

**Exploring the development of ‘verb gram’ sequences and grammar patterns in L2 writing:  
A case study of L1 Korean speakers**

**Dong Ok Lim**

**Hughes Hall**

**October 2022**

**This thesis is submitted for the degree of Doctor of Philosophy**

## **Preface**

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared it is specified in the text.

It is not substantially the same as any work that has already been submitted before for any degree or other qualification except as declared in the preface and specified in the text.

It does not exceed the prescribed word limit for the relevant Degree Committee.

**Exploring the development of ‘verb gram’ sequences and grammar patterns in L2 writing:  
A case study of L1 Korean speakers**

Dong Ok Lim

**Abstract**

Usage-based approaches claim that language acquisition is sensitive to frequency. Thus, learners can utilize and produce more schematic language structures as they gain more exposure to the language. This understanding has made a significant contribution to learner corpus research. To date, a plethora of studies has explored various linguistic structures, such as constructions, in terms of usage. However, some known research gaps should be addressed. First, most of the studies utilized a smaller learner corpus. Second, studies used a top-down approach where the analysis was based on pre-selected constructions. However, a bottom-up approach could shed a valuable understanding of the kind of language structures that the learners can produce at different developmental trajectories. Lastly, such a study has not been conducted in the L1 Korean context. Therefore, this research aimed to explore the use of verb grammar patterns across L1 Korean speakers through the Cambridge Learner Corpus (CLC). The CLC is the largest corpus to date and contains written exam scripts from six different proficiency levels.

The current study implemented a bottom-up approach to explore the verb 4-gram sequences in L1 Korean speakers. This study is quasi-longitudinal, where the proficiency levels were used as a proxy to measure language development. To do so, an L1 Korean speaker sub-corpus was created from the CLC, which consisted of 1,973 learners and 914,343 words. The 4-gram verb sequences were extracted from A1 to C2 proficiency levels, and the examples from the key sequences were categorized based on pattern grammar. The findings were compared with other L1 speakers from the CLC and L1 English speakers from the written academic portion of the British National Corpus (BNC).

The findings reveal that there is a clear transition towards a greater variety of complex grammar and a wider range of meanings that become more prevalent in higher proficiency levels. Furthermore, idiomatic expressions, phrasal verbs, and collocations were present across all levels, although they become more prevalent with higher proficiency levels. The comparison

with other L1 speakers and L1 English speakers shows a similarity in the use of most frequent verb 4-gram sequences, although they tend to diverge as the structures become less frequent.

Overall, this study proposes that there is a clear development pattern in L2 usage among L1 Korean speakers and that the teachers and curriculum decisions should be made based on such findings. The study also calls for further research into different parts of speech, such as nouns and adjectives.

## Acknowledgements

First and foremost, my greatest gratitude goes to my supervisor, Dr. Pascual Pérez-Paredes, for his unwavering patience, support, and guidance throughout my journey. He has offered many inspirations during the past four years, constantly challenging me to think beyond the scope of this study. I am greatly indebted to him for his expertise in the field.

I would like to express my gratitude to my advisors, Dr. Michael Evans and Dr. Linda Fisher, for their feedback on my earlier drafts. I would also like to thank Dr. Junkyu Lee for his kind support. I am thankful to my viva examiners, Professor Neil Mercer and Dr. Niall Curry, for their feedback. And major thank you to the members of the Faculty of Education and Hughes Hall for the administrative support throughout this journey.

My sincere gratitude goes to the Cambridge University Press for giving me access to the Cambridge Learner Corpus for this study.

I am thankful for the people who have supported me through this journey. To Dami So- Thank you for your loving support throughout the years. We've laughed, cried, and celebrated all the various stages together, and I'm truly blessed to have you in my life. To Stephanie Adeyemo- Thank you for being my rock, supporting me through prayers, and with your kind words. To Winnie Pui—Thank you for supporting me through thick and thin. You are so amazing! Finally, to my friends in the United States- Thank you for all your support. I'm grateful for your visits, cards, packages, and emotional support.

Lastly, I am forever indebted to my family. To my sister- Thank you for being there for me through it all. To my parents- You two have been my biggest cheerleaders of all. Even when I had doubted myself, thank you for believing in me and cheering me on. You have been patient with me, and I cannot thank you enough.

## Table of Contents

<b>Preface</b>	i
<b>Abstract</b>	ii
<b>Acknowledgements</b>	iv
<b>Table of Contents</b>	v
<b>List of Tables</b>	ix
<b>List of Figures</b>	xii
<b>Abbreviations</b>	xiii
<b>Chapter 1 Introduction</b>	1
1.1 Key Theories and Research Gap of the Study	2
1.1.1 Usage-Based Approaches to Language learning	2
1.1.2 Pattern Grammar	4
1.1.3 Learner Corpora	5
1.1.4 Clarification About the Term “verb 4-gram sequences”	7
1.2 Purpose of the Study	7
1.3 Outline of the Thesis	8
<b>Chapter 2 Literature Review</b>	10
2.1 Introduction	10
2.2 Usage-Based Perspective on Language Learning	10
2.2.1 Cognitive Grammar in Language Development	15
2.2.2 Usage-Based Grammar: (Verb) Constructions and Argument Structure	17
2.3 Francis and Hunston’s Pattern Grammar	21
2.3.1 Comparison of Construction and Pattern Grammar	22
2.3.2 Verb Patterns	25
2.3.3 Using Parts-of-Speech (POS) n-grams to Explore Patterns	28
2.4 Dynamic Perspective of L2 Development	28
2.4.1 Roles of Interlanguage and Its Impact on L2 Writing Development	30
2.4.2 Empirical Studies on Measurements of L2 Writing Development	31

2.5 Learner Corpora and L2 Research	33
2.5.1 Use of Learner Corpora in L2 Developmental Studies	35
2.5.2 Contrastive Interlanguage Analysis in Learner Corpora	38
2.6 Conceptual Framework of the Study	40
2.6 Summary	42
<b>Chapter 3 Methodology</b>	<b>44</b>
3.1 Introduction	44
3.2 Research Questions	44
3.3 Methodology for the Investigation of Sub-Corpora	45
3.3.1 Study Design	44
3.3.2 Ethical Considerations	47
3.3.3 Instruments	48
3.3.3.1 Cambridge Learner Corpus (CLC)	49
3.3.3.1.1 CLC in Sketch Engine	50
3.3.3.1.2 Common European Framework of Reference for Languages (CEFR)	52
3.3.3.1.3 Exams	53
3.3.3.1.4 Style/Format	54
3.3.3.2 L1 Korean Speaker Sub-Corpus	57
3.3.3.3 Other L1 Speaker Sub-Corpus	59
3.3.3.4 The British National Corpus (BNC)	62
3.3.3.5 Collins WordBanks Online	63
3.4 Procedures	67
3.4.1 Identifying Verb 4-gram Sequences and Node Verbs	68
3.4.2 Identifying Grammar Patterns	73
3.4.3 Comparing L1 Korean Speakers with Other L1 speaker and L1 English Speakers	76
3.4.3.1 Comparing with Other L1 Speaker Sub-Corpus	77
3.4.3.2 Comparing with L1 English Sub-Corpus	77
3.6 Reliability	78
3.7 Data Analysis	79

3.7.1 Analysis of the Node Verbs	79
3.7.2 Analysis of the Core and Emerging Sequences	80
3.7.3 Analysis of the Grammar Patterns	80
3.8 Potential Methodological Issues with the Study Design	81
3.9 Summary	83
<b>Chapter 4 Results</b>	<b>85</b>
4.1 Introduction	85
4.2 Most Frequently Used Verb 4-gram Sequences Among L1 Korean Speakers	87
4.2.1 What are the ‘core’ verb 4-gram sequences?	89
4.2.2 What are the ‘emerging’ verb 4-gram sequences?	91
4.3 Most Frequently Used Node Verbs for the ‘core’ and ‘emerging’ Verb 4-gram Sequences	94
4.4 Verb 4-gram Sequences to Pattern Grammar	99
4.4.1 Among the core Sequences, what grammar patterns emerge in the L1 Korean sub-corpus?	92
4.4.2 Among the emerging sequences, what grammar patterns emerge in the L1 Korean sub-corpus?	115
4.5 How do verb 4-gram sequences differ between L1 Korean speakers and other speakers of L1?	129
4.5.1 A Comparison of the Core Sequences Between L1 Korean Speakers and Other L1 speaker	136
4.5.2 A Comparison of the Emerging Sequences Between L1 Korean Speakers and Other L1 speaker	149
4.6 How do the verb 4-gram sequences differ between L1 Korean speakers and L1 English speakers?	155
4.6.1 A Comparison of the Core Sequences Between L1 Korean Speakers and L1 English Speakers	159
4.6.2 A Comparison of the Emerging Sequences Between L1 Korean Speakers and L1 English Speakers	162
4.7 Summary	165
<b>Chapter 5 Discussion</b>	<b>167</b>
5.1 Introduction	167

5.2 L1 Korean Speakers	168
5.2.1 Ranking of Verb 4-gram Sequences	168
5.2.2 Node Verbs	172
5.2.3 Categorization of Sequences to Pattern	175
5.3 L1 Korean Speakers vs. Other L1 speaker	181
5.3.1 Comparison of Verb 4-gram Sequences	181
5.3.2 Comparison of Patterns	183
5.4 L1 Korean Speakers vs. L1 English Speakers	184
5.4.1 Comparison of Verb 4-gram Sequences	184
5.4.2 Comparison of Patterns	185
5.5 Summary	186
<b>Chapter 6 Conclusion</b>	188
6.1 Introduction	188
6.2 Summary of Major Findings	188
6.3 Critical Reflection on the Overall Study	189
6.4 Implications of the Present Study	191
6.5 Contribution of the Study	193
6.5.1 Methodological Contribution of the Study	193
6.5.2 Educational Contribution of the Study	193
6.6 Limitations	194
6.7 Recommendations for Future Research	195
<b>References</b>	197
<b>Appendices</b>	212
Appendix 1	214
Appendix 2	212
Appendix 3	215
Appendix 4	217

## List of Tables

- Table 2.1** Examples to demonstrate that patterns are also constructions (Hunston, 2019)
- Table 2.2** Patterns of verbs (Francis et al, 1996; Hunston & Francis, 2000)
- Table 2.3** List of major learner corpora
- Table 3.1** Size of the corpora used in this study
- Table 3.2** Example of test take metadata provided in the CLC
- Table 3.3** The CEFR English profile
- Table 3.4** Description of exams in the CLC
- Table 3.5** Different types of writing styles included in the exams
- Table 3.6** Different types of writing formats included in the exams
- Table 3.7** Number of participants per proficiency level in the other L1 speaker sub-corpus
- Table 3.8** Word per proficiency level in the other L1 speaker sub-corpus
- Table 3.9** Distribution of exam types in the L1 Korean speaker sub-corpus
- Table 3.10** Number of participants per proficiency level in the L1 Korean speaker sub-corpus
- Table 3.11** Word per proficiency level in the other L1 speaker' sub-corpus
- Table 3.12** Distribution of exam types in the other L1 speaker sub-corpus
- Table 3.13** Top 20 first languages per proficiency level (word frequency/number of learners)
- Table 3.14** Classification of subject domains in the BNC written sub-corpus
- Table 3.15** Text types of the Collins WordBanks Online
- Table 3.16** Percentage distribution of format across the six proficiency levels in the L1 Korean sub-corpus
- Table 4.1** Top 25 verb sequences for L1 Korean speakers
- Table 4.2** Ranking and normalized frequency (per mil. words) of “core” sequences
- Table 4.3** Ranking and normalized frequency (per mil. words) of “emerging” sequences
- Table 4.4** Percentage of the top node verbs from each proficiency level in the “VV DT NN IN” sequence
- Table 4.5** Percentage of the top node verbs from each proficiency level in the “VV IN DT NN” sequence
- Table 4.6** Percentage of the top node verbs from each proficiency level in the “VVN IN DT NN”, “VV DT JJ NN”, “VVG IN DT NN” sequences

**Table 4.7** Percentage of the top node verbs from each proficiency level in the emerging sequences

**Table 4.8** Distribution of grammar pattern in the “VV DT NN IN” sequence among L1 Korean speakers

**Table 4.9** Categorization of the “VV DT NN IN” sequence among L1 Korean speakers

**Table 4.10** Categorization of the **V n** pattern from the “VV DT NN IN” sequence among L1 Korean speakers

**Table 4.11** Distribution of grammar patterns in the “VV IN DT NN” sequence among L1 Korean speakers

**Table 4.12** Categorization of the **V *preposition* n** pattern in the “VV IN DT NN” sequence among L1 Korean speakers

**Table 4.13** Distribution of grammar patterns in the “VVN IN DT JJ” sequence among L1 Korean speakers

**Table 4.14** Categorization of the **V *preposition* n** pattern in the “VVN IN DT JJ” sequence among L1 Korean speakers

**Table 4.15** Distribution of grammar patterns in the “VVD DT NN IN” sequence among L1 Korean speakers

**Table 4.16** Categorization of the **V n *preposition* n** pattern in the “VVD DT NN IN” sequence among L1 Korean speakers

**Table 4.17** Categorization of the **V n** pattern in the “VVD DT NN IN” sequence among L1 Korean speakers

**Table 4.18** Top 25 rankings of A2 levels in the other L1 speaker sub-corpus

**Table 4.19** Top 25 rankings of B1 levels in the other L1 speaker sub-corpus

**Table 4.20** Top 25 rankings of B2 levels in the other L1 speaker sub-corpus

**Table 4.21** Top 25 rankings of C1 levels in the other L1 speaker sub-corpus

**Table. 4.22** Top 25 rankings of C2 levels in the other L1 speaker sub-corpus

**Table 4.23** Comparison of core and emerging sequence rankings in the L1 Korean and other L1 sub-corpora

**Table. 4.24** Categorization of the “VV DT NN IN” sequence in A2 level (L1 Korean vs. other L1 speaker)

**Table 4.25** Meaning groups of the “VV DT NN IN” sequence in A2 level (L1 Korean vs. other L1 speaker)

**Table. 4.26** Categorization of the “VV DT NN IN” sequence in B2 level (L1 Korean vs. other L1 speaker)

**Table 4.27** Meaning groups of the “VV DT NN IN” sequence in B2 level (L1 Korean vs. other L1 speaker)

**Table 4.28** Categorization of the “VV DT NN IN” sequence in C2 level (L1 Korean vs. other L1 speaker)

**Table 4.29** Meaning groups of the “VV DT NN IN” sequence in C2 level (L1 Korean vs. other L1 speaker)

**Table 4.30** Categorization of the “VVN IN DT JJ” sequence in B2 level (L1 Korean vs. other L1 speaker)

**Table 4.31** Categorization of the “VVN IN DT JJ” sequence in C2 level (L1 Korean vs. other L1 speaker)

**Table 4.32** Comparison of the top 25 verb 4-gram sequences in the BNC and C2 level in the L1 Korean sub-corpus and other L1 speaker

**Table 4.33** Categorization of the “VV DT NN IN” sequence in C2 level (BNC vs. L1 Korean speakers)

**Table 4.34** Meaning groups of the “VV DT NN IN” sequence in C2 level (BNC vs. L1 Korean speakers)

**Table 4.35** Categorization of the “VVN IN DT JJ” sequence in C2 level (BNC vs. L1 Korean speakers)

**Table 5.1** Normalized frequency distribution of the top 10 sequences among L1 Korean speakers

**Table 5.2** 50 most frequently used verbs in the EFCAMDAT learner corpus (Römer, 2019)

## List of Figures

- Figure 2.1** Taxonomic model of L2 complexity (Bulté & Housen, 2012)
- Figure 2.2** Concordance lines for the lemma ‘happiness’ in the BNC corpus
- Figure 2.3** Framework of the study
- Figure 3.1** Screenshot of the CLC in *Sketch Engine*
- Figure 3.2** Top 20 distribution of L1 in the L2 learner sub-corpus (normalized to 1-million words)
- Figure 3.3** Main page of verb patterns in Collins Dictionary
- Figure 3.4** Example from ‘simple patterns’ under verbs
- Figure 3.5** Example with ‘Verb with complement’ pattern
- Figure 3.6** Different groups belonging to the **V n** group
- Figure 3.7** Description of the ‘say’ group
- Figure 3.8** Data extraction and analysis procedures in the study
- Figure 3.9** Example of top 25 4-slot verb 4-gram sequences in the CLC
- Figure 3.10** Process of identifying core and emerging sequences
- Figure 3.11** Node verb ranking of the VV DT NN IN sequence
- Figure 3.12** Example from the verb 4-gram sequence, “VV DT NN IN”
- Figure 3.13** Example of the verb “arrange” in Collins Dictionary
- Figure 3.14** Procedure for **V n** identification
- Figure 3.15** Procedure for **V preposition** identification
- Figure 4.1** Outline of the results section
- Figure 4.2** Distribution of normalized frequency of the core sequences from A2 to C2 level
- Figure 4.3** Distribution of normalized frequency of the emerging sequences from B1 to C2 level
- Figure 4.4** Core sequences: BNC vs. L1 Korean speakers
- Figure 4.5** Emerging sequences: BNC vs. L1 Korean speakers

## Abbreviations

**BNC:** British National Corpus

**CEFR:** Common European Framework Reference of Languages

**CIA:** Contrastive Interlanguage Analysis

**CoE:** Council of Europe

**CAF:** complexity, accuracy, and fluency

**CQL:** Contextual Query Language

**EFCAMDAT:** English First-Cambridge Open Language Database

**ESL:** English as a Second Language

**EFL:** English as a Foreign Language

**L1:** first language

**L2:** second language

**LCR:** learner corpus research

**POS:** parts-of-speech

## Chapter 1 Introduction

Since the inception of second language (L2) writing research in the early 1900s (Matsuda, 2003), a myriad of studies has been dedicated to disentangling how L2 learners become proficient writers by investigating the factors that contribute to their L2 writing development. Empirical studies have looked at writing development from the perspective of grammatical complexity (Biber, Gray, & Poonpon, 2011; Lan, Liu, & Staples, 2019), lexical complexity (Kim, 2014), fluency (Palviainen, Kalaja, & Mäntylä, 2012), and accuracy (Kim & Emeliyanova, 2021) to name a few. These studies have made a significant contribution to our understanding of L2 writing development. However, traditional indices of absolute complexity (e.g., mean length of clause) do not align with some approaches to language learning, such as usage-based theories (Ellis, 2002; Kyle & Crossley, 2017), where language is seen as a dynamic process (Tomasello, 2009), rather than absolute. That is, language reshapes through usage events and continues to restructure as users gain more experience with it.

The main motivation for this study comes from the widely acknowledged notion that English academic writing continues to pose many difficulties for L2 learners (Kilmova, 2014; Kim, 2008; Reid, 2002; Silva, 1993). Such is particularly the case in languages such as Korean, where there are many typological differences between the two languages. Research to date has illustrated that L1 Korean speakers struggle when it comes to L2 English writing (Cho, 2014). With such cognizance, studies have offered a plethora of insight regarding learner errors and what learners perceive to be most challenging in L2 writing. However, what is still lacking in research is a comprehensive understanding of what the learners *can* produce at different trajectories of language development. Such knowledge can be particularly valuable to both educators and curriculum designers as they will have clearer evidence of learners' linguistic capabilities at different stages of their L2 writing development. Rather than overloading learners with too much information with regards to what "good" academic writing is to resemble, the teachers will be able to take a step-by-step approach by providing learners with a bite-sized amount of information at a time. Thus, this research is an attempt to provide English educators of L1 Korean speakers with a way to cater to their students as a way to facilitate a better learning experience for the students.

The current study is situated in usage-based theories and attempts to make sense of L2 English writing development among the first language (L1) Korean speakers using a bottom-up approach. As opposed to a top-down approach, where the learner data is compared with that of native English speakers, in a bottom-up approach, the data is drawn “from a learner corpus and using the learners’ own production as a starting point for error correction and gradual enrichment” (Osborne, 2004, p. 251). This approach is innovative in that many studies to date have chosen specific linguistic structures (e.g., constructions) as a starting point. However, this study uses a bottom-up research design by broadly exploring the use of verb part-of-speech (POS) sequences instead of a pre-selected linguistic feature as a starting point. A bottom-up research design is particularly useful because it could potentially provide a new understanding of the processes of L2 writing development. Furthermore, it could bring to the surface what the learners can produce at different stages of the learning trajectory. The following sections further discuss the approaches adopted in this study.

This chapter sets the stage for this study by discussing the main theories that shape the framework. It then addresses the gaps in the literature, terms selected for this study, and finally, the aims of the study.

## **1.1 Key Theories and Research Gaps of the Study**

The main purpose of this study is to understand the ways in which language evolves and changes among L1 Korean speakers as they become more proficient L2 writers. To better understand the goals of this study, the key theories and the respective research gaps are addressed.

### **1.1.1 Usage-based Approaches to Language Learning**

Usage-based approaches to language acquisition (Ellis, Römer, & O’Donnell, 2016) has their roots in cognitive linguistics. Its main emphasis is on the notion that as learners gain more exposure to the language, their use of language also expands and develops through means of restructuring the form-meaning mapping (Pérez-Paredes et al., 2020; Tomasello, 2003). Based on such a claim, several assertions can be made. First, language is dynamic (Kemmer & Barlow, 2000) in that as users engage with the language both productively and receptively throughout their lifetime, their language use will also change according to their speech community. Second,

language is sensitive to frequency (Kemmer & Barlow, 2000, Tyler & Ortega, 2016), where simple and most commonly used words and phrases are acquired first. That is, given that simple words and the linguistic feature will be used much more frequently than that more complex and less common, they will be entrenched in the learner's linguistic system before the learners are further exposed to more complex vocabulary and grammar. Third, usage-based approaches stress the emergence of schematic constructions as learners gain more experience with the language (Francis, Hunston, & Manning, 1996). In other words, the meaning of constructions is embedded not in the individual lexical items but at the level of phrases and sentences (Goldberg, 1995). Another critical claim in usage-based theories is that abstract linguistic representation of learners' grammar is built gradually from specific lexical content (Tomasello, 2009). That is, through repetition and encounter of similar usage instances, the learners are able to generalize the syntactic patterns. This study sets out to analyze the emergence of such patterns, and in order to do so, some original methodological approaches in corpus-based research are proposed.

Taking the above claims as a basis for language development, this study attempts to make sense of the restructuring process and how various linguistic features emerge as learners become more proficient in the L2. To date, studies have explored L2 writing from a usage-based perspective. For instance, Kyle and Crossley (2017) used computational indices to measure the frequency of verb argument constructions. They found that human raters may be influenced by both relative frequencies as well as the strength between the verb and the argument construction. Furthermore, Verspoor, Schmid, and Xu (2012) explored variables at the word, phrase, and sentence level to illustrate the dynamic usage of language in L2 writing in texts written by L1 Dutch speakers. The results were drawn from 64 different measures (e.g., sentence level, phrase level, verb level). The findings revealed that measurements such as sentence length, errors, and constructions were some of the factors that distinguished between different proficiency levels. However, it was pointed out that these indices are not absolute and can change depending on the learner. The empirical research on usage-based approaches to language learning has provided an in-depth understanding of the ways in which various linguistic features emerge. Although these studies have offered tangible insights, there are several gaps in the literature that this study seeks to fill. First, the current study takes on an exploratory approach, as opposed to starting with a prescribed set of constructions or lexical items, in understanding language usage. In other words, rather than pre-selecting the linguistic features to be investigated, the current study uses a

bottom-up approach to illustrate the ways in which language use evolves with proficiency. Additionally, research in learner language to date tended to focus on identifying patterns of overuse and underuse of linguistic features and comparing them with a native corpus (Paquot & Granger, 2012). Again, the current study takes an innovative approach to usage-based studies by illustrating what learners *can* produce at different trajectories of proficiency.

