefficient*	*SE*	*t*	*p*
Intercept	17.826			
Time	0.240	0.167	1.440	.161
Proficiency	0.014	0.045	-0.300	.767
Random Effect	Name	*sd*		
Learner	Intercept	0.154		
	Slope	0.093		

Table 6.10 Marginal and Conditional R^2s for the three verb-noun collconstruction linear mixed effects models

	$R^2_{marginal}$	$R^2_{conditional}$
VN_Tok	.029	.029
VN_NA_Typ	.091	.091
VN_NC_Typ	.048	.078

6.2.3   Adverb-Adjective Collconstructions

Adverb-adjective collconstruction normalized entropy, N/A token proportion scores, and

non-collexeme token proportion scores were entered into three separate mixed effects models

with time and proficiency entered in as fixed effects and random intercepts and slopes calculated

for the random effect of learners. The coefficients, standard errors, t-scores, and p-values for the fixed effects in each of the 3 models are shown in Tables 6.11 to 6.13. The model for adverb-adjective normalized entropy frequency found no significant main effects for time or proficiency. Additionally, no significant main effects for time or proficiency were found for N/A token proportion scores and non-collexeme token proportion scores. The values for $R^2_{marginal}$ (Table 6.14) show that the fixed effects of time and proficiency only accounted for between 5.45% and 12.44% of the variance in the 3 adverb-adjective collconstruction indices. Taken together, these findings indicate that there was little measurable change in the learners' use of adverb-adjective collconstructions over time, and that changes in their proficiency as measured by their TOEFL test scores had no significant effect on their adverb-adjective collconstruction production.

Table 6.11 Results of linear mixed effects model for adverb-adjective collconstruction normalized entropy

Fixed Effects	coefficient	SE	t	p
Intercept	-0.030			
Time	-0.007	0.005	-1.284	.226
Proficiency	0.002	0.001	1.538	.148
Random Effect	Name	sd		
Learner	Intercept	0.256		
	Slope	0.006		

Table 6.12 Results of linear mixed effects model for adverb-adjective collconstruction N/A token proportion

Fixed Effects	coefficient	SE	t	p
Intercept	-30.802			
Time	-0.179	0.136	-1.311	.195
Proficiency	0.104	0.037	2.816	.007
Random Effect	Name	sd		
Learner	Intercept	0.000		
	Slope	0.000		

Table 6.13 Results of linear mixed effects model for adverb-adjective collconstruction non-collexeme token proportion

Fixed Effects	coefficient	SE	t	p
Intercept	-2.894			
Time	-0.838	0.460	-1.822	.089
Proficiency	0.208	0.142	1.466	.160
Random Effect	Name	sd		
Learner	Intercept	13.247		
	Slope	0.267		

Table 6.14 Marginal and Conditional R^2s for the three verb-noun collconstruction linear mixed effects models

	$R^2_{marginal}$	$R^2_{conditional}$
RJ_Ent	.055	.187
RJ_NA_Tok	.124	.124
RJ_NC_Tok	.067	.276

6.3   Summary and Discussion

This chapter investigated the longitudinal development of productive adjective-noun, verb-noun, and adverb-adjective collconstruction knowledge in the writing of six ESL students over the course of one year. Growth in multiple indices of collconstruction frequency, diversity, and collexeme strength was examined through the use of linear mixed effects models that included time and proficiency level as fixed effects and learners as a random effect. Across all nine tested models, no significant effect of time was found, demonstrating that there was no measurable change in the learner's use of collconstructions over the course of the study period. The models also found no significant effect for proficiency level, suggesting that the learner's growth in overall language proficiency did not correspond to a change in their productive collconstruction in writing. Comparing these results with findings from previous studies, the results in this chapter seem to contradict those found in Crossley and Salsbury (2011). In their study, the authors found that the same group of learners became more accurate in their spoken bigram use over the course of the year. This difference in findings may be due to the fact that

Crossley and Salsbury looked at these learners' spoken data and included all bigrams, regardless of their parts of speech, in their analysis. This might, then, suggest that spoken phraseological knowledge development, at least in the beginning stages of L2 development, might occur more quickly than written phraseological knowledge development. It also might suggest that, for beginning learners, productive knowledge of other types of phraseological units (grammatical collocations, lexical bundles) develops earlier than that for adjective-noun, verb-noun, and adverb-adjective collconstructions. The results of this study also seem to confirm those of Bestgen and Granger (2014), who found little development in the use of collgrams (i.e. bigrams assigned MI and t-scores) in intermediate and advanced student writing over the course of one semester of study.

It is possible, however, that these non-significant findings are the result of high levels of interindividual variation in the production of the three structural collconstruction categories. In all but one of the models (the model for verb-noun N/A type proportions), the standard error or standard deviations for the random intercept or the random slope for learners was at least 50% of the value of the coefficients for either the intercept or the fixed effect of time. These rather high standard errors and standard deviations suggest that the individual learners in this group are experiencing such different developmental trajectories in their collconstruction production that it is difficult for a significant group trajectory to be created. To further examine this possibility, Figures 6.1 and 6.2 plot the individual trajectories for each of the learners for adjective-noun non-collexeme token proportion scores and verb-noun N/A type proportion scores. These indices were chosen because they produced two of the models that were closest to finding a significant effect of time. As can be seen in the figures, all of the learners' developmental trends are highly different from one another. For instance, in Figure 6.1, EunHui's adjective-noun non-collexeme

token proportions increase from 33.33% at Week 3 to 100% at Week 50. In contrast, Takako's adjective-noun non-collexeme proportions show only a modest increase from Week 1 (7.69%) to Week 50 (14.29%). Similarly, in Figure 6.2, it can be seen that two of the learners increase their verb-noun N/A type proportion scores from Week 1 to Week 50, three of them decrease their scores, and one remains at about the same level. In addition to interindividual variability, the developmental trajectories in Figures 6.1 and 6.2 also display a strong amount of intraindividual variability. Looking again at EunHui's developmental trajectory in Figure 6.1, it is clear that her adjective-noun non-collexeme token proportions scores, while increasing from Week 1 to Week 50, fluctuate throughout. Her scores increase from Week 6 to Week 15, decrease from Week 15 to Week 34 (with a slight increase at Week 26), then increase again in Weeks 43 and 50. This pattern of nonlinear development holds for all six learners' individual trajectories represented in the two figures below.
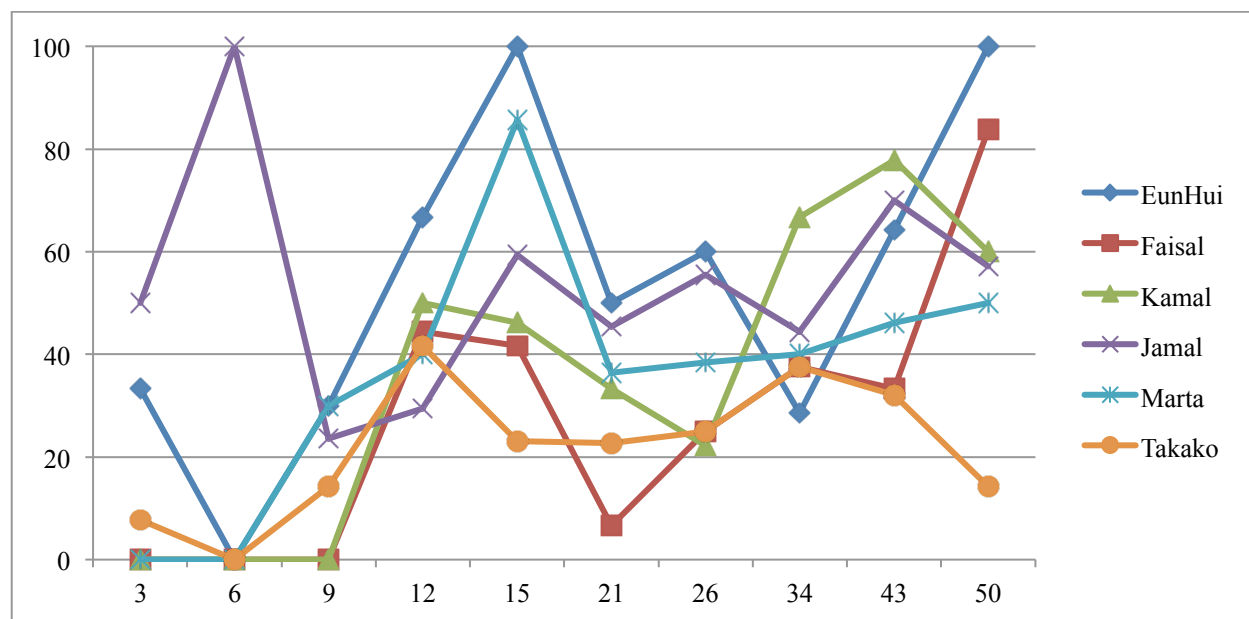

Figure 6.1 Individual trajectories for adjective-noun non-collexeme token proportion scores
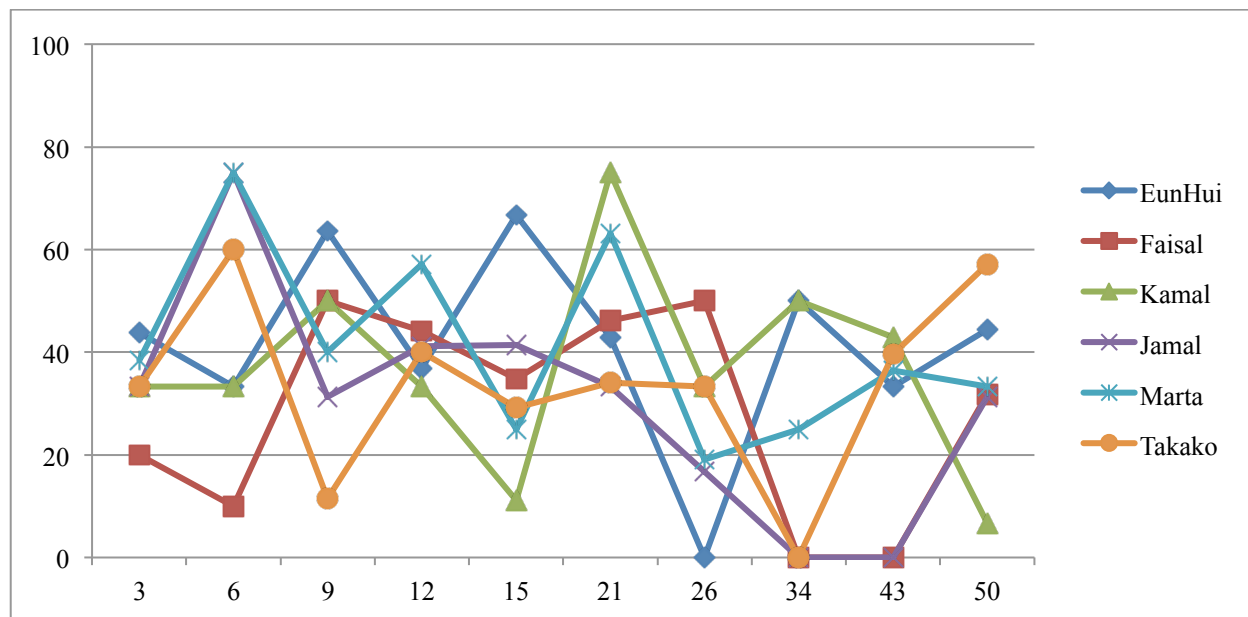
Figure 6.2 Individual trajectories for verb-noun N/A type proportion scores

The two main limitations of this analysis were the small sample size of the corpus and the small number of collconstruction tokens in the texts. The small number of learners represented in this corpus makes it difficult to find significant quantitative changes in collconstruction production for the entire group as well as limits the generalizability of the results. All of the texts in this corpus, similar to those in the cross-sectional corpus, contained a limited number of collconstructions tokens, which may have led to high degrees of variation in the proportion score indices. This was strongest for adverb-adjective collconstructions. These items never occurred more than 3 times per 100 words in a text and did not occur at all in 13 of the 60 texts in this analysis.

In conclusion, this chapter has shown that, even over a year of intensive English study, there is a little significant development in the productive knowledge of collconstructions for beginner ESL learners when studied as a group. However, it has also shown that these non-significant findings may be due to interindividual and intraindividual variability in collconstruction frequency, diversity, and collexeme indices for the different texts. This may

hide the fact that some learners are experiencing significant development in their collconstruction production despite the non-significant findings for the group as a whole. The next chapter will further examine this possibility by analyzing growth in frequency, diversity, and collexeme measures of collconstruction production for individual learners utilizing graphical methods from Dynamic Systems Theory (DST; Verspoor, 2015).

7   DYNAMIC DEVELOPMENT OF COLLCONSTRUCTIONS

In the previous chapter, I examined the longitudinal development of productive collconstructional knowledge through the calculation of linear mixed effects models. The results from these models indicated that there was no significant development of productive collconstructional knowledge over the course of one year for six ESL learners at the group level. However, a subsequent examination of the developmental trajectories for each of the learners revealed that there was a great deal of inter- and intraindividual variability in collconstruction development. Some learners showed strong growth in collconstruction production while others showed much less development. These findings highlight the importance of analyzing longitudinal changes for individual learners in order to effectively understand how productive collconstruction development occurs.

To that end, the current chapter investigates the development of productive collconstructional knowledge for the learners in the Salsbury written corpus (Salsbury, 2000) from a Dynamic Systems Theory (DST; Verspoor, 2015) perspective. DST conceptualizes L2 development as an emergent process driven by a learner's interactions with the target language. As such, this perspective towards L2 development shares many of the same basic assumptions towards language and L2 development as Usage-based Second Language Acquisition (UBSLA; Bybee, 2010), Complexity Theory (CT; Larsen-Freeman, 2010), and Language as a Complex Adaptive System (LaCAS; Ellis & Larsen-Freeman, 2009). The analyses in this chapter utilized graphical methods common to DST research that allow for an in-depth investigation of individual developmental trajectories, the role of variability in L2 development, and how different types of productive collconstruction knowledge interact in the developmental process.

All of this was done in order to address research question 3: How does the use of collconstructions in L2 writing develop in individual learners over time?

7.1   Dynamic Systems Theory and L2 Development

According to Dynamic Systems Theory, language is a dynamic and complex system composed of smaller linguistic subsystems such as the phonetic system, grammar system, and the lexical system. Each of these subsystems also contains further nested subsystems; for instance, the lexical subsystem consists of different types of interacting lexical knowledge (i.e. receptive, productive, semantic associations). Language, then, exists as a constantly changing self-organized system that emerges from the interaction between these different subsystems (van Geert & Verspoor, 2015). In terms of L2 development, DST holds that an individual's interlanguage emerges through the dynamic interaction of their limited linguistic resources (e.g. interlanguage, L1) and internal and external resources (e.g. motivation, working memory, interlocutors) (de Bot et al., 2007). Additionally, interlanguage subsystems interact and can either share learning resources or compete for them. Subsystems sharing resources are called "connected growers" and form supportive relationships that require fewer resources between them for development to occur (van Geert, 1991). An example of a supportive relationship is the one between reading comprehension and lexical knowledge. As lexical knowledge increases, a learner's reading comphrehension abilities also increase. This increased reading ability allows the learner to pay more attention to unknown words and increases their ability to learn them through context. On the other hand, subsystems can also form competive relationships. In these situations, learners may be forced to choose between subsystems and direct their resources to one at the expense of another (Verspoor & van Dijk, 2011).

Because L2 development depends on the dynamic interaction of a learner's current interlanguage and internal and external resources, this process is marked by high degrees of interindividual and intraindividual variability (van Geert & Verspoor, 2015). Regarding the former, all learners begin the learning process at different initial states. These initial states consist of, among other things, their experiences with the language to be learned, their first language knowledge, and their available resources. According to DST, differences in initial states will cause learners to have unique growth trajectories that may be quite different from one another in the long-term. In addition, individual learners are, consciously or unconsciously, constantly making choices in how they allocate their limited resources (van Geert & Verspoor, 2015). Some learners may, at certain stages in their learning, choose to focus on grammar instead of vocabulary, leading to significant growth in the former but not in the latter. Others may choose accuracy over complexity, leading to accurate, yet simplistic language use. All of these factors lead to individual learners following their own unique developmental trajectories. Intraindividual variability often occurs before periods of rapid language development as learners explore new communicative strategies in a process of trial and error (Verspoor, 2015). Using the resources of their immediate context, learners attempt to employ new strategies in order to achieve a communicative goal. If the learner is unsuccessful, they may alternate between older and newer strategies or alter their use of the new strategy. Successful attempts, on the other hand, are tallied by the language user, leading to continuing use of the strategy and its stabilization in their interlanguage. However, this stability is temporary, with variability increasing again as the learner attempt to employ new strategies or adjust the use of old ones. As a result, language development from a DST perspective occurs in a discontinuous, step-wise fashion, with

alternating periods of stability, high degrees of variability, and stability at higher ability levels (Baba & Nitta, 2014).

Overall, DST studies of L2 development have tended to support claims for the role of variability in language development. Verspoor, Lowie, and van Dijk (2008), in their case study of one advanced English learner's lexical development and sentence complexity development over the course of three years, found that mean word length exhibited the largest amount of growth over the three-year period. They also found that this growth was marked by alternating periods of stability and variability. Scores in initial stages were relatively stable, exhibited extreme variability as mean word length increased in the intermediate stages, and eventually re-stabilized by the end of the study. Focusing on the beginning of the language development process, Spoelman and Verspoor (2010) traced the development of morphological, phrasal, and sentence-level complexity for one L1 Dutch beginning learner of Finnish. Results indicated that all three complexity measures exhibited varying degrees of variability in the intermediate stages of the study before eventually stabilizing as the learner produced longer and more complex nouns, noun phrases, and sentences. Zheng (2016) investigated the development of single word and lexical bundle production in the writing of upper-intermediate L1 Chinese university students over the course of a year. Group results indicated that while lexical sophistication and diversity increased during the year, lexical density remained flat, and lexical bundle development tended to follow a u-shaped curve. Looking at data for individual learners, it was found that one learner's lexical sophistication and density showed wide variation early on, but became more stable as the learner developed. Similar patterns were also found for another learner's lexical diversity scores, with periods of variability giving way to increasing and more stable scores.

Findings from each of the above reviewed studies also support the DST notion of interconnected development between linguistic subsystems. Results from Verspoor et al. (2008) demonstrated that sentence length and lexical diversity were in a competitive relationship throughout the study. In weeks where the learner increased his sentence length, it came at the expense of his lexical diversity and vice versa. Spoelman and Verspoor (2010) found that their beginning L2 Finnish learner exhibited problems in attending to both noun phrase and sentence complexity in her writing, with one decreasing in weeks when the other increased. However, the relationship between these indices became less competitive as the learner's interlanguage developed. They also found supportive relationships for word complexity and sentence complexity and for word complexity and noun phrase complexity. In Zheng (2016), supportive relationships were found between lexical sophistication, diversity, and density measures. However, the strength of these supportive relationships fluctuated over time, with the weakest correlations observed during the intermediate stages of the study. Regarding single-word and lexical bundle production, non-significant or negative correlations between lexical bundle and all three single word indices indicated a trade-off in single word and multi-word sequence development.