### **1.1.2 Pattern Grammar**

Pattern Grammar (Francis, 1993; Hunston & Francis, 1999; Hunston, 2015) was initially developed to capture the grammatical behavior derived from lexemes in a learner dictionary (Sinclair, 1995). According to Hunston and Francis (1999), language can be described in many patterns. That is, words have a tendency to a particular pattern that they appear. These patterns are typically observed through a large collection of data. Recent discoveries have been made that aspects of constructions and patterns are interchangeable (Hunston, 2019; Hunston & Su, 2019). The main difference between construction grammar and pattern grammar derives from its methodological origin. Constructions originated from the traditions of cognitive linguistics, while pattern grammar grew out of corpus linguistics (Hunston & Su, 2019). Traditionally, the discussion around form-meaning mapping in the usage-based approaches has involved constructions rather than patterns. More recently, however, it is becoming more relevant that patterns can indeed be constructions and vice versa (Hunston & Su, 2019). In the study by Hunston and Su (2019), the authors demonstrated that the same form-meaning mapping found in constructions could also be used as patterns. Furthermore, pattern grammar can offer a valuable contribution to our current understanding of usage-based approaches.

The scope of this study does not concern the similarities and dissimilarities between construction grammar and pattern grammar. Rather, pattern grammar was chosen over constructions for several reasons. Firstly, due to the vastness of constructions, identifying all of them would be an arduous task (Hunston & Su, 2019). On the other hand, pattern grammar comprises a manageable number of patterns (~200 patterns), which can be systematically categorized. Furthermore, applying a pattern grammar taxonomy rather than constructions in the current study is due to the nature of this study. While the current study recognizes that the meaning of words is not embedded in the individual lexical items, it does not attempt to illustrate the ways in which language is represented in our cognitive system. Instead, the current study

strives to illustrate language patterns from an observational standpoint. In that sense, pattern grammar was considered a more appropriate approach for the study.

### **1.1.3 Learner Corpora**

Learner corpora, defined as “electronic collections of texts produced by language learners” (Granger, 2008, p. 259), have made a significant contribution to second language acquisition theories. Among many of its merits, learner corpora are valuable for their ability to illustrate frequency information regarding various linguistic features in L2 production and to compare across and between different corpora (Gablasova, Brezina, & McEney, 2017). One issue when comparing data between other corpora, however, is to do with variability that is not due to L1 influence such as age, gender, and socioeconomic background (Ädel, 2015). Thus, in order to make inferences about interlanguage differences, the corpora should only vary in terms of the variable(s) being investigated but should be similar otherwise (Leech, 1998). The Cambridge Learner Corpus (CLC), the learner corpus used in this study, is suitable in the respect that although it does not filter for learner variables such as age and gender, it is categorized by proficiency, which is systematic. Furthermore, the main goal of this study is to understand L2 writing development among all L1 Korean speakers, irrespective of their background differences. In that sense, larger data would encapsulate the most accurate representation of L1 Korean speakers. Additionally, the current study also compares the findings from L1 Korean speakers with that of other L1 speakers. The ability to draw upon two large data sets from the same learner corpus is valuable in that not only are the findings generalizable individually, but it also warrants a more reliable comparison across the data as well.

As one of the pioneering studies in usage-based theories which utilize a large learner data, this study fills some of the prominent gaps that currently exist in learner corpora research. Primarily, of the biggest contributions of this study is both the immensity and robustness of the data. The current study uses the Cambridge Learner Corpus (CLC), a large learner corpus consisting of exam scripts. This corpus contains over 55.5 million words with learners from various contexts. Moreover, the data is categorized by the Common European Framework of Reference for Languages (CEFR) levels, a world-recognized system of measuring learner proficiency. Empirical research has shown that although studies have used L2 proficiency as a proxy for development, findings are often difficult to interpret due to the various interpretation of

the different proficiency levels (Gablasova, Brezina & McEnery, 2017). However, the corpus used in some studies (e.g., Kim, 2014) has liberally adopted the term “CEFR” to illustrate that the data has been categorized under a reliability proficiency measure. This could be misleading because it cannot be made certain that the raters were trained examiners and that all data has been assigned the CEFR level based on a systematic approach that often takes place in high-stake exams.

In a similar vein, the CLC is an immense written learner corpus, which includes learners from a wide range of contexts. Many existing learner corpora are smaller in scale (Aarts & Granger, 1998; Paquot & Granger, 2012) and often consist of data that were collected in a particular setting. To date, there are only a handful of written learner corpora consisting of L1 Korean speakers (e.g., The Gachon Learner Corpus; The International Corpus Network of Asian Learners of English (ICNALE); The Neungyule Interlanguage Corpus of Korean Learners of English (NICKLE); The Seoul National University Korean-speaking English Corpus (SKELC); The Yeonsei English Learner Corpus (YELC); The Korean Learner Corpus). Aside from the ICNALE and YELC, the proficiency level in the corpora lacked systematicity. The more significant issue, however, was the setting in which the data was collected. The YELC and SKLEC, for example, were a collection of essays written by university students. Others did not provide learners’ metadata, making it difficult to discern how the data was collected. Thus, the inclusiveness of the CLC makes it possible to generalize the data across a broader range of contexts.

Lastly, as mentioned above, only a few studies have used a bottom-up approach to examine L2 learner language (Aarts & Granger, 1998; Gilquin, 2018). There are even fewer studies that use part of speech (POS) tags to explore L2 writing development. The bottom-up approach could be particularly valuable in language research as it draws on the big picture about what the learners *can* do at different stages of language trajectory. While a top-down approach provides insight into trends such as misuse, overuse, and underuse of the linguistic feature under investigation, a bottom-up approach would illustrate *how* different linguistic features characterize L2 writing.

#### **1.1.4 Clarification About the Term “verb 4-gram sequence”**

This study adopted the term “verb 4-gram sequence,” which comprises a node verb followed by three POS tags (“V.\*” [] [] []). The term distinguishes it from others, such as “constructions,” which could be misleading since the current study draws upon pattern grammar. The motivation for choosing “verb 4-gram sequences” is first to illustrate that the “sequences” drawn out from the corpus are 4-grams, headed by a node verb. Additionally, the term “sequences” denote that the 4-grams identified are sequences of words that are most commonly used among the learners. Using other terms such as “chunks” could be misleading since the study is not identifying lexical chunks. Thus, acknowledging that the “4-grams” drafted could be a random sequence of words (e.g., “went there. I”). The study utilizes “verb 4-gram sequences” throughout.

### **1.2 Purpose of the Study**

The main purpose of this study is to understand the L2 writing development of L1 Korean speakers using a quasi-longitudinal method. By doing so, it seeks to offer the ways in which language evolves and restructures as learners become proficient in the L2. In order to achieve this goal, the study is broadly divided into three parts. First and foremost, it attempts to identify the changes in the use of verb 4-gram sequences among L1 Korean speakers using a large learner corpus. To do so, it first identifies the top most frequently occurring sequences in each proficiency level. Then, the verb 4-gram sequences are ranked according to their normalized frequencies and compared across levels to observe the changes in the use of the sequences as learners become more proficient in the L2. Grounded in the theories of usage-based approaches to language acquisition, doing so would allow us to see the various verb 4-gram sequences that characterize learner language at the different developmental trajectories.

Another purpose of this study is to regroup some of the identified samples from the verb 4-gram sequences according to the pattern grammar taxonomy (Hunston & Francis, 2000). By regrouping the lexical realizations, it seeks to illustrate the form groups of the verbs and understand how the form-meaning relationships change across the proficiency levels. While the verb 4-gram sequences can provide an overview of the sequences used by learners at various proficiency levels, it lacks the ability to illustrate the *kind* of language used by learners. Thus, by

re-categorizing some of the candidate sequences by grammar pattern and then their respective meaning groups, the findings further probe into the variation in language use.

Lastly, the study seeks to compare the findings from L1 Korean speakers with two different cohorts: other L1 speakers and L1 English speakers. The purpose of comparing data across the various groups is firstly to identify the interlanguage differences. More specifically, the study seeks to confirm whether the verb 4-gram sequences and verb patterns identified in the study are specific to L1 Korean speakers or across other L1 speakers as well. In that sense, the other L1 speaker sub-corpus did not discriminate for any specific L1; rather, it included all other L1 speakers in the CLC. Furthermore, comparing the findings among L1 Korean speakers at the near-native level (e.g., C2) with that of L1 English speakers is to be conducted as a way to illustrate the extent to which L1 Korean speakers mirror the language use of L1 English speakers at the highest proficiency level. Overall, contrasting the findings from L1 Korean speakers will help to understand better the characteristics of L2 writing development among L1 Korean speakers.

### **1.3 Outline of the Thesis**

The first chapter of this thesis sets the stage by briefly discussing the overarching theories and terms used in this study. Then, the aims of the study and the research gaps are addressed. The remaining chapters are as follows:

Chapter 2 provides a comprehensive review of the relevant literature, staging the theoretical and conceptual framework which will be utilized to address the issues and fulfill the research gaps. It discusses the research gaps and the ways in which this study contributes to addressing some of the problems. Furthermore, the chapter provides a thorough definition and explanation of the terms and concepts used in the study.

Chapter 3 provides an overview of the methodology of the study. It first presents the research questions. Then, it gives a detailed explanation of the corpora and various instruments used in the study. Methodological gaps are also addressed, along with a detailed outline of the procedure and data collection process. The procedure for data analysis is also provided.

Chapter 4 discusses the results of the study. The findings are presented according to the research questions presented in Chapter 3. It first presents two types of verb 4-gram sequences identified among L1 Korean speakers: “core” and “emerging” sequences. The examples from

these sequences are then regrouped according to pattern grammar. The findings are then compared among other L1 speakers and L1 English speakers.

Chapter 5 reports the findings from Chapter 4 and discusses them in light of the theoretical framework and results from the previous study. It addresses the ways in which the findings provide an in-depth understanding of L2 writing development from the perspective of usage-based approaches to language learning.

Chapter 6 concludes the thesis by summarizing the main findings of the study, as well as discussing the methodological contributions and pedagogical implications of the study. It also offers recommendations for future research.

## **Chapter 2 Literature Review**

### **2.1 Introduction**

This chapter provides a critical review of the literature, presenting a detailed overview of the theories, frameworks, and empirical research which identify the gaps and formulate a unique framework for this study. It first begins by broadly discussing usage-based approaches to language acquisition. Then, the following section transitions to Francis & Hunston's pattern grammar, drawing its relevance to usage-based approaches. The next section delves into the second language (L2) development by discussing relevant areas such as the impact of interlanguage on L2 writing development and how L2 writing development has previously been measured. Then, a modern approach to measuring syntactic development using a usage-based perspective is presented. Finally, the chapter concludes by exploring the past and current trends of incorporating learner corpora in exploring part-of-speech (POS) patterns and the theoretical framework adopted in this study.

### **2.2 Usage-Based Perspective on Language Learning**

Usage-based approaches to language, a term that was first coined by Langacker (1987, 1988), is grounded on the notion that an individual's linguistic system emerges from usage and experience with it (Behrens, 2009; Bybee, 2010; Ellis, O'Donnell, & Römer, 2013; Tomasello, 2003). The usage-based perspective on language acquisition was first observed with respect to first language acquisition among children. The proponents of usage-based theories claim that children acquire their first language from intentional communication and actual usage events (Lieven & Tomasello, 2008). Initially, children are introduced to the language through individual lexical items and chunks such as "apple" or "what's that?". In the early stages of language acquisition, when the child has yet to acquire the grammar rules, language would be a simple association between words. For example, if someone asked a child the question "What's that?", they would respond accordingly with an utterance, but this is most likely due to the caretaker's continued use of the expression, which would have been stored as a form of a request for information in the child's brain. Tomasello (2000) addresses a few fundamental facts about child language acquisition. During the early stages of language development, children tend to reproduce whole adult utterances rather than words. As these strings of words are used

frequently by the caretaker, they then become entrenched in the child's repository of language (Lieven, 2014). Thus, much of a child's words may be fixed and item-based expressions. As the child grows, however, so does his or her use of language in a way where the child moves away from fixed expressions to more abstract use of language. This could be seen as their attempt at trying to create novel utterances by replacing "slots" in adult utterances with different words. At this stage, the child may produce errors as they try to internalize the grammar rules. Evidently, usage-based approaches do not consider language to be a natural phenomenon that babies are born with. Rather, through meaningful interactions with the caretaker, children *learn* to use language in sophisticated ways without being taught.

In a similar vein, Kemmer and Barlow (2000) define the central properties of usage-based models with respect to nine characteristics. Firstly, the authors claim that there is a close relationship between linguistic structures and usage. In other words, the linguistic system is a mental system that has a close link with usage events where the ways in which language is used are built upon the speaker's experience with the language. Expanding on the claims made by Tomasello (2000), abstract linguistic representations of one's grammar are built gradually and start with specific lexical content. With repetition and instances of similar usage events, which could be at the level of phonemes, morphemes, and syntactic patterns, the general representation of patterns and schemas<sup>1</sup> emerges. That being said, usage events of language are at the crux of the dynamic and constant restructuring of the linguistic system throughout life. As interlocutors continue to engage with language, their language will also reshape. Thus, language is not a "set phenomenon" and is subject to continuous change.

Secondly, Kemmer and Barlow (2000) emphasize the importance of frequency. As stated above, if exposure to language can impact how it is used, then the frequency of exposure must be a fundamental factor in how language becomes entrenched in the learner's mind. Supporting such a claim, Lieven and Tomasello (2008) claim that the "entrenchment of lexically specific strings, as a result of high frequency, is of considerable theoretical importance in linguistic theory" (p. 176). Thus, the most frequently used language units or patterns result in a greater chance of being conventionalized in the speaker's mind. The link between the ability to construct

---

<sup>1</sup> Defined as "a cognitive representation comprising a generalization over perceived similarities among instances of usage. Schemas arise via repeated activation of a set of co-occurring properties and are used to produce and understand linguistic expressions." (Kemmer & Barlow, 2000, p.xxiii)

language and human cognitive sensitivity to frequency has been supported by many studies (e.g., Tyler & Ortega, 2016). For instance, if humans use language based on what they hear, then the frequency of the input should provide compelling evidence for language acquisition. Then, it is evident that simple sentence patterns and words with high frequency are learned and processed faster than abstract sentence patterns and words with low frequency (MacWhinney, 2001). Frequency is also in line with Zipfian distribution (Zipf, 1935), which follows that the probability of occurrence of prototypical forms (e.g., canonical collocations, conversational structures, idioms, lexical items, morphemes, phrases, non-canonical collocations) will be much greater than others, which will further elicit the process of implicit language acquisition (Larsen-Freeman, 2013). The Zipfian distribution is an experimental law, also known as Zipf's law, which claims that language has a more substantial reliance on frequency (Piantadosi, 2014). Thus, shorter words, which are often more straightforward and easier to remember than long and complex words, are favored over words with uncommon usage. As will be discussed further, regarding second language learning, this is to say that simple and frequent words are acquired first.

Third, “comprehension and production are integral, rather than peripheral, to the linguistic system” (Kemmer & Barlow, 2000, p. xi). This view rejects the traditional idea that “competence” and “performance” were seen as two distinct entities. Instead, the speaker's ability to perform and use the language is highly dependent upon their linguistic competence. Fourth, both learning and experience are at the crux of language acquisition. If language is acquired through usage and experience, then it is evident that language should be learned. Thus, it is not an innate mechanism that one is born with. On this end, “linguistic representations [are] emergent, rather than stored as fixed entities (Kemmer & Barlow, 2000, p. xii). In other words, language is not stored as a fixed set of rules but as linguistic patterns. These patterns continue to evolve and emerge with exposure. Thus, usage-based is grounded in patterns (Barlow, 2000; Biber, 2000) and constructions (Ellis, Römer, & O'Donnell, 2016; Goldberg, 2006) in that language can be described into form-meaning pairings which are entrenched in the learner's mind (Goldberg, 1995).

Similar to the points mentioned above, usage, synchronic variation, and diachronic change are very closely related (Kemmer & Barlow, 2000). As our speech community continues to change, the way we speak also adjusts accordingly. In that sense, it is evident that the

linguistic system is closely linked with our non-linguistic systems, where our language competence is not only the work of our linguistic ability but also our knowledge about the cultural events and the interaction in particular scenarios (Fillmore, 1977). Lastly, Kemmer and Barlow (2000) state that context plays a crucial role in the linguistic system. The language usage always occurs with communicative intentions with the interlocutor. Thus, language is influenced by the context in which it is used.

So far, the discussion of usage-based approaches to language learning has been observed at the general level. However, over the past 30 years, a myriad of studies have taken a similar approach in illustrating that the same usage-based theories about acquiring a first language can be applied to acquiring a second language (Bybee, 2008; Ellis, 2008, 2009; Ellis, O'Donnell Römer, 2013; Eskildsen, 2008; Kim & Sung, 2019; Kyle & Crossley, 2015; Pérez-Paredes, Mark, & O'Keeffe, 2020, Shin, 2017, Verspoor, Schmid, Xu, 2012). For instance, Ellis (2008) explored features associated with both L1 and L2. He claims that some of the similarities between L1 and L2 acquisition include aspects such as frequency, contingency relationship, cue competition, and salience. However, he notes that some aspects of usage-based approaches are more specific to L2, which may help explain why L2 learners are much less successful than L1 speakers in their language learning. One of the crucial points made is the notion of input and intake (Corder, 1967). The term "input" refers to the exposure provided to the learners. In contrast, "intake" refers to the amount of the L2 input that the learners are able to process and build understanding successfully and thereby allowing for learners to use the language. While the input in the L2 may be similar to L1 in that the learners are immersed in a naturalistic environment, L2 learners differ significantly from L1 with respect to their intake. Other studies have also explored L2 learning from a usage-based perspective. Eskildsen and Cadierno (2007) looked at the use of negation patterns in speaking by Mexican learners of English. More recently, Shin (2017) conducted a study among L1 Korean speakers to investigate the learners' developmental pattern of English argument structure constructions. In the experiment, participants were taught six English argument structure types over three month period. Followed by instruction, the participants were given a wide range of tasks such as grammaticality preference tasks, writing tasks, and free-writing tasks in order to measure their frequent use of argument constructions. In brief, the results manifested that after exposure to the argument structure constructions, the participants demonstrated their ability to use target constructions but

preferred two-argument constructions over the three-argument constructions. Based on these studies, it can be derived that there is some evidence of the ways in which usage-based approaches to L2 learning are similar to that of L1 acquisition, but there is still much more to be discovered.

Taking the claims made with regard to usage-based approaches to language acquisition, Tyler and Ortega (2018) proposed five tenets to usage-inspired L2 instruction. Firstly, the authors claimed that language learning is meaning-based. As mentioned, the usage-based perspective rejects the traditional view. Instead, it advocates that the meaning does not lie within individual words but in the constructions (Goldberg, 1995, 2006). On that note, the syntactic patterns such as abstract and complex constructions carry meaning (Langacker, 1991). Thus, language in usage-based learning is exemplar-based, and “grammar is what results when formulas are re-arranged, or dismantled and reassembled in different ways” (Hopper, 1987, p. 145). It is not an innate system that one is born with but what is a result of exposure and usage. The second tenet is that just as “meaning is grounded in the physical world and is embodied” (Tyler & Ortega, 2018, p. 7), so is language and language learning. Again, language is formed through interaction and engagement with the world. It is a representation of our conceptual and cognitive perception of the world and is subject to change as our community continues to change during our lifetime.

Thirdly, Tyler and Ortega (2018) emphasize that “language and language learning are critically situated in contextualized social interactions” (p. 8). Context is what drives the language, and thus, the user’s linguistic choice will differ accordingly. Fourth, the emergence of language is not distinct from any other aspects of “learning.” Instead, learning a language requires the same cognitive mechanisms as learning to master a skill (e.g., riding a bike). Just as it requires practice and repetition in order to be good at something, the same rule applies to language. That is, frequency of exposure is at the heart of usage-based learning, where the amount of input dictates the learning outcome. This is a plausible assumption when considering how L2 learners pick up simple and frequent words at a much faster rate than difficult and complex words and constructions. For L2 learners, this may mean that language learning is both implicit and explicit. While they internalize complex and abstract constructions through usage, there also needs to be some level of exposure for their attention to be drawn to the various forms

and meanings. This is particularly the case for idiomatic expressions and phrases. Lastly, language is not fixed and is entitled to change over the span of our lives.

It is evident that the five tenets proposed by Tyler and Ortega (2018), in conjunction with the general theories of usage-based approaches to language learning, emphasize the importance of exposure and interaction. Language *can* be learned through exposure, but by no means are humans born with innate knowledge about the language. In sum, it might be beneficial to revisit the definition of usage-based as proposed by Langacker (1987):

“Substantial importance is given to the actual use of the linguistic system and a speaker’s knowledge of this use; the grammar is held responsible for a speaker’s knowledge of this use; the grammar is held responsible for a speaker’s knowledge of the full range of linguistic conventions, regardless of whether these conventions can be subsumed under more general statements. [It is a] nonreductive approach to linguistic structure that employs fully articulated schematic networks and emphasizes the importance of low-level schemas” (p. 494).

As a final note, the theoretical framework of this study is situated in the view that language is learned through usage events. Thus, frequency of exposure plays a pivotal role in L2 development. The current study is an attempt to understand the ways in which language evolves and develops with development.

### **2.2.1 Cognitive Grammar in Language Development**

There has been a long debate involving the two divergent linguistic paradigms of language development: generative and cognitive approaches. In the broader sense, generative approaches to language learning are foregrounded in the notion that “language is not learned” (Chomsky, 1980) and that human beings are born with innate grammatical structures. The proponent figure behind generative approaches to language is Chomsky (1957, 1959, 1965), who argues that the complexity of human language and our ability to generate an infinite number of expressions at all levels of abstraction can be explained only through the scrutinization of the individual’s genetically determined linguistic system. He proves this claim through the illustration of how children learn to talk (Chomsky, 1959). When babies begin to interact with their caretakers, they are initially exposed to simple baby utterances such as “*More milk*” or “*Give me*” (Braine, 1976); however, as children grow and their linguistic system evolves, they become capable of constructing complex and abstract utterances such as “*Could I have some*

*more milk?*” without any exposure to such input. This brings to an important discussion of *how* and *what* allows children to be able to produce such language. The Poverty of the Stimulus argument (Schwartz & Sprouse, 2013) supports the notion that there is not rich enough exposure to linguistic environments in which children are able to acquire every aspect of the language. The only way in which this can be explained, then, is to acknowledge that there must be some “built-in knowledge” (e.g., Universal Grammar) that accounts for children’s ability to utilize abstract linguistic features without instances of exposure.

On the other hand, cognitive linguists (e.g., Croft & Cruse, 2004; Langacker, 2000; Lakoff, 1987; Robinson & Ellis, 2008b; Tomasello, 2003) view language acquisition in ways that are different from the generative approaches to language acquisition. The most noticeable difference between the two views is that cognitive linguists do not view language to be grounded in our innate language system but rather as something that grows and evolves along with our experience in the world (Ellis, Römer, & O’Donnell, 2006). That is, as our perception of the world continues to grow and change, our use of language changes as well. Thus, cognitive linguistics can be broadly expressed in relation to three elements that are closely intertwined together: language, communication, and cognition (Ellis & Robinson, 2008). Language is a primary means of communication, a vehicle for information delivery, and how we interact with and respond to the world around us. Through our interaction with language, we are able to determine structures, which involves a wide range of aspects of cognition, such as “remembering of the utterances and episodes, the categorization of experience, the determination of patterns among and between stimuli, the generalization of conceptual schema and prototypes from exemplars” (Ellis & Robinson, 2008, p. 3). Whereas theories have viewed basic units of language at the level of individual lexical items, cognitive linguistics view basic language representation as constructions, form-meaning mappings that are entrenched in the learners’ repository system (Ellis & Robinson, 2008).

With such divergent perspectives on language development, there has also been ample discussion of both generative and cognitive approaches to language learning (e.g., Lakoff, 1991; Domaradzki, 2007; Bybee, 2010). It should be recognized, however, that the contrasting views of cognitive and generative grammar do not pose the superiority of one approach over the other. Instead, they are two mutually exclusive theories of standpoints which we foreground our understanding of the complex phenomenon of language development. For example, Langacker

(2000) uses contrasting terms to provide a fundamental distinction between generative and cognitive grammar. He associates the nature of cognitive grammar with words such as “maximalist,” “non-reductive,” and “bottom-up.”

On the contrary, he claims that words such as “minimalist,” “reductive,” and “top-down” are more closely linked with generative theory. Interestingly, Langacker’s choice of terms points to an essential distinction between the two approaches- in cognitive grammar, the grammar is idiosyncratic in nature and is dependent upon the experience and its usage, whereas in generative grammar, the grammar is “stripped down” and “economical” (Kemmer & Barlow, 2000). Thus, generative grammar and cognitive grammar offer contrasting views about language learning in which different linguistic theories are foregrounded. The perspective of usage-based theories, however, takes on the view of cognitive grammar in that language and grammar emerges from events of usage, and it is through usage that the language users are able to formulate abstract language without exposure to it.

In lieu of the discussion in this section, the development is also observed from the perspective of verb 4-gram sequences (see chapter 1) rather than from individual lexical items. By doing so, it seeks to capture the various ways in which language use becomes abstract as learners become proficient users and how L1 Korean speakers’ use of language differs from that of other L1 and L2 users of English.

### **2.2.2 Usage-Based grammar: (Verb) Constructions and Argument Structure**

The usage-based approaches to language learning stress the notion that schematic constructions are from the users’ experience with the language (Bybee, 2010; Ellis, 2012). More recently, however, studies have shown many overlapping areas between construction grammar and pattern grammar (Hunston & Su, 2017; Hunston, 2019). While neither the discussion of construction nor drawing a link between pattern grammar and construction grammar falls within the scope of this study, this section discusses the ways in which constructions have played a central role in usage-based theories.

The notion of Construction Grammar (Filmore, Kay, & O’Conner, 1988; Lakoff, 1987; Goldberg, 1995) argues that “constructions” carry meaning that is independent of the meaning of individual lexical items. More specifically, Goldberg (1995) claims that “C is a CONSTRUCTION iff<sub>def</sub>C is a form-meaning pair  $\langle F_i, S_i \rangle$  such that some aspect of  $F_i$  or some

aspect of  $S_i$  is not strictly predictable from  $C$ 's component parts or from other previously established constructions" (p.4). Based on this assertion, constructions refer to the pairing of a word from its abstract meaning, which can occur between word levels or at sentence levels. It is a *unit of knowledge* where the meaning is non-predictable based on individual lexical items or a combination of them. For example, idiomatic expressions such as "piece of cake" or "call it a day" do not carry meaning across individual lexical items. But instead, the meaning is constructed in the phrase. Thus, constructions unite both vocabulary and grammar, considering it as a single entity (Sethuraman, 2002).

One fundamental position of construction grammar in theories of cognitive linguistics is that language can be conceptualized as constructions (Goldberg, 1995, 2006), establishing form-meaning mapping in learners' minds (Croft & Cruse, 2004; Goldberg & Casenhiser, 2008; Robinson & Ellis, 2008a), which are then entrenched in the learner's repository system (Goldberg, 1995). It should be noted, however, that the establishment of the form-meaning connection is not restricted to words (Goldberg, 1995, 2006). Form-function pairings can occur between various aspects of language, including idiomatic expressions and abstract grammar constructions (Ellis & Larson-Freeman, 2009; Ellis, Römer, & O'Donnell, 2016).

The term "construction" and its varying degrees of abstraction is at the heart of many usage-based studies on language acquisition. Depending on how frequently the word occurs between the interlocutors in a conversation, for example, learners can discover the candidate construction within the speech stream, from which they are able to identify the grammatical categories (Tyler & Ortega, 2016). By doing so, learners posit simple syntactic structures, which they can then recall at their disposal. Additionally, theories of cognitive linguistics claim that schematic construction becomes apparent from language users' experience with and memories of particular exemplars (Bybee, 2010; Goldberg, 2006). Thus, constructions and usage-based theories are closely related in that constructions are learned while being engaged in communication (Barlow & Kemmer, 2000; Hopper, 1998).

In line with the acquisition of individual words in usage-based theories, constructions are also learned in a progressive manner (Sethuraman, 2002). In other words, the acquisition of basic syntactic patterns can allow learners to acquire more complex structures. The acquisition of constructions can be better elaborated by discussing how first language acquisition takes place. For instance, when a child first learns a language, he or she starts by using individual words,

which then extend to more complex and abstract constructions. In a study by Tomasello (1992), he observed his English-speaking daughter's use of multi-words in her speaking. At first, his daughter's speech started with using specific verbs to convey her thoughts. With the progressions of her development, Tomasello noticed that his daughter was using each verb with its own unique set of constructions. An interesting finding from this study was that each verb appeared to be used at its own unique utterance-level schemas, which did not show any correlation with how other verbs were used. In other words, the acquisition of one type of construction with a specific verb did not mean that the child was able to use the same construction with another verb.

Although the study by Tomasello (1992) provided a tangible insight into the relationship between verbs and constructions, studies have shown that the language used by children was not fully adult-like (e.g., Bowerman, 1976; Tomasello, 2000). Furthermore, such a result cannot be generalized to second language acquisition (SLA) for the reason that, unlike first language (L1) acquisition, second language learners have already acquired an L1. For instance, Kim and Sun (2019) examined L1 Korean speakers' use of resultative constructions in argumentative writing. They used a corpus consisting of 117 samples with 30,305 words. The results indicated that learners with higher proficiency were about to not only utilize less frequent resultative constructions but were also able to use a wider range of verbs as well. Year and Gordon (2009) also conducted a study in the Korean context to investigate children's use of ditransitive constructions in an ESL setting. The participants in the six classes were assigned to either a skewed frequency group or a balanced frequency group. In the skewed frequency group, the input was heavy on the word *give*, whereas in the balanced frequency group, there was no skewness in the input. The study proposed that skewed input did not facilitate better construction learning, but instead, balanced construction learning incorporating a wide range of verbs seemed to be much more effective. Both of these studies were conducted in the Korean context; however, the focus in these studies was on a specific type of construction (e.g., ditransitive construction, resultative construction), which does not provide a holistic understanding of how the wide range of constructions<sup>2</sup> are used by learners. While a close zoomed-in observation of a single construction is valuable, looking at it from a zoomed-out perspective can also provide valuable pedagogical implications.

---

<sup>2</sup> Note: this study does not seek to observe constructions. Rather, it seeks to understand the L2 writing development of patterns. More explanation about this is provided later.

The discussion now transitions to verbs, more specifically, verb-argument constructions. Empirical research has demonstrated that the meaning of constructions can be identified with respect to their verbal distribution (Goldenberg, Casenhiser, & Sethuramann, 2004; Stefanowitsch & Gries, 2003). Additionally, languages are typically used to discuss the occurrences of events with participants, in which the function is largely informed by verbs (Perek, 2015). Thus, verbs do not occur in isolation but typically in conjunction with other words, which are identified as verb “arguments.” What this claim entails about verbs is in line with constructions in that meaning is carried out in both the lexicon as well as in syntactic constructions. Below is an example of eight ways in which the verb “kick” can mean differently based on its syntactic constructions (Goldberg, 1995).

- a) Pat **kicked** the wall.
- b) Pat **kicked** Bob black and blue.
- c) Pat **kicked** the football into the stadium.
- d) Pat **kicked** at the football.
- e) Pat **kicked** his foot against the chair.
- f) Pat **kicked** Bob the football.
- g) The horse **kicks**.
- h) Pat **kicked** his way out of the operating room.

Notably, the meaning of the verb “kick” is identical in all of the sentences above; however, the differences in the meanings are in the constructions. The above example also confirms that argument realization cannot be fully informed by individual words (Perek, 2015). Instead, it is through the observation of both individual words, together with their constructions, that learners’ full knowledge can be informed.

Previous studies have shown that children’s use of verbs is highly sensitive to their mothers’ usage (e.g., Goldberg, Casenhiser & Sethuramann, 2004). Furthermore, it is claimed that “The most important pattern are those of the verbs” (Hornby, 1974, p. v) and that “all language provides ways to talk about events and their participants; this function is typically assumed in great part by verbs. It is precisely for this reason that, more so than other content words, verbs are rarely uttered in isolation but are usually accompanied by certain other words,

called the “arguments” of the verb” (Perek, 2015, p.1). Notably, verbs are an essential part of speech and should most certainly be observed with respect to other words surrounding them.

### **2.3 Francis and Hunston’s Pattern Grammar**

Pattern grammar is “an approach to lexis and grammar based on the concept of phraseology and of language patterning arising from work on large corpora” (Hunston & Francis, 2000, cover blurb). The idea of “pattern grammar” was first introduced in the 1980s and 1990s, where it was used for annotating dictionary entries in the *Collins COBUILD English Dictionary* (Sinclair *et al.*, 1995). Pattern grammar has implications at both theoretical and practical levels. At the theoretical level, the pattern grammar attempts to “generalize from Sinclair’s ideas about units of meaning and about the connection between form and meaning to achieve a more systematized approach to the grammar of lexicon” (Hunston, 2019, p. 328). At the practical level, pattern grammar attempts to “arrive at a means of coding the behaviour of individual word sense in a way that would be suitable for a monolingual dictionary of English for learners” (Hunston, 2019, p. 328). Thus, pattern grammar is grounded in the notion that, essentially, all words belong to a pattern (Hunston & Francis, 2000). That is, “there is a strong tendency for sense and syntax to be associated” (Sinclair, 1991, p. 65). Grammar patterns are particularly useful in identifying the grammatical behavior of the words and thereby making it possible to generalize existing patterns in learner language.

As with any approach, however, pattern grammar is subject to potential issues. Firstly, identifying patterns is done manually in accordance with the guidelines available online. Thus, it is subject to variability depending on the coder. Secondly, in order to identify the correct pattern, concordance lines will have to be consulted. This may bring much inconvenience, mainly when using large data, as it can be time-consuming to sift through individual concordance lines. Next, there is not a fine line between a pattern, and non-pattern and frequent occurrence does not indicate that the sequence of words is a pattern (Hunston & Francis, 2000). As will be shown later, the same is also the case in this study, where some of the frequently occurring sequences did not belong to any particular pattern. Thus, careful judgment is in demand when categorizing and interpreting patterns.

In recent years, pattern grammar has been used in various language-based studies (e.g., Hunston & Su, 2019) and has illustrated a plethora of pedagogical implications (Hunston, 2019).

Despite such caveats, pattern grammar is particularly useful in identifying patterns in a systematic way. Moreover, evidence from corpus alone can be used to generalize patterns or ‘constructions’ (Hunston, 2019). The following section discusses the ways in which constructions and pattern grammar diverge and overlap.

### 2.3.1 Comparison of Construction and Pattern Grammar

Thus far, constructions and pattern grammar have been discussed separately. More recently, however, the discussion of the similarities and the compatibility between the two approaches have been acknowledged (e.g., Ellis, Römer & O’Donnell., 2016, Stefanowitsch & Gries, 2003, Hunston & Su, 2017, Hunston, 2019). This section discusses areas where the two approaches share similarities and differences. It also provides the rationale for situating the study in the framework of pattern grammar rather than construction grammar.

There are two recent studies that have specifically explored the alignment of pattern grammar and construction grammar (Hunston, 2019) and proposed the ways in which meaning-pattern combinations can also be identified as constructions (Hunston & Su, 2017). In Hunston (2019), the author suggests that the meaning group of grammar patterns can be used to identify potential constructions. She also highlights that based on the literature, it is possible to claim that patterns and constructions are coterminous. The example in **Table 2.1** demonstrates her claim.

Example	Pattern nomenclature	Construction nomenclature
<i>... talked his way into a top job</i>	verb-way- prepositional phrase	The way construction
<i>... dismissed the idea as rubbish</i>	Verb-noun-as-noun	The appositive as construction
<i>... fooled them into believing that...</i>	Verb-noun-into-ing	The causative into construction
<i>... told them a story...</i>	Verb-noun-noun	The ditransitive construction
<i>... nibbled at a piece of cheese</i>	Verb-at-noun	The conative at construction

**Table 2.1 Examples to demonstrate that patterns are also constructions (Hunston, 2019)**

As shown in the table, the example "...talked his way into a top job" illustrates the pattern "verb-way-prepositional phrase" represented by a verb followed by "way" and preposition phrase. Similarly, from the perspective of construction, the same example would be identified under the 'way construction'. Another example is "...told them a story", which follows

a "verb-noun-noun" pattern, which would be identified as the ditransitive construction. However, although patterns can often be used to identify constructions, it may not be that one pattern is equivalent to one construction. But through the meaning groups in the grammar pattern, the constructions can be identified at a consistent level. Lastly, she claims that regrouping the identified groups into broader categories could potentially prepose the hierarchy of the constructions.

Hunston and Su (2019) also propose that many meaning-pattern combinations of adjective complementation can also be identified as constructions. The authors demonstrate this by taking 44 adjective complementation patterns, comprising four patterns: “1) adjectives followed by a that-clause, to-infinitive clause, wh-clause, or -ing clause, 2) adjectives followed by a prepositional phrase, 3) patterns with *it* and 4) patterns with *there*” (see Hunston & Su, 2019 p. 573-574). The patterns were also comprised of three meaning groups: 1) the ‘nervous’ group, 2) the ‘angry’ group, and 3) the ‘good’ group. Each of the 44 adjective complementation patterns was then annotated with respect to its patterns. Interestingly, the authors were able to identify constructions that were aligned by pattern as well as sets of adjectives that were used in patterns in some cases. For example, the authors proposed the construction *it* v-link ADJ clause for a group of patterns plus its set of adjectives. The authors do, however, note that the alignment of pattern and construction is debatable. Still, the study offers an insightful view of the extent to which constructions and pattern grammar are aligned.

Most recently, Green (2019) conducted a study drawing on the frameworks of both pattern grammar and construction grammar to describe the lexicogrammar of academic vocabulary in two academic wordlists: Academic Word List (Coxhead, 2000) and Secondary Vocabulary List (Green & Lambert, 2018). The grammatical patterns which are most frequently associated with academic vocabulary are extracted using collocation analysis, a type of collocation analysis that is commonly used in construction grammar. The study also points out that in construction grammar, mid-level constructions are essentially grammar patterns that are associated with the lexical item. The study attempted to enrich both the Academic Word List and Secondary Vocabulary List with patterns in hopes of providing implications that would add more depth to vocabulary knowledge.

The discussion now transitions to some similarities and differences shared by pattern grammar and constructions. Firstly, neither construction grammar nor pattern grammar considers

syntax and lexis as distinct entities. Rather, as viewed among usage-based linguists (e.g., Langacker, 1987, Bybee, 1996, Fillmore, 1989, Goldberg, 1995, Croft, 2001, Tomasello, 2003), vocabulary and grammar are dependent where meaning is not embedded in individual lexical entities, but in a meaningful combination of morphemes, words, and phrases. Similarly, both construction grammar and pattern grammar emphasize the inter-connectedness of form and function. Both construction grammar and pattern grammar are also similar with respect to the status of grammar rules where “grammatical generalizations do not rest on a rigid foundation but are the accumulation of the patterns of hundreds of individual words and phrases.” (Sinclair, 1991, p. 100). Then, it can be seen that grammar is essentially bottom-up rather than top-down.

One of the significant differences between construction grammar and pattern grammar is in its origin. Construction grammar is born out of cognitive linguistics (Fillmore, 1989; Goldberg, 1995; Croft, 2001), whereas corpus linguistics derives from corpus linguistics (Hunston & Su, 2017). Although both are usage-based models of language, construction grammar defines constructions as ‘mental constructs’ and argues that grammar emerges from its use through the grammaticalization of lexical items. This entails that if a word is used frequently enough, then it has the ability to turn vocabulary into ‘grammatical information’ (Green, 2019). Pattern grammar, on the other hand, is observation-based and has an interest in the output (Hunston, 2014). Thus, it does not model the mental representation of languages, such as the emergence of abstract language phenomena and schematization of the language, as it does in constructing grammar.

This section has provided some existing discussions involving the intersection of constructions and patterns. More specifically, studies have demonstrated the ways in which patterns can inform constructions and vice versa. Because the nature of constructions and patterns are different, although corpora are becoming more widely used for constructions (Hunston & Su, 2017), whether to prioritize either constructions or patterns may be difficult to say. However, it is certain that the incorporation of both construction grammar and pattern grammar could offer rich evidence to learners’ language development in the usage-based sense. The study observes grammar patterns rather than constructions for the following reasons: 1) Currently, there is not enough detailed description of the wide range of constructions (Hunston, 2019). This could be particularly problematic in terms of pedagogical implications. 2) The current study is a corpus-based study, which attempts to look at various emerging verb

sequences. As descriptive research, rather than prescriptive research, one of the main aims of this study is to observe the emerging sequences through frequency. It takes on a bottom-up approach without a list of constructions to be drafted. One of the main reasons for doing so is to be able to explore the wide range of patterns and verb sequences without any restrictions. 4) Lastly, the study hopes to make a contribution to pedagogy to inform both teachers and students of the various verb sequences that are most prevalent in different proficiency levels. For these reasons, the study looks at patterns rather than constructions.

### 2.3.2 Verb Patterns

Verb patterns consist of a verb followed by other words. These words could range from a wide range of groups such as a noun group, an adjective group, a prepositional phrase, an adverb group, or a finite or non-finite clause (Francis, Hunston & Manning, 1996).

**Table 2.2** below shows a list of the verb patterns, accompanied by examples. The complementation of verbs is spread over 12 chapters (see *Collins Dictionary for a comprehensive list*), encompassing a detailed description of all possible English verb patterns. Broadly speaking, the patterns can be observed in accordance with the following categories: simple patterns, simple patterns with prepositions and adverbs, complex patterns, complex patterns with prepositions and adverbs, and it-clauses. As noted, all the pattern starts with ‘V’, representing that the pattern is headed by a verb. Other worthwhile abbreviations include ‘to-inf’ (to-infinitive clause), ‘pron-refl’ (reflexive pronoun), ‘that’ (that clause), and ‘pl-n’ (plural noun). The words in italics represent the actual words that appear in the pattern. Taking the pattern ‘V *about* n,’ for example, denotes a pattern that is headed by a verb, followed by the preposition ‘about’ and a noun. The sentence ‘Don’t bother *about* cleaning up’ is an example of this pattern.

It should be noted that not all verb patterns start with a verb. An example showing this is the pattern ‘pl-n V *together*,’ where the pattern includes the word preceding the verb ‘pl-n,’ followed by the word *together*. The sentence ‘The cells *clump* together’ is an example of this pattern.

Patterns	Examples
<b>SIMPLE PATTERNS</b>	
V	The meeting has <i>ended</i> .
V n	I <i>broke</i> my left leg.
V pl-n	The research <i>compares</i> two drugs.
V pron-refl (reflexive pronoun)	I <i>enjoyed</i> myself.
V amount	Two and two <i>make</i> four.
V adj	He <i>escaped</i> unhurt.
V-ing	She <i>started</i> walking.
V to-inf	John <i>began</i> to laugh.
V inf (bare infinitive)	I <i>helped</i> save these animals.
V that	We <i>agreed</i> that she was not to be told.
V wh	A passer-by <i>inquired</i> why the television cameras were there.
V wh-to-inf (to-infinitive clause introduced by a wh-word)	I <i>have forgotten</i> what to say.
V with quote	'hello', he <i>said</i> .
V so/not	I <i>think</i> so.
V as if as though	You <i>look</i> as if you've seen a ghost.
V and v	I'll <i>go</i> and see him.
<b>SIMPLE PATTERNS WITH PREPOSITIONS AND ADVERBS</b>	
V prep/adv, V adv/prep	He <i>ran</i> across the road.
V adv	Sarah has fair skin that <i>burns</i> easily.
Pl-n V together	The cells <i>clump</i> together
V prep	She <i>chewed</i> on her pencil
V about n, V across n, V after n, V against n, V around/round n	Don't bother <i>about</i> cleaning up She <i>cut across</i> the grass I <i>circled around</i> the building.
V as adj	A large number of plants <i>qualify</i> as medicinal.
V as n	His wife <i>works</i> as a designer.
V as to wh	We <i>advise</i> as to whether the group has a good legal case or not.
V at n	The unemployment rate <i>peaked</i> at 11 per cent.
V between pl-n	I <i>liaise</i> between these groups.
V by amount	The overall number of jobs <i>decreased</i> by 1000.
V by -ing	She <i>began</i> by telling me what the exhibition was about.
V for n, V from n, V in n, V in favour of n, V into n, V like n, V of n, V off n, V on to n, V onto n, V out of n, V over n, V	She <i>could pass</i> for a man. The majority have argued <i>in favour of</i> waiting. The plane broke <i>into</i> pieces. All the components <i>can run</i> off battery power. <i>Don't fret</i> over things you can't change. She <i>sailed</i> through her exams.

*through* n, V *to* n, V  
*towards/toward* n, V  
*under* n, V *with* n

---

### COMPLEX PATTERNS

---

<b>V n n</b>	I <i>wrote</i> him a letter.
<b>V n adj</b>	The darkness <i>could drive</i> a man mad.
<b>V n -ing</b>	I <i>kept</i> her waiting.
<b>V n to-inf</b>	My advisers <i>counselled</i> me to do nothing.
<b>V n inf</b>	She <i>heard</i> the man laugh.
<b>V n that</b>	I <i>told</i> her that there had been an accident.
<b>V n wh</b>	He <i>showed</i> me where I should go.
<b>V n wh-to-inf</b>	I'll <i>show</i> you how to do it.
<b>V n with quote</b>	'We'll do it', she <i>promised</i> him.
<b>V n -ed (the past participle form of another verb)</b>	I <i>had</i> three wisdom teeth extracted.

---

### COMPLEX PATTERNS WITH PREPOSITION AND ADVERBS

---

<b>V n prep/adv, V n adv/prep</b>	Andrew <i>chained</i> the boat to the bridge. <i>Stir</i> the sugar in.
<b>V n with adv</b>	He <i>switched</i> the television on. He <i>switched</i> on the television.
<b>V pl-n with together</b>	He <i>lashed</i> her hands together.
<b>V way prep/adv</b>	She <i>ate</i> her way through a pound of chocolate.
<b>V n as adj</b>	We <i>accept</i> this premise as fundamental.
<b>V n about n, V n against n, V n at n, V n by n, V n for n, V n from n, V n in n, V n into n, V n of n, V n off n, V n on n, V n onto n, V n on to n, V n out of n, V n over n, V n to n, V n towards/toward n, V n with n</b>	I <i>warned</i> him about the danger. I <i>saw</i> the question as crucial. I <i>borrowed</i> the money from my father. He <i>couldn't get</i> any more information out of Ted. They contributed \$3 <i>toward</i> costs. He <i>has confused</i> fact with fiction.
<b>V n as to wh</b>	I <i>informed</i> him as to what his legal rights were.
<b>V n between/among pl-n</b>	I <i>divided</i> his money among his children.
<b>V n into -ing</b>	She <i>nagged</i> him into cutting his hair.

---

### IT CLAUSES

---

<b>It V clause</b>	It <i>doesn't matter</i> what you think.
<b>It V to n clause</b>	It <i>sounds</i> to me as if you don't want to help her.
<b>It V prep clause</b>	It <i>came</i> to light that the plane had not been insured.
<b>It V n clause</b>	It <i>struck</i> me that the story would make a good film.
<b>It V adj clause</b>	It <i>feels</i> good to have finished a piece of work.