Three other studies have provided additional evidence on how different linguistic subsystems interact and influence each other's development. Larsen-Freeman (2006) investigated the development of grammatical complexity, lexical complexity, and fluency in the narrative writing of five female high-intermediate Chinese learners of English over a period of six months. Group results showed that participants wrote more fluently and accurately with a greater degree of grammatical and lexical complexity over time. However, examining individual results revealed that participants took different paths in their development. For example, one

learner made greater gains in lexical complexity at the expense of her grammatical complexity. On the other hand, the other participants experienced the opposite trend, increasing their grammatical complexity at the expense of their lexical complexity. Lastly, Caspi and Lowie (2010, 2013) presented two case studies that examined the relationships between different types of receptive and productive vocabulary knowledge over time. In both studies, advanced L2 learners took multiple versions of a test measuring active recall, active recognition, and controlled word production and wrote essays that served as free production data. Overall, both studies found that the relationship between the two productive vocabulary knowledge dimensions were more competitive than that for the receptive vocabulary knowledge dimensions. According to the authors, these findings indicate that productive vocabulary knowledge may take longer to develop and stabilize than receptive vocabulary knowledge.

7.2   Methods

7.2.1   Data

The analysis in this chapter focuses on three of the six learners in the Salsbury corpus: EunHui, Marta, and Takako. These learners were chosen because they produced the greatest number of texts and had the smallest amount of large breaks (>3 weeks) between freewrite submissions. Furthermore, a subcorpus was created for each learner that was comprised of a sample of texts they wrote over the course of one year. For Marta and Takako's subcorpora, intervals between texts ranged from three to five weeks. For EunHui's subcorpus, intervals between texts typically ranged from four to five weeks, although the break between her last two texts was 6 weeks. Different intervals were used for the three learners in order to ensure that the intervals between each text were as even as possible. An overview of the learner subcorpora is presented in Table 7.1.

Table 7.1 Overview of the three learner's subcorpora used in this analysis

EunHui		Marta		Takako	
Week	Words	Week	Words	Week	Words
1	295	1	193	1	313
6	120	5	184	5	138
11	633	9	458	10	1,075
15	153	12	244	15	436
20	775	15	278	20	836
25	589	19	123	25	766
30	929	23	262	30	652
34	225	27	621	34	228
39	2,327	31	345	39	1,397
43	246	34	228	44	397
50	226	38	631	50	244
		41	561		
		43	377		
		47	152		
		50	231		
Total	**6,518**		**4,888**		**6,482**
Words/week	**592.55**		**325.87**		**589.27**

7.2.2 Data Analysis

In order to examine the development of collconstruction production in the learner data and the role of variability in that development, moving min-max graphs (van Dijk, Verspoor, & Lowie, 2011) were created for selected indices. Min-max graphs calculate and plot minimum and maximum scores for subsets of data using moving windows that shift one data collection point at a time. For example, a moving window of five observations would first plot the minimum and maximum scores for the first five observations in the data set. The next moving window would plot minimum and maximum scores for the second to sixth data points, with this process continuing until the final data point. Once complete, min-max graphs allow researchers to visually inspect variation in observed scores over time and see how this variation relates to long-term changes in minimum and maximum scores. For the current study, a moving window of three time points was utilized. In addition to plotting the actual developmental trend lines and

minimum and maximum scores, second order polynomial trend lines will be plotted in order to plot the general developmental trend (van Dijk, Verspoor, & Lowie, 2011).

In order to investigate the interconnected nature of longitudinal development for the multiple adjective-noun, verb-noun, and adverb-adjective collconstruction indices, the current study also makes use of moving correlation coefficients between variables (Verspoor & van Dijk, 2011). Similar to moving min-max graphs, moving correlation coefficients measure the association for two variables over a set window of data collection points, most often five. Positive correlations indicate a supportive relationship, while negative correlations indicate a competitive one. By moving the window one point at a time over the entire data set, longitudinal changes in these relationships can be observed. For the current study, a moving correlation window of three time points was utilized. Moving correlations were calculated for indices both within collconstruction categories (e.g. verb-noun frequencies correlated with verb-noun collexeme scores) and across collconstruction categories (e.g. verb-noun frequencies correlated with adjective-noun frequencies).

7.2.3   Learner and Variable Selection

Rather than look at development for all three learners for all indices in each structural collconstruction category, the analysis in this chapter focuses on a selection of indices for one learner per structural collconstruction category. In order to do this, I followed the approach laid out by Baba and Nitta (2004) for selecting individual learners. Each learner's scores for all of the indices were correlated with time (in weeks). Because a strong correlation indicates the greatest amount of growth, the learner showing the strongest average correlation between time and the adjective-noun, verb-noun, or adverb-adjective collconstruction indices was selected for that collconstruction category. Concerning the selection of indices, token frequencies were chosen

over type frequencies because they had stronger correlations with time across structural

collconstruction categories. Similarly, N/A type proportion and non-collexeme type proportion

were chosen over the other collostructional indices due to their stronger correlations with time.

7.3   Results

7.3.1   Adjective-Noun Collconstructions

For the analysis of adjective-noun collconstruction development, I chose to focus on

Marta because the average correlation between her adjective-noun token frequencies ($r = .332$),

normalized entropy scores ($r = .150$), N/A type proportion scores ($r = -.481$), and non-collexeme

type proportion ($r = .435$) were most strongly correlated with time. Figures 7.1 and 7.2 show the

moving min-max graphs (moving window of three data points) for Marta's token frequency and

normalized entropy indices. Starting with token frequency, it can be seen in Figure 7.1 that the

trajectory for Marta's production of adjective-noun collconstruction tokens approximates a u-

shaped pattern as it decreases over time before experiencing an increase near the end of the study

period. However, this increase is small, as her Week 50 token frequency is only 1 token per 100

words greater than her Week 1 frequency. The greatest change in Marta's adjective-noun token

frequencies was the degree of variability from week to week. These changes in variability

occurred in three phases. In the initial phase of the study period, which lasted from Week 1 to

Week 19, her token frequencies experienced a high degree of variability as they fluctuated

between 5.978 (Week 5) and 1.626 (Week 19) tokens per 100 words. In phase two, which lasted

from Week 23 to 34, her use of adjective-noun collconstructions becomes more stable as the

bandwidth between minimum and maximum frequencies narrowed. During this time, her token

frequencies fluctuated between 1.159 and 3.060 tokens per 100 words. The final phase of

Marta's adjective-noun collconstruction token development began in Week 38 and was marked

by a high degree of stability. Except for Week 47, Marta's token frequencies remained fairly stable between 3.170 and 3.448 tokens per 100 words. The comparatively low token frequency in Week 47 was likely due to the fact that this text was much shorter than the other texts written during this period. Focusing on Marta's normalized entropy scores (Figure 7.2), it can be seen that, except for Week 15, her scores tend to fluctuate between a high of 1.000 and low of 0.928 with very little development occurring towards a more Zipfian distribution of adjective-noun collconstructions.
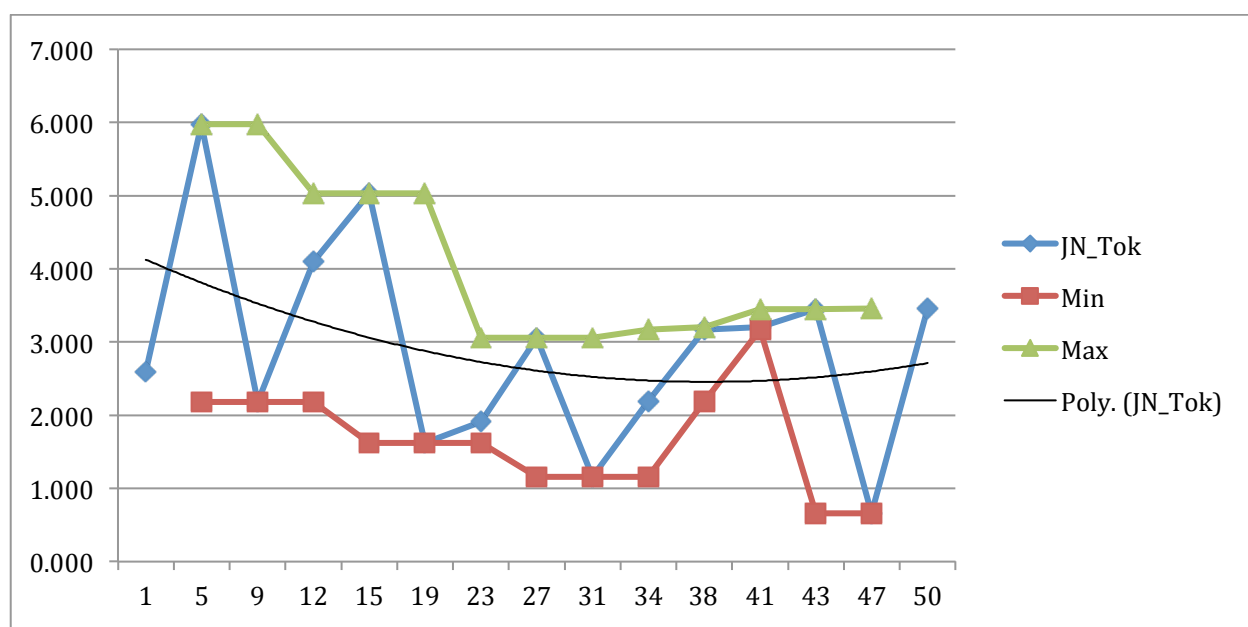

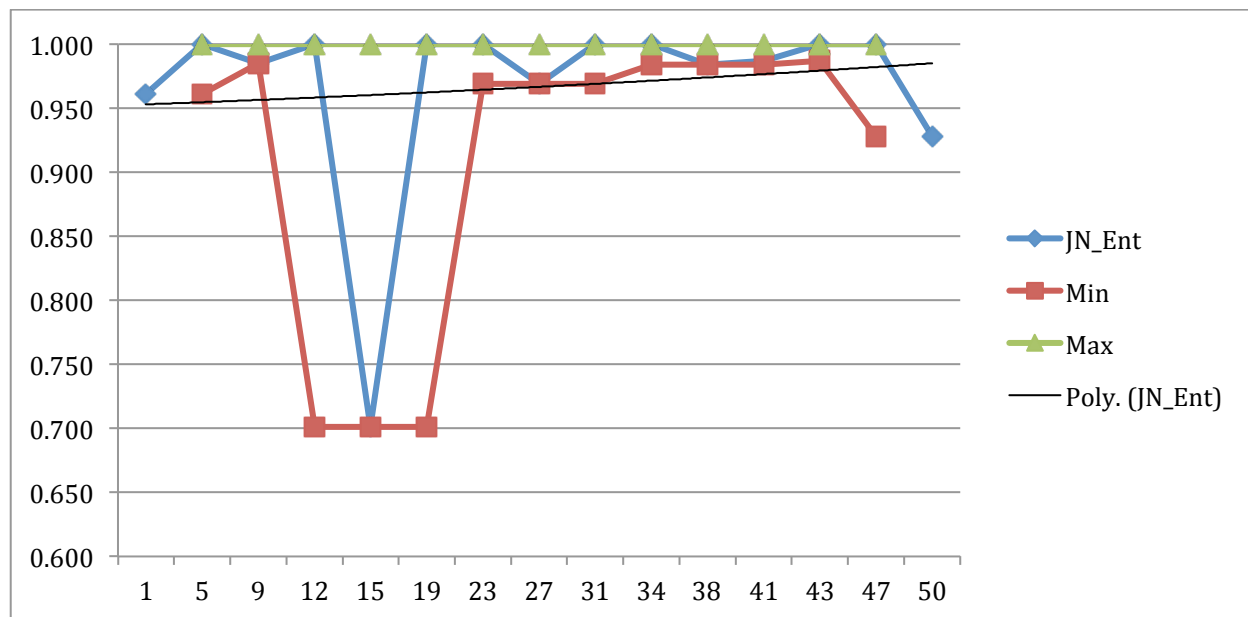*Figure 7.1 Marta's adjective-noun collconstruction token frequency development*

Figure 7.2 Marta's adjective-noun collconstruction normalized entropy development

The min-max graphs for Marta's production of adjective-noun N/A types and non-collexeme types are shown in Figures 7.3 and 7.4. In contrast to her token frequency and normalized entropy scores, both of these indices show clear developmental trajectories as Marta decreases her reliance on collconstructions not present in the reference corpus and begins to use those that are positively, but not strongly, associated (i.e. collexeme scores between 0 and 2.5). For the N/A types (Figure 7.3), this decreasing use occurs in three phases. The first phase, from Week 5 to Week 15, saw a linear decrease in N/A type proportions from 72.22% (4.348 types per 100 words) to 20.00% (0.360 types per 100 words). This initial period of linear decrease in scores was followed by a period of high variability in scores in the second phase, lasting from Week 19 to 38. During this phase, scores fluctuated between peaks of 37.50% (0.966 types per 100 words) and 60.00% (1.316 types per 100 words) and minimum scores of 0.00% in two weeks. In Week 41, there is a phase transition leading to the final phase in which N/A type proportion scores again experience a mostly linear decrease to Week 47 before increasing again at Week 50. It is important to note, however, that the N/A type proportion score of 0.00% in the

Week 47 text is due to the fact that only one adjective-noun collconstruction occurred in that text and it was classified as non-collexeme. If this text was removed from the graph, Marta's N/A type proportions would show a linear decrease during this final phase of the study period. The results for Marta's N/A type frequencies mirror those for her N/A type proportion scores as they declined from 0.891 types per 100 words in Week 41 to 0.433 types per 100 words in Week 50.
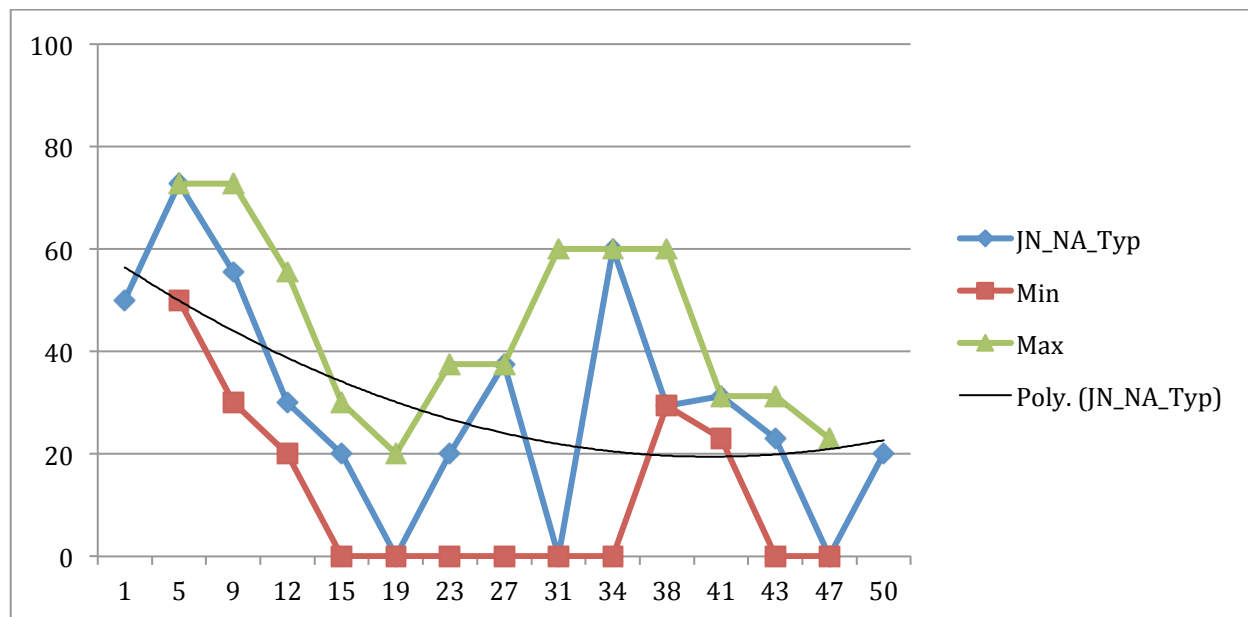

Figure 7.3 Marta's adjective-noun collconstruction N/A type proportion development

Figure 7.4 shows a similar, yet opposite trajectory for adjective-noun non-collexeme type scores. There is an overall increase in these scores from 25.00% (0.518 types per 100 words) in Week 1 to 40.00% (0.866 types per 100 words) in Week 50. This change in non-collexeme proportion scores occurs in two phases. In phase 1, which lasted from Week 5 to Week 15, proportions scores increased fairly linearly from 9.09% (0.543 types per 100 words) to 60.00% (1.079 types per 100 words). In phase two, which started in Week 19 and lasted to the end of the study period, proportion scores showed a high degree of variability with mostly wider bandwidths between minimum and maximum scores. Nevertheless, the scores during this phase never drop below Marta's initial proportion score of 25.00% and usually stay above 30.00%.

This is also the case for her non-collexeme type frequencies, with the lowest frequency during this phase (0.658 types per 100 words) still above her frequencies in Weeks 1 and 5.
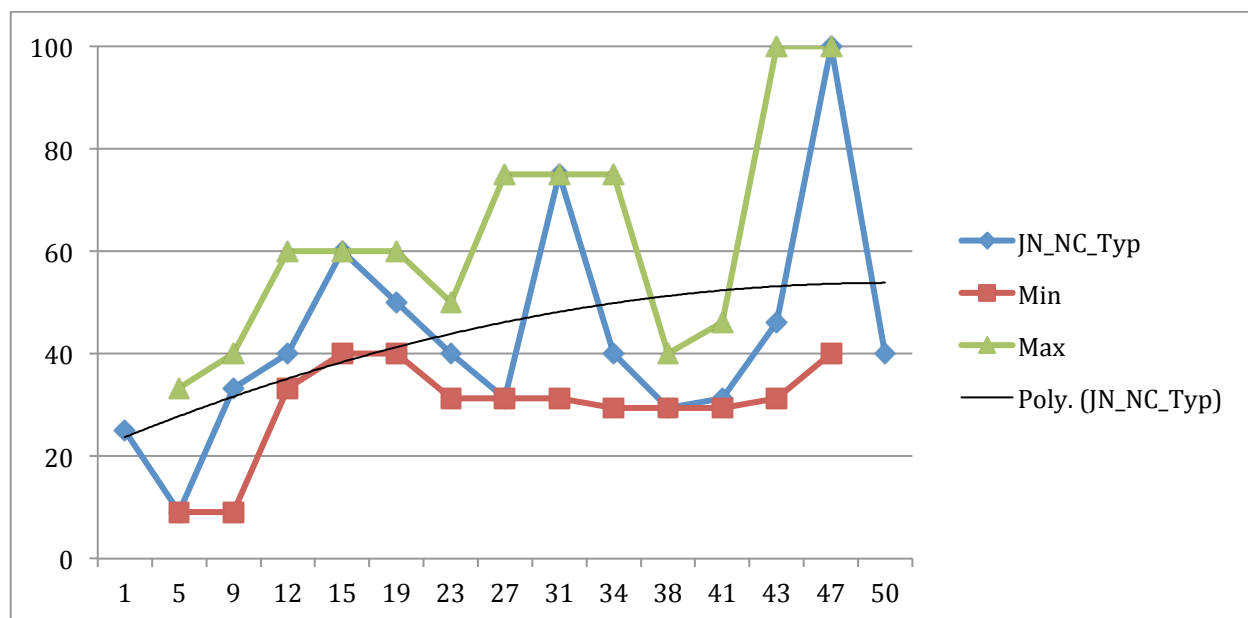

Figure 7.4 Marta's adjective-noun collconstruction non-collexeme proportion development

To further illustrate these findings, Table 7.2 displays the adjective-noun collconstructions in Marta's first (Week 1) and last (Week 50) texts. We can see that in the Week 1 text, Marta's collconstructions consist of one non-collexeme, one negatively associated, and two N/A collconstruction types. In contrast, her Week 50 text, which only contains one more type than her Week 1 text, contains one N/A adjective-noun, two non-collexeme, and two collexeme combinations. So, even though her Week 50 text non-collexeme type proportion score (40.00%) is only slightly better than her Week 1 score (25%), her ability to use positively associated adjective-noun collconstructions has increased, as evidenced by her use of more non-collexeme and collexeme collconstructions.