<b>V it clause</b>	I <i>hate</i> it when she's away
<b>V it to n clause</b>	I <i>owe</i> it to my parents to work hard.
<b>V it as n/adj clause</b>	He <i>would take</i> it as an insult if I left. He <i>regards</i> it as significant that the Government is suggesting cuts.
<b>V it n clause</b>	They <i>felt</i> it in their duty to visit her in hospital.
<b>V it adj clause</b>	I <i>think</i> it best if you tell them the truth.
<b>It V</b>	It <i>snowed</i> all afternoon.
<b>It V adj</b>	It <i>was</i> very windy.
<b>It V adj prep/adv</b>	It's nice here.
<b>It V n</b>	It's <i>blowing</i> a gale.
<b>It V to n</b>	It <i>got</i> to the point where we couldn't bear to be in the same room as each other.
<b>It V prep/adv that</b>	It <i>says</i> here that they have live music.
<b>V it</b>	They <i>didn't</i> make it.
<b>V it prep/adv</b>	My family <i>hate</i> it in Southampton.

**Table 2.2** Patterns of verbs (Francis et al., 1996; Hunston & Francis, 2000)

### 2.3.3 Using Parts-of-Speech (POS) n-grams to Explore Patterns

N-grams, where *N* stands for a number of linguistic units, is a string of words that can be extracted using corpora (Capelle & Grabar, 2016). For instance, a 2-gram, also known as a bi-gram) is a sequence of two words such as “an apple” or “the bear.” N-grams have been particularly useful for both quantitative and qualitative analyses to inform overuse, underuse, and misused learner language when compared with other corpora. Furthermore, taking such an approach can enhance the representativeness and generalizability of the data. A wide range of studies has used n-grams to extract lexical bundles (e.g., Allen, 2009; Chen & Baker, 2010; Juknevičiennė, 2009; Ping, 2009), collocation (e.g., Groom, 2009; Granger & Bestgen, 2014), formulaic sequence (e.g., Götz & Schilk, 2011), and clausal sequences (e.g., De Cock, 2007).

### 2.4 Dynamic Perspective of L2 Development

From the point in which babies start to babble to when they are able to form concrete sentences, the developmental trajectory of the first language (L1) is predictable and can often be traced. With respect to L2, however, tracing development is much more complex due to its dynamic nature and the numerous interdependent variables (e.g., cognitive, social, and environmental factors) that contribute to learners' ability to require a second language (De Bot, Lowie & Verspoor, 2007). Because of the multifaceted nature of L2 acquisition, where there is

no ‘one size fits all’ way to explain the complicated phenomenon of L2 development, understanding learners’ whole trajectory of language learning based on a single factor would be inaccurate.

In addition to the unpredictable nature of L2 development, it takes a highly nonlinear position in that learners show signs of both progress and regress. Verspoor, Lowie, & Van Dijk (2008) conducted a case study with an advanced English learner where the researchers collected 18 academic writing samples over the course of three years. The student’s development was measured through vocabulary use and complexity. During the three-year period, the student demonstrated the ability to utilize a wider range of words and complex sentence constructions. An in-depth analysis of the student’s vocabulary (measured by word length, type-token ratio, and degree of word use in the Academic Word list) and sentences (measured by the length of noun phrases and the number of words per finite verb) showed a competitive relationship between the two indices where there appeared a tradeoff between more varied words and longer sentences. The analyses of the learner’s writing development over the course of a long period of time have revealed that although it is evident that the learner demonstrated development, such showed a nonlinear pattern where there were signs of both progress and regress.

The theory that encompasses the phenomenon of the complex system of language system development can be understood from the complex dynamic systems theory (Fogal & Verspoor, 2020). The main argument from the dynamic systems perspective is that a number of factors are constantly competing with one another, especially when the learners’ language system is unstable. Thus, fluctuation is normal, and in order to gain a full scope of the holistic development of when and how the language system changes, it is necessary to observe the different patterns and variations in extensive developmental data (Verspoor, Lowie, and Van Dijk, 2008). The current study is an attempt to capture the varying degrees of linguistic development by using large data. There are, of course, caveats such as not being able to capture the individual characteristics and behavior; however, such is very difficult to achieve with a large set of data. Instead, the study strives to understand language development through means of generalization.

### 2.4.1 Roles of Interlanguage and Its Impact on L2 Writing Development

It has been widely acknowledged that acquiring an L2 is not, and should not, be treated in the same way that L1 is acquired. Interlanguage, in addition to many other variables, contributes to L2 learners' successful language learning experience. It is debatable as to whether L2 development is similar to that of L1 development. For instance, Cancino, Rosansky, and Schumann (1978) conducted a longitudinal study among 6 L2 Spanish-speaking participants between the ages 5 and 33 to test whether or not L2 development was similar to that of early L1 development. Interestingly, they found that although L2 learners vary, both with respect to between and within participants, the general learning trajectory was similar to that of L1. More recently, Kim (2016) explored L1 Korean speakers' learning of L2 English resultative construction. Resultative constructions are different in English and Korean, both syntactically and semantically, and is an area that many L1 Korean speakers have little understanding of. The result from the acceptability judgment task revealed that L1 Korean speakers revealed a much lower score in both comprehension and production when compared to native speakers. The authors claimed that L1 Korean speakers' understanding of resultative constructions was influenced by their L1. A piece of supporting evidence to Kim (2016) could be that “*reconstructing* a language is more complex than its initial induction because, during development, L2 constructions are in direct competition with those of the learners' L1, and these may represent alternative ways of construing the same reality” (Ellis & Cadierno, 2009, p. 112). The cross-linguistic influence may, in fact, hinder learners' ability to become proficient.

Supporting Kim (2016), there is much evidence that L1 can influence L2 learning. For instance, Aarts and Granger (1998) used tag sequences in learner corpora to lend support to interlanguage. More specifically, they found that when compared to native writers, learners overused patterns that started with a connective while underusing patterns with prepositions. The authors also demonstrated that these overused and underused patterns were specific to learners' L1. There is a solidified understanding that “language processing is rational in that it is exquisitely sensitive to prior usage” (Ellis, 2008, p. 373).

Owing to the empirical research involving interlanguage and the role of L1 in L2 learning, we are now aware that L2 writing must be particularly challenging. For instance, L2 writers may lack awareness of the register difference between speech and writing prose (Gilquin & Paquot, 2008). According to Flowerdew (2001), there are three areas that are particularly

difficult for L2 learners in academic writing: collocational patterning, pragmatic appropriacy, and discourse features.

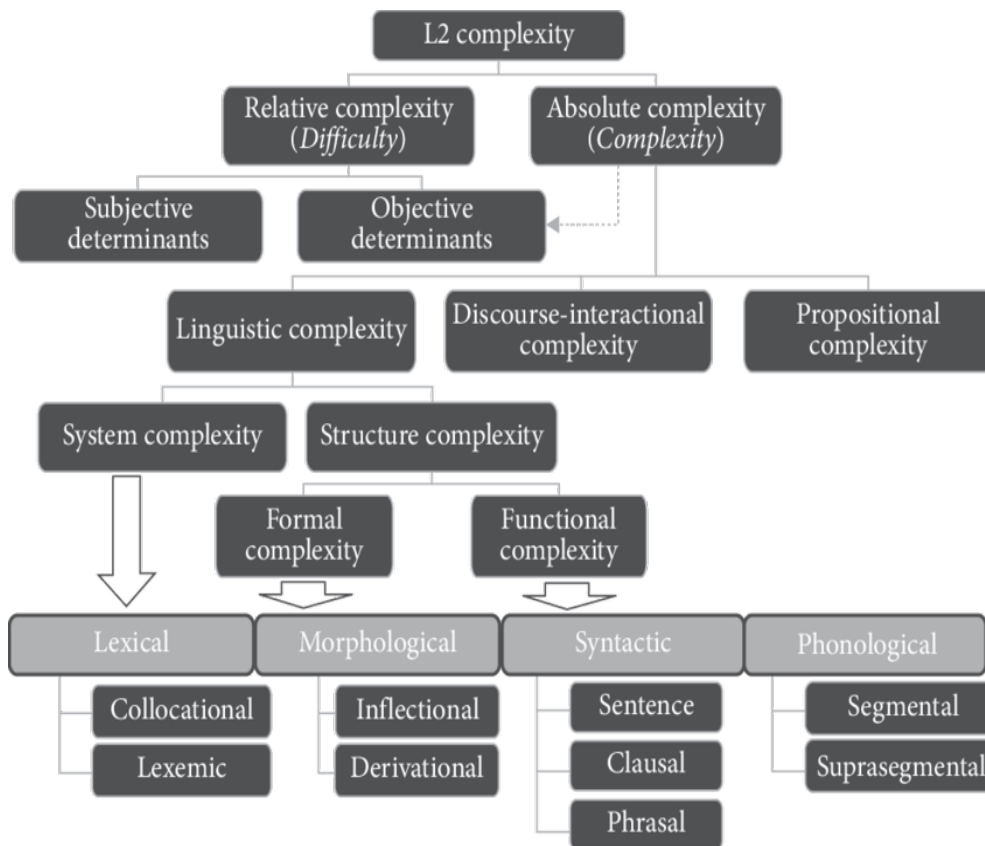
Although it is clear that cross-linguistic influence is evident, much research has looked at relatively small areas to manifest L1 influence on L2 learning. In particular, studies mentioned above have shown that certain constructions are much more complex to attain than others or that they are either overused or underused based on the language users' L1. However, what is missing in the literature is an exploration of how these patterns are overused or underused over a period. Doing so will elicit a further understanding of what the learners *can* produce at different stages of the developmental trajectory.

#### **2.4.2 Empirical Studies on Measurements of L2 Writing Development**

The development of L2 writing can be measured in numerous ways. Most notably, complexity, accuracy, and fluency (CAF) measures (Wolf-Quintero, Inagaki, & Kim, 1998) have been one of the most commonly used measures of L2 writing development across many studies (e.g., Beers & Nagy, 2011; Becker, 2010; Hunt, 1970; Grants & Ginther, 2000; Otto, 2006; Verspoor, Schmid, & Xu, 2012; Verspoor & Scmiskova, 2012). These indices have gained much merit in L2 writing development. While CAF measures are not within the scope of this study, it might be worth looking at the different ways in which development has been and can be measured.

In linguistics and second language studies, the term complexity has received much attention (Dahl, 2004; Pallotti, 2015), which has ranged from as broad as second language acquisition (Housen & Kuiken, 2009) to as confined as second language writing (Ortega 2003; Polio 2001; Wolfe-Quintero, Inagaki, & Kim 1998). Complexity, in linguistics, is associated with the capability of using more sophisticated language (Ellis 2009), indicating its close relationship with L2 development; as learners become more proficient in the language, they are expected to be able to use more complex structures of the language. Thus, complexity has been used widely as an indicator of both a learner's proficiency and language development (Bulté & Housen, 2012). In particular, research has shown that with the increment of age in children, they tend to use more syntactically complex structures in both their speech and writing (Otto, 2006). Similarly, as they get older, their attempt to express more complex ideas also increases partly due to more exposure (Fang, Schleppegrell, & Cox, 2006).

L2 complexity is a complex construct (Norris & Ortega 2009) and is defined differently across various studies. Bulté & Housen (2012) has proposed a taxonomic model of L2 complexity, demonstrating the multidimensional nature of L2 complexity (**Figure 2.1**). This taxonomy examines L2 complexity in a wide range of dimensions, which also provides a comprehensive picture of L2 writing development in terms of complexity. As shown in **Figure 2.1**, L2 complexity is broken down into multiple facets. Conclusively, Bulté & Housen (2012) divided L2 complexity with respect to lexical, morphological, syntactic, and phonological complexity. Lexical complexity entails the complexity of words in relation to lexeme and collocation. Morphology and syntactic complexity both deal with the grammatical structure, seeking to observe complexity at the sentence, clausal, and phrasal level on the use of derivational and inflectional morpheme. Phonological complexity is relevant to speaking since it observes complexity in relation to segmental and suprasegmental units.



**Figure 2.1.** Taxonomic model of L2 complexity (Bulté & Housen, 2012)

Grammatical complexity in writing is commonly defined in terms of using more elaborate language and a variety of syntactic patterning (Foster & Skehan, 1996). It demonstrates a learner's ability to compose with more significant grammatical variations and sophistication (Wolfe-Quintero, Inagaki, & Kim, 1998), making it one of the most commonly used measures of L2 writing development. In a similar vein, a more proficient writer is able to recall a variety of grammatical structures and sophisticated sentences, clauses, and phrases at their disposal. On the contrary, a less proficient writer would only be able to make use of a limited variety of such structures.

Previous studies have used grammatical complexity as an index of both language development and progress (e.g., Bulté & Housen, 2014; Crossley & McNamara, 2014; Ortega, 2003) and have extensively used ratio- and mean-based measures, including mean length of T-unit, mean # of clause per T-unit (see Wolfe-Quintero, Inagaki, & Kim, 1998). However, these units are more suitable for capturing characteristics of conversation rather than writing (Biber, Gray, & Poonpon, 2011). Thus more recently, a more corpus-based approach has been taking into account observing register variations (e.g., Biber, 1992; Biber & Gray, 2010; Biber, Gray & Poonpon, 2011) by operationalizing grammatical complexity based on the standardized rate in which certain grammatical structure occurs (e.g., verbs, adjectives per 1,000 words). In this study, the term "complexity" is conceptualized with respect to the usage-based approaches. In other words, language use is considered complex depending on the frequencies and idiomaticity of the words and phrases. Comparing measures of complexity according to the CAF measures is not within the scope of this study. Instead, the current study uses an innovative approach to identifying the process of language development from a bottom-up approach. Rather than using specific indices to calculate complexity, it uses frequency and pattern grammar as a way to show *how* language changes and evolves, an approach that cannot be measured by the existing complexity indices.

## **2.5 Learner Corpora and L2 Research**

The use of corpora in research has gained merit over the past few decades for its authenticity in analyzing actual language patterns as they appear in natural contexts, as well as the ability to achieve great generalizability from a large set of data (Biber, 2009). In particular,

corpus-based studies have been of much value in L2 studies as an ideal approach to usage-based analyses (Kemmer & Barlow, 2000).

Before delving into how corpora have been used in L2 developmental studies, it would be appropriate to first define the term. In a broader sense, a corpus (plural “corpora”) is a “body of text which is carefully sampled to be maximally representative of a language or language variety” (McEnery & Wilson, 1996, p. 87). Each corpus is unique in that it is constructed based on strict linguistic criteria that have been designed to fulfill the purpose of the research (Atkins, Clear, & Ostler, 1992; Sinclair, 1996). The data extracted from a corpus can be observed in a wide range of ways, such as compiling a frequency list and/or looking at the concordance lines. A frequency list is a ranking of words, phrases, patterns, or constructions which then can be used to show 1) the most and least frequent use and 2) their distribution across different registers (Carter & McCarthy, 2004). Concordance lines, on the other hand, are lines of texts taken from the corpus which contains the target word(s). Below is an example of a set of 10 concordance lines for the word ‘happiness’ from the British National Corpus (BNC).

1	<input type="checkbox"/>	ⓘ	Written miscell...	perience: on freedom, justice, relationships, morality, suffering and	<b>happiness</b>	' (6). </s><s> In this regard, it is recognised that while a faith backg
2	<input type="checkbox"/>	ⓘ	Spoken context-...	do you think of Romeo taking Juliet away from you? </s><s> Well	<b>happiness</b>	then what I can give her. </s><s> Okay, thank you very much. </s>
3	<input type="checkbox"/>	ⓘ	Written books a...	geois family could maintain the illusion of a harmonious, hierarchic	<b>happiness</b>	, surrounded by the material artefacts which demonstrated it and r
4	<input type="checkbox"/>	ⓘ	Written books a...	</s><s> shift their weight and softly kiss, </s><s> burning with their	<b>happiness</b>	... </s><s> Our hosts go back along the jetty, </s><s> with their stic
5	<input type="checkbox"/>	ⓘ	Written books a...	me well, though neither of us knew </s><s> The way to the other's	<b>happiness</b>	</s><s> Surely now, </s><s> Sophisticated in another's life, she c
6	<input type="checkbox"/>	ⓘ	Spoken context-...	she excelled herself and managed magnificently when the family's	<b>happiness</b>	depended on it. </s><s> I went blind in me right eye overnight. </s>
7	<input type="checkbox"/>	ⓘ	Spoken context-...	more to say except I hope that their year of office will be filled with	<b>happiness</b>	and wish them well with all their endeavours on behalf of Council. <
8	<input type="checkbox"/>	ⓘ	Written books a...	ity had to prevail, Himmelfarb noted." The principle of the greatest	<b>happiness</b>	was as inimical to the idea of liberty as it was to the idea of rights."
9	<input type="checkbox"/>	ⓘ	Written books a...	ive: it is not to champion the interests of the majority - the greatest	<b>happiness</b>	of the greatest number - that this new technology is being summon
10	<input type="checkbox"/>	ⓘ	Written books a...	hat was his duty, after all. </s><s> But her heart and emotions and	<b>happiness</b>	held little interest for him. </s><s> After her mother's death, Sara c

**Figure 2.2** Concordance lines for the lemma ‘happiness’ in the BNC corpus

As shown in **Figure 2.2**, in the center of each of the ten sentences in red is the word ‘happiness’, the target word that is being investigated. Furthermore, surrounding ‘happiness’ are words that come before and after the target word. Concordance lines are particularly useful when looking at different usage of the target word as well as identifying idiomatic and formulaic language patterns. Owing to the systematicity and the robustness of corpora, where results are based on actual figures as manifested in the data, it has gained much attention in second language studies. For instance, Ellis and Cadierno (2009) claim that “the realization of the primacy of language usage to language acquisition necessitates a commitment to corpus linguist

methods and to study the contextualized functional discourse” (p. 112). Thus, it is evident that corpus-based methods are of great value in language research.

The corpus linguistics approach has been used extensively in a variety of language studies ranging from data-driven learning (see Boulton & Cobb, 2017) to second language vocabulary learning (e.g., Daskalovska, 2014; Lin & Liou, 2009; Sun & Wang, 2003; Yunxia, Min & Zhuo, 2009), second language writing (e.g., Babanoglu, 2014; Bolton, Nelson & Hung, 2002; Field & Yip, 1992; Lu, 2011), and second language speaking (e.g., Xu & Xu, 2007). However, perhaps the most valuable finding from research in corpus linguistics in the last 40 years is that language is patterned (Römer, 2009). This has been manifested through extensive research on collocation, formulaic sequences, and phraseology. The following section explores learner corpora as well as the ways in which empirical research has shed further knowledge on L2 development.

### 2.5.1 Use of Learner Corpora in L2 Developmental Studies

Learner corpus research is a branch of corpus linguistics that began to emerge in the late 1980s and early 1990s (Granger, 2008). What sets learner corpora apart from other types of existing corpus types is that the data contains language produced by L2 learners (Gilquin, Granger & Paquot, 2007). Learner corpora have made a tremendous contribution to language research, particularly regarding the ways in which learner language changes and evolves as learners become more proficient in the L2.

**Table 2.3** illustrates the major learner corpora to date.

Corpus	First language	Text type	Size (words)
Arab Learner English Corpus (ALEC)	Arabic	Essays written by freshman students as part of first level college writing course	Analysis: 184,749 Narrative: 67,527 Synthesis: 66,015 Argumentation: 192,298
The Advanced Learner English Corpus (ALEC)	Mainly Swedish	Essays written by university students of English	1.3 million

		linguistics and English literature	
The BATMAT Corpus	Swedish Finnish	BA/MA dissertations	3 million
The Bilingual Corpus of Chinese English Learners (BICCEL)	Chinese	In-class assignments	2 million
The British Academic Written English Corpus (BAWE)	Mainly L1 speakers but may also include data produced by L2 speakers	ESP papers	6.5 million
The Cambridge Learner Corpus (CLC)	Various	Exam scripts	50 million
The Chinese Learner English Corpus (CLEC)	Chinese	N/A	1 million
Corpus and Repository of Writing (CROW)	24 languages (mainly Chinese and Arabic)	Analysis, narrative, literature review, argument, empathy writing, proposal, reflection	9 million
The Gachon Learner Corpus	Korean (+ a few Chinese & Spanish speakers)	Written journal assignments	2.5 million
The Hong Kong University of Science & Technology (HKUST) learner corpus	Chinese-mostly Cantonese	Untimed assignments written for EFL courses and school leaving exams	25 million
The International Corpus of Crosslinguistic Interlanguage (ICCI)	Various	Essays (20-min in-class tasks)	9,000 essays
The International Corpus of Learner English (ICLE)	Various	Argumentative and literary essays	3 million
The Longman Learners' Corpus	Various	Essays and exam scripts	10 million
The Michigan Corpus of Upper-level Student Papers (MICUSP)	Mix of both native and non-	ESP papers; A-grade papers;	2.6 million

	native speakers of English	undergrad papers that have been assessed and accepted (e.g., research proposals) but not published	
The NUS Corpus of Learner English (NUCLE)	Several East Asian languages, predominantly Chinese	Student essays on a wide range of topics (e.g., pollution, healthcare)	1 million
The Scientext English Learner Corpus	French	Academic argumentative texts	1.1 million
The Seoul National University Korean-speaking English Learner Corpus (SKELC)	Korean	Student essays	900,000
The SILS Learner Corpus of English	Various (mainly Japanese)	Student essays	3.2 million
The Taiwanese Corpus of Learner English (TLCE)	Chinese	Journals and essays (descriptive, narrative, expository, argumentative)	2 million
The Ten-Thousand English Compositions of Chinese Learners (TECCL)	Chinese	Essays (various topics). Includes collaborative writing samples	1.8 million
The Uppsala Word Reference Corpus	Various	Forum posts	38 million
The Yonsei English Learner Corpus (YELC)	Korean	Yonsei University English Diagnostic Tests (Part 1: Descriptive task, max. 100 words; Part 2: Argumentative test, max. 300 words)	1 million

**Table 2.3** List of major learner corpora

Although each of these corpora was constructed to fulfill specific research purposes, there are some noticeable caveats that make them either unsuitable or less ideal for this study: 1) Many of these corpora are restricted in terms of L1 (e.g., Arabic, Swedish). 2) There is bias in text types. Some are collected in a single setting (e.g., diagnostic tests for university entrance) and lack variety in registers (e.g., in-class assignments, journal assignments). Among these, the most concerning issue has to do with the lack of information provided about the learners' proficiency level. Some of them have adopted generic terms such as "advanced" or "intermediate," whereas others have used an in-house rubric that does not provide a full scope of the learner's proficiency level. The Cambridge Learner Corpus (CLC) and The Yeonsei English Learner Corpus (YELC) were two exceptions where both corpora used the Common European Framework of Reference for Languages (CEFR) to benchmark proficiency levels. Although YELC was large in size (~1 million words), it was considered unsuitable due to its limited registers of descriptive and argumentative essays. On the other hand, the CLC is both unique and robust in that it includes learners from over 100 different countries, not discriminating against specific regions or socioeconomic backgrounds. No other corpus known to date meets these criteria.

Criteria in choosing the two sub-corpora for this study are as follows:

1. One of the sub-corpora must include L1 Korean speakers and/or at least five different L1 speakers.
2. The corpus must contain more than 500,000 words.
3. Must consist of a wide range of text types and registers.
4. Must consist of a detailed description of proficiency levels (e.g., cannot use age or grade level as proxy).

Thus, the CLC in this study addresses many of the existing gaps in learner corpus research, such as vague use of the term with regards to proficiency levels, skewness of the data to a single cohort of the group, and small size.