Table 7.2 Adjective-noun collconstructions in Marta's Week 1 and Week 50 texts

Week 1		Week 50	
Collconstruction	Category	Collconstruction	Category
big country	NC	bad custom	NA
different building	NG	good reason	CO
valuable university	NA	passive smoking	CO
wonderful city	NA	public place	NC
		valuable thing	NC

Note. NA, NG, NC, and CO stand for N/A, negative, non-collexeme, and collexeme collconstruction categories.

In light of the results shown in Table 7.2, I created a new collostructional category that contains all positively associated adjective-noun collconstruction types in learner texts regardless of whether or not they meet the threshold set for collexeme classification. I then created a min-max graph for their proportions in Marta's text over the course of the study. This graph is shown in Figure 7.5. Compared to the earlier figure for her non-collexeme type proportion scores (Figure 7.4), this figure shows a clearer trajectory in Marta's use of positively associated adjective-noun collconstructions. Her scores exhibit an overall increase from 25.00% (0.518 types per 100 words) in Week 1 to 80.00% (1.732 types per 100 words) in Week 50. This trajectory, similar to the one for N/A type proportions, occurred in three phases. In phase 1, which lasted from Week 1 to Week 12, her proportion scores increased linearly from 25.00% (0.518 types per 100 words) to 70.00% (2.869 types per 100 words). Phase 2, which lasted from Week 15 to Week 34, is marked by a high degree of variability with wide bandwidths between minimum and maximum scores. During this phase, scores fluctuated between two instances of 40.00% in Weeks 23 (0.763 types per 100 words) and 34 (0.877 types per 100 words) and peaks scores of 100.00% (1.626 types per 100 words) in Week 19 and 60.00% (0.870 types per 100 words) in Week 31. This phase of high variability was followed by another linear increase in scores with a relatively smaller bandwidth between minimum and maximum scores in phase 3. During this phase, scores increased from 47.06% (1.268 types per 100 words) to a final score of

80.00% (1.732 types per 100 words). The proportion scores of 100.00% (0.658 types per 100 words) is a result of that text containing only one adjective-noun collconstruction (*sad history*).
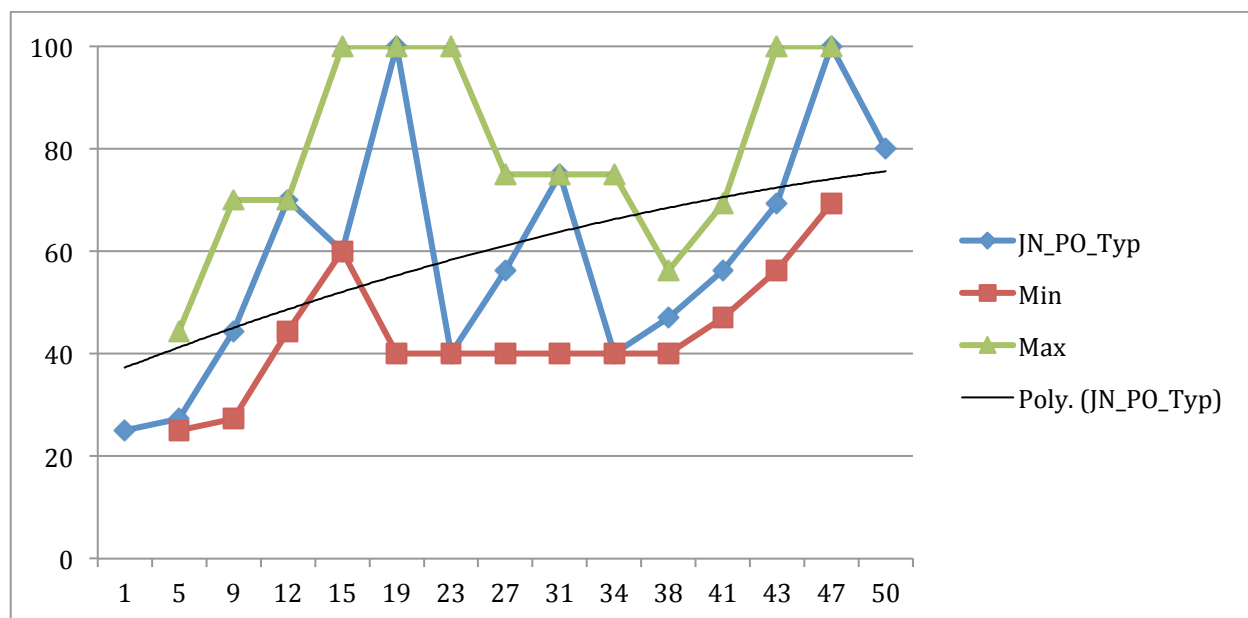

Figure 7.5 Marta's positively associated adjective-noun collconstruction type proportions over time

7.3.2 Verb-Noun Collconstructions

For the analysis of verb-noun collconstruction development, I chose to focus on EunHui. Her token frequencies (r = .752), normalized entropy scores (r =.226), N/A type proportion scores (r = -.521), and non-collexeme type proportion scores (r = .617) showed the strongest average correlations with time for this structural collconstruction category. The moving min-max graph for EunHui's verb-noun collconstruction token frequency is shown in Figure 7.6. The figure suggests that, over the course of one year, EunHui increased her use of verb-noun collconstructions in her English freewriting. Her token frequencies increased from 4.068 tokens per 100 words to 4.425 tokens per 100 words. Although her final token frequency may not be that much larger than her initial token frequency, it is still higher than her lowest frequency of 2.614 in Week 15. As can be seen in Figure 7.6, changes in EunHui's verb-noun token frequency occurred nonlinearly in three phases. These three phases showed alternating patterns of decline

and increase in verb-noun token frequencies. The initial phase, lasting from Week 1 to Week 15, was marked by a decrease in token frequency from 4.068 to 2.614 tokens per 100 words. This initial decline in token frequency was followed by the second phase (Weeks 20-34) in which her token frequencies increased linearly. Token frequencies during this phase increased from 2.839 to 6.222 tokens per 100 words. In the final phase, which began in Week 39 and lasted to the end of the study period, EunHui's verb-noun token frequencies again experienced a decline, ending at 4.425 tokens per 100 words in Week 50.
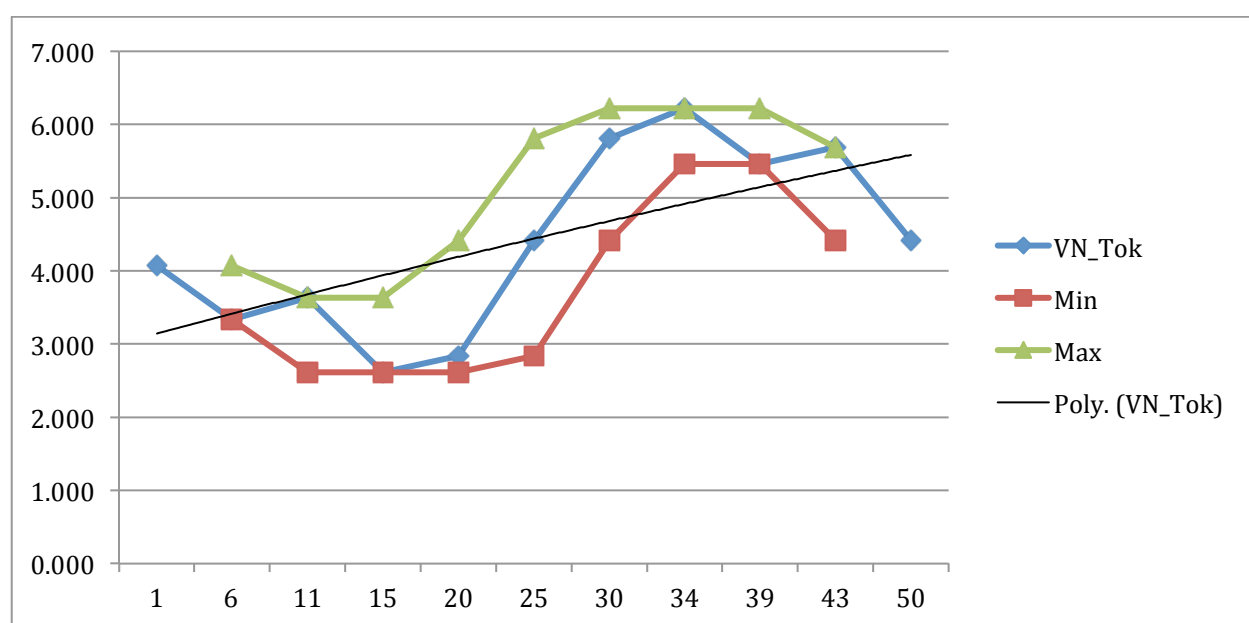

Figure 7.6 EunHui's verb-noun collconstruction token frequency development

Figure 7.7 illustrates the development of EunHui's normalized entropy scores for verb-noun collconstructions. From Weeks 1 to 20, her normalized entropy scores alternate between a perfect 1.000 and 0.946. This period of high variability is followed by a period in which her scores become more stable, with the bandwidth between her minimum and maximum normalized entropy scores narrowing in Weeks 20 through 34. The bandwidth stays relatively narrow as, starting in Week 39, her entropy scores begin to decrease. These findings suggest that, over the

course of one year, EunHui's ability to repeat some collconstructions within the same text increased and became more stable.
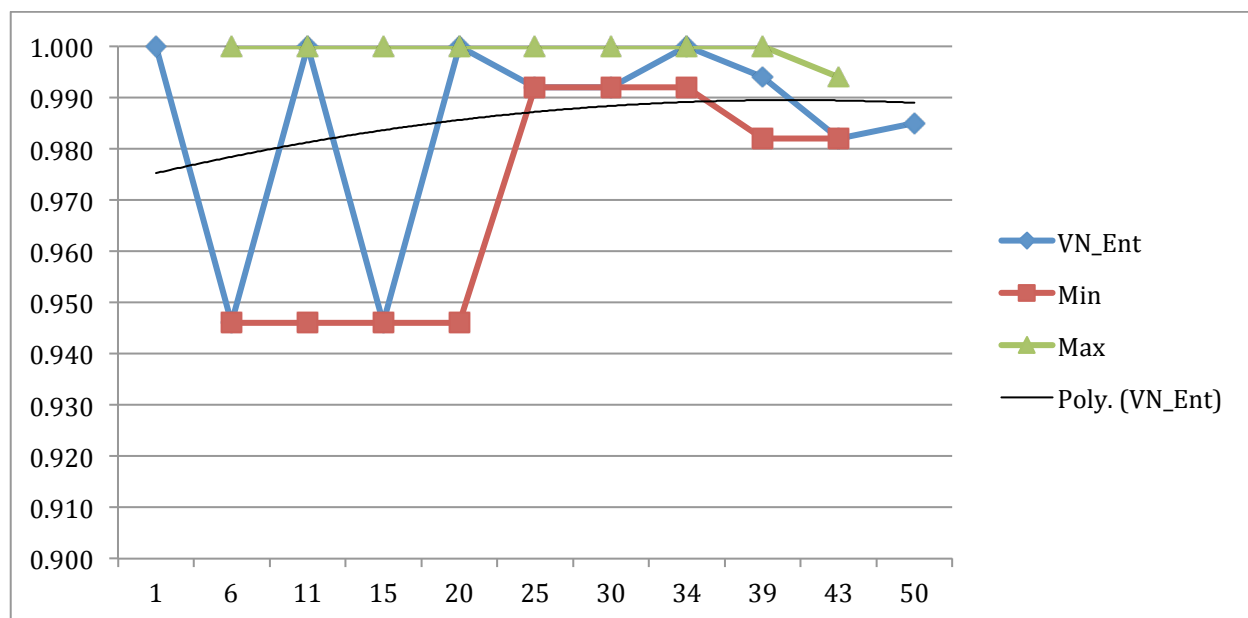

Figure 7.7 EunHui's verb-noun collconstruction entropy development

Figures 7.8 and 7.9 present the trajectories for EunHui's N/A type and non-collexeme proportion scores over the course of one year of study. Looking at Figure 7.8, there appears an overall decline in EunHui's reliance on N/A verb-noun collconstruction types. This decline occurs in three phases throughout the study period. The first phase lasts from Week 1 to Week 15. During this phase, EunHui's N/A verb-noun type proportion scores fluctuate between 83.33% (3.390 types per 100 words) in Week 1, 33.33% (0.833 types per 100 words) in Week 6, and 66.67% (1.307 types per 100 words) in Week 15. In phase two (Weeks 20-30), these scores experience a linear decrease from 59.09% (1.677 types per 100 words) to 38.78% (2.045 types per 100 words). In the final phase, which lasted from Week 34 to the end of the study period, EunHui's N/A type proportion scores became increasingly stable. During this period, her use of N/A types entered an attractor state as her scores fluctuated between 50.00% (3.111 types per 100 words) and 33.33% (2.192 types per 100 words).
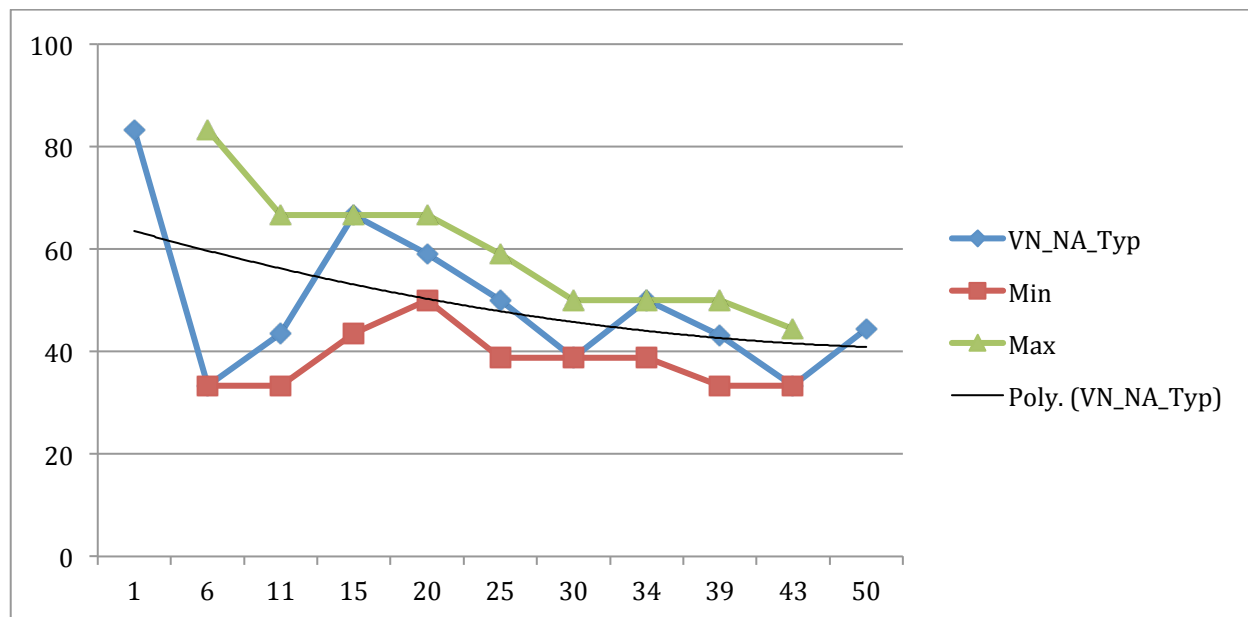
Figure 7.8 EunHui's verb-noun collconstruction N/A proportion development

The min-max graph in Figure 7.9 shows that EunHui's reliance on non-collexeme verb-noun collconstructions increases over the study period. From Week 1 to Week 50, her proportion scores increase from 8.33% (0.339 types per 100 words) to 22.22% (0.885 types per 100 words). This change in her use of non-collexeme verb-noun collconstructions occurs in three phases. In the first phase, which lasted from Week 1 to Week 11, EunHui's non-collexeme type proportions are fairly stable as they fluctuate between 0.00% and 13.04% (0.474 types per 100 words). The second phase starts at Week 15 and continues until Week 30 as her proportion scores increase and peak at 38.77% (2.045 types per 100 words). Following this phase of progress in non-collexeme proportion scores, her scores enter another phase of relatively strong stability, with scores remaining between a maximum of 22.22% (0.885 types per 100 words) and 16.66% (0.813 types per 100 words). Even though the scores during this phase are lower than the peak score attained in Week 30, they are all still higher than her initial proportion score.
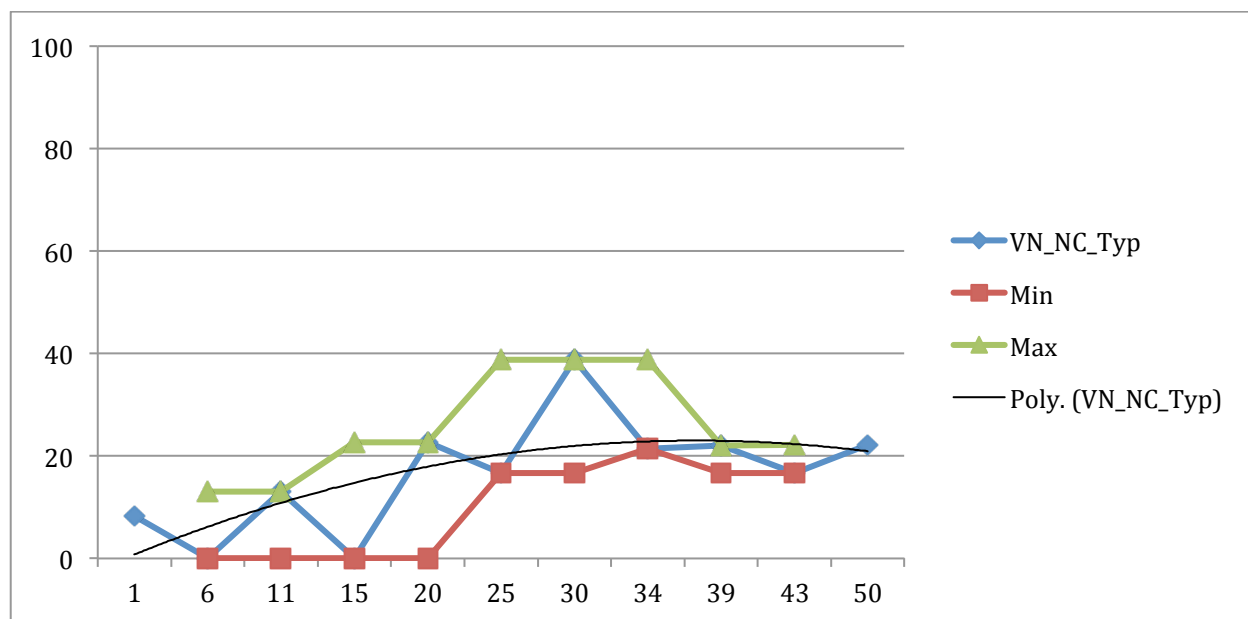
Figure 7.9 EunHui's verb-noun non-collexeme collconstruction proportion development

To further illustrate these findings, Table 7.3 displays the verb-noun collconstructions in EunHui's first (Week 1) and last (Week 50) texts. In Week 1, EunHui produced 12 verb-noun collconstructions, 9 of which were not found in the academic reference corpus. In her Week 50 text, this proportion of N/A types shrinks as only 3 of 9 collconstructions are not attested in the academic reference corpus. Although she still produces the same number of non-collexeme types, EunHui also produces two significantly associated types (*smoke cigarette, watch movie*). So, even though it may appear that her increasing proportion of positively associated collconstructions is simply the result of her producing fewer types overall, there is evidence that her knowledge of positively associated verb-noun collconstructions has in fact increased over the course of the year.

Table 7.3 Verb-noun collconstructions in EunHui's Week 1 and Week 50 texts

Week 1		Week 50	
Collconstruction	Category	Collconstruction	Category
accept information	NA	agree opinion	NA
answer living	NA	breath air	NA
have examination	NA	breath smoke	NA
have plan	NG	cause feeling	NG
hit racquetball	NA	cause pollution	NC
like Autumn	NA	have effect	NC
make friend	NC	have reason	NG
need money	NC	smoke cigarette	CO
open account	NA	watch movie	CO
read weather	NA		
receive bow	NA		
ruffle temper	NA		

Note. NA, NG, NC, and CO stand for N/A, negative, non-collexeme, and collexeme collconstruction categories.

Similar to the analysis for Marta's adjective-noun collconstructions, I created a combined category for EunHui's positively associated verb-noun collconstruction types composed of types from the non-collexeme and collexeme categories. I then charted the trajectory of the proportion scores for this category in a min-max graph, shown in Figure 7.10. As can be seen in the graph, EunHui shows an increasing reliance on positively associated verb-noun collconstruction types. From Week 1 to Week 50, her positive type proportion types scores increase from 8.33% (0.339 types per 100 words) to 44.44% (1.770 types per 100 words). However, the trajectory of this increase exhibits a high degree of variability throughout the study period. The variability is the highest in the initial phase of the study period, which lasted from Week 1 to Week 15. Scores during this time fluctuated between 0.00% and 30.44% (1.106 types per 100 words). This phase is followed by a phase in which the bandwidth between minimum and maximum scores narrows and her proportion scores increase to a peak of 40.82% (2.153 types per 100 words) in Week 30. From Week 34 to Week 43, EunHui's scores enter a phase of increased stability, with scores fluctuating between 16.67% (0.813 types per 100 words) and 29.66% (1.504 types per 100

words). Similar to the results for the non-collexeme type proportion scores during these weeks, these scores are all above EunHui's initial positive type proportion score. This phase ends with her positive type proportion score again increasing at Week 50 to 44.44% (1.770 types per 100 words).
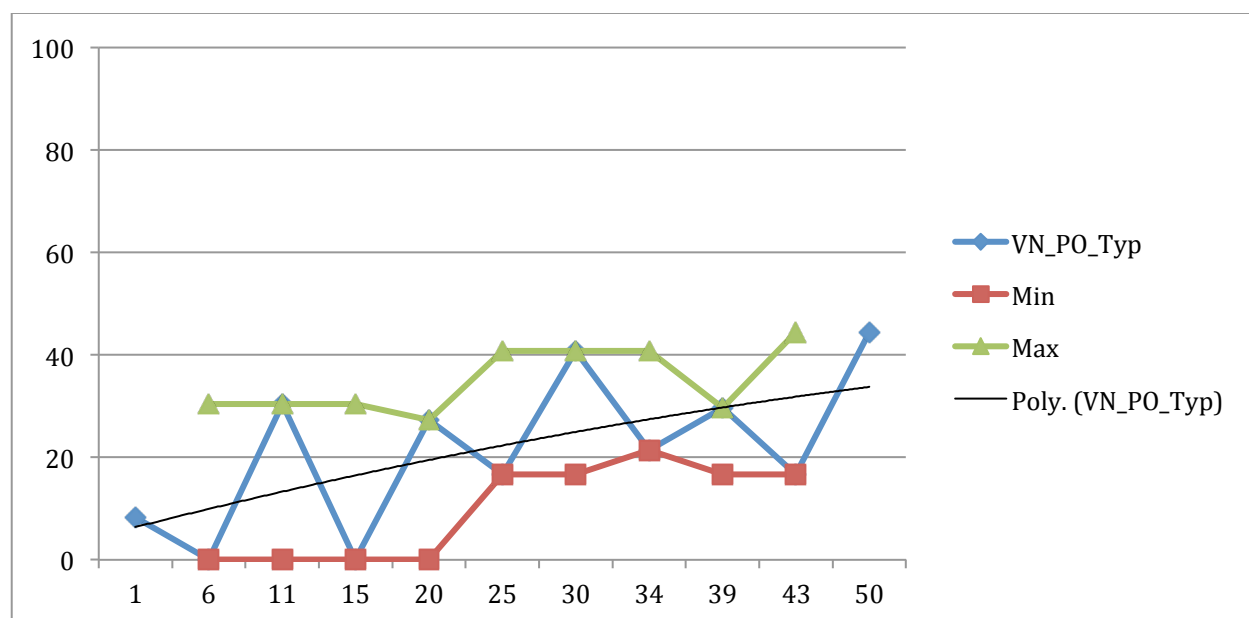

Figure 7.10 EunHui's positive verb-noun collconstruction type proportion score development

7.3.3 Adverb-Adjective Collconstructions

For the analysis of adverb-adjective collconstruction development, I chose to focus on Takako as his token frequencies ($r = .131$), normalized entropy scores ($r = -.080$), N/A type proportion scores ($r = -.338$), and non-collexeme type proportion scores ($r = .651$) showed the strongest average correlations with time for this structural collconstruction category. The min-max graph for Takako's adverb-adjective token frequencies over the course of one year is shown in Figure 7.11. To begin, it can be seen in the graph that adverb-adjective collconstructions, compared to adjective-noun and verb-noun collconstructions, did not occur very frequently in the learner data. In 5 of Takako's 11 texts, adverb-adjective token frequencies were below 1 token per 100 words. In terms of the overall trajectory, Takako's use of adverb-adjective showed little

growth over the study period. In Week 1, his token frequency was 0.319 tokens per 100 words. This only increased slightly over the study period, ending at 0.820 tokens per 100 words. Between those points, there was a high degree of variability with large bandwidths between minimum and maximum token frequencies. His token frequencies ranged from peaks of 2.174 tokens per 100 words in Week 5 and 3.070 tokens per 100 words in Week 34 to minimum scores of 0.688 tokens per 100 words in Week 15 and 0.716 tokens per 100 words in Week 39. Similarly, the min-max graph in Figure 7.12 demonstrates that there was little development in Takako's adverb-adjective normalized entropy scores over the course of one year. Outside of Week 5 and 50 when his scores were 0.000 (due to his use of only one adverb-adjective type [*little sad* in Week 5 and *very bad* in Week 50]), his normalized entropy scores fluctuated between 1.000 and .894 throughout the study period.
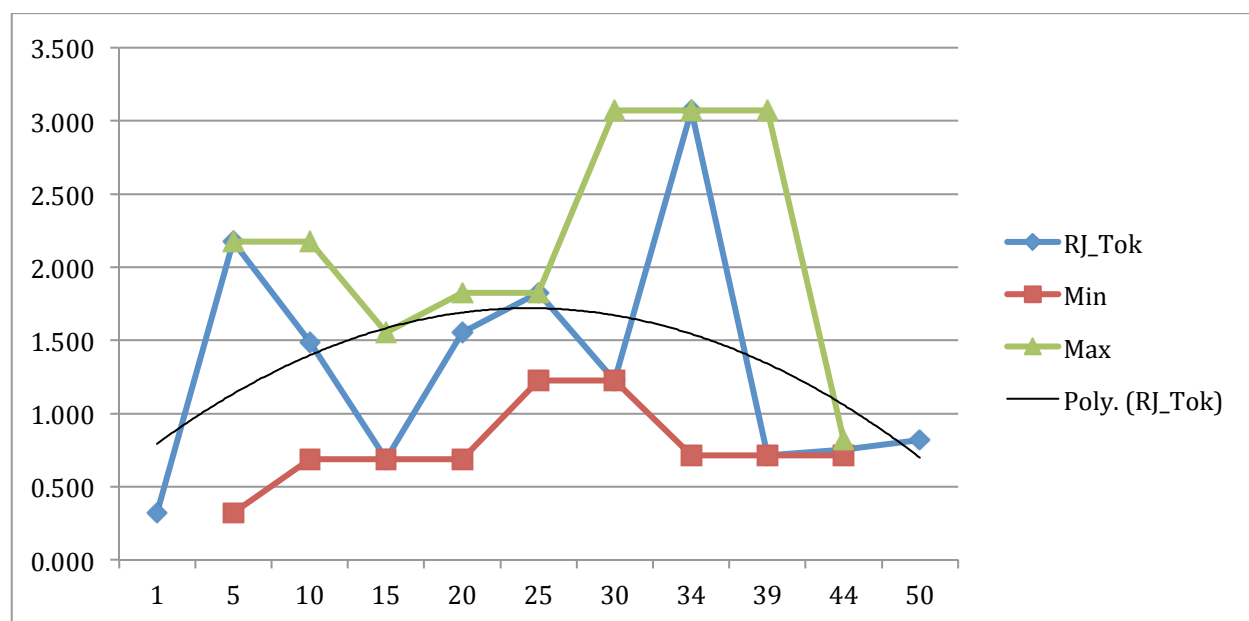

Figure 7.11 Takako's adverb-adjective collconstruction token frequency development
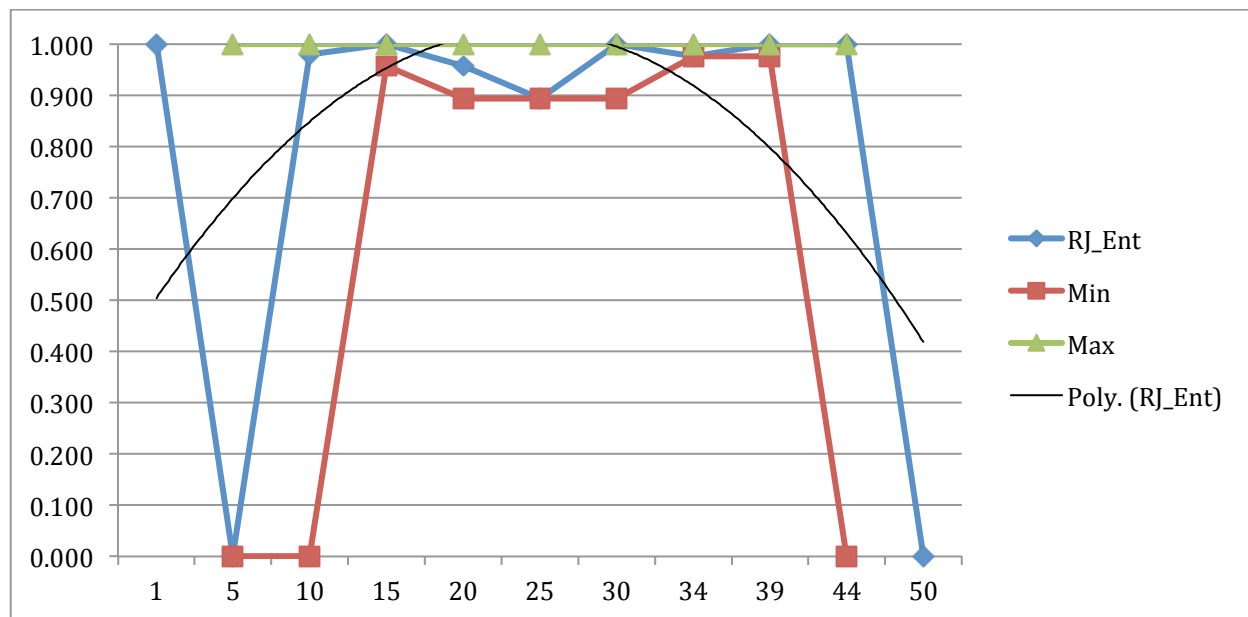
Figure 7.12 Takako's adverb-adjective collconstruction normalized entropy development

Figures 7.13 and 7.14 show the min-max graphs for Takako's N/A type and non-collexeme type proportion scores over the course of one year. While the polynomial trend lines in both graphs suggest that there is growth over time for both indices, these results should be treated with caution. These trend lines are most likely the result of extreme outliers in each graph. In figure 7.13, for instance, the N/A proportion score in Week 5 is 100.00% while the scores for all other weeks never reach above 30% and fluctuate throughout the study period. This extreme outlier is due to the fact that Takako produced only one adverb-adjective collconstruction type (*little sad*). In figure 7.14, the non-collexeme type scores in Weeks 1 and 5 scores are at 0.00%, while all other scores are above 60% and also fluctuate between 60% and 100% throughout the study period. Again, these outliers were most likely the result of Takako producing only one collconstruction type in each of those weeks. Taken together with the normalized entropy results, it appears that adverb-adjective collconstructions did not occur frequently enough in the learner texts for any strong changes to occur for these beginner learners.
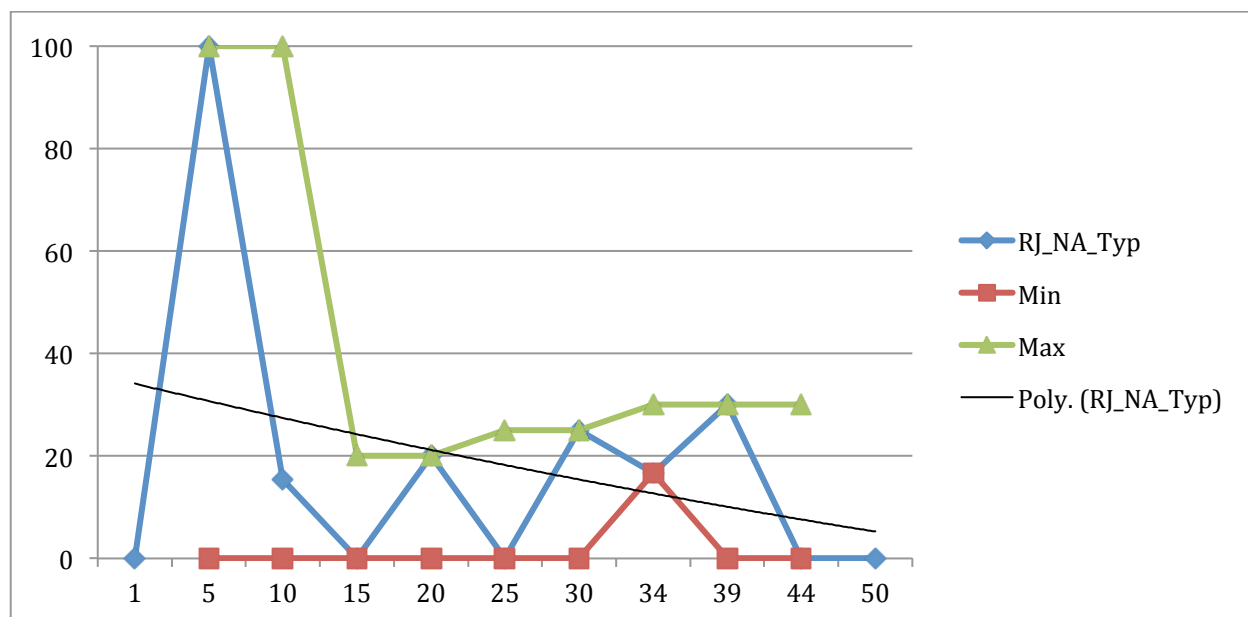
Figure 7.13 Takako's N/A adverb-adjective collconstruction type proportion score development
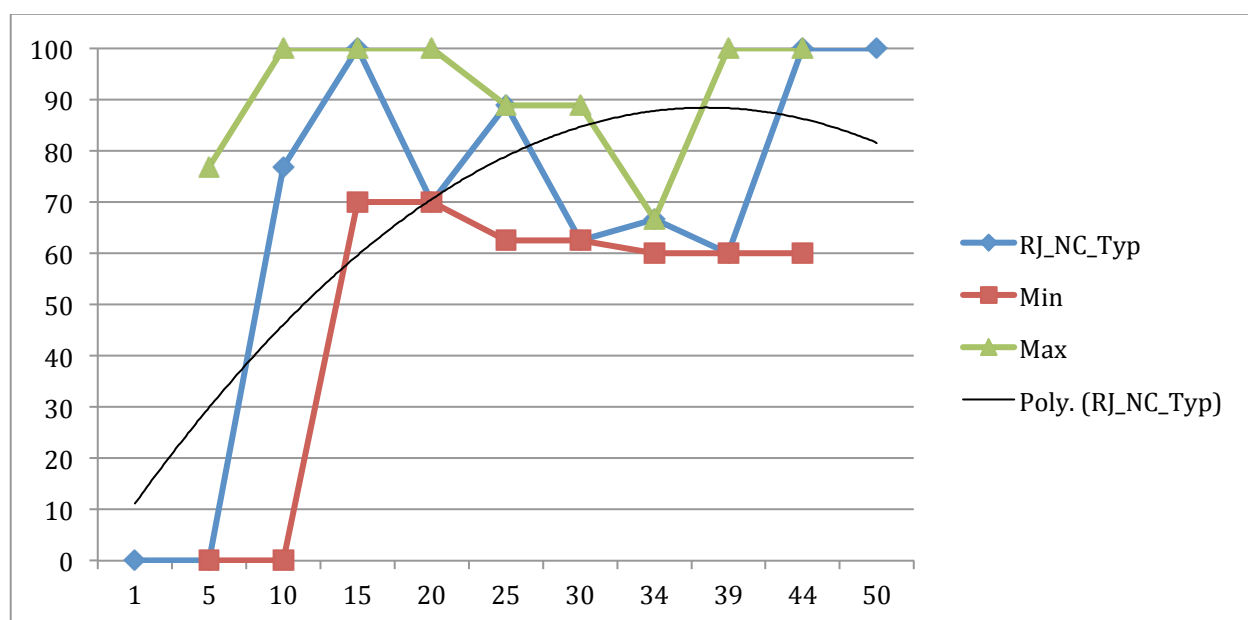
Figure 7.14 Takako's non-collexeme adverb-adjective collconstruction type proportion score development

7.3.4 Interconnected Development of Collconstructions

In this section, I examine interconnected development for the different indices of productive collconstructional knowledge both within and across structural collconstruction categories. I do this by calculating and charting moving correlations (window of three time

points) between the following indices: (1) Marta's adjective-noun token frequencies and non-collexeme type proportion scores, (2) EunHui's verb-noun token frequencies and non-collexeme type proportion scores, (3) EunHui's verb-noun and adjective-noun token frequencies, and (4) Marta's adjective-noun and verb-noun N/A type proportion scores. These charts allow for a visual examination of how different types of productive collconstructional knowledge either support each other's development or compete for resources (Caspi & Lowie, 2006, 2010; Larsen-Freeman, 2006; Spoelman & Verspoor, 2010; Verspoor et al., 2008; Zheng, 2016).