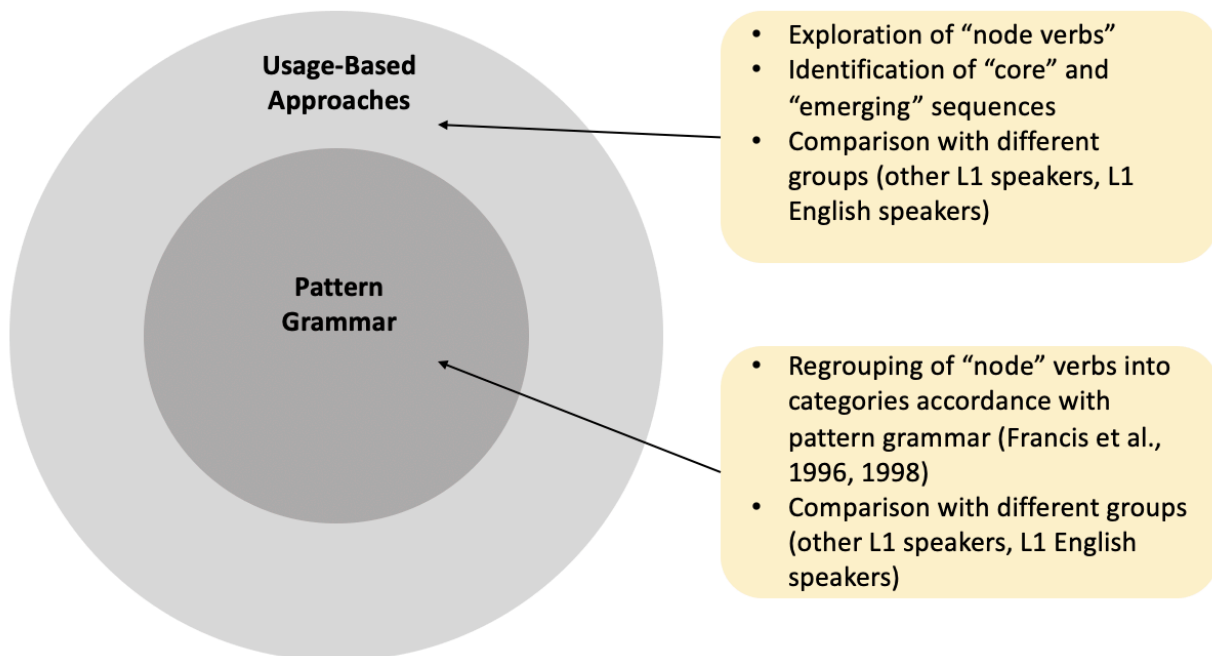
### **2.5.2 Contrastive Interlanguage Analysis in Learner Corpora**

As illustrated thus far, corpus-based approaches have shed much light on how we understand language use. This is particularly the case in learner corpus research (LCR), which

has allowed a systematic and in-depth investigation into how L2 language users become proficient both at a certain time in their developmental trajectory as well as throughout their history of the language learning journey. In addition to the wide range of approaches and methodologies that have shed insights deriving from a single corpus, there are also ways to examine interlanguage between different groups. This approach, also known as Contrastive Interlanguage Analysis (CIA), is a term that was coined by Granger (1996). In this methodology, linguistic features from a learner corpus can be compared with 1) native speakers or 2) another learner corpus. The comparison between the L1 vs. L2 comparison has made a significant contribution to understanding patterns such as misuse, underuse, and overuse of various linguistic features both at the lexical level as well as the phrasal level (Granger, 2015). In the case of L2 vs. L2 comparison, it makes it possible to identify L1 features and interlanguage differences between learners (e.g., Ädel, 2008). While the traditional view of the CIA had received a lot of criticism, Granger (2015) proposed a new version of the methodology where the ‘varieties’ in language and interlanguage were taken into consideration. The most noticeable difference, thus, is that “there are a large number of different reference points again which learner data can be set” (Granger, 2015, p. 17) and that these different points can be in reference or learner varieties.

This study also takes on the methodologies claimed by the CIA, where it seeks to explore interlanguage differences in two areas. First, the difference is observed at different proficiency levels as a way to trace how language develops over time. As a quasi-longitudinal study, the study refers to the six proficiency levels in accordance with the Common European Framework of Reference for Languages (CEFR), and comparisons are made across L1 Korean speakers. Next, the interlanguage differences are observed with respect to two different learner groups: other L1 speakers and L1 English speakers. The comparison with other L1 speakers is to identify patterns of verb 4-gram sequences that are specific to L1 Korean speakers. Finally, the comparison with L1 English speakers is a way to probe into the extent to which the language used among L1 Korean speakers mirrors that of L1 English speakers.

## 2.6 Conceptual Framework of the Study



**Figure 2.3** Framework of the study

In consideration of the theories discussed in this chapter, a unique conceptual framework has been formulated for this study. This section revisits the core theories and discusses the key aims the study seeks to achieve.

First of all, cognitive linguistics posits that language, communication, and cognition are inextricable (Ellis & Robinson, 2008). Language is used to facilitate meaningful communication between interlocutors, and it is through these interactions we can expand and build on our existing knowledge. Then how exactly is language learned through communication? Usage-based approaches to language use look at *how* language is shaped by social interactions (Ellis, Römer, & O'Donnell, 2016) through the exploration of the specific communicative events in which language is used and learned (Tomasello, 2000a). Because language and cognition are co-created by each other through our unique experiences, there is significant variation between individuals; language is essentially a dynamic system that accumulates over the course of our lifetime (Bybee & McClelland, 2005). This is not to say, however, that language acquisition is unpredictable. Despite its dynamic process, language exhibits structures and patterns that can be identified as it continues to evolve and change over time (Bybee, 2010). These patterns can be understood from the theories of construction grammar (Goldberg, 1996; 2006), which claim that

language can be conceptualized through constructions and form-meaning mappings that become more entrenched through use. It is through our experience and memories of exemplars that schematic constructions emerge even without prior exposure (Bybee, 2010; Goldberg, 2006). While it may be true that learners infer and ultimately learn ‘constructions,’ one of the caveats of construction grammar is that the current literature does not offer an exhaustive description of potential constructions (Hunston, 2019). Alternatively, Hunston (2019) offers a different way of interpreting ‘constructions’ through the lenses of pattern grammar (Francis, 1993; Hunston & Francis, 2000).

Pinned by the theories of cognitive linguistics and corpus linguistics, the current study is conceptualized around the framework facilitated by usage-based theories and pattern grammar. Within the realm of these two theories, the study attempts to provide a bottom-up, data-driven understanding of 1) the process of L2 writing development through means of adjusting verb 4-gram sequences and 2) what the emerging patterns are based on the identified core and emerging sequences. Below is a detailed explanation of how the two theories are used in this study.

Usage-based theories hold that linguistic competence is shaped by meaningful interactions held throughout learners’ history of language usage (Bybee, 2010; Ellis, O’Donnell, & Römer, 2013; Robinson & Ellis, 2008; Tomasello, 2003). Language is experience-driven, and therefore, the frequency of meaningful communication plays a crucial role in learners’ language production (Barlow & Kemmer, 2000; Tomasello, 2000a). Additionally, it is claimed that grammar is a “product of usage” (Langacker, 1987, 2000; Perek, 2015), and this provides an important understanding of how the linguistic system continues to restructure itself as complex and abstract constructions emerge in the learners’ repository of linguistic knowledge.

Pattern grammar (Francis, 1993; Hunston & Francis, 2000) was originally derived for the purpose of annotating lexical items in the *Collins COBUILD English Dictionary* (Sinclair, 1995). Similar to the theories of construction grammar, pattern grammar also attempts to generalize the grammatical behavior of individual words. It is also underpinned by the idea that words are conceptualized beyond words rather than in lexical items. Currently, research supports that meaning-pattern combinations (e.g., grammar patterns) can also be identified as construction (Hunston & Su, 2019). While constructions and patterns may seem similar in nature, the main difference between them, however, is that construction grammar is derived from the theories of cognitive linguistics while pattern grammar arose from corpus linguistics. Thus, the focus of

construction grammar is on the mental representation of the form-meaning mappings, whereas pattern grammar is grounded in observations that can be made based on evidence from corpora.

At the heart of this study are the theories of usage-based approaches to language acquisition and pattern grammar. Through these theories, this study seeks to trace the development of L2 writing among L1 Korean speakers using a quasi-longitudinal corpus. The study is divided into two parts. The first part broadly explores the usage of verb-fronted sequences through the exploration of the normalized frequency of the verb 4-gram sequences. From the ranked sequences, two types of sequences (“core” and “emerging” sequences) emerge. Through the scrutinization of the frequency of the verb 4-gram sequence in each proficiency level and how they change across the six proficiency levels, the study seeks to trace and understand how the usage of verb sequences evolves with the development of learners’ language competency.

In the second part, a more thorough investigation of the core and emerging verb 4-gram sequences are conducted to tap further into the sequence by categorizing them in accordance with pattern grammar (Francis, 1993; Hunston & Francis, 1999; Hunston, 2015). The regrouping of the node verbs will help to formulate a conceptualization of the hierarchy of the constructions that emerge with proficiency. Once the grammar patterns are identified, the data is then compared with reference to other L1 speakers and L1 English speakers as a way to explore potential L1 factors.

In summary, **Figure 2.3** illustrates a visual representation of the framework of this study. As can be seen, this study is conceptualized around the framework of these two theories to both trace and show L2 writing development in L1 Korean speakers using a bottom-up approach. By combining these two approaches, this study will shed a new understanding of the L2 writing development trajectory among L1 Korean speakers.

## **2.7 Summary**

This section discussed in-depth the theories that shape the conceptual framework of this study. Simply put, the current study takes an innovative approach to looking at L2 writing development. Firstly, it takes on the view that language is usage-based and that it is shaped through the user’s unique experience and meaningful interaction with the language. Thus, frequency is at the heart of usage-based approaches to language learning in that the most

frequent and simple words tend to be learned prior to more complex words. In a similar vein, as learners become more proficient in the language, they are able to construct complex and abstract sentences. This can be explained by the ways in which language continues to restructure and evolve in the user's cognitive system as they gain more experience with it.

Using usage-based approaches as an underpinning of the current study, it seeks to identify the patterns that are associated with the frequency of the “verb 4-gram” sequences that are identified from the learner data. That is, deriving from the understanding that the meaning of lexical items is embedded in the grammar, words alone cannot entail much about the learner's language ability. Thus, the study uses pattern grammar to identify unique L2 writing developmental patterns among L1 Korean speakers.

## **Chapter 3 Methodology**

### **3.1 Introduction**

This study aimed to explore L2 writing development among L1 Korean speakers with respect to their use of English verb 4-gram sequences. This research was a quasi-longitudinal study, also known as pseudo-longitudinal (Johnson & Johnson, 1999), where the CEFR level was used as a proxy for language development. By doing so, this study hoped to shed light on the underlying processes of L2 writing development that L2 learners undergo as they become more proficient in their L2 writing.

In order to fulfill the aims of this study, the research was broadly divided into four parts. In the first part, the frequencies of the verb 4-gram sequences were compared across the six CEFR proficiency levels (A1 to C2) to identify 1) sequences that occur most frequently across all proficiency levels (henceforth “core” sequences) and 2) sequences that occur less frequently in lower levels but occur more frequently as learners become more proficient in the L2 (henceforth “emerging” sequences). Then, the “node” verbs, or main verbs, from each of the core and emerging sequences were compared across the proficiency levels. Next, an in-depth analysis was conducted by categorizing the selected examples from the core and emerging sequences according to pattern grammar. Finally, the identified patterns were compared across other speakers of L1 and L1 English speakers to identify possible display of interlanguage and their prevalence across various proficiency levels. Furthermore, it sought to identify the extent to which L1 Korean speakers at the highest proficiency level deviate from or converge with L1 English speakers with respect to their use of verb 4-gram sequences.

The chapter begins with the research questions and then discusses the methodology used for the four parts mentioned above.

### **3.2 Research Questions**

This section provides a brief explanation of each of the research questions in this study.

The first research question explored the frequency of distribution of the verb 4-gram sequences in each proficiency level in order to identify sequences that are 1) constantly ranked high across the various proficiency levels and 2) ranked below the top 10 in lower proficiency

levels but surfaced to within top 10 ranking in the highest level. The first set of research questions was designed as a primary step in identifying the patterns of language use.:

1. What are the most frequent verb 4-gram sequences from each proficiency level?
  - 1-1. What are the “core” verb 4-gram sequences?
  - 1-2. What are the “emerging” verb 4-gram sequences?

The second and third research questions were an in-depth exploration of the core and emerging verb 4-gram sequences from the first research question. The second research question explored the “node” verbs from the core and emerging verb sequences. Through this, it sought to illustrate that meaning of a word does not exist in the lexeme alone but in chunks. It was also designed to explore whether advanced language users use phrases that are more idiomatic in nature. The third research question re-categorized the sequences identified in the first research question according to grammar patterns. By doing so, it hoped to identify usage patterns among L1 Korean speakers in a more systematic way.

2. What are the most frequently used “node” verbs from each verb 4-gram sequences?
3. From sequence to pattern: What grammar patterns emerge from the selected core and emerging sequences?

The fourth and fifth research questions explored two areas. Firstly, it looked at the L1 impact on L2 learning and its prevalence in different L2 proficiency levels. Furthermore, the differences in the use of verb 4-gram sequences in L1 English speakers and L1 Korean speakers at the near-native proficiency (C2 level) were explored.

4. How do verb 4-gram sequences differ between L1 Korean speakers and other L1 speaker?
5. How do verb 4-gram sequence differ between L1 Korean speakers and L1 English speakers?

### **3.3 Methodology for the Investigation of Sub-Corpora**

This section describes the methods of the study. It begins with the design of the study, followed by some ethical considerations. It then provides an in-depth description of the Cambridge Learner Corpus (CLC), the British National Corpus (BNC), and the Collins WordBanks Online Corpus, as well as the two sub-corpora drawn from the CLC. Finally, the

section concludes with a discussion of the Collins English Dictionary, the source which was consulted for the categorization of the examples from the sequences into pattern grammar.

### 3.3.1 Study Design

This study was designed to fulfill the research gaps stated in Chapter 1 (see 1.1). To reiterate, the purpose of this study was to explore how L1 Korean speakers' use of verb 4-gram sequences changes across different proficiency levels. By doing so, it sought to understand what learners are capable of producing at different learning trajectories. Thus, a bottom-up approach design is implemented to illustrate how the use of verb 4-gram sequences evolves and changes as L1 Korean speakers become more proficient in the L2.

This study also took on the methodologies claimed by the Contrastive Interlanguage Analysis (CIA) (see 2.5.2 for an in-depth discussion), where it aimed to explore interlanguage differences in two areas. First, the difference was observed at different proficiency levels as a way to trace how language develops over time. As a quasi-longitudinal study, the study referred to the six proficiency levels in accordance with the CEFR, and comparisons were made across L1 Korean speakers. Next, the interlanguage differences were observed with respect to two different learner groups: other L1 speakers and L1 English speakers. The comparison with other L1 speakers intended to identify patterns of verb 4-gram sequences that are specific to L1 Korean speakers. Finally, the comparison with L1 English speakers was a way to probe into the extent to which the language used among L1 Korean speakers mirrors that of L1 English speakers.

The observation was made from several angles based on two learner corpora and a reference corpus. In order to select the most appropriate corpora for the study, the following points were taken into consideration:

- 1) The research context for this study should broadly cover a wide representation of L1 Korean speakers. In other words, the learner corpus should not be skewed with respect to any particular demographics such as age, gender, education level, and socioeconomic background. Thus, the only discriminating factor should be learners' proficiency level, which is used to benchmark the developmental trajectory in this study.

- 2) While the findings from L1 Korean speakers alone could provide a wealth of insight into what learners are capable of producing at different proficiency levels, a comparison with other L1 speakers would provide a richer understanding as to whether the characteristics of the frequently used verb 4-gram sequences are specific to L1 Korean speakers or are prevalent across all L2 learners.
- 3) In addition to the comparison with other L1 speakers, a comparison with L1 English speakers would illustrate to what extent L1 Korean speakers' use of verb 4-gram sequences deviate from that of English speakers in the context of academic writing. Thus, the reference should mirror the learning corpora adapted for this study.

Taking these points into consideration, this study drew from the following three corpora: the Cambridge Learner Corpus (henceforth, CLC), the British National Corpus (henceforth, BNC), and the WordBanks Online Corpus. Two sub-sections from the CLC were drawn. Firstly, the L1 Korean speaker sub-corpus (see 3.3.4.2) consisted of all learners that have indicated Korean as their L1, regardless of ethnicity. Next, the other L1 speaker sub-corpus (see 3.3.4.3) was composed of all other speakers of L1 in the CLC. In other words, this portion of the sub-corpus consisted of the rest of the CLC, excluding L1 Korean speakers. From the BNC (see 3.3.4.4), the academic part of the corpus was drawn. While finding a reference corpus can be challenging, the BNC was chosen for its immensity, thereby considering to represent language use as much as possible. A more detailed rationale for using the BNC will be provided in the following section. Lastly, the WordBanks Online Corpus was consulted for the node verbs analysis. The node verbs from the L1 Korean speaker sub-corpus were ranked according to the frequency in the WordBanks Online Corpus. More information about this corpus will be provided in a later section.

### **3.3.2 Ethical Considerations**

A few crucial ethical considerations were taken into account in this study. First, all exam scripts in the CLC are labeled with a unique document ID. To maintain the anonymity of the data, the document IDs were disclosed in the current study. Any concordance lines taken from the study were labeled with the CEFR level attained, learner's L1, the exam taken from, and the year of the study. Second, where appropriate, any names in the script were replaced with a

pseudonym. Lastly, all collected data were used for the sole purpose of this study. All the information from the corpora remained confidential.

### 3.3.3 Instruments

To reiterate, the current study is carried out using three different corpora. Firstly, the CLC was consulted to draw the two sub-corpora used for this study (L1 Korean speakers and other L1 speakers). To briefly explain, the L1 Korean speaker sub-corpus was used to 1) draft the verb 4-gram sequences and 2) draft the most commonly used node verbs from each of the selected verb 4-gram sequences. The other L1 speaker sub-corpus was also used to draft the verb 4-gram sequences, and the extracted data was used to compare with that of the L1 Korean speakers. Then, the academic portion of the BNC was used in order to understand how the use of verb 4-gram sequences compared between L1 Korean speakers and L1 English speakers. Next, the Collins WordBanks Online corpus was consulted to compare the frequency ranking of the node verbs from the selected sequences in the L1 Korean sub-corpus. Finally, the Collins English Dictionary was used to categorize the examples from the verb 4-gram sequences according to pattern grammar. The use of these tools is discussed briefly to illustrate their purpose, and further description will be provided in the procedures.

The examples from the sequences were then categorized according to the grammar patterns from the Collins English Dictionary. **Table 3.1** illustrates the size of the three main corpora used in this study, as well as the three sub-corpora taken from the CLC and the BNC.

Name of the corpus	Size	Content
Cambridge Learner Corpus (CLC)	55.5 million words	Written exam scripts from English learners
L1 Korean speaker sub-corpus taken from the CLC	914,363 words	Written exam scripts from English learners
Other L1 speaker sub-corpus taken from the CLC	53.8 million words	Written exam scripts from English learners
British National Corpus (BNC)	100 million words	Written and spoken modern English by native speakers and advanced speakers of English
Written Academic sub-corpus taken from the BNC	84 million words	Written samples taken from a wide range of sources
Collins WordBanks Online	550 million words	Written and spoken English

**Table 3.1** Size of the corpora used in this study

The following sections discuss these three types of corpora as well as other instruments used in this study.

### **3.3.3.1 Cambridge Learner Corpus (CLC)**

The CLC is a 55.5-million-word corpus of English learners consisting of over 260,000 open-ended written exam scripts collected between 1993 and 2012. The corpus contains learners from over 200 countries with around 140 different first languages. It is one of the largest learner corpora created to date. The CLC is part of the Cambridge English Corpus, which is composed of two corpora: L1 English and the CLC. The L1 English speaker corpus contains both written and spoken materials that were provided by British and American L1 English speakers. CLC, on the other hand, consists of learner exam scripts and is limited to written samples. The corpus is currently under the co-ownership of Cambridge University Press and Cambridge Assessment and is used for a variety of purposes, including both in-house research and course and materials development.

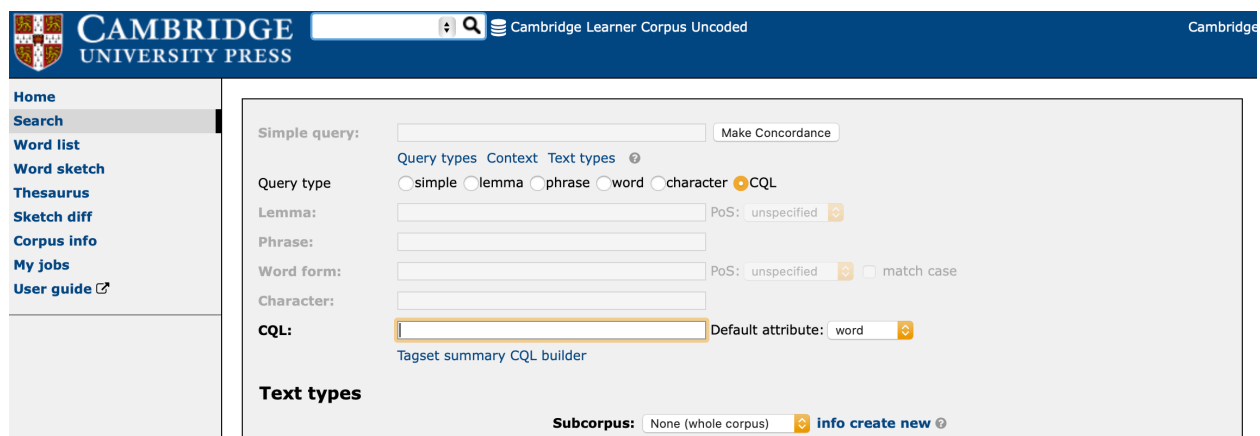
Within the CLC, two types of corpora can be found: the CLC coded and CLC uncoded. The CLC coded corpus is composed of 29 million words, and learners' errors are coded according to an in-house coding system. The coded corpus is particularly useful when observing various learner errors across proficiency levels as well as different L1 speakers. The uncoded corpus is much larger in size, with around 55 million words, but it is not annotated learner errors. Since errors were not within the scope of this study, the CLC uncoded corpus was chosen.

Among the various learner corpora that are readily available, the CLC was chosen for the current study for two main reasons. First and foremost, it is a mega corpus containing over 20 different types of exams across a large number of test takers belonging to various L1 groups, ethnicity, and socioeconomic backgrounds. It was particularly of importance that the corpus chosen for this study is representative of 1) all L1 Korean speakers and 2) other L1 speakers. Because the purpose of this study is to generalize the trajectory of L2 writing development across all L1 Korean speakers, a learner corpus collected in a particular setting (e.g., a single university) would not necessarily represent all L1 Korean speakers. Secondly, the measurement of proficiency level in many learner corpora has posed issues (see Gablasova, Brezina & McEnery, 2017). One of the biggest issues is that many studies describe proficiency levels with

respect to subjective terms such as “beginning,” “intermediate,” and “advanced,” which can, in fact, vary drastically depending on how proficiency was measured. Some studies have also used age as a proxy, which can be problematic since it cannot be claimed that learners are at the same L2 developmental trajectory based on age. In the case of the CLC, however, it uses the Common European Framework of Reference for Languages (CEFR), an international standard that is widely accepted to describe language proficiency. In order to conduct a quasi-longitudinal study, it was crucial that the corpus contains a clear distinction about the learners’ proficiency, and thus, the CLC was chosen as the strongest candidate for this study.

### 3.3.3.1.1 CLC in Sketch Engine

The CLC is accessible through *Sketch Engine* (Kilgarriff et al., 2014), an online text analysis software. **Figure 3.1** shows a screenshot of the CLC as presented on *Sketch Engine*. As can be seen, there are a plethora of tools to choose from, including but not limited to lemma, phrases, words, and Contextual Query Language (CQL). CQL is a special code or query language which can be used to search for a wide range of complex grammatical and lexical structures. For instance, the CQL [**word=**”.”] would search for all one-letter words. Texts can also be searched with respect to words, phrases, or tags. The parts-of-speech (POS) tags in the CLC follows the Penn Treebank Tagset (see **Appendix 4**). Each word in the CLC is assigned a POS tag. For example, the word “and” is tagged for CC, which stands for coordinating conjunction. Using these POS tags, a CQL can be built on the platform to filter the search by a single POS tag or a series of tags.



**Figure 3.1** Screenshot of the CLC in *Sketch Engine*

In addition to the wide range of query types, the CLC also offers an exhaustive list of text types to choose from. These include basic information such as the test taker’s first language, nationality, age, education level, and gender to information about the exams, CEFR level attempted, CEFR level achieved, and the year in which the exam was attempted. There are more exam-specific categories such as style (e.g., advice, survey, argumentative), format (e.g., article, composition, story), and register (e.g., formal, informal), catering researchers with a wide range of information to choose from.

Each document in the CLC is labeled with its own unique document number. As shown in **Table 3.2**, a list of metadata information is provided for each test taker. A full script is also accessible by clicking document numbers; question paper, however, is restricted and, therefore, cannot be accessed.

---

<b>Token number</b> XXX
<b>Document number</b> XXX
<b>doc.id</b> XXXXX
<b>First Language</b> Korean
<b>Nationality</b> Korea, South
<b>Exam</b> CPE
<b>CEFR level exam</b> C2 MASTERY
<b>CEFR level student performance</b> C2 MASTERY
<b>Pass/Fail</b> at level
<b>Year</b> 1999
<b>Age</b> 20-30
<b>Gender</b> F
<b>ALTE level</b> 5
<b>Education level</b> College or University
<b>Written grade</b> B
<b>Wordcount</b> 709
<b>Question number</b> 1
<b>Question ID</b> XXX
<b>Question paper</b> (link provided)
<b>Format</b> Composition/essay
<b>Register</b> Formal
<b>Style</b> Descriptive/creative autobiographical

---

**Table 3.2** Example of test taker metadata provided in the CLC

### 3.3.3.1.2 Common European Framework of Reference for Languages

The Common European Framework of Reference for Language (CEFR) was originally designed by the Council of Europe (CoE) with its inception in the early 1990s. The first version in English was published in 2001. The CEFR provides “a common basis for the elaboration of language syllabuses, curriculum guidelines, examinations, and textbooks across Europe. It describes in a comprehensive way what language learners have to learn in order to use a language for communication and what knowledge and skills they have to develop so as to be able to act effectively” (Council of Europe, 2001, p. 1). Currently, there are over 30 translated languages available, and the framework is recognized worldwide. More recently, using the CEFR as a standardizing measure in learner corpus research has been acknowledged to be particularly crucial for comprehensiveness and consistency (Díez-Bedmar, 2012; Tono & Díez-Bedmar, 2014; Römer, 2019).

**Table 3.3** illustrates the breakdown of the CEFR into three categories: proficient user, independent user, and basic user. Within each of these three levels, there are two sub-levels from A1 to C2. Each level includes its distinct Cambridge exams, which are included in the CLC. (See **Appendix 1** for a comprehensive description of each level).

	CEFR levels		Cambridge Exams
<b>Proficient user</b>	<b>C2</b>	Mastery	CPE
	<b>C1</b>	Effective operational proficiency	CAE
<b>Independent user</b>	<b>B2</b>	Vantage	FCE
	<b>B1</b>	Threshold	PET
<b>Basic user</b>	<b>A2</b>	Waystage	KET
	<b>A1</b>	Breakthrough	

**Table 3.3** The CEFR English profile

As illustrated in this section, not only is the CEFR one of the most commonly used measures for language proficiency, but it is a scale that is robust enough to explain the learners’

language competency at the global level. Unlike studies that use rather vague terms such as “intermediate” level without providing a thorough rubric about what the term entails, the current study utilizes the CEFR to make assertions about the learners’ developmental trajectory using a quasi-longitudinal method (see Díez-Bedmar, 2018). Thus, a robust indicator of the learners’ proficiency level was particularly crucial in this study.

### 3.3.3.1.3 Exams

As mentioned above, the CLC consists of over 20 different types of language exams. Broadly speaking, the tests are divided into two types: business English and English language skills (see **Table 3.4**). Business English exams are designed to prove English language skills to employers in their capability to succeed in an international workplace. On the other hand, English language skills are primarily geared towards proving English proficiency in educational settings.

<b>Business English</b>	
BEC (1/2/3)	Business English Certificate
BECP	Business Preliminary (B1 qualification)
BECV	Business Vantage (B2 qualification)
BECH	Business Higher (C1 qualification)
BULATS	Business Language Testing Service
<b>General and Higher Education</b>	
IELTS ac/gt	International English Language Testing System (academic/general testing)
ICFE	International Certificate in Financial English (discontinued in 2016)
ILEC	International legal English Certificate (discontinued in 2016)
KET/KETfS	Cambridge English: Key/Key for School (A2 qualification)
PET/PETfS	Cambridge English: Preliminary/Preliminary for School (B1 qualification)
FCE	Cambridge English: First (B2 qualification)
CAE	Cambridge English: Advanced (C1 qualification)
CPE	Cambridge English: Proficiency (C2 qualification)
CELSP	Certificate in Communicative Skills in English: Preliminary
CELSV	Certificate in Communicative Skills in English: Vantage
CELSH	Certificate in Communicative Skills in English: Higher
SFLL (1/2/3)	Standard for Foreign Language Learning
SfLE (1/2/3)	State Foreign Language Exam

**Table 3.4** Description of exams in the CLC

It should also be noted that some of the exams are geared specifically to certain qualifications. The PET exam, for instance, awards a B1 qualification; therefore, this exam is only found among learners that have attained a B1 level proficiency. Furthermore, exams such as ICFE and ILEC have been discontinued since 2016.

### 3.3.3.1.4 Style/Format

The writing tasks in the CLC vary greatly across styles and formats depending on the exams. The exams are categorized into eight styles in the CLC (see **Table 3.5**), ranging from giving advice to expressing an opinion and formulating a response based on a given task.

	Style																	
	BEC (1/2/3)	BECP	BECV	BECH	BULATS	CELSH	CELSP	CELSV	IELTS ac/gt	ICFE	ILEC	KET/KETf	PET/PETfS	FCE	CAE	CPE	SFLL (1/2)	SfLE
Advice	•	•	•	•		•	•	•		•			•	•	•	•	•	•
Application/response	•	•	•	•		•		•	•					•	•		•	•
Argumentative/opinion	•		•	•	•	•	•	•	•				•	•	•	•	•	•
Business	•	•	•	•	•	•	•	•	•	•	•		•	•	•	•	•	•
Complaint/apology/res ponse	•	•	•	•		•	•	•	•			•	•	•	•	•	•	•
Critical	•		•	•		•		•						•	•	•		•
Descriptive/creative autobiographical	•	•	•	•		•	•	•	•				•	•	•	•	•	•
Informative/news	•	•	•	•	•	•	•	•	•			•	•	•	•	•	•	•

**Table 3.5** Different types of writing styles included in the exams

The listed styles are in accordance with the CLC. Because access to the exam scripts or the prompts was not granted, it was difficult to gauge the decision behind categorizing the tests into different styles. Also, the CLC is also comprised of various types of writing formats. **Table 3.6** shows a list of different types of writing formats present in each of the exams. All written tasks in the CLC are open-ended and thus, do not include tasks such as fill-in-the-blanks.

	Format																	
	BEC (1/2/3)	BECP	BECV	BECH	BULATS	CELSH	CELSP	CELSV	IELTS ac/gt	ICFE	ILEC	KET/KETS	PET/PETS	FCE	CAE	CPE	SFLL (1/2/3)	SFLE (1/2/3)
Article	•		•	•		•		•						•	•	•	•	
Composition/essay	•		•	•		•		•	•					•	•	•	•	
Informative/instructional text	•	•	•	•		•	•	•	•			•	•	•	•	•	•	•
Letter/reference	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Note/email/memo	•	•	•	•		•	•	•	•		•	•	•	•	•	•	•	•
Proposal	•		•	•		•		•						•	•	•	•	•
Report	•		•	•	•	•		•	•	•				•	•	•	•	•
Review	•		•	•		•		•						•	•	•	•	
Speeches/spoken response	•		•					•						•		•		
Story	•	•	•				•	•					•	•		•		
Survey/questionnaire/form	•	•	•	•		•	•	•					•	•	•		•	

**Table 3.6** Different types of writing formats included in the exams

As evidenced by the styles and formats, the exams differ significantly with respect to their types. Below are some examples of the written tasks taken from various exams found online:

**Business:**

- 1) You are going to attend an engineering exhibition in Frankfurt soon. Write an email to your assistant: 1) explaining why you will be away; 2) letting her know the dates you will be away; 3) saying what work she should do while you are away. *Write between 30-40 words* (BECP, B1 level, Format: email)
- 2) You are a regional sales manager for an international company. You have been asked to go to a meeting at your company's head office. You cannot go, so somebody else will go in your place. Write an email to Erica Young, who is organising the meeting: 1) apologising for not being able to go to the meeting; 2) explaining why you cannot go; 3) saying who will go. *Write 40-50 words* (BECV, B2 level, Format: email)
- 3) The bar chart below (*not shown*) shows the cost of buying three different photocopiers, the cost of warranty on each machine, and their expected running cost for the first two years. Using the information from the chart, write a short report comparing the costs for the three machines. *Write 120-140 words*. (BECH, C1 level, Format: report)

### General and Higher Education:

- 1) Read the email from your English friend, Alex.  
“It’s great you can come to my house this evening to watch a DVD. What time can you come? Which DVD do you want to watch? What would you like to eat?”  
Write an email to Alex and answer the questions. Write 25-35 words. IKET, A2 level, Format: email)
- 2) This is part of a letter you received from your pen friend.  
“I have to give a presentation to my English class about either a successful sportsperson or a musician from your country. Who should I choose? What information could I include?”  
Now write a letter answering your friend’s questions. Write your letter in about 100 words. (PET, B1 level, Format: letter)
- 3) Your English teacher has asked you to write a story. Your story must begin with this sentence: “Tim felt angry as he got off the train”. Write your story in about 100 words. (PET, B1 level, Format: story)
- 4) Last year a new sport and leisure centre opened in your hometown. You have recently received the following letter from the manager of the centre: “I am writing to all members of the Active Leisure Club to thank you for your custom over the last year. We want to make the facilities even better this year! Please send a short report to us, telling us what you think are the best and worst aspects of the club. We want to hear your suggestions for new facilities.” Write your report to the manager of the club. Write between 140-190 words. (FCE, B2 level, Format: report)
- 5) You see this advertisement in an international student magazine:  
“Volunteers needed  
We are looking for volunteers to help out at a famous, international sporting event. We are looking for friendly, respectful people with good language skills, good team skills and a ‘can-do’ attitude. We need people to welcome delegates, provide customer service and solve problems. If you think you have what it takes, apply now.”  
Write an application to become a volunteer. Mention: 1) your language skills; 2) your personal qualities; 3) examples of times when you have demonstrated team skills; 4) any relevant work experiences. Write 220-260 words. (CAE, C1 level, Format: letter)
- 6) Your local newspaper is asking its readers to write a review of a restaurant that they have eaten at recently and would recommend to others:  
“Which restaurant would you recommend?”  
Write your review of the restaurant and the meal that you ate there, and describe some or all of the following points: the atmosphere, attitude of the staff, the quality of the service and whether the restaurant is good value for money. Write the review for the newspaper readership. Write around 280-320 words. (CPE, C2 level, Format: review)

### 3.3.3.2 L1 Korean Speaker Sub-Corpus

The L1 Korean speaker sub-corpus was taken from the CLC. Broadly, it consist of a total of 1,973 test takers<sup>3</sup>. As shown in **Table 3.7**, the majority of the learners are between A2 and C1 levels. Among them, the B1 proficiency level group contains the largest number of learners, while the C2 level group contains the fewest number of learners.

L1 Korean Speakers					
A1	A2	B1	B2	C1	C2
133	266	711	559	222	82

**Table 3.7** Number of participants per proficiency level in the L1 Korean speaker sub-corpus

Despite there being fewest learners in the C2 level, the number of words is greater than both A1 and A2 level groups combined (see **Table 3.8**). The entire sub-corpus consists of 914,363 words across the six proficiency levels. It should be noted that the A1 level is omitted from all analyses due to its small data. Furthermore, the concordance lines suggest that the learners at the A1 level struggled to produce concrete sentences and, therefore, tended to copy parts of the prompt in their writing. This was visible because the majority of the sentences were similar across all learners. Therefore, the A1 level data was excluded from analyses.

L1 Korean Speakers					
A1	A2	B1	B2	C1	C2
6,491	37,866	273,392	276,168	140,337	65,019

**Table 3.8** Words per proficiency level in the L1 Korean speaker sub-corpus

As shown in **Table 3.9**, the L1 Korean speaker sub-corpus consists of 24 different exam types (see **3.3.3.1.3** for the explanation of the exams). The IELTS ac/gt exams were the most common in the B1, B2, and C1 levels, and this is to be expected since often, foreigners will often have to meet language requirements to attend school or work in English-speaking countries. As

<sup>3</sup> All figures related to the CLC reported henceforth are from the time the data was collected (December 2017) and thus may be different than the current version.

elaborated above (see 3.3.3.1.3), many of the exams are level-specific, and thus, they were only available for specific levels. For instance, KET, the English qualification exam for A2 level, is predominantly found in A1 and A2 levels. On the other hand, a certificate in English language skill exams such as CELSV and CELSP is for B2 and C1 levels, respectively.

Exam	A1	A2	B1	B2	C1	C2
BEC1	-	-	670	-	-	-
BEC2	-	-	-	685	-	-
BEC3	-	-	-	-	398	-
BECH	-	-	-	2,282	-	-
BECP	-	5,913	2,180	141	-	-
BECV	-	-	8,023	-	423	-
CAE	-	-	-	33,674	40,284	7,165
CELSP	-	-	2,799	-	-	-
CELSV	-	-	-	748	-	-
CPE	-	-	-	-	40,395	53,207
FCE	-	-	50,587	73,021	15,914	-
KET	4,720	2,646	339	-	-	-
KETfS	1,134	-	41	-	-	-
ICFE	-	-	-	624	-	-
IELTS ac	-	-	157,257	142,045	31,334	4,000
IETLS gt	-	-	35,886	20,197	9,685	647
ILEC	-	-	-	-	553	-
PET	-	23,904	14,835	1,129	-	-
PETfS	-	4,806	-	370	-	-
SfLE1	637	-	-	-	-	-
SfLE2	-	597	-	-	-	-
SfLE3	-	-	775	-	-	-
SFLL1	-	-	-	1,252	-	-
SFLL2	-	-	-	-	1,351	-

**Table 3.9** Distribution of exam types in the L1 Korean speaker sub-corpus

On the other hand, the BEC exams are divided into three different levels (1, 2, 3), where the "1" corresponds to the B1 level, 2 to the B2 level, and 3 to the C1 level. Thus, the data can only be found at these levels. The BECP (Business Preliminary) and BECV (Business Vintage) are also business qualification exams designed to test an individual's ability to function in international business. Some of the existing exams have changed their names over the years (e.g., CAE, Cambridge English: Advanced -> C1 Advanced), but these data were not eliminated in this study.

### 3.3.3.3 Other L1 Speaker Sub-Corpus

From the CLC, another sub-corpus was drafted. This sub-corpus consisted of all L1 speakers but with the exception of L1 Korean speakers. As can be predicted, the other L1 speaker sub-corpus is much larger than that of the L1 Korean speaker sub-corpus, consisting of 231,701 learners from 121 nationalities (see **Table 3.10**).

<b>Other L1 speaker (-L1 Korean Speakers)</b>					
<b>A1</b>	<b>A2</b>	<b>B1</b>	<b>B2</b>	<b>C1</b>	<b>C2</b>
61,711	74,962	37,897	27,695	17,644	11,792

**Table 3.10** Number of participants per proficiency level in the other speakers' L1 sub-corpus

In total, there are 53,819,650 words across the six proficiency levels (see **Table 3.11**). Similar to the L1 Korean sub-corpus, the other L1 speaker sub-corpus is also concentrated heavily between A2 and C1 levels. Since the purpose of the L1 speaker sub-corpus is to supplement the L1 Korean speaker sub-corpus, all analyses also excluded the A1 level from the other L1 speaker sub-corpus.

<b>Other L1 speaker (-L1 Korean Speakers)</b>					
<b>A1</b>	<b>A2</b>	<b>B1</b>	<b>B2</b>	<b>C1</b>	<b>C2</b>
2,879,139	9,946,316	9,449,530	12,486,631	10,182,073	8,875,961

**Table 3.11** Words per proficiency level in the other L1 speaker' sub-corpus

**Table 3.12** shows the distribution of the exams in the other L1 speaker sub-corpus. In the A1 proficiency level, the business exam, BECP, and the basic level English qualification exam, KET, are the most common. The A2 level is also concentrated heavily on the basic level English qualification exams. Business English exams such as BEC, BECH/P/V, and BULATS are predominantly visible in B1 to C2 levels, with the exception of BECP, which is most prevalent in A1. As with the L1 Korean speaker sub-corpus, IELTS exams were the most commonly found in the B1 and C1 levels.

Exam	A1	A2	B1	B2	C1	C2
BEC1	-	-	109,796	47,549	-	-
BEC2	-	-	-	36,042	-	-
BEC3	-	-	-	-	75,982	-
BECH	-	-	-	1,249,841	399,550	506,284
BECP	2,444,794	-	489,638	137,806	-	-
BECV	-	-	2,040,821	905,188	473,000	-
BULATS	30	-	4,653	15,793	4,536	-
CAE	-	-	-	723,758	3,215,062	1,114,235
CELSH	-	-	-	-	123,484	-
CELSP	-	-	333,884	226,996	-	-
CELSV	-	-	-	-	-	-
CPE	-	-	-	-	2,338,146	6,531,407
FCE	-	-	1,197,770	2,591,028	1,081,421	-
ICFE	-	-	-	868	22,131	-
IELTS ac	-	-	1,868,964	2,160,210	1,169,416	300,652
IELTS gt	-	-	1,019,922	1,314,821	805,747	423,383
ILEC	-	-	-	54,139	291,764	-
KET	2,484,796	718,961	90,125	68,192	-	-
KETfS	306,049	26,842	69,562	44,774	-	-
PET	-	4,997,763	1,921,083	990,372	-	-
PETfS	-	1,637,660	156,627	167,020	-	-
SfLL1	-	-	-	123,385	-	-
SfLL2	-	-	-	-	181,834	-
SfLE1	88,264	-	-	-	-	-
SfLE2	-	120,296	-	-	-	-
SfLE3	-	-	146,685	103,667	-	-

**Table 3.12** Distribution of exam types in the other L1 speaker sub-corpus

As mentioned, the other L1 speaker sub-corpus consists of various L1. **Table 3.13** shows the rankings of the top 20 first languages per proficiency level in the corpus, along with the number of learners and word frequency (see **Appendix 2** for a comprehensive list of the L1s included in the other L1 speaker sub-corpus). In all six proficiency levels, Spanish-Latin American is among the most common, with Greek, German, Portuguese, French, Chinese, Italian, and Spanish-European being other languages that ranked in the top.

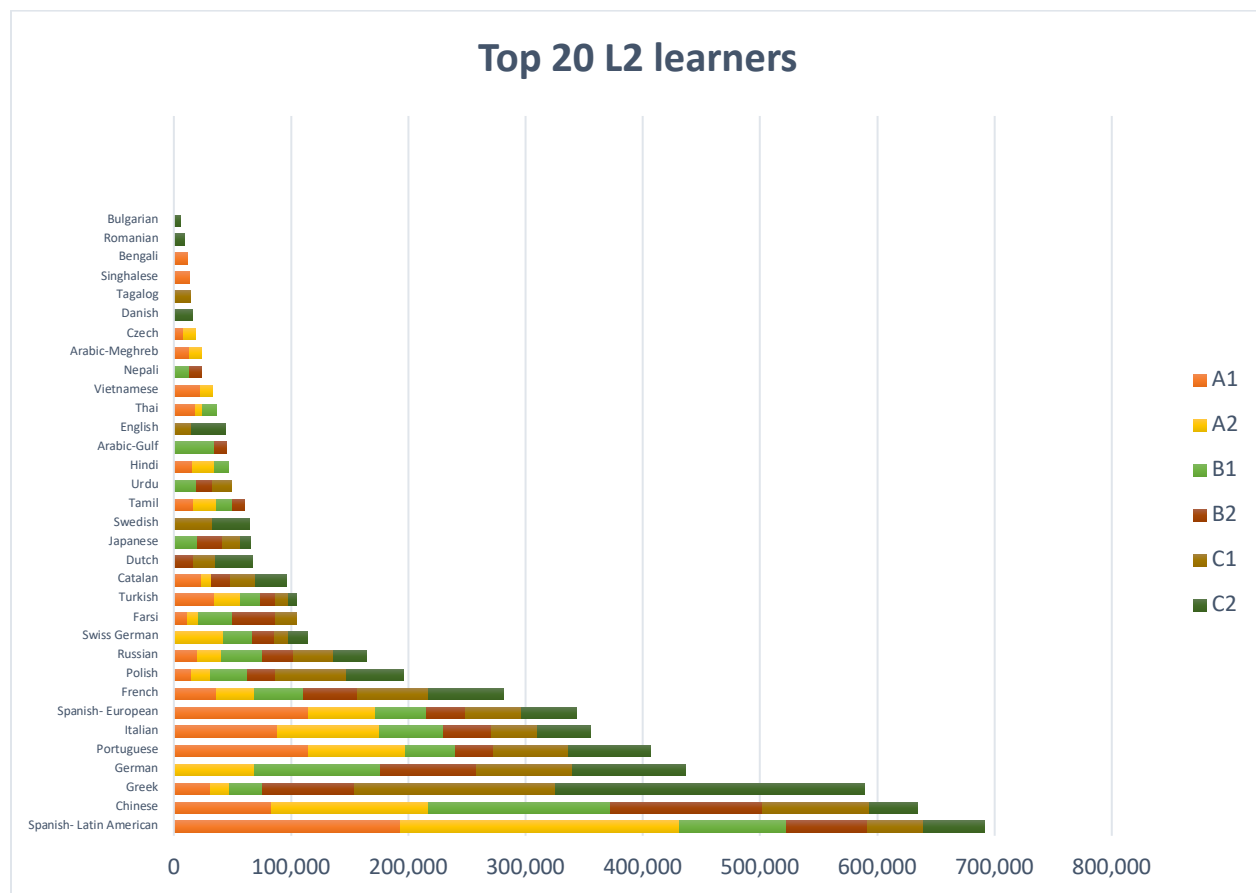
Rank	A1 (121 L1)	A2	B1	B2	C1	C2
1	Spanish- Latin American (556,308/12,702)	Spanish- Latin American	Chinese (1,466,168/5,126)	Chinese (1,623,676/3,134)	Greek (1,740,847/2,719)	Greek (2,341,618/2,973)

		(2,368,247/17,426)				
2	Portuguese (331,568/7,525)	Chinese (1,336,137/9,209)	German (1,019,632/4,895)	German (1,023,794/2,331)	Chinese (929,245/2,115)	German (858,480/1,157)
3	Spanish-European (330,287/7,141)	Italian (865,769/7,756)	Spanish- Latin American (862,150/4,542)	Greek (986,390/2,182)	German (828,257/1,649)	Portuguese (621,571/773)
4	Italian (254,508/5,850)	Portuguese (818,055/6,535)	Italian (519,380/2,956)	Spanish-Latin American (862,150/4,542)	Portuguese (655,616/969)	French (565,857/705)
5	Chinese (238,709/5,058)	German (681,544/4,676)	Spanish-European (413,342/1,801)	French (572,934/1,264)	Polish (618,585/991)	Spanish- Latin American (464,014/602)
6	French (104,509/2,071)	Spanish-European (568,049/4,686)	Portuguese (403,721/2,363)	Italian (501,690/1,549)	French (612,545/1,349)	Polish (429,600/559)
7	Turkish (99,263/2,070)	Swiss German (420,048/2,941)	French (400,357/1,625)	Farsi (456,888/883)	Spanish-European (484,481/739)	Spanish-European (420,833/522)
8	Greek (89,924/2,027)	French (321,327/2,518)	Russian (330,162/1,372)	Spanish-European (413,342/1,801)	Spanish-Latin American (483,070/823)	Italian (394,926/504)
9	Catalan (68,885/1,479)	Turkish (224,183/1,665)	Arabic-Gulf (325,324/835)	Portuguese (403,721/2,363)	Italian (406,703/685)	Chinese (358,643/597)
10	Vietnamese (66,070/1,190)	Russian (202,976/1,687)	Polish (297,759/1,146)	Russian (330,162/1,372)	Russian (345,557/663)	Dutch (279,002/359)
11	Russian (58,607/1,207)	Tamil (196,909/1,396)	Farsi (270,640/625)	Polish (297,759/1,146)	Swedish (339,279/590)	Swedish (277,867/362)
12	Thai (52,012/1,025)	Hindi (185,119/1,328)	Greek (266,207/837)	Japanese (267,660/576)	Catalan (221,818/325)	English (259,008/409)
13	Tamil (47,819/996)	Greek (159,675/1,434)	Swiss German (236,345/1,191)	Swiss German (236,345/1,191)	Dutch (193,998/324)	Russian (247,302/330)
14	Hindi (46,976/1,008)	Polish (158,770/1,157)	Japanese (190,280/547)	Dutch (206,072/478)	Farsi (185,617/330)	Catalan (235,974/288)
15	Polish (43,886/758)	Vietnamese (100,729/856)	Urdu (179,177/344)	Catalan (196,523/421)	Urdu (159,373/254)	Danish (144,735/181)
16	Arabic- Meghreb (39,134/877)	Czech (99,423/700)	Turkish (161,049/532)	Urdu (179,177/344)	Japanese (155,641/256)	Swiss German (144,529/226)
17	Singhalese (37,371/743)	Arabic-Meghreb (95,735/683)	Tamil (126,409/608)	Turkish (161,049/532)	English (154,060/255)	Romanian (79,979/129)
18	Farsi (34,078/585)	Farsi (94,151/560)	Nepali (125,917/228)	Arabic-Gulf (129,422/265)	Tagalog (144,995/242)	Japanese (77,993/100)
19	Bengali (32,911/581)	Catalan (82,817/691)	Thai (118,183/304)	Tamil (126,409/608)	Swiss German (120,395/284)	Turkish (57,873/73)
20	Czech (24,076/527)	Thai (61,862/400)	Hindi (113,981/444)	Nepali (125,917/228)	Turkish (112,305/182)	Bulgarian (53,057/74)

**Table 3.13** Top 20 first languages per proficiency level (word frequency/number of learners)

**Figure 3.2** displays the ranking of the top 20 L1 in the other L1 speaker sub-corpus. Evidently, some obvious differences are in Greek speakers, where the data is clustered among B2 to 2 levels. On the contrary, the majority of Spanish-Latin American speakers are in the lower

levels from the A1 to B1 levels. Overall, however, it is evident that there is a good distribution of a wide range of different L1s across all six proficiency levels.