Figure 7.15 shows the moving correlations for Marta's adjective-noun token frequencies and non-collexeme type proportions. The figure shows that these two types of adjective-noun collconstructional knowledge alternate between periods of competing and supportive development. In Weeks 5 and 9, these two indices are negatively correlated. This indicates that, initially, there was a trade-off between growth in Marta's ability to produce positively associated adjective-noun combinations and her ability to produce adjective-noun collconstructions frequently. This state of competing development decreases, however, and in Weeks 12 through 19, the two collconstruction knowledge types are in a more supportive relationship in which both indices experience high degrees of variability. This supportive relationship again gives way to one of competing development in Weeks 23 through 38, suggesting that, similar to Weeks 5 and 9, Marta's production of more non-collexeme collconstruction types during this time came at the expense of her producing fewer adjective-noun tokens overall and vice versa. This relationship between non-collexeme type proportions and token frequencies becomes more supportive in Week 41, but then returns to a more competitive one in Week 47.
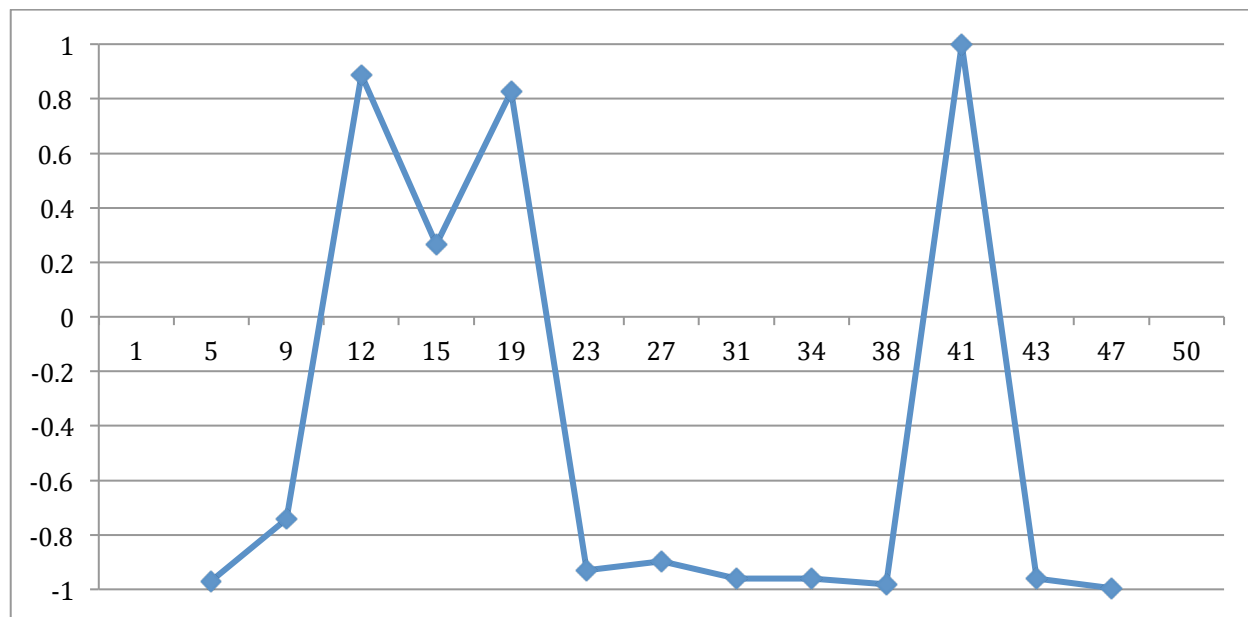
Figure 7.15 Moving correlations between Marta's adjective-noun collconstruction token frequency and non-collexeme type proportion

Figure 7.16 displays the moving correlations between EunHui's verb-noun collconstruction token frequency and non-collexeme type proportion scores. As can be seen in the figure, these two types of collconstructional knowledge exist in a mostly supportive relationship during the year. The positive correlations between the two in Weeks 6 through 30 and again in Week 39 demonstrate that as EunHui increased production of verb-noun collconstructions, she increased her use of positively associated verb-noun combinations. These two types of knowledge did, however, enter a competitive state of interconnected development in Week 34 and again in Week 43. The low correlation ($r = -.07$) in Week 34 indicates that the level of competition between these knowledge types at that time was minimal, although this competition becomes stronger in Week 43.
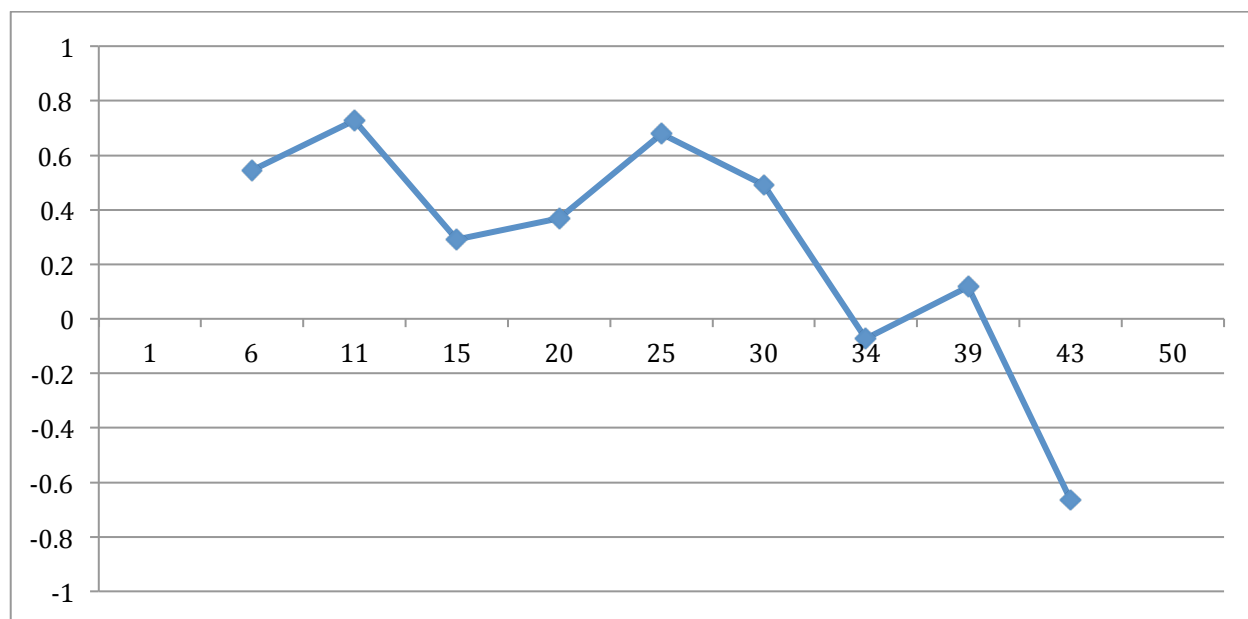
Figure 7.16 Moving correlations between EunHui's verb-noun collconstruction token frequency and non-collexeme type proportion scores

Shifting focus to interconnected development across structural collconstruction categories, the moving correlations between EunHui's verb-noun and adjective-noun token frequencies are shown in Figure 7.17. These two indices initially exist in a supportive relationship, as indicated by the positive correlations in Weeks 6 through 15. This supportive relationship between the two token frequency indices gives way to a more competitive one in Weeks 20 and 25. There is a brief return to a supportive relationship in Week 30, before returning to a more competitive one in Weeks 34 and 39. To allow for further investigation of the causes behind this shifting relationship between adjective-noun and verb-noun collconstruction token frequency development, Figure 7.18 displays the token frequencies for adjective-noun and verb-noun collconstructions in EunHui's freewrites. If we examine the frequencies in Figure 7.18, we can see that the initial positive relationship between adjective-noun and verb-noun token frequencies is the result of both indices fluctuating up and down in the initial phase of EunHui's freewriting. The negative correlations between the two indices in Weeks 20 and 25 appear to be the result of differences in when growth begins for each of the indices. Verb-noun

collconstructions begin increasing in frequency in Week 25, while adjective-noun collconstruction token frequencies do not begin to increase until Week 30. Once both token frequencies begin to increase, they enter the positive and supportive relationship exhibited by the high positive correlation in Week 30. This positive relationship ends in Week 34 as verb-noun collconstruction token frequency reaches its peak and begins to level off while adjective-noun collconstruction token frequency continues to develop. The positive relationship returns, however, in Week 43 as both token frequencies increase from Week 39 to Week 43 and then decline in Week 50.
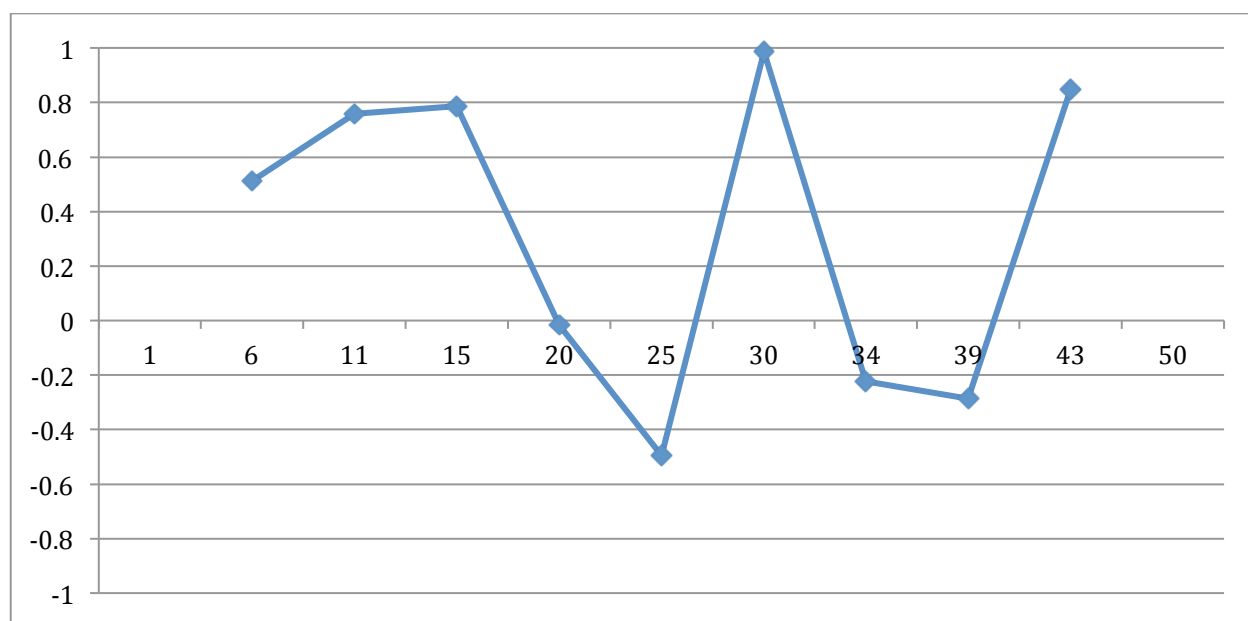

Figure 7.17 Moving correlations between EunHui's adjective-noun collconstruction and verb-noun collconstruction token frequencies
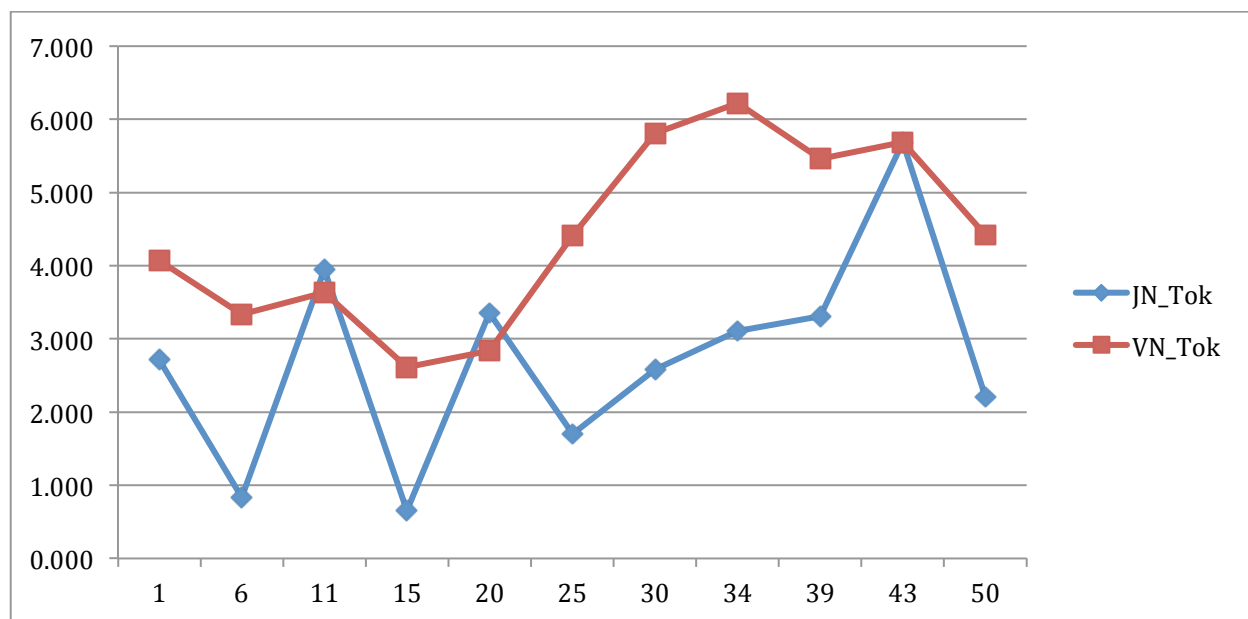
Figure 7.18 EunHui's adjective-noun and verb-noun collconstruction token frequencies

Figure 7.19 illustrates the moving correlations between Marta's adjective-noun N/A type proportion scores and her verb-noun N/A type proportion scores, while Figure 7.20 displays their values over the course of the year. The correlations suggest that, for most of the year, these two types of productive collconstruction knowledge are positively associated and support each other's development. At 9 of 13 time points where moving correlations were calculated, the correlations were positive, indicating that these two indices either increased or decreased together. In Week 9, the correlation was small ($r = .020$), most likely due to the fact that, in Week 12, adjective-noun N/A type proportion decreased while verb-noun N/A type proportion increased. The four time points at which the moving correlations were negative were Week 19 and Weeks 34 through 41. The negative correlation in Week 19 occurred as Marta's verb-noun N/A type proportion scores increased from Week 15 to 23 while her adjective-noun N/A type proportion scores fluctuated between 20.00% and 0.00% during those three time points. Concerning the negative correlations in Weeks 34 through 41, the data in Figure 18 suggests that this is due to the fact that Marta's adjective-noun N/A type production is at a stage of high

variability, while her verb-noun N/A type production is fairly stable. From Week 31 to 43, her verb-noun N/A type proportion scores range from 25.00% to 36.36%. During that same time span, her adjective-noun N/A type proportion scores increase from 0.00% in Week 31 to 60.00% in Week 34. This increase is then followed by a decrease to 28.57% in Week 38 and a slight increase to 31.25% in Week 41. This finding suggests that productive collostructional knowledge of adjective-noun and verb-noun collconstructions predominantly support each other in their development, although there may be points in which one knowledge type enters a period of relative stability while the other one continues to experience variability.
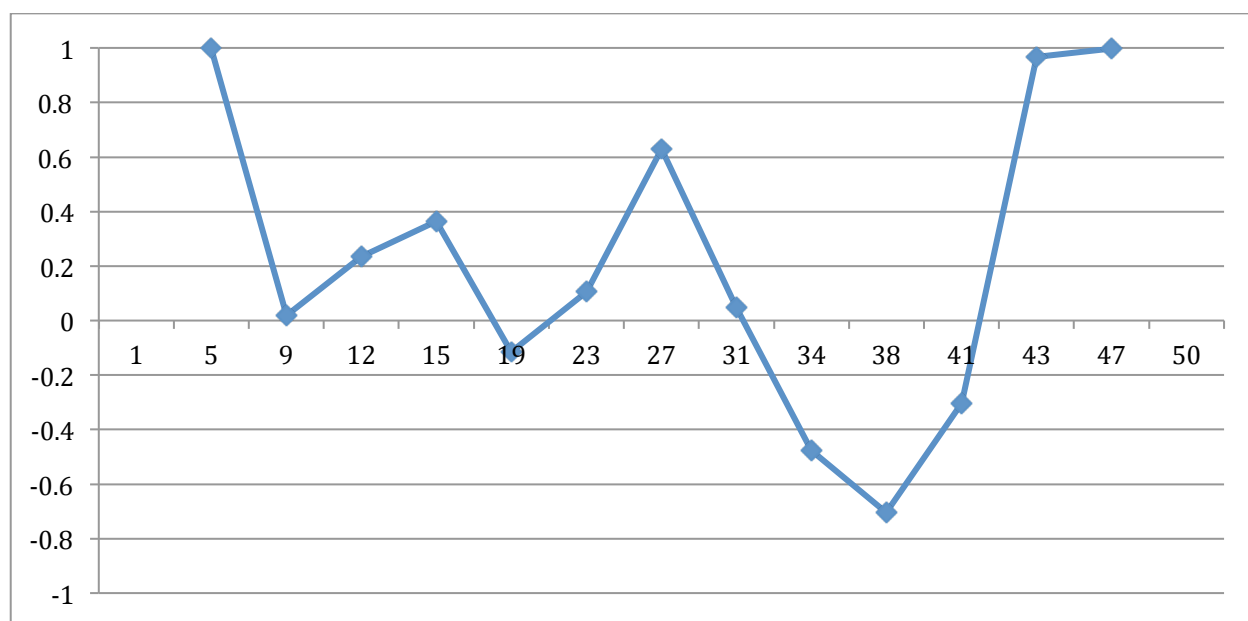
Figure 7.19 Moving correlations between Marta's adjective-noun collconstruction and verb-noun collconstruction N/A type proportions
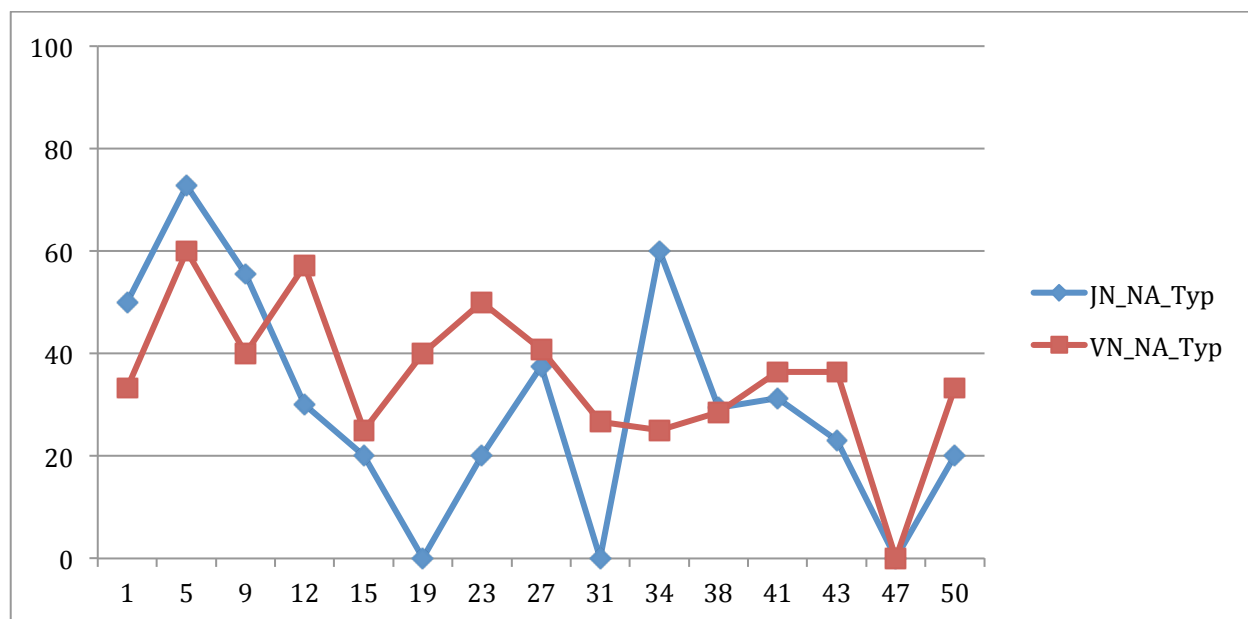
Figure 7.20 Marta's adjective-noun and verb-noun collconstruction N/A type proportion scores

7.4   Summary and Discussion

This chapter explored how productive adjective-noun, verb-noun, and adverb-adjective collconstruction knowledge develops in three individual ESL learners over the course of one year. By charting growth trajectories for individual learners, it was revealed that two of the three learners developed their ability to produce the different categories of collconstructions in their writing. However, the results also showed that the learners differed in both what specific types of collconstruction knowledge showed the strongest development and how these knowledge types developed. Marta, for instance, greatly increased the proportion of positively associated adjective-noun collconstructions she produced in her writing. There was, however, a trade-off in Marta's adjective-noun collconstruction production, as the growth in proportion scores came at the expense of her producing fewer adjective-noun tokens. During the year, she also developed her use of positively associated verb-noun collconstructions, as evidenced by the mostly positive correlations between the two proportion score indices. Together, these findings indicate that, for Marta, productive collconstruction knowledge development mostly consisted of an increase in

her ability to produce nativelike collconstructions. Development for EunHui, on the other hand, mostly occurred in her use of verb-noun collconstructions. In addition, unlike Marta, whose development in adjective-noun non-collexeme type proportion scores came at the expense of her adjective-noun token frequencies, EunHui managed to develop her ability to produce a greater proportion of verb-noun non-collexeme types while simultaneously increasing how frequently she produced verb-noun collconstructions.

Taken together, these results provide some support for those from Li and Schmitt (2009) and Li and Schmitt (2010). Similar to the findings of the current analysis of beginning learner writing, both studies showed that advanced L2 writers take different trajectories in developing their productive knowledge of lexical phrases and adjective-noun bigrams. The authors also showed that phraseological unit development often occurs non-linearly. In Li and Schmitt (2010), one learner showed growth in all indices of academic adjective-noun bigrams, one showed a decline in all indices of bigram production, and another showed growth in some indices and decline in others. In Li and Schmitt (2009), the authors found that their subject's use of lexical phrases, while showing overall development, fluctuated from text to text.

The results presented in this chapter also support findings from DST studies focusing on the role of variability and interconnected development in longitudinal L2 development. Regarding the former, these studies have demonstrated that language development usually occurs in a step-wise fashion, with phases characterized by a high degree of variability often preceding phases of strong growth (Spoelman & Verspoor, 2010; Verspoor et al., 2008; Zheng, 2016). Similarly, it was found in the study presented in this chapter that phases of growth in collconstruction production were preceded by phases of high variability. For instance, EunHui's verb-noun N/A type proportion scores initially existed in a state of high variability before

entering phases of decline and eventual stability. The results presented in this chapter also

highlighted the step-wise nature of development according to a DST perspective. For several

indices, change occurred as they alternated between phases of mostly growth or decline and

phases of high variability. For example, Marta's growth in positive adjective-noun

collconstruction types occurred as her scores underwent an initial phase of growth, then entered a

phase marked by a high degree of variability, and then entered another phase of mostly linear

growth. Concerning interconnected development, DST research has shown how at different

stages of the developmental process, some types of linguistic knowledge interact to either

support each other's development or compete for resources (Caspi & Lowie, 2006, 2010; Larsen-

Freeman, 2006; Spoelman & Verspoor, 2010; Verspoor et al., 2008; Zheng, 2016). In this

chapter, it was shown that the same occurs for different types of collconstruction knowledge,

with learners experiencing growth in one index of collconstruction production (e.g. collexeme

strength) at the expense of another (e.g. token frequency).

Despite the strength of these results, it is important to note that texts analyzed in this

chapter did not contain a very large amount of collconstruction tokens. For instance, Marta's

texts never contained more than 6 adjective-noun collconstruction tokens per 100 words.

Similarly, verb-noun collconstructions never occurred more than 6.25 times per 100 words.

These small token frequencies limit the strength of these results and require a level of caution

when interpreting the results for the proportion scores, as small changes in frequency for a

collexeme category could lead to large changes in proportion scores for that category.

In conclusion, this chapter has shown how productive collconstruction knowledge

develops for individual learners over the course of one year. It has specifically highlighted the

individualized, highly variable, and multi-faceted nature of the development of productive

collconstruction knowledge for beginner L2 English learners. As such, this chapter has also provided strong support for the argument that future L2 development research should adopt a DST perspective towards longitudinal research in order to fully capture a true representation of the language development process.

8   CONCLUSION AND OUTLOOK

The goal of this dissertation was to refine current approaches to phraseology research in learner corpora by defining lexical collocations as constructions ("collconstructions") within a Construction Grammar perspective and analyzing their variation across proficiency levels and development over time. In order to accomplish this, a definition of collconstructions was proposed that conceptualized lexical collocations (Granger and Paquot, 2008) as constructions (Goldberg, 1995, 2006, 2013) that exist within the constructicons of fluent language users. A framework for analyzing collconstruction production in learner language was created and utilized to examine variation in a cross-sectional corpus of L1 Korean EFL learners and development in longitudinal corpus of freewrites produced by ESL learners. Variation was observed for different types of collconstruction knowledge in different structural and functional collconstruction categories across proficiency levels. Changes in collconstruction production were also observed for individual learners over time. A summary of the findings and a discussion of their contributions and implications for future research are provided below.

8.1   Summary of Findings

8.1.1   Research Question 1: Cross-sectional Collconstruction Development

The qualitative analysis of functional collconstruction subcategories across learner proficiency levels appeared to show that high-beginner, low-intermediate, and high-intermediate learners differed in their use of certain functional subcategories. These results suggest that for some functional collconstruction categories, such as [positive] + *experience*, more proficient learners were less reliant on general purpose adjectives and adverbs and produced a wider range collconstructions than less proficient learners. For another functional collconstruction category, the [acquire] + *money* collconstruction category, more advanced learners seemed to produce a

more restricted and more semantically coherent range of collconstructions. In both cases, the collconstructions more frequently used by the high-intermediate learners were more targetlike than those produced by the high-beginner learners. These findings support those from previous studies of productive constructional knowledge development that have shown how learners become more schematic and targetlike in their use of constructions in their L2 (Ellis & Ferreira-Junior, 2009a; Eskildsen, 2009, 2012, 2015, Mellow, 2006; Myles et al., 1998, Römer & Garner, under review).

The results of the ordinal logistic regression indicated that L1 Korean EFL writers judged to be more proficient by trained human raters were more likely to produce fewer adverb-adjective collconstructions in their writing. The results also showed that more proficient writers were more likely to produce a more restricted range of verb-noun collconstructions that is more Zipfian in its distribution. None of the collexeme proportion score indices for any of the structural collconstruction categories were found to be predictive of human judgments of writing quality. These results seem to go against the findings in Garner, Crossley, and Kyle (in press), who found significant differences in bigram association scores across the same levels of proficiency. Given that the current study focused specifically on adjective-noun, verb-noun, and adverb-adjective collconstructions, these findings suggest that, between beginner and intermediate stages of L2 writing, development in the use of strongly associated phraseological units may be occurring for other types of phraseological units (e.g. grammatical collocations, phrasal verbs) but less so for the units under analysis here.

8.1.2   Research Question 2: Group Level Longitudinal Collconstruction Development

The results of this analysis indicated that there was no significant change in the ESL learners' use of adjective-noun, verb-noun, and adverb-adjective collconstructions over the

course of one year at the group level. No significant effects on collconstruction production were found for overall language proficiency scores, indicating that growth in language proficiency did not correspond to change in collconstruction production. Although this finding confirms similar findings from previous longitudinal studies of bigram association strength development in L2 writing (Bestgen & Granger, 2014), these results must be taken with some caution. Follow-up graphical analysis of growth trajectories for individual learners revealed significant inter- and intraindividual variability in the use of collconstructions over the study period. Combined with the low sample size, this may have resulted in models that suggest that no development is occurring when it actually is occurring for some learners.

8.1.3 Research Question 3: Longitudinal Collconstruction Development for Individual Learners

The results indicated that, over the course of one year, individual learners made gains in their production of different structural categories of collconstructions. These results also showed that growth occurred differently for each of the learners. For example, growth for one learner occurred as she increased her use of more strongly associated and targetlike collconstructions, with the strongest gains seen for adjective-noun collconstructions. This growth in association strength for collconstructions, however, came at the expense of her producing fewer adjective-noun collconstruction types. For another learner, the greatest growth came in her verb-noun collconstruction production as she increased how frequently she produced them and how strongly associated they were. These findings support those from Li and Schmitt (2009, 2010) who found that development in productive phraseological unit knowledge for advanced L2 writers varies from learner to learner. Additionally, periods of strong change in collconstruction production for both learners in this analysis were often preceded by phases in which

collconstructions varied widely from text to text. These findings support those from previous DST studies that have demonstrated the strong role variability can have in longitudinal L2 development and how different types of linguistic knowledge influence each other in the development (Caspi & Lowie, 2010, 2013; Larsen-Freeman, 2006; Spoelman & Verspoor, 2010; Verspoor et al., 2008; Zheng, 2016).

8.2   Contributions

This dissertation makes several important contributions to learner corpus research on L2 phraseology development. The first major contribution this dissertation makes is that it provides a method for analyzing collocations in learner language that combines both covarying collexeme analysis and analysis of the lexical functions of collocations. This approach has the benefit of providing a framework for including both functional and constructional aspects when analyzing collocation production by L2 learners. Additionally, this approach provides a way for accounting for the strong associations lexemes may have with certain grammatical constructions (Stefanowitsch & Gries, 2009) as well as their colligational primings (Hoey, 2005). By defining collocations as constructions, researchers can analyze co-occurrence patterns for adjectives, nouns, verbs, and adverbs as they fulfill specific lexical functions and within the grammatical constructions in which they frequently co-occur. In doing so, this approach allows researchers to investigate productive collocational knowledge in a more comprehensive manner that more closely approximates how this knowledge is likely represented in the mental lexicon. By focusing on the association strength of words within the grammatical structures they frequently co-occur in, this approach also has the benefit of removing a lot of the noise when calculating statistical association measures.

Another contribution this dissertation has made to research on L2 phraseology is that it has highlighted the need to take a multi-faceted approach to phraseology. This approach includes examining different structural categories of phraseological units, different types of productive knowledge for phraseological units, and different functional subcategories of each structural category. Although cross-sectional changes for functional subcategories were not examined systematically, results from this part of the analysis did reveal that change was ongoing for some functional subcategories of verb-noun and adverb-adjective collconstructions. Similarly, although there were no changes in the association strength for verb-noun collconstructions, the more proficient learners were shown to use a more restricted range of these collconstructions than lower proficiency learners. Lastly, although a majority of indices for all collconstructions within a structural category showed no significant change across proficiency levels, variation was observed for certain functional subcategories. Thus, it is important for future research to examine phraseological unit production in L2 speech or writing from multiple perspectives in order to fully examine how the development occurs.

The results from the cross-sectional analysis have also shown how a multi-faceted approach to phraseology entails that we consider the possibility that different categories of collconstructions develop along different trajectories. For instance, normalized entropy results for two functional subcategories in the cross-sectional analysis showed contrasting trends in change across proficiency levels. For the adjective-noun collconstruction category [positive] + *experience*, entropy scores were higher in the high-intermediate subcorpus as the learners in this corpus used a wider range of targetlike adjectives. In contrast, these same learners produced a more restricted range of verbs in the verb-noun collconstruction [acquire] + *money* than the high-beginner learners. These findings, although based on small token frequencies for only two

functional subcategories, provide some preliminary evidence to suggest that different structural types of collconstructions may take different paths of development in learner language. It may be the case that, in verb-noun collconstructions, the noun restricts the number of targetlike verbs it can co-occur with to express a specific lexical function. For learners, the process of acquiring a specific verb-noun collconstruction token would, then, be a process whereby learners start with a wide variety of verb choices and gradually narrow their choices to only the most targetlike verbs. For adjective-noun collconstructions, the opposite may be true. In this case, learners may start with only one or two adjectives that are likely very general in their semantics and occur with a wide range of nouns (e.g. *good, bad*). As learners develop, their range of adjectives in the functional category increases as they acquire more adjectives and become more aware of which adjectives are most strongly associated with a target noun.

Another possibility is that different collconstruction categories proceed along the same path of development, but at different rates. In this scenario, development in collconstructional knowledge would occur along an upside down u-shaped trajectory. From beginning to intermediate stages of development, learners would go from relying on fixed exemplars to using constructions more schematically. From intermediate to advanced, the opposite would occur as learners begin to restrict their selection of lexemes in the construction to only those that are most idiomatic and targetlike. From this perspective, the results of the cross-sectional analysis suggest that verb-noun collconstructions begin to develop earlier in learner language than adjective-noun collconstructions. At the intermediate stages of language proficiency, learners may have already moved from relying on fixed verb-noun combinations to using them more schematically and are beginning to restrict their range of verb-noun combinations to those that are most targetlike. In contrast, intermediate learners might still be developing their ability to use a wide range of

adjectives in adjective-noun combinations. However, given the small number of collconstruction tokens in the dataset, more research with larger corpora is needed to more fully investigate these possibilities.

The longitudinal analyses in this dissertation also make several contributions to L2 phraseology research. First, they have shown how a multi-faceted approach to investigating L2 phraseology can provide more detailed information concerning the longitudinal development of productive phraseological knowledge for individual learners. Although the quantitative analysis found no significant changes over time at the group level, the results for the third research question demonstrated that different learners take different paths in their productive collconstructional knowledge development. Learners were found to differ in both the types of collconstructional knowledge (e.g. frequency, association strength) and the structural collconstruction categories (e.g. adjective-noun, verb-noun) that showed the greatest longitudinal development. These results confirm those from Li and Schmitt (2010), who found differences in how advanced L2 learners developed their collocational knowledge. In addition to providing support for multi-faceted approaches to L2 phraseology research, they provide support for approaches that examine individual development alongside development for groups of learners.

Second, the longitudinal analyses in this dissertation have provided support for emergentist theories (e.g. Complexity Theory, Dynamic Systems Theory) of language development that conceptualize the language learning process as dynamic and complex (Larsen-Freeman, 2012; van Geert & Verspoor, 2015). By charting the development of productive collconstructional knowledge for individual learners, it was shown that phases of high variability in the use of collconstructions were often immediately followed by periods of linear development. This confirms findings from previous studies, such as Spoelman and Verspoor

(2010), Verspoor et al. (2008), and Zheng (2016), who found similar patterns in grammatical and lexical development. The results also showed how collconstruction development, similar to development for other types of linguistic knowledge, develops in a step-wise fashion, with phases of relative growth or stability alternating with phases of high variability. Lastly, the results of this analysis showed how different types of collconstructional knowledge interact in the developmental process. Throughout the study period, it was found that, for one learner, frequency and association strength alternated between competing for resources and supporting each other's growth. Similar results were found for the same types of collconstructional knowledge (i.e. frequency, association strength) across structural categories. These results confirm those from previous studies that have shown how different types of single word knowledge interact in the developmental process (Caspi & Lowie, 2010, 2013; Zheng, 2016). Taken together, these results suggest that, similar to other linguistic subsystems (e.g. grammatical, lexical), L2 phraseological knowledge develops dynamically as multiple types of productive phraseological knowledge interact to produce a self-organizing phraseological system.

8.3   Implications and Future Directions

The findings of this dissertation have important implications for future L2 phraseology research, usage-based second language acquisition, and L2 phraseology instruction.

8.3.1   L2 Phraseology Research

This dissertation has provided an approach to examining collocations in learner language that more closely resembles the nature of word associations in the mental lexicon by analyzing lexical co-occurrence patterns within specific grammatical constructions as well as when they fulfill certain lexical functions. By defining lexical collocations as collconstructions, researchers

can examine how learners develop these constructions, the word associations that exist within them, and their use in fulfilling a wide range of lexical functions. Although the studies included in this dissertation focused only on three categories of lexical collconstructions (i.e. verb-noun, adjective-noun, adverb-adjective), future research could extend this framework to other types of phraseological units. These could include grammatical collocations, such as verb-preposition (e.g. *depend on*) and adjective-preposition (e.g. *interested in*). Extending the collconstructional framework to grammatical collocations is especially interesting given the contrasting findings of the cross-sectional analysis in this dissertation and findings from Garner et al. (in press) that showed significant differences in bigram association strength from high-beginner to high-intermediate L2 writing.

Future studies should also expand the analysis of different functional subcategories of collconstructions. Due to the relatively small sample size in this study, only one functional subcategory per structural category was conducted. These subcategories were also limited in that they only included one controlling lexeme (e.g. *experience* in the [positive] + *experience* collconstruction). Therefore, future research should use larger learner corpora to investigate more comprehensive categories of lexical functions as well as the whole range of lexical functions expressed through collconstructions in learner texts. Regarding the former, this research would involve identifying collconstructions that fulfill a specific standard lexical function, such as "positive evaluation" or "support" regardless of controlling lexeme (Mel'čuk, 1998, 2007). Concerning the analysis of a range of functional collconstructions, this research would classify all collconstructions within a structural category into categories of standard lexical functions.

There also exist two avenues for further development of how association strength for collconstructions is investigated. First, future research could include other measures of association between collocating words. One promising measure for future inclusion is ΔP, a measure of contingency between collocating words (Ellis, 2006a; Gablasova et al., 2017; Gries & Ellis, 2015). By including this measure, researchers can examine how the use of one lexeme in a collconstruction predicts the use of other lexemes and how learners develop knowledge of such relationships. Further support for the use of this measure of association strength can be found in two studies by Garner et al. (in press, under review), who found that bigram ΔP was the strongest predictor of human judgments of L2 writing proficiency among a range of n-gram indices.

Second, future collconstructional research could benefit from doing away with the distinction between non-collexeme and collexeme collconstructions. In the individual-level longitudinal analyses, graphs for the development of positively, yet not strongly, associated collconstructions gave the impression that learners made only small improvements in their use of positively associated collconstructions. However, subsequent analyses revealed that the learners also improved their use of strongly associated collconstructions and that collapsing these categories into one single category more accurately reflected changes in their collconstruction production. Thus, future collconstruction research may be better served by getting rid of threshold values for collexeme status and instead simply compare the use of collconstructions that are too infrequent for association score calculation, negatively associated collconstructions, and positively associated collconstructions.