**Figure 3.2** Top 20 distribution of L1 in the other L1 speaker sub-corpus (normalized to 1-million words)

### 3.3.3.4 The British National Corpus (BNC)

The British National Corpus (BNC) is a 100-million-word corpus comprised of written and spoken modern English by native speakers and advanced speakers of English. The written section of the BNC takes up about 90% of the entire corpus and is a compilation of extracts from a wide range of texts such as academic books, journals, newspapers, published and unpublished letters and memoranda, and essays from schools and universities (see **Appendix 3** for a comprehensive list of BNC written genres).

For the purpose of this study, only the written part of the BNC was considered. The BNC written sub-corpus consists of 99,065,905 tokens and 84,770,970 words. Unlike the CLC, the

BNC is a sample corpus containing text samples that generally do not exceed over 45,000 words. It is not limited to any particular subject, register, or genre and contains texts from a wide range of domains (see **Table 3.14**).

	Texts	Words	%
Imaginative	476	16,496,420	18.75
Applied science	370	7,174,152	8.15
Arts	261	6,574,857	7.47
Belief & thought	146	3,037,533	3.45
Commerce & finance	295	7,341,163	8.34
Leisure	438	12,237,834	13.91
Natural & pure science	146	3,821,902	4.34
Social science	526	14,025,537	15.94
World affairs	483	17,244,534	19.60

**Table 3.14** Classification of subject domain in the BNC written sub-corpus

The domain is broadly divided into two: imaginative and informative. Imaginative texts account for roughly around 20% of the corpus and consist of fiction and creative text.

Informative texts, on the other hand, account for 80% and are subject-specific. Among these texts, 57.18% are books, while 32.52% come from periodicals. Miscellaneous published work accounts for 4.81%, and miscellaneous unpublished work makes up 4.02% of the corpus.

The texts in the BNC corpus were collected from the 1960s to the 1970s. The authors' ages ranged from those belonging to the 0-14 group to the 60+ group. While the domicile of more than 50% of the authors is unknown, the ones that have been identified are residences of the UK, Ireland, Commonwealth, Continental Europe, and the USA.

### 3.3.3.5 Collins WordBanks Online

The Collins WordBanks Online is a 550-million word corpus (Tokens: 551,531,292; Words: 462,302,063; Sentences: 29,337,326; Paragraphs 3,436,439; Documents: 22,620) consisting of written and spoken English dating back from 1972 to 2005, although the majority of the texts were collected between 2001 and 2005. One unique aspect of this corpus is that it accounts for eight different varieties of English, such as British English, American English, and Canadian English. The corpus was used for the creation of the Collins dictionaries and is

available to users upon subscription. **Table 3.15** illustrates the various text forms included in the WordBanks corpus.

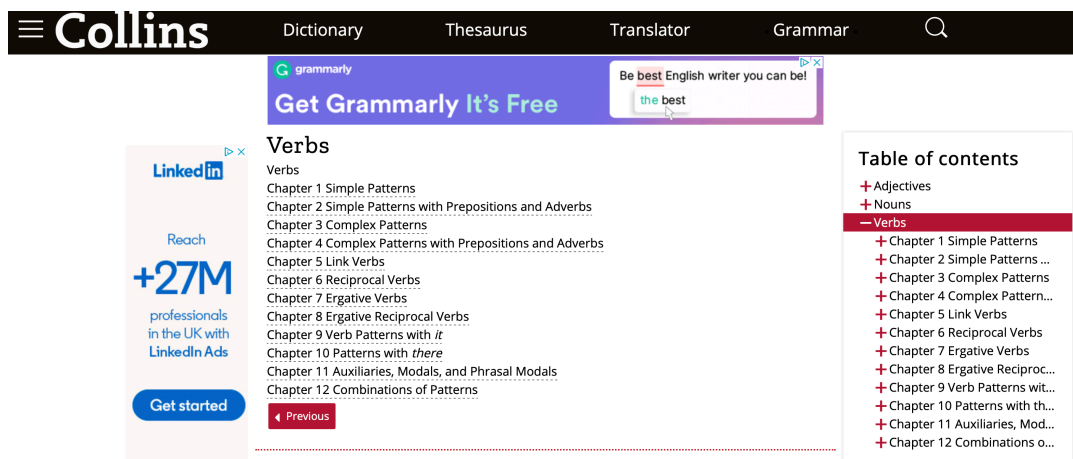
DOC. Text form	Tokens	%	Documents
Book	152,322,145	27.54	1,413
Ephemera	8,867,835	1.60	5,165
Magazine	43,780,298	7.91	3,992
Newspaper	286,543,293	51.80	8,335
Report	149,568	0.03	138
Spoken	61,508,350	11.12	3,577

**Table 3.15** Text types of the Collins WordBanks Online

### 3.3.3.6 Collins English Dictionary

The Collins English dictionary was first published on 31 December 2011 online (*collinsdictionary.com*). It contains a wide range of useful resources for both language learners and teachers. In addition to the dictionary, the Collins English dictionary also contains other resources such as English usage, grammar patterns, and translator, to name a few.

The Collins grammar pattern is a sub-section of the Collins dictionary, which includes a comprehensive listing of grammar patterns used in English adjectives, nouns, and verbs. **Figure 3.3** shows the different sections in which the verbs are divided. As can be seen, the verbs are divided into different “chapters,” ranging from simple patterns to more complex patterns to the combination of different patterns.



**Figure 3.3** Main page of verb patterns in Collins English Dictionary

Clicking on each of the chapters would reveal further information about the patterns. For instance, under “Chapter 1 Simple pattern”, are a wide array of verbs such as “V n” – verb followed by a noun group (see **Figure 3.4**).

## Chapter 1 Simple Patterns

### Chapter 1 Simple Patterns

In this chapter we describe simple verb patterns. The simplest of these is V, where the verb may be used on its own. In all the other patterns, the verb is followed by one other element, such as a noun group, an '-ing' clause, or a that-clause. Patterns in which the verb is followed by a prepositional phrase or adverb group are described in Chapter 2.

[V](#)  
[V n](#)  
[V pl-n](#)  
[V pron-refl](#)  
[V amount](#)  
[V adj](#)  
[V -ing](#)  
[V to-inf](#)  
[V inf](#)  
[V that](#)  
[V wh](#)  
[V wh-to-inf](#)  
[V with quote](#)  
[V so/not](#)  
[V as if, V as though, V like](#)  
[V and v](#)

Next ▶

**Figure 3.4** Example from ‘simple patterns’ under verbs

In the **V n** pattern, for example, are the different structures such as “verb with complement,” “verb with an object,” and “verb with adjunct” followed by sample sentences (see **Figure 3.5**).

## V n

### V n

The verb is followed by a noun group. The passive pattern is **be V-ed**.

This pattern has three structures:

Structure I: Verb with Complement

*He was my friend.*

Structure II: Verb with Object

*The thieves broke a window.*

Structure III: Verb with Adjunct

*Children don't talk that way.*

[Structure I: Verb with Complement](#)

[Structure II: Verb with Object](#)

[Structure III: Verb with Adjunct](#)

[Other related patterns](#)

**Figure 3.5** Example of ‘Verb with Complement’ pattern

Within each of the structures are various groups. As shown in **Figure 3.6**, the ‘verb with object’ pattern is then further broken into different groups. Each of the examples belongs to one of the groups, albeit the list is not comprehensive and only includes the top 400 more frequently used verbs listed in the Collins WordBanks corpus.

- II.1 The 'kill', 'eat', and 'fix' groups
  - II.2 The 'bring', 'buy', and 'operate' groups
  - II.3 The 'cover', 'follow', and 'record' groups
  - II.4 The 'build' group
  - II.5 The 'change', 'concern', and 'arrange' groups
  - II.6 The 'watch', 'break a record', and 'approve' groups
  - II.7 The 'form' group
  - II.8 The 'start' and 'stop' group
  - II.9 The 'do' and 'take', 'turn a corner', and 'open your eyes' groups
  - II.10 The 'face', 'take three days', and 'next week sees' groups
  - II.11 The 'top', 'develop', and 'include' groups
  - II.12 The 'show', 'allow', and 'identify' groups
  - II.13 The 'hear', 'interest', and 'give an impression' groups
  - II.14 The 'say', 'describe', and 'call' groups
- Structure information: Verb with Object
- 

**Figure 3.6** Different groups belonging to the **V n** group

Within each of the groups is a description of verbs belonging to the following groups, followed by a list of verbs belonging to the group (see **Figure 3.7**).

## V n

The 'say' group

The Object indicates what is said or written. The verb and the Object together indicate the process of speaking or writing. This includes:

verbs which themselves indicate communication e.g. *ask, say, speak*

verbs which indicate communication only when used with particular noun groups e.g. *draw, make, pass*

*Mothers **come out with** remarks like that and there's not a thing sons can do about it.*

*He **gave out** a scream of pain.*

*Some letters are either full of praise or downright rude but **don't make** any constructive suggestion.*

*On that day, the Hopi leaders arrived at a deserted strip of coastline near Laguna, California, to gather salt and **offer** their prayers.*

*When booking the flight, **put in** your request for high-protein low-fat meals.*

*At one stage the newspaper, which had been particularly vitriolic, **ran** a story which was riddled with inaccuracies.*

*Shirley took me into a separate room where, clipboard in hand, she **ran through** various personal details she'd need.*

*I took some of my drink. I **didn't say** anything.*

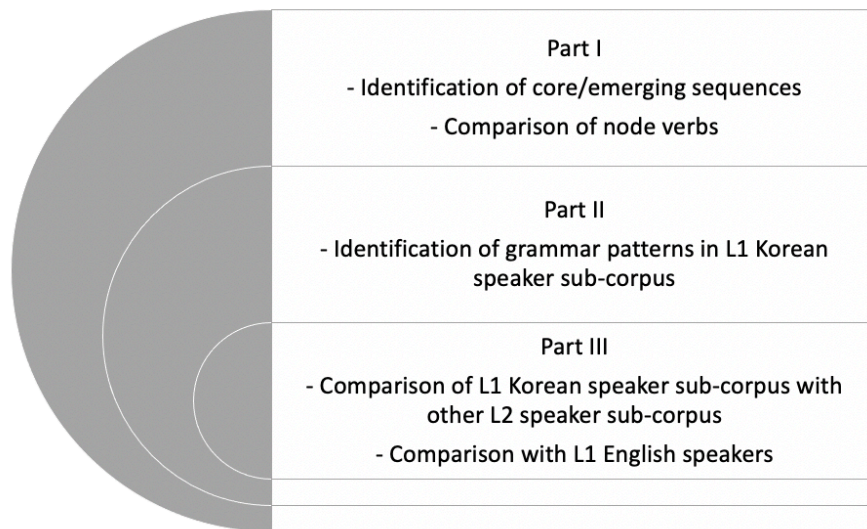
*I can't believe that Paul **wrote** that letter.*

**Figure 3.7** Description of the 'say' group

Notably, the identification of the meaning groups can be laborious and subjective. Despite this caveat, the categorization of the verbs in the verb sequences can offer valuable insight into learners' language usage in a systematic way.

### 3.4 Procedures

The procedure for this study is in three parts (see **Figure 3.8**). The first part discusses the process of identifying the "core" and "emerging" verb 4-gram sequences in the CLC, followed by how the data were then compared across the six proficiency levels in the L1 Korean speaker sub-corpus. Additionally, the procedure for comparing the "node verbs," also known as main verbs or head verbs, is explained. Part two focuses on how the grammar patterns were classified among the "core" and "emerging" sequences in the two sub-corpora and the reference corpus. The last part elaborates on the process of comparing the L1 Korean speaker sub-corpus with the other L2 speaker sub-corpus as well as the L1 English corpus.



**Figure 3.8** Data extraction and analysis procedures in the study

### 3.4.1 Identifying Verb 4-gram Sequences and Node Verbs in the CLC

The first and foremost step in collecting data was to establish the two sub-corpora from the CLC. The first sub-corpus consisted of L1 Korean speakers and the second sub-corpus consisted of all other L2 speakers in the CLC, exclusive of L1 Korean speakers. The two sub-corpora were created and stored in *Sketch Engine* using the following steps:

- 1) In the CLC, click ‘create sub-corpus’.
- 2) Create a name for the sub-corpus (e.g., L1 Korean) and ensure that the sub-corpus only consisted of learners who have indicated that they speak Korean as their L1.
- 3) Repeat Step 2 for the other L1 speaker sub-corpus.

Once the two sub-corpora were created, the next step was to search for the verb 4-gram sequences using Corpus Query Language (CQL). To reiterate, CQL is a language function that uses regular expressions, which then allows the search of complex grammar or lexical patterns within the corpus. Simply put, it is a way to query the corpus to search for the desired language pattern. In the case of this study, the interest was in the verb 4-gram sequences. The verb 4-gram sequences are 4- slots, or a consecutive sequence of four words, headed by a verb (henceforth “node verb”). The rationale for using a 4-gram window is supported by empirical studies looking at lexical bundles (Bal, 2010), syntactic patterning (Ellis, 2012), and verb argument constructions (Ninio, 2011). Since too few words have the potential of losing the sense of shedding tangible

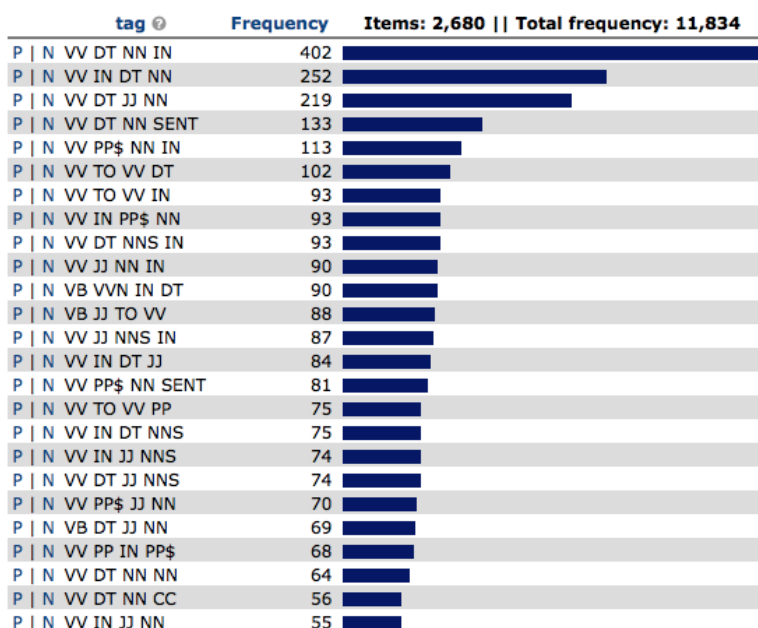
insight into the use of sequences, while too many words may be too lengthy and have the potential to miss important patterns, the 4-slot window was also chosen to be suitable for the current study.

The 4-slot verb 4-gram sequences were identified using the following CQL:

**“V.\*”/[]/[]/[]**

In the above, “V” stands for verbs, followed by “.\*”, which indicates to search for all verbs (e.g., past tense, modal verbs). The three [] [] [] denote the three slots following the verb, irrespective of parts of speech or punctuation. For example, “v.\*” [][] [] could indicate results such as “eat a red apple” or “play in the playground.”

To search for the verb 4-gram sequences, the above CQL was entered in the respective sub-corpus in Sketch Engine, which was filtered by learners’ performance level. For instance, the CQL “V.\*” [][] [] was followed by the filter “A1 CEFR learner performance,” which would display all verbs followed by three slots. **Figure 3.9** shows an example of the top 25 4-slot verb 4-gram sequences, which have been arranged by node tags. The tags are labeled in accordance with the Penn Treebank POS tagset (see **Appendix 4** for a comprehensive list), but as an example, the tag “VV DT NN IN” indicates verb-determiner-noun-preposition.



**Figure 3.9** Example of top 25 4-slot verb 4-gram sequences in the CLC

The same procedure was repeated six times in the L1 Korean sub-corpus, once for each of the six CEFR proficiency levels. Then, the same procedure was replicated for the other L1 speaker sub-corpus. Once the verb 4-gram sequences for each proficiency level from the two sub-corpora in the CLC were drafted, the sequences were downloaded into Microsoft Excel. Careful planning and recording took place to ensure that all data was kept separate for L1 Korean speakers and other L1 speakers, as well as between the six proficiency levels.

From the recorded data, the next step was to identify the sequences that ranked in the top 25 from each proficiency level. To do so, the verb 4-gram sequences were first ranked according to their frequencies and then normalized to one-million words. Note in **Figure 3.11** that although the raw frequencies may have been different, some of the sequences ranked the same once the frequencies were normalized. The ranked verb 4-gram sequences in the L1 Korean sub-corpus were then compared across the six proficiency levels. As can be seen in **Figure 3.11**, the core and emerging sequences are identified. The process is as such:

- 1) Look at the top 10 sequences and identify the verb 4-gram sequences that are ranked at the top. In this case, the sequence VV DT NN IN appears first in three of the six levels and second in two of the six. Also, it was evident that the top five sequences in the C2 level were consistent from B2 to C1, with a few appearing in lower levels such as A2 and A1 levels. These were then labeled as ‘core’ sequences, meaning they occurred most frequently across all proficiency levels (labeled in bold).
- 2) Look at the remaining sequences in the top 10 in the C2 level. Then identify where these sequences ranked from A1 to C1 levels. It was clear that some of the verb 4-gram sequences ranked lower in the C1 level (e.g., VVN IN JJ NNS) and even lower in the B2, B1, and A2 levels. Six sequences were identified in such a way and were labeled as “emerging” sequences, sequences that ranked lower in lower proficiencies but ranked higher in advanced levels (labeled in red font).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1	A1				A2				B1				B2				C1			C2				
2	Rank	Tag	Freq.	Norm.	Rank	Tag	Freq.	Norm.	Rank	Tag	Freq.	Norm.	Rank	Tag	Freq.	Norm.	Rank	Tag	Freq.	Norm.	Rank	Tag	Freq.	Norm.
3	1	VBP VVG TO VV	17	3006.72	1	VBP VVG TO VV	79	2394.81	1	VV DT NN IN	394	1654.25	1	VV DT NN IN	425	1766.48	1	VVN IN DT NN	230	1881.25	1	VV DT NN IN	95	1677.17
4	2	VV IN PP SENT	13	2299.26	2	VV DT NN IN	62	1879.47	2	VV IN DT NN	258	1083.24	2	VVN IN DT NN	327	1359.15	2	VV DT NN IN	203	1660.41	2	VVN IN DT NN	81	1430.01
5	3	VV IN PPS NN	10	1768.66	3	VV IN DT NN	53	1606.64	3	VVN IN DT NN	196	822.93	3	VV IN DT NN	274	1138.86	3	VV IN DT NN	175	1431.39	3	VV IN DT NN	68	1200.50
6	4	VV TO VV IN	8	1414.93	4	VV PP IN PPS	46	1394.45	4	VV DT JJ NN	196	822.93	4	VV DT JJ NN	232	964.29	4	VV DT JJ NN	158	1292.34	4	VV DT JJ NN	65	1147.54
7	4	VV DT NN SENT	8	1414.93	5	VV DT JJ NN	44	1333.82	4	VVG IN DT NN	175	734.76	5	VVG IN DT NN	207	860.38	5	VVG IN DT NN	106	867.01	5	VVG IN DT NN	52	918.03
8	5	VVP TO VV IN	7	1238.06	5	VZ RB JJ SENT	44	1333.82	5	VBP VVG TO VV	160	671.78	6	VZ DT JJ NN	171	710.75	6	VVN IN DT JJ	104	850.65	5	VZ DT JJ NN	52	918.03
9	5	VVP PP VVD PPS	7	1238.06	6	VV DT NN SENT	43	1303.50	6	VV DT NN SENT	152	638.19	7	VV DT NN SENT	144	598.52	7	VZ DT JJ NN	91	744.32	6	VVN IN DT JJ	49	865.07
10	5	VVP IN PPS NN	7	1238.06	7	VV PP RB SENT	40	1212.56	6	VZ RB JJ IN	152	638.19	8	VVP DT NN IN	128	532.02	8	VZ RB JJ IN	77	629.81	7	VVD DT NN IN	48	847.41
11	5	VVG TO VV IN	7	1238.06	8	VV PPS NN SENT	39	1182.25	7	VZ DT JJ NN	132	554.22	9	VV PPS NN SENT	122	507.08	9	VVD DT NN IN	76	621.63	8	VVN IN PPS NN	40	706.18
12	5	VV DT NN IN	7	1238.06	9	VVP TO VV IN	35	1060.99	8	VV PPS NN SENT	125	524.83	10	VBP VVG TO VV	121	502.93	10	VBP VVG TO VV	71	580.73	9	VV PPS NN IN	37	653.21
13	6	VV PPS NN IN	6	1061.20	9	VVP PP MD VV	35	1060.99	9	VV IN PPS NN	124	520.63	11	VV IN DT NN	119	494.61	11	VV IN DT NN	69	564.38	10	VVN IN JJ NNS	34	600.25
14	6	VV PP RB SENT	6	1061.20	10	VZ RB JJ IN	33	1000.36	10	VV IN DT NN	121	508.03	12	VVD IN DT NN	116	482.14	11	VV IN DT JJ	69	564.38	10	VVD DT JJ NN	34	600.25
15	6	VV IN CD NN	6	1061.20	11	VV PP SENT PP	32	970.05	10	VV PPS NN IN	121	508.03	13	VV TO VV DT	115	477.99	12	VVG DT NN IN	68	556.20	10	VBD DT JJ NN	34	600.25
16	6	VV IN CD	6	1061.20	12	VVD IN DT NN	31	939.74	11	VV DT NN IN	118	495.44	14	VVN IN JJ NN	112	465.52	12	VB VVN IN DT	68	556.20	11	VVG DT JJ NN	31	547.29
17	6	VHP VVN PPS NN	6	1061.20	13	VV TO VV DT	30	909.42	12	VV JJ NN IN	105	440.85	14	VV DT NN IN	112	465.52	13	VV PPS NN IN	64	523.48	11	VVO IN DT NN	31	547.29
18	6	VZ PPS JJ NN	6	1061.20	14	VVG TO VV DT	28	848.79	13	VV TO VV DT	103	432.46	15	VVN IN DT JJ	110	457.21	13	VZ JJ TO VV	64	523.48	11	VZ DT NN IN	31	547.29
19	6	VBP VVG IN DT	6	1061.20	14	VV PPS NN IN	28	848.79	14	VV PPS NN IN	101	424.06	15	VV TO VV IN	110	457.21	13	VBP VVG TO VV	64	523.48	12	VVG DT NN IN	30	529.63
20	7	VVP IN DT NN	5	884.33	14	VV IN PPS NN	28	848.79	15	VZ JJ TO VV	100	419.86	15	VBP RB JJ IN	110	457.21	13	VV JJ TO VV	64	523.48	12	VBD RB JJ IN	30	529.63
21	7	VVD IN PPS NN	5	884.33	15	VBP PP SENT PP	27	818.48	16	VVN IN JJ VV	97	407.27	16	VZ JJ TO VV	105	436.42	14	VVN IN PPS NN	63	515.30	13	VV IN DT JJ	28	494.32
22	7	VV PPS NN SENT	5	884.33	16	VHP DT JJ NN	26	788.17	16	VVD DT NN IN	97	407.27	17	VV IN PPS NN	102	423.95	15	VBD VVN IN DT	60	490.76	13	VJ TO VV	28	494.32
23	7	VV PP NN SENT	5	884.33	17	VVP IN DT NN	25	757.85	17	VVP TO VV IN	96	403.07	17	VV DT NNS IN	102	423.95	16	VV DT NNS IN	58	474.40	14	VB JJ TO NN	28	494.32
24	7	VV PP IN PP	5	884.33	17	VVD PP IN PPS	25	757.85	18	VVD IN DT NN	92	386.27	18	VV JJ TO VV	101	419.80	17	VVP IN DT NN	57	466.22	14	VV DT NN SENT	25	441.36
25	7	VV IN DT NN	5	884.33	18	VVP TO VV PP	23	697.22	18	VZ RB JJ SENT	92	386.27	19	VVZ DT NN IN	97	403.17	18	VVP DT NN IN	55	449.86	14	VV DT NN IN DT	25	441.36
26	7	VZ RB JJ IN	5	884.33	18	VV TO VV DT	23	697.22	19	VVG DT NN IN	90	377.87	20	VVG DT NN IN	95	394.86	19	VBD DT JJ NN	54	441.69	15	VV IN PPS NN	24	423.71
27	7	VZ DT NN IN	5	884.33	19	VV TO VV PP	22	666.91	19	VV TO VV PP	90	377.87	20	VVP PP NN SENT	95	394.86	20	VV PPS JJ NN	53	433.51	15	VV DT NN IN	24	423.71
28	7	VBP PP SENT PP	5	884.33	19	VV IN NN IN	22	666.91	20	VV DT NNS IN	89	373.68	21	VV TO VV IN	93	386.55	20	VZ DT NN IN	53	433.51	16	VZ RB JJ IN	23	406.05
29	8	VVP VV PPS NN	4	707.46	20	VV IN DT NN	21	636.60	20	VB JJ TO VV	89	373.68	21	VVN IN JJ NNS	93	386.55	21	VVN IN JJ NNS	52	425.33	16	VB VVN IN DT	23	406.05
30	8	VVP TO VV DT	4	707.46	20	VV IN PP SENT	21	636.60	21	VVP TO VV DT	88	369.48	21	VV JJ NN IN	93	386.55	21	VV RB JJ IN	52	425.33	17	VBP RB JJ IN	22	388.40
31	8	VVP PPS NN IN	4	707.46	20	VV DT NN IN	21	636.60	22	VV DT NN IN	84	352.68	22	VB VVN IN DT	91	378.23	22	VB DT JJ NN	49	400.79	18	VV TO VV IN	21	370.74
32	8	VVP DT NN SENT	4	707.46	21	VV RB PP VVP	20	606.28	23	VZ RB JJ IN	83	348.48	23	VV JJ NNS IN	90	374.08	23	VVN IN DT NNS	47	384.43	19	VVP IN DT NN	20	359.09
33	8	VVD PPS NN IN	4	707.46	21	VV PP IN DT	20	606.28	24	VZ RB JJ NN	81	340.09	23	VBP JJ TO VV	90	374.08	23	VVN IN DT JJ	47	384.43	19	VVN TO VV DT	20	359.09
34	8	VV PPS NN NN	4	707.46	21	VZ DT JJ NN	20	606.28	25	VVD DT JJ NN	80	335.89	23	VBD DT JJ NN	90	374.08	24	VV TO VV IN	46	376.25	19	VZ RB JJ TO	20	359.09
35	8	VV PP IN NN	4	707.46	22	VVP TO VV PPS	19	575.97	26	VVN IN DT JJ	78	327.49	24	VVD DT JJ NN	89	369.92	24	VV PP IN DT	46	376.25	19	VZ RB JJ TO VV	20	359.09
36	8	VV PP DT NN	4	707.46	22	VVP IN PPS NN	19	575.97	27	VV TO VV PP	77	323.29	24	VV TO VV PP	89	369.92	25	VV IN PPS NN SENT	45	368.07	20	VBD RB JJ CC	19	335.09
37	8	VV IN DT JJ	4	707.46	22	VBP VVG IN DT	19	575.97	28	VV DT NN NN	76	319.09	24	VV IN DT JJ	89	369.92	25	VV IN PPS NN	45	368.07	21	VVN DT JJ NN	18	317.78
38	8	VV DT NN	4	707.46	23	VBP RB JJ IN	18	545.65	29	VV IN DT NNS	74	310.70	25	VVD DT NN IN	85	353.30	26	VVP TO VV IN	44	359.89	21	VV PPS NN SENT	18	317.78
39	8	VZ RB JJ SENT	4	707.46	24	VVP RB VV TO	17	515.34	30	VVN IN CD NN	73	306.50	26	VV DT JJ NNS	80	332.51	26	VVG DT JJ NN	44	359.89	21	VV PP TO VV	18	317.78
40	8	VZ IN PPS NN	4	707.46	25	VVG TO VV IN	16	485.02	31	VV IN NN SENT	72	302.30	27	VV PP IN PPS	79	328.36	26	VV JJ NNS IN	44	359.89	21	VV DT NN	18	317.78
41	8	VZ DT JJ NN	4	707.46	25	VBP RB JJ SENT	16	485.02	31	VZ DT NN IN	72	302.30	27	VV IN DT NNS	79	328.36	26	VV DT NN CC	44	359.89	21	VBD JJ TO VV	18	317.78
42	9	VVP TO VV PP	3	530.60	26	VVG TO VV PP	15	454.71	32	VVZ DT NN IN	70	293.90	28	VVP TO VV DT	78	324.20	27	VV DT JJ NNS	43	351.71	22	VVP DT NN IN	17	300.13
43	9	VVP TO VV NN	3	530.60	26	VVG IN PPS NN	15	454.71	32	VVN IN PPS NN	70	293.90	28	VZ DT NN IN	78	324.20	27	VV DT JJ NNS	43	351.71	22	VVP DT NN IN	17	300.13
44	9	VVP TO VV JJ	3	530.60	26	VVD DT JJ NN	15	454.71	32	VZ DT JJ NN	70	293.90	29	VVP DT JJ NN	77	320.04	28	VVP TO VV DT	42	343.53	22	VV PP IN DT	17	300.13
45	9	VVP RB VH DT	3	530.60	26	VV PPS JJ NN	15	454.71	33	VVP DT JJ NN	69	289.70	29	VZ RB RB JJ SENT	77	320.04	28	VVD DT JJ NN	42	343.53	22	VBD DT NN IN	17	300.13
46	9	VVP PP MD VV	3	530.60	26	VZ RB RB JJ CC	15	454.71	33	VVG TO VV IN	69	289.70	29	VB DT JJ NN	77	320.04	28	VV TO VV DT	42	343.53	22	VV DT NN IN	17	300.13
47	9	VVP JJ NN SENT	3	530.60	27	VVD DT NN IN	14	424.40	33	VV IN PP SENT	69	289.70	30	VV IN JJ NNS	76	315.89	28	VZ RB RB JJ TO	42	343.53	23	VV PPS JJ NN	16	282.47
48	9	VVP IN DT JJ	3	530.60	27	VV PP DT NN	14	424.40	33	VV IN DT JJ	69	289.70	31	VV PPS JJ NN	74	307.57	29	VBP RB JJ IN	41	335.35	23	VV IN DT NNS	16	282.47

**Figure 3.10** Process for identifying core and emerging sequences

Once the “core” and “emerging” sequences were identified, the next step was to look at which node verbs in the selected sequences were most frequent across the proficiency levels among L1 Korean speakers. To do so, each of the sequences was individually searched again in the CLC. For instance, the core sequence “VV DT NN IN” (verb-determiner-noun-preposition) was searched, but the corpus was filtered for A2 levels only in the L1 Korean sub-corpus. Then, the data was downloaded onto an excel file, and the frequencies of the node verbs were counted. **Figure 3.11** illustrates how the frequencies of the top node verbs for the VV DT NN IN sequence were stored. This step was repeated for each of the core and emerging sequences in the L1 Korean speaker sub-corpus.

	A2		B1		B2		C1		C2						
	62		394		425		203		95						
3	give	6	9.7%	make	31	7.9%	get	26	6.1%	make	12	5.9%	take	9	9.5%
4	make	6	9.7%	get	21	5.3%	make	24	5.6%	take	10	4.9%	see	6	6.3%
5	see	6	9.7%	see	21	5.3%	take	21	4.9%	learn	9	4.4%	get	4	4.2%
5	write	5	8.1%	give	18	4.6%	give	17	4.0%	find	8	3.9%	find	3	3.2%
7	buy	4	6.5%	take	14	3.6%	learn	15	3.5%	provide	6	3.0%	enjoy	2	2.1%
8	attend	3	4.8%	use	13	3.3%	see	15	3.5%	get	5	2.5%	experience	2	2.1%
9	find	3	4.8%	change	12	3.0%	find	12	2.8%	give	5	2.5%	give	2	2.1%
0	take	3	4.8%	learn	11	2.8%	enjoy	10	2.4%	improve	5	2.5%	improve	2	2.1%
1	change	2	3.2%	spend	11	2.8%	buy	8	1.9%	see	5	2.5%	save	2	2.1%
2	get	2	3.2%	reduce	10	2.5%	keep	7	1.6%	read	4	2.0%	turn	2	2.1%
3	read	2	3.2%	find	9	2.3%	use	7	1.6%	ask	3	1.5%	achieve	1	1.1%
4	wear	2	3.2%	know	9	2.3%	change	6	1.4%	increase	3	1.5%	assume	1	1.1%
5	access	1	1.6%	buy	7	1.8%	experience	6	1.4%	recommend	3	1.5%	build	1	1.1%
6	agree	1	1.6%	keep	7	1.8%	increase	6	1.4%	reduce	3	1.5%	captivate	1	1.1%
7	arrange	1	1.6%	solve	7	1.8%	know	6	1.4%	share	3	1.5%	cause	1	1.1%
8	build	1	1.6%	do	6	1.5%	pay	6	1.4%	spend	3	1.5%	consider	1	1.1%
9	complete	1	1.6%	attend	5	1.3%	reduce	6	1.4%	tell	3	1.5%	create	1	1.1%
0	do	1	1.6%	provide	5	1.3%	write	5	1.2%	attend	2	1.0%	describe	1	1.1%
1	improve	1	1.6%	show	5	1.3%	feel	4	0.9%	become	2	1.0%	discuss	1	1.1%
2	increase	1	1.6%	discuss	4	1.0%	follow	4	0.9%	build	2	1.0%	double	1	1.1%
3	keep	1	1.6%	enjoy	4	1.0%	improve	4	0.9%	cause	2	1.0%	drive	1	1.1%
4	lesten	1	1.6%	pay	4	1.0%	offer	4	0.9%	cover	2	1.0%	earn	1	1.1%
5	live	1	1.6%	save	4	1.0%	save	4	0.9%	discuss	2	1.0%	endure	1	1.1%
6	put	1	1.6%	send	4	1.0%	spend	4	0.9%	earn	2	1.0%	enhance	1	1.1%
7	send	1	1.6%	write	4	1.0%	understand	4	0.9%	enhance	2	1.0%	enter	1	1.1%
8	stay	1	1.6%	bring	3	0.8%	attend	3	0.7%	experience	2	1.0%	execute	1	1.1%
9	suggest	1	1.6%	leave	3	0.8%	avoid	3	0.7%	feel	2	1.0%	explain	1	1.1%
0	use	1	1.6%	obtain	3	0.8%	do	3	0.7%	forget	2	1.0%	express	1	1.1%
1	watch	1	1.6%	play	3	0.8%	earn	3	0.7%	hear	2	1.0%	face	1	1.1%
2	work	1	1.6%	put	3	0.8%	examine	3	0.7%	keep	2	1.0%	feel	1	1.1%

Figure 3.11 Node verb ranking of the VV DT NN IN sequence

Following the identification of the node verbs, the next step was to draft examples from each of the designated verb 4-gram sequences. For this part of data extraction, each of the proficiency levels was consulted separately, and the same procedure was repeated six times. Careful planning and data labeling took place to prevent data mixing. Taking the core sequence VV DT NN IN from the B2 level as an example (see Figure 3.12), a random selection of fifty examples was downloaded and stored in Microsoft Excel.

Query **V, B2 VANTAGE, Korean, Korea, South 11,834** > Positive filter **VV, DT, NN, IN 402** (6.30 per million) ⓘ

Page 1 of 21 Go Next Last

doc#1142	Guidebook when I visited Lampton. So I want to <b>correct the information about</b> Lampton and give y
doc#1218	before you can easily notice. </p><p> I would like to <b>see the change in</b> an advantage and a
doc#1584	you'd never seen in your country. Also, you can <b>take a photograph of</b> the local people. You
doc#1584	of English and Italian. I'm about to <b>learn another language as</b> well. </p><p> The
doc#2012	investigate about some animal but if you can't <b>find those animal around</b> you, you can go to a
doc#2184	machine for visitors. Finally it is required to <b>increase the number of</b> receptionists who ar
doc#2291	20:15. When the show finished, it was too late to <b>catch the train to</b> home. I had to take
doc#2789	your advertisement. If I am employed, I would <b>share the house with</b> other camp participa
doc#3139	. The market opens every Wednesday and you can <b>find a lot of</b> antiques, very intere
doc#3273	them ignore their bodies. </p><p> First of all, we would <b>see a lot of</b> adolescents in a pari
doc#3284	, the college principal has asked me to <b>write a report on</b> what studen's compl
doc#3284	say, the development of transportation could <b>make the earth into</b> one-day-living zone.
doc#3595	again because of darkness. </p><p> Dear Sir, I've just <b>read a report on</b> Monday's newspaper
doc#3595	some error in your review. So I'm writing to <b>inform the truth on</b> the festival and give
doc#3831	now. </p><p> 9th Dec.1997 To. the editor of the Times To <b>see the report in</b> your newspaper on t
doc#4317	in a bookshop? I think it's really good idea to <b>get a job in</b> the bookshop. </p>
doc#4317	. It is easy to correct letters and to <b>save a lot of</b> information. We just
doc#4481	you lose your money". </p><p> Dear Helen. </p><p> It is so nice to <b>hear the news from</b> you that I have won
doc#4481	outdoor clothes, and also if it is necessary to <b>bring some money for</b> me. </p><p> Thanl
doc#5031	of the student tourist guidebook </p><p> I am writing to <b>give some information about</b> Lampton. </p><p>

Figure 3.12 Example from the verb 4-gram sequence, “VV DT NN IN”

Again, this procedure was repeated for 1) each core and emerging sequences, 2) proficiency levels, and 3) the L1 Korean speaker sub-corpus and other L1 speaker sub-corpus.

### 3.4.2 Identifying Grammar Patterns

Once the core and emerging verb 4-gram sequences were identified, the next step was to conduct in-depth analyses with the selected verb 4-gram sequences and the 50 random examples drafted from each of them. The process in its entirety took place by looking into the grammar pattern in Collins Dictionary (*grammar.collinsdictionary.com*). The website contains a comprehensive listing of all the grammar patterns that are used in English in adjectives, nouns, and verbs.

Due to the immensity of the data, only the top two core and emerging sequences were consulted from each proficiency level. The starting point for identifying the grammar patterns was to first pull out the 50 verbs from each of the two core and emerging sequences. To reiterate, this step was repeated for each proficiency level as well as for each sub-corpus (L1 Korean speakers and other L1 speakers). To ensure that the classifications are as accurate as possible, the concordance lines were kept with the verb in the case that meaning needed to be consulted within the context to gain a full understanding of how the word was used. For instance, the word ‘give’ in ‘give information about’ would belong in the ‘describe’ group where the speaker is providing more information about something. On the other hand ‘give some present for’ would belong in the ‘do’ and ‘take’ group where the subject is providing a gift to someone else. By looking at the word ‘give’ alone, no such speculation could have been made since there is not enough context to indicate how the verb was used. Thus, it was crucial that all concordance lines were kept in order to provide the context in which the verb was used.

Once the verbs were identified, the next step was to search for the words in the dictionary. Taking the verb ‘arrange’ for instance, the result would provide as follows (see **Figure 3.13**).

## Definition of 'arrange'

# arrange

Collins COBUILD

Word Frequency



(əˈrɛɪndʒ )

Word forms: 3rd person singular present tense **arranges** , present participle **arranging** , past tense, past participle **arranged**

### 1. VERB

If you **arrange** an event or meeting, you make plans for it to happen.

*She arranged an appointment for Friday afternoon at four-fifteen.* [*VERB noun*]

*This time it was a friend ringing to try to arrange a fishing trip in Scotland.* [*VERB noun*]



*The prime minister threw the carefully arranged welcome into chaos.* [*VERB-ed*]

Synonyms: plan, agree, prepare, determine [More Synonyms of arrange](#)

### 2. VERB

If you **arrange** with someone **to** do something, you make plans with them to do it.

*I've arranged to see him on Friday morning.* [*VERB to-infinitive*]

*It was arranged that the party would gather for lunch in the Royal Garden Hotel.* [*be*

*VERB-ed that*]

*He had arranged for the boxes to be stored until they could be collected.* [*VERB + for*]

[*Also VERB that*]

### 3. VERB

If you **arrange** something **for** someone, you make it possible for them to have it or to do it.

*I will arrange for someone to take you round.* [*VERB + for*]

*The hotel manager will arrange for a baby-sitter.* [*VERB + for*]

*I've arranged your hotels for you.* [*VERB noun*]

*Transport is not included but can be arranged.* [*be VERB-ed*]

**Figure 3.13** Example of the verb “arrange” in Collins Dictionary

It should be noted that the starting point for identifying the grammar patterns did not take place by immediately consulting the list of patterns and meaning groups (see **Figures 3.4, Figure 3.5, Figure 3.6**). The main reason for this was that as the grammar pattern only contained the top 400 most frequent verbs in the Collins WordBanks corpus, therefore, many infrequent verbs (e.g., verbs outside of the top 400 most frequent) were not listed. This could be problematic because many verbs could potentially be excluded not necessarily because they are not a grammar pattern but simply because they were less frequent. Furthermore, since there was a wide range of sections for each verb, clicking through the sections based on subjective judgment could not only be inaccurate but take a long time. To eliminate the possibility of an inaccurate

classification process, the identification of the pattern took place in the Collins Dictionary rather than Collins Grammar Patterns.

In **Figure 3.13**, three definitions for the word ‘arrange’<sup>4</sup> are presented. Followed by each of the example sentences are the grammar patterns associated with the verb ‘arrange’ in the sentence. Once the corresponding definition of the word was located, then each of the example sentences was consulted to find the usage of the word as used in the corpus. The next step was to identify the verb on the grammar pattern page. For example, in the sentence “I’ve arranged your hotel for you.”, the **V n** pattern was consulted to find its respective meaning group. In the **V n** page, the first step was to identify its structure (see **Figure 3.6**).

Under the structure are different meaning groups that would have to be identified based on the researcher’s intuition. To find the meaning group that the word ‘arrange’ most closely corresponds to, guesses had to be made based on the provided description. In the example above, the verb ‘arrange’ would belong in the ‘arrange’ meaning group where the description provided is as follows:

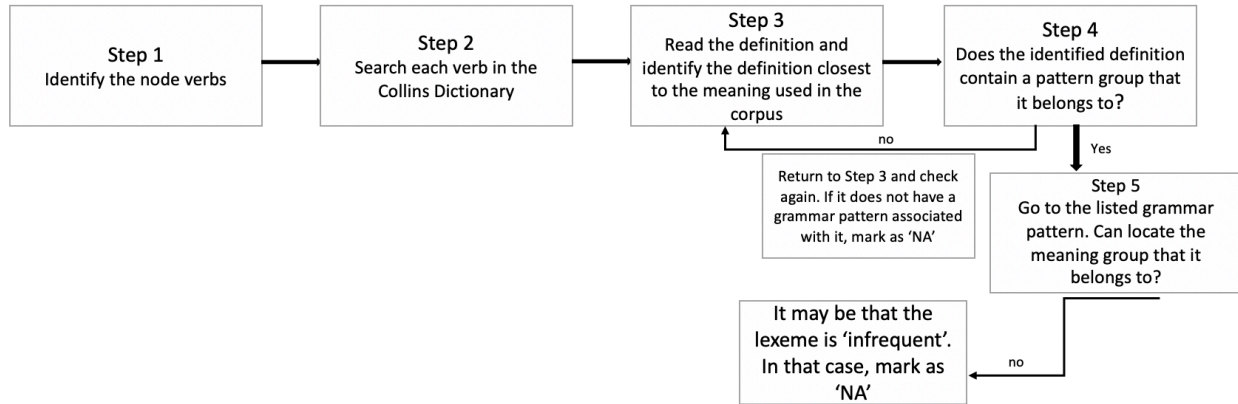
“These verbs are concerned with making different not the object itself but the circumstances surrounding the object. This includes:

- arranging for someone to take up a job or a position, or to leave a job or a position (e.g., admit, elect, fire)
- providing a home, money, or education for something or someone (e.g., house, keep)
  - Making arrangements about something such as the time of an event (e.g. fix [a date], move [an event])

This procedure took place for all node verbs in the **VV DT NN IN** and