This dissertation has also demonstrated how taking a multi-faceted approach to L2 productive collconstructional knowledge can provide a more comprehensive view of development across proficiency levels and over time. The studies included in this dissertation

have shown that some types of collconstructional knowledge may develop for some learners while other types of knowledge remain more static. These findings point to two avenues of further development in future learner corpus research on L2 phraseology. First, future research should include more than a single measure of frequency or association strength. Instead, research should include both frequency and association strength as well as others types of indices, such as diversity, contingency, or even native speaker judgments of appropriateness. Second, future studies of collocation development should, similar to the current project, include more than one structural type of collocation in their analysis. This could include verb-noun collocations, adjective-noun collocations, and multiple types of grammatical collocations (e.g. verb-preposition, adjective-preposition).

This dissertation also highlights the benefits of examining longitudinal L2 phraseological unit development from a Dynamic Systems Theory perspective. By charting development for indices of collconstruction production in multiple texts produced over an extended period of time, this dissertation was able to provide a detailed picture of the processes learners go through when developing their productive collconstructional knowledge. Additionally, the calculation of moving correlations between different indices allowed for an examination of how different types of collconstructional knowledge interact over time. Future L2 phraseology research could continue to utilize these approaches in order to examine other types of productive phraseological knowledge as well as development for other types of phraseological units. In doing so, this research could provide a more comprehensive analysis of learners' phraseological knowledge and its development over time. Such studies could also examine the interaction between learners' phraseological knowledge and other types of linguistic knowledge (e.g. lexical, grammatical), showing how these types of knowledge interact in interlanguage development.

In terms of learner corpus design, the longitudinal studies in this dissertation point to the need for larger and denser corpora of learner speech and writing over time. The group-level longitudinal analysis found no measurable difference in collconstruction productive over time. One possible reason for these non-significant differences was the amount of interindividual differences in development. With only six learners in this dataset, interindividual differences possibly had a strong effect on the statistical analyses. Although individual differences will remain in larger longitudinal learner corpora, these differences may not have as strong an effect on quantitative measures of development when more data is available from more learners. However, most large-scale longitudinal learner corpora only contain a small amount of data points for each learner. This makes it difficult to follow up group-level analyses with more detailed analyses of individual development. For instance, the corpus in Bestgen and Granger (2014) contains 171 texts produced by 57 L2 writers, with each writer only contributing three texts each. Therefore, researchers should work towards creating more large-scale longitudinal learner corpora that contain texts written by a large group of learners with each learner contributing texts more frequently (e.g. every two to three weeks instead of beginning, middle, and end of study period). These learner corpora would prove beneficial in that they would allow researchers to more reliably examine longitudinal development at both the group and individual levels for the same learners.

8.3.2 Usage-based Second Language Acquisition

The findings in this dissertation have at least two important implications for future usage-based studies of second language acquisition. First, the results of the cross-sectional analysis support the use of normalized entropy as a measure of diversity. Results indicated that more proficient learners were more likely to produce a more restricted range of verb-noun

collconstructions than their less proficient counterparts. The qualitative cross-sectional analyses also showed that higher proficiency learners' use of the [acquire] + *money* collconstruction was more restricted and targetlike with a lower entropy score. These findings provide some support for those from Römer and Garner (under review), who found lower normalized entropy scores for verb-argument constructions in the spoken production of advanced learners compared to intermediate learners. The entropy scores for the advanced L2 speakers were also closer to entropy scores found in the British National Corpus than entropy scores for the intermediate learners (BNC Consortium, 2007). In addition, these findings provide some support for those from Ellis and O'Donnell (2014) that showed lower entropy scores signifying a higher degree of semantic cohesion within a construction. Future research into L2 construction acquisition could continue to use normalized entropy in order to not only examine the diversity of lexemes within constructions, but also examine how the distribution of lexemes within constructions are shaped. This feature or normalized entropy is another advantage of this diversity measure over the more commonly used type-token ratio. Future research should also examine the diversity of lexemes in construction production by L1 language users in order to better understand nativelike distributions of lexemes within constructions and provide a baseline for comparison with L2 language production.

Second, the findings from this dissertation provide some preliminary support for the need to consider the possibility that development in constructional knowledge is more complex than previously thought. For the adjective-noun functional category [positive] + *experience*, more proficient learners produced a greater range of adjectives, a finding that is consistent with other studies of constructional development (Ellis & Ferreira-Junior, 2009a; Eskildsen, 2009, 2012, 2015, Mellow, 2006; Myles et al., 1998). On the other hand, the opposite trend was observed for

verb-noun collconstructions, with the more proficient learners producing a more limited range of [acquire] + *money* collconstruction tokens. These findings tentatively suggest that different types of constructions experience different trends in development, with some increasing in schematicity while others decrease in their lexical diversity. Alternatively, it may be the case that constructional development approximates an upside down u-shaped pattern, with learners first moving towards greater constructional schematicity and then later moving towards relying on only the most idiomatic and targetlike lexemes in the construction. To investigate these possibilities, future research into constructional development should investigate these possibilities across a range of constructions and across beginner, intermediate, and advanced learner proficiency levels and across a range of L1 backgrounds.

8.3.3   Phraseology Pedagogy

This dissertation has tentative implications for phraseology instruction in the L2 writing classroom. The findings from the cross-sectional analysis suggest that, when producing certain functional collconstruction categories (e.g. [positive] + *experience*), more proficient L2 writers choose collconstructions that are more targetlike (e.g. *valuable experience*). Given these results, it would be beneficial for L2 writing instruction to include opportunities for learners to notice targetlike exemplars of different functional collconstruction categories. One possibility for increasing opportunities for noticing targetlike collconstructions is the use of Data-Driven Learning (DDL; Johns, 1991; Leńko-Szymańska & Boulton, 2015). In this approach, language learners are given direct access to language corpora or corpus-derived learning materials in order to investigate language use on their own. According to Flowerdew (2015), DDL activities increase opportunities for noticing, and possibly uptake, because concordance lines are a type of enhanced input that focuses learners' attention to recurrent word patterns. Concerning

collconstruction instruction, DDL activities would have the benefit of focusing learners' attention to a range of targetlike word combinations that fulfill specific lexical functions. Several studies have provided support for these claims by showing that learners studying collocations using corpus-based activities make strong gains in collocational knowledge and use collocations more correctly in their writing (Chan and Liou, 2005; Daskalovska, 2015; Wu, Witten, and Franken, 2010).

One specific type of DDL activity that could accomplish the goal of increasing learners' noticing of targetlike collconstructions is Kennedy and Miceli's (2010) "pattern-hunting" task. In this activity, learners working on a specific topic would first brainstorm a list of key words they would need to write about the topic. Learners would then search for these words in a corpus and examine concordance lines for the most frequently occurring word combinations for expressing their desired meaning. For instance, a student wanting to discuss the types of positive experiences students could have doing part-time jobs could search the corpus for instances of *experience* occurring with an adjective. Upon examining the resulting concordance lines, the student may notice strongly associated collconstructions tokens such as *valuable experience, positive experience,* and *meaningful experience*.

Additionally, opportunities for noticing targetlike collconstructions outside of focused DDL tasks could be strengthened through enhanced input, input flooding (Wood, 2015), or explicit discussions (Littlemore, 2009). Enhanced input, a technique in which target collconstructions tokens would be highlighted in a text, has been shown to lead to greater gains in phraseological unit knowledge (Choi, 2017; Peters, 2012). Studies of collocation acquisition through input flooding have shown that embedding and repeating target phraseological units in a text can lead to gains in collocational knowledge (Madlener, 2015; Sonbul & Schmitt, 2012;

Szudarski & Carter, 2015; Webb, Newton, & Chang, 2013). Lastly, according to Littlemore

(2009), explicit discussions focusing on encyclopedic knowledge of single words could help

learners build up their collocational knowledge for those words. Regardless of the approach

taken, the results of this dissertation suggest that collconstruction instruction should focus on

pushing students past the use of general-purpose tokens (e.g. *good experience, get money*) and

more towards the use of tokens that are more strongly associated (e.g. *valuable experience, earn*

money).

8.4   Limitations

Despite the overall strengths of the studies conducted in this dissertation, there are several

limitations that need to be noted. First, the corpus in the cross-sectional analysis contained an

unbalanced number of essays at each level, with the high-beginner and low-intermediate

subcorpora being twice the size of the high-intermediate subcorpus. Future studies would benefit

from a more balanced representation of texts from multiple levels of writing proficiency. These

studies could also include texts written by advanced learners, thus allowing a more complete

view of collconstruction production across proficiency levels.

The second limitation concerns the size and scope of the cross-sectional corpus. The

cross-sectional corpus only contained writing from learners from a single L1 background

(Korean), limiting the generalizability of these results to only that L1 context. The cross-

sectional corpus was also small and contained texts that were relatively short, with most texts

being under 300 words. These features of the cross-sectional corpus limited the amount of

different collconstruction types produced by the learners. As a result, a more systematic analysis

of variation in the functional collconstruction subcategories across proficiency levels was not

possible. The length of the texts also affected the number of collconstruction tokens in each text,

which may have led to higher levels of variability in the proportion scores. For instance, two texts containing 5 tokens each would have a difference in non-collexeme proportion scores of 20% simply because one text contained 3 non-collexeme collconstructions while the other contained 2. This was especially true for the adverb-adjective collconstruction category, as they occurred between 1-3 times per 100 words and did not occur at all in 69 of the 276 texts. Lastly, all of the texts in the cross-sectional corpus were written in response to the same prompt ("part-time jobs"). While collconstructions that occurred in the prompt were removed (e.g. "have job" and "part-time job"), the use of a single prompt may have narrowed the range of collconstructions types the learners would produce. As such, the analysis of cross-sectional variation in collconstruction production only offers a small glimpse of these learners' collconstructional knowledge. Altogether, these limitations of the cross-sectional corpus restrict the generalizability of these results and necessitate further research using corpora that include a larger number of learners from more L1 backgrounds producing longer texts in response to multiple prompts.

Sample size was also a limitation of the longitudinal corpus. While the small size of the corpus did allow for an examination of development for individual learners, it made it difficult to examine significant quantitative changes in collconstruction production for the entire group. It also limits the generalizability of the results that can be made. Future longitudinal learner corpus research on L2 phraseology should strive to use larger and denser corpora that allow for the analysis of group development as well as development for individual learners. Similar to the cross-sectional corpus, the longitudinal corpus also suffered from the limited number of collconstruction tokens in each text, limiting the strength of these results and possibly leading to high degrees of variation in the proportion score indices. Again, this was strongest for adverb-

adjective collconstructions, as they never occurred more than 3 times per 100 words in a text and did not occur at all in 13 of the 60 texts in the linear-mixed effects models.

Another limitation of the longitudinal analyses was the lack of control over the topics the learners wrote about. The texts in this corpus were freewrites in that learners were allowed to write about a range of topics about everyday life, school life, and current events. This may have affected the types of collconstructions the learners produced in each text. It also made an analysis of functional collconstruction subcategories difficult as some subcategories could appear frequently in one text but never appear in another one simply because of topic differences. Future longitudinal L2 collconstruction research should strive to better control the texts learners produce in terms of text types and topics.

Another limitation of this dissertation concerns the reference corpus used. Although previous research has shown that n-gram indices derived from COCA Academic are predictive of human judgments of L2 writing proficiency, this corpus might not be the best approximation of the learner's target collconstruction production. Additionally, the use of an academic reference corpus for beginner and intermediate learners and learners writing non-academic texts (i.e. some of the freewrites in the Salsbury corpus) may have obscured developmental trends that could have occurred. For example, learners possibly produced more strongly associated collconstructions commonly produced in more general purpose writing (e.g. newspapers, magazines, fiction). Also, some collconstructions could have different strengths of association based on the reference corpus (i.e. certain word associations are likely to vary in frequency across registers). While it is believed that this does not invalidate the findings from these analyses, it does necessitate that future research explore using collconstruction indices derived

from other reference corpora in order to determine which indices are most useful for examining collconstruction development at beginner and intermediate proficiency levels.

8.5   Concluding Thoughts

Over the past several decades, the field of corpus linguistics has taken a greater interest in the phraseological nature of the English language. This has also led to a greater interest in understanding how L2 English learners acquire productive phraseological knowledge in speech and writing. This research has been incredibly helpful in revealing the nature of L2 productive phraseological knowledge and its development over time and across proficiency levels (Ebeling & Hasselgård, 2015; Paquot & Granger, 2012). The current dissertation adds to this growing body of research by showing how operationalizing collocations as constructions within a Construction Grammar framework can help better connect the lexical and the grammatical in phraseology research and more closely represent how language users employ collocations in their language production. It also adds to existing research by showing how a multi-faceted approach to phraseology research can provide a more complete view of the L2 phraseological system and its development in L2 writing. While there exist clear avenues for future development, it is hoped that this dissertation serves as a first step in more closely aligning L2 phraseology research with Construction Grammar perspectives and in investigating the complex nature of productive phraseological knowledge in L2 usage.

REFERENCES

Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word-combinations. In A.P. Cowie (Ed.), *Phraseology: Theory, Analysis, and Applications* (pp. 101-122). Oxford: Oxford University Press.

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*, 81-92.

Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word-combinations. In A.P. Cowie (Ed.), *Phraseology: Theory, Analysis, and Applications* (pp. 101-122). Oxford: Oxford University Press.

Anthony, L. (2015). TagAnt (Version 1.2.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/

Anthony, L. (2014). AntConc (Version 3.4.3) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/software

Appel, R. & Wood, D. (2016). Recurrent word combinations in EAP test-taker writing: Differences between high- and low-proficiency levels. *Language Assessment Quarterly, 13*(1), 55-71.

Arnon, I. & Cohen Priva, U. (2013). More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech, 56*(3), 349-371.

Arnon, I. & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Language and Memory, 62*, 67-82.

Baba, K. & Nitta, R. (2014). Phase transition in development of writing fluency from a complex dynamic systems perspective. *Language Learning, 64*(1), 1-35.

Bartoń, K. (2016). MuMIn: Multi-Model Inference [Software]. Available from https://CRAN.R-project.org/package=MuMIn

Bestgen, Y. & Granger, S. (2014). Quantifying the development of phraseological competence in L2 English writing: An automated approach. *Journal of Second Language Writing, 26*, 28-41.

Bencini, G. & Goldberg, A.E. (2000). The contribution of argument structure constructions to sentence meaning. *Journal of Memory and Language, 43*(4), 640-651.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(3), 371-405.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education.

Caspi, T. & Lowie, W. (2010). A dynamic perspective on L2 lexical development in academic English. In R. Chacon-Beltran, C. Abello-Contesse, & M Torreblanca-Lopez (Eds.), *Insights into Non-native Vocabulary Teaching and Learning* (pp. 41-58). Bristol: Multilingual Matters.

Caspi, T. & Lowie, W. (2013). The dynamics of L2 vocabulary development: A case study of receptive and productive knowledge. *Revista Brasileria de Linguistica Aplicada, 13*(2), 437-462.

Chan, T.P. & Liou, H.C. (2005). Effects of web-based concordancing instruction on EFL students' learning of verb-noun collocations. *Computer Assisted Language Learning, 18*(3), 231-250.

Chen, M. (2013). Phrasal verbs in a longitudinal learner corpus: Quantitative findings. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *Twenty Years of Learner Corpus Research: Looking Back, Moving Ahead* (pp. 49-59). Louvain-la-Neuve: Presses Universitaires de Louvain.

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

Chen, Y.H. & Baker, P. (2016). Investigating criterial discourse features across second language development: Lexical bundles in rated learner essays, CEFR B1, B2, and C1, *Applied Linguistics, 6*(1), 849-880.

Choi, S. (2017). Processing and learning of enhanced English collocations: An eye movement study. *Language Teaching Research, 21*(3), 403-426.

Conklin, K. & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics, 29*(1), 72-89.

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Council of Europe. (2003). *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment.* Strasbourg: Language Policy Division.

Crossley, S.A. & Salsbury, T. (2011). The development of lexical bundle accuracy and production in English second language speakers. *International Review of Applied Linguistics, 49*, 1-26.

Daskalovska, N. (2015). Corpus-based versus traditional learning of collocations. *Computer Assisted Language Learning, 28*(2), 130-144.

Davies, M. (2009). The Corpus of Contemporary American English: 450 million words, 1990-present. Available online at: http://corpus/byu.edu/coca

de Bot, K., Lowie, W., & Verspoor, M. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and Cognition, 10*(1), 7-21.

De Cock, S. (2004). Preferred sequences of words in NS and NNS speech. *BELL: Belgian Journal of English Language and Literatures,* 225-246.

De Cock, S. (1998). A recurrent word combination approach to the study of formulae in the speech of native and non-native speakers of English. *International Journal of Corpus Linguistics, 3*(1), 59-80.

Deshors, S.C. (2016). Inside phrasal verb constructios: A co-varying collexeme analysis of verb-particle combinations in EFL and their semantic associations. *International Journal of Learner Corpus Research 2*(1), 1-30.

Durrant, P. (2015). Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Applied Linguistics*, Advance Access published April 19, 2015. doi: doi: 10.1093/applin/amv011

Durrant, P. & Doherty, A. (2010). Are high-frequency collocations psychologically real? Investigating the thesis of collocational priming. *Corpus Linguistics and Linguistic Theory, 6*(2), 125-155.

Durrant, P. & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics, 47*, 157-177.

Ebeling, S.O. & Hasselgård, H. (2015). Learner corpora and phraseology. In S. Granger, G. Gilquin & F. Meunier (Eds.), *The Cambridge Handbook of Learner Corpus Research* (pp. 207-229). Cambridge: Cambridge University Press.

Ellis, N.C. (2002). Frequency effect in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24*, 143-188.

Ellis, N.C. (2006a). Language acquisition as rational contingency learning *Applied Linguistics, 27*(1), 1-24.

Ellis, N.C. (2006b). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics, 27*(2), 164-194.

Ellis, N.C. & Ferreira-Junior, F. (2009a). Construction learning as a function of frequency, frequency distribution, and function. *The Modern Language Journal, 93*(3), 370-385.

Ellis, N.C. & Ferreira-Junior, F. (2009b). Constructions and their acquisition: Islands and the distinctiveness of their occupancy. *Annual Review of Applied Linguistics, 7*, 187-220.

Ellis, N.C., Frey, E., & Jalkanen, I. (2009). The psycholinguistic reality of collocation and semantic prosody (1): Lexical access. In U. Römer & R. Schulze (Eds.), *Exploring the Lexis-Grammar Interface* (pp. 89-114). Amsterdam: John Benjamins.

Ellis, N.C. & Larsen-Freeman, D. (Eds.). (2009). *Language as a Complex Adaptive System. Language Learning, 59*(Supplement 1).

Ellis, N.C. & O'Donnell, M.B. (2014). Construction learning as category learning: A cognitive analysis. In T. Herbst, H.J. Schmid, & S. Faulhaber (Eds.), *Constructions Collocations Patterns.* Berlin: De Gruyter Mouton, 71-97.

Ellis, N.C., O'Donnell, M.B., & Römer, U. (2014a). The processing of verb-argument constructions is sensitive to form, function, frequency, contingency, and prototypicality. *Cognitive Linguistics, 25*(1), 55-98.

Ellis, N.C., O'Donnell, M.B., & Römer, U. (2014b). Second language verb-argument constructions are sensitive to form, function, frequency, contingency, and prototypicality. *Linguistic Approaches to Bilingualism, 4*(4), 405-431.

Ellis, N.C., & Wulff, S. (2014). Usage-based approaches to SLA. In B. VanPatten & J. Williams (Eds.), *Theories in Second Language Acquisition: An Introduction* (2nd Edition ed., pp. 75-93). NY: Routledge.

Erman, B. & Warren, B. (2000). The idiom principle and the open choice principle. *Text, 20*, 29-62.

Eskildsen, S.W. (2009). Constructing another language -- Usage-based linguistics in second language acquisition. *Applied Linguistics, 30*(3), 335-357.

Eskildsen, S.W. (2012). L2 negation constructions at work. *Language Learning, 62*(2), 335-372.

Eskildsen, S.W. (2015). What counts as a developmental sequence? Exemplar-based L2 learning of English questions. *Language Learning, 65*(1), 33-62.

Evert, S. (2009). Corpora and collocations. In A. Lüdeling & M. Kytö (Eds.), *Corpus Linguistics. An International Handbook* (pp. 1211-1248). Berlin: Mouton de Gruyter.

Flowerdew, L. (2015). Data-driven learning and language learning theories: Whither the twain shall meet. In A. Leńko-Szymańska, & A. Boulton (Eds.), *Multiple Affordances of Language Corpora for Data-Driven Learning* (pp. 15-36)*.* Amsterdam: John Benjamins.

Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning, 67*(S1), pp. 155-179.

Garner, J.R. & Crossley, S.A. (forthcoming). A latent curve model approach to studying L2 n-gram development. *The Modern Language Journal.*

Garner, J.R., Crossley, S.A., & Kyle, K. (forthcoming). Beginning and intermediate L2 writer's use of n-grams: An association measures study. *International Review of Applied Linguistics in Language Teaching.*

Garner, J.R., Crossley, S.A., & Kyle, K. (under review). N-gram measures and L2 writing proficiency.

Gilquin, G. (2015). The use of phrasal verbs by French-speaking EFL learners. A constructional and collostructional corpus-based approach. *Corpus Linguistics and Linguistic Theory, 11(1)*, 51-88.

Goldberg, A.E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure.* Chicago: University of Chicago Press.

Goldberg, A.E. (2006). *Constructions at Work: The Nature of Generalization in Language.* Oxford: Oxford University Press.

Goldberg, A.E. (2013). Constructionist approaches. In T. Hoffmann & G. Trousdale (Eds.), *Oxford Handbook of Construction Grammar* (pp. 15-31). Oxford: Oxford University Press.

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, 52*(3), 229-252.

Granger, S. & Paquot, M. (2008). Disentangling the phraseological web. In S. Granger & F. Meunier (Eds.), *Phraseology: An Interdisciplinary Perspective* (pp. 27-49). Amsterdam: John Benjamins.

Gries, S.Th. (2013). 50-something years of work on collocations: What is or should be next... *International Journal of Corpus Linguistics, 18*(1), 137-165.

Gries, S.Th. & Ellis, N.C. (2015). Statistical measures for Usage-Based Linguistics. *Language Learning, 65*(Supplement 1), 228-255.

Gries, S.Th. & Stefanowitsch, A. (2004). Extending collostructional analysis: A corpus-based perspectives on 'alternations'. *International Journal of Corpus Linguistics 9*(1), 97-129.

Gries, S.Th. & Wulff, S. (2005). Do foreign language learners also have constructions? *Annual*

Review of Cognitive Linguistics, 3, 182-200.

Gries, S.Th. & Wulff, S. (2009). Psycholinguistic and corpus-linguistic evidence for L2 constructions. *Annual Review of Cognitive Linguistics, 7*, 163-186.

Hasko, V. (2013). Capturing the dynamics of second language development via learner corpus research: A very long engagement. *The Modern Language Journal, 97*(S1), 1-10.

Hilpert, M. (2014). *Construction Grammar and Its Application to English.* Edinburgh: Edinburgh University Press.

Hoey, M. (2005). *Lexical priming: A new theory of words and language*. New York: Routledge.

Hsu, J.Y. (2007). Lexical collocations and their relation to the online writing of Taiwanese college English majors and non-English majors. *Electronic Journal of Foreign Language Teaching, 4*(2), 192-209.

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*.

Hunston, S. (2002). *Corpora in Applied Linguistics.* Cambridge: Cambridge University Press.

Hyland, K. (2008). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics, 18*(1), 41-62.

Ishikawa, S. (2013). The ICNALE and sophisticated contrastive interlanguage analysis of Asian Learners of English. In S. Ishikawa (Ed.), *Learner corpus studies in Asia and the world, 1* (pp. 91-118). Kobe, Japan: Kobe University.

Janssen, N. & Barber, H.A. (2012). Phrase frequency effects in language production. *PLoS One, 7*(3), e33202. doi: 10.1371/journal.pone.0033202

Jiang, N. & Nekrasova, T. (2007). The processing of formulaic sequences by second language speakers. *The Modern Language Journal, 91*(3), 433-445.

Juknevičienė, R. (2008). Collocations with high-frequency verbs in learner English: Lithuanian learners vs native speakers. *KALBOTYRA, 59*(3), 119-127.

Juknevičienė, R. (2009). Lexical bundles in learner language: Lithuanian learners vs. native speakers. *KALBOTYRA, 61*(3), 61-72.

Kennedy, G. (2003). Amplifier collocations in the British National Corpus: Implications for English Language Teaching. *TESOL Quarterly, 37*(3), 467-487.

Kuznetsova, A., Brockhoff, P.B., & Christensen, R.H.B. (2015). LmerTest: Tests in Linear Mixed Effects Models [Software]. Available from https://CRAN.R-project.org/package=lmerTest

Kumar, U., Kumar, V., & Kapur, J.N. (1986). Normalized measures of entropy. *International Journal of General Systems, 12*(1), 55-69.

Kyle, K. & Crossley, S.A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application *TESOL Quarterly, 49*(4), 757-786.

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*(4), 590-619.

Larsen-Freeman, D. (2010). Having and doing: Learning from a complexity theory perspective. In P. Seedhouse, S. Walsh, & C. Jenks (Eds.), *Conceptualizing 'Learning' in Applied Linguistics* (pp. 52-68). Basingstoke, UK: Palgrave Macmillan.

Laufer, B. & Waldman, T. (2011). Verb-Noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning, 61*(2), 647-672.

Leńko-Szymańska, A. (2014). The acquisition of formulaic language by EFL learners: A cross-sectional and cross-linguistic perspective. *International Journal of Corpus Linguistics, 19*(2), 225-251.

Leńko-Szymańska, A., & Boulton, A. (2015). *Multiple Affordances of Language Corpora for Data-Driven Learning.* Amsterdam: John Benjamins.

Li, J. & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing, 18*, 85-102.

Li, J. & Schmitt, N. (2010). The development of collocation use in academic texts by advanced L2 learners: A multiple case study perspective. In D. Wood (Ed.), *Perspectives on Formulaic Language: Acquisition and Communication* (pp. 23-45). London: Continuum.

Littlemore, J. (2009). *Applying Cognitive Linguistics to Second Language Learning and Teaching.* Basingstoke, English: Palgrave Macmillan.

Madlener, K. (2015). *Frequency Effects in Instructed Second Language Acquisition.* Berlin: De Gruyter Mouton.

Martinez-Garcia, M.T. & Wulff, S. (2012). Not wrong, yet not quite right: Spanish ESL students' use of gerundial and infinitival complementation. *International Journal of Applied Linguistics, 22*(2), 225-244.

Mel'čuk, I. (1998). Collocations and lexical functions. In A. P. Cowie (Ed.), *Phraseology: Theory, Analysis, Applications* (pp. 23-53). Oxford: Oxford University Press.

Mel'čuk, I. (2007). Lexical Functions. In H. Burger, D. Dobrovol'skij, P. Kühn, & N. Norrick (eds.), *Phraseology: An International Handbook of Contemporary Research,* 119-131. Berlin: Mouton de Gruyter.

Mellow, J.D. (2006). The emergence of second language syntax: A case study of the acquisition of relative clauses. *Applied Linguistics, 27*(4), 645-670.

Myles, F., Hooper, J., & Mitchell, R. (1998). Rote or rule? Exploring the role of formulaic language in classroom foreign language learning. *Language Learning, 48*(3), 323-363.

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics, 24*(2), 223-242.

O'Donnell, M.B., Römer, U., & Ellis, N.C. (2013). The development of formulaic sequences in first and second language writing: Investigating effects of frequency, association, and native norm. *International Journal of Corpus Linguistics, 18*(1), 83-108.

Ohlrogge, A. (2009). Formulaic expressions in intermediate EFL writing assessment. In R. Corrigan, E.A. Moravcsik, H. Ouali & K.M. Wheatley (Eds.), *Formulaic Language (Volume 2): Acquisition, Loss, Psychological Reality, and Functional Explanations* (pp. 375-385). Amsterdam: John Benjamins.

Paquot, M. (2017). The phraseological dimension in interlanguage complexity research. *Second Language Research Special Issue on Linguistic Complexity,* 1-25.

Paquot, M. & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics, 32*, (130-149).

Pawley, A. & Syder, F.H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and Communication* (pp. 191-225). London: Longman.

Peters, E. (2012). Learning German formulaic sequences: The effect of two attentiondrawing techniques. *Language Learning Journal, 40*(1)*,* 65-79.

Qi, Y. & Ding, Y. (2011). Use of formulaic seqeunces in monologues of Chinese EFL learners. *System, 39,* 164-174.

R Core Team (2015). R: A language and environment for statistical computing [Software]. Vienna: R Foundation for Statistical Computing. Available from https://www.R-project.org/

Reznicek, M. & Zinsmeister, H. (March 2013). *Why learner texts are easy to tag.* Paper presented at DGfS Jahrestagung, Potsdam, Germany. Available at: https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/mitarbeiter-

innen/marc/ReznicekZinsmeister130313_DGfS_POSInKobalt.pdf. (last accessed on January 30, 2018).

Römer, U. (2009). The inseparability of lexis and grammar: Corpus linguistic perspectives. *Annual Review of Applied Linguistics, 7*, 140-162.

Römer, U. & Garner, J.R. (under review). The emergence of verb constructions in spoken learner English: Tracing effects of usage and proficiency.

Salsbury, T. (2000). *The Acquisitional Grammaticalization of Unreal Conditionals and Modality in L2 English: A Longitudinal Perspective*. Indiana University.

Schmid, H. (1994): Probabilistic Part-of-Speech Tagging Using Decision Trees. *Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK.

Sinclair, J.M. (1991). *Corpus Concordance Collocation*. Oxford: Oxford University Press.

Siyanova-Chantura, A. & Martinez, R. (2015). The idiom principle revisited. *Applied Linguistics, 36*(5), 549-569.

Sonbul, S. & Schmitt, N. (2012). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning, 63*(1), 121-159.

Spoelman, M. & Verspoor, M. (2010). Dynamic patterns in development of Accuracy and Complexity: A longitudinal case study in the acquisition of Finnish. *Applied Linguistics, 31*(4), 532-553.

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*, 214-225.

Stefanowitsch, A. & Gries, S.Th. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics, 8*(2), 209-243.

Stefanowitsch, A. & Gries, S.Th. (2005). Covarying collexemes. *Corpus Linguistics and Linguistic Theory, 1*(1), 1-43.

Stefanowitsch, A. & Gries, S.Th. (2009). Corpora and grammar. In A. Lüdeling & M. Kytö (Eds.), *Corpus Linguistics. An International Handbook* (933-952). Berlin: Mouton de Gruyter.

Stokes, S. & Surendran, D. (2005). Articulatory complexity, ambient frequency, and functional load as predictors of consonant development in children. *Journal of Speech, Language, and Hearing Research, 48,* 577-591.

Stubbs, M. (2007). An example of frequent English phraseology: Distributions, structures, and

functions. In R. Facchinetti (Ed.), *Corpus Linguistics 25 Years on* (pp. 89-105). Amsterdam: Rodopi.

Szudarski, P. & Carter, R. (2015). The role of input flood and input enhancement in EFL learners' acquisition of collocations. *International Journal of Applied Linguistics, 26*(2), 245-265.

Thouësny, S. (March 2009). *Increasing the reliability of a part-of-speech tagging tool for use with learner language*. Paper presented at Pre-conference (AALL'09) workshop on automatic analysis of learner language: From a better understanding of annotation needs to the development and standardization of annotation schemes, Tempe, AZ. Available at: https://calico.uoregon.edu/wp-content/uploads/2016/12/thouesny.pdf. (last accessed on January 30, 2018).

Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning, 61*(2), 569-613.

Tribushina, E. (2013). Spatial adjectives in Dutch child language: Towards a usage-based model of adjective acquisition. In C. Paradis, J. Hudson, & U. Magnusson (Eds.), *The Construal of Spatial Meaning: Windows into Conceptual Space.* Oxford: Oxford University Press.

Tribushinina, E. & Gillis, S. (2017). Advances and lacunas in usage-based studies of first language acquisition. In J. Evers-Vermeul & E. Tribushinina (Eds.), *Usage-based approaches to language acquisition and language teaching* (pp. 13-46). Berlin: De Gruyter Mouton.

van Dijk, M., Verspoor, M.H., & Lowie, W. (2011) Variability and DST. In M.H. Verspoor, K. de Bot, & W. Lowie (Eds.), *A Dynamic Approach to Second Language Development* (pp. 55-84). Amsterdam: John Benjamins.

van Geert, P. (1991). A dynamic systems model of cognitive and language growth. *Psychological Review, 98*(1), 3-53.

van Geert, P. & Verspoor, M.H. (2015). Dynamic systems and language development. In B. MacWhinney & W. O'Grady (Eds.), *The Handbook of Language Emergence* (pp. 537-555). Oxford: Wiley-Blackwell.

Verspoor, M.H. (2015). The dynamics of a usage-based approach. In J. Daems, E. Zenner, K. Heylen, D. Speelman & H. Cuyckens (Eds.), *Change of Paradigms - New Paradoxes: Recontextualizing Language and Linguistics* (pp. 29-43). Berlin: De Gruyter Mouton.

Verspoor, M.H., Lowie, W., & Van Dijk, M. (2008). Variability in second language development from a dynamic systems perspective. *The Modern Language Journal, 92*(2), 214-231.

Verspoor, M. & van Dijk, M. (2011). Visualizing interaction between variables. In M. Verspoor, K. de Bot, & W. Lowie (Eds.), *A Dynamic Approach to Second Language Development* (pp. 85-98). Amsterdam: John Benjamins.

Vidakovic, I. & Barker, F. (2010). Use of words and multi-word units in Skills for Life Writing examinations. *Cambridge ESOL: Research Notes, 41*, 7-14.

Webb, S., Newton, J., & Chang, A.C.S. (2013). Incidental learning of collocation. *Language Learning, 63*(1), 91-120.

Wood, D. (2015). *Fundamentals of Formulaic Language: An Introduction.* London: Bloomsbury.

Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

Wu, S., Witten, I.H., & Franken, M. (2010). Utilizing lexical data from a web-derived corpus to expand productive collocation knowledge. *ReCALL, 22*(1), 83-102.

Wulff, S. & Gries, S.T. (2011). Corpus-driven methods for assessing accuracy in learner production. In P. Robinson (ed.), *Second Language Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance* (61-87)*.* Amsterdam: John Benjamins.

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.

Zheng, Y. (2016). The complex, dynamic development of L2 lexical use: A longitudinal study on Chinese learners of English. *System, 56*, 40-53.

Zinsmeister, H. & Breckle, M. (2010). Starting a sentence in L2 German – Discourse annotation of a learner corpus. In M. Pinkal, I. Rehbein, S. Schulte im Walde, & A. Storrer (Eds.), *Semantic Approaches in Natural Language Processing: Proceedings of the Conference on Natural Language Processing.*

APPENDIX: CEFR Writing Rubric

(Council of Europe, 2003)

	Overall	Range	Coherence	Accuracy
C2	Can write clear, *highly accurate and smoothly flowing complex texts in an appropriate and effective personal style conveying finer shades of meaning. Can use a logical structure which helps the reader to find significant points.*	Shows great flexibility in *formulating ideas in differing linguistic forms to convey finer shades of meaning precisely, to give emphasis and to eliminate ambiguity. Also has a good command of idiomatic expressions and colloquialisms.*	Can create coherent and cohesive texts making full and appropriate use of a variety of organisational patterns and a wide range of connectors and other cohesive devices.	Maintains consistent *and highly accurate grammatical control of even the most complex language forms. Errors are rare and concern rarely used forms.*
C1	Can write clear, well-structured *and mostly accurate texts of complex subjects. Can underline the relevant salient issues, expand and support points of view at some length with subsidiary points, reasons and relevant examples, and round off with an appropriate conclusion.*	Has a good command of a broad range of language allowing him/her to select a formulation to express him/herself clearly in an appropriate style on a wide range of general, academic, professional or leisure topics without having to restrict what he/she wants to say. *The flexibility in style and tone is somewhat limited.*	Can produce clear, smoothly flowing, well-structured text, showing controlled use of organisational patterns, connectors and cohesive devices.	Consistently maintains a high degree of grammatical accuracy; *occasional errors in grammar, collocations and idioms.*
B2	Can write clear, detailed *official and semi-official texts on a variety of subjects related to his field of interest, synthesising and evaluating information and arguments from a number of sources. Can make a distinction between formal and informal language with occasional less appropriate expressions.*	Has a sufficient range of language to be able to give clear descriptions, express viewpoints on most general topics, using some complex sentence forms to do so. *Language lacks, however, expressiveness and idiomaticity and use of more complex forms is still stereotypic.*	Can use a number of cohesive devices to link his/her sentences into clear, coherent text, though there may be some "jumpiness" in a longer text.	Shows a relatively high degree of grammatical control. Does not make errors which cause misunderstandings.
B1	Can write straightforward connected texts on a range of familiar subjects within his field of interest, by linking a series of shorter discrete elements into a linear sequence. *The texts are understandable but occasional unclear expressions and/or inconsistencies may cause a break-up in reading.*	Has enough language to get by, with sufficient vocabulary to express him/herself with some circumlocutions on topics such as family, hobbies and interests, work, travel, and current events.	Can link a series of shorter discrete elements into a connected, linear text.	Uses reasonably accurately a repertoire of frequently used "routines" and patterns associated with more *common situations. Occasionally makes errors that the reader usually can interpret correctly on the basis of the context.*
A2	Can write a series of simple phrases and sentences linked with simple connectors like "and", "but" and "because". *Longer texts may contain expressions and show coherence problems which makes the text hard to understand.*	Uses basic sentence patterns with memorized phrases, groups of a few words and formulae in order to communicate limited information mainly in everyday situations.	Can link groups of words with simple connectors like "and", "but" and "because".	Uses simple structures correctly, but still systematically makes basic mistakes. *Errors may sometimes cause misunderstandings.*
A1	Can write simple isolated phrases and sentences. *Longer texts contain expressions and show coherence problems which make the text very hard or impossible to understand.*	Has a very basic repertoire of words and simple phrases related to personal details and particular concrete situations.	Can link words or groups of words with very basic linear connectors like "and" and "then".	Shows only limited control of a few simple grammatical structures and sentence patterns in a memorized repertoire. *Errors may cause misunderstandings.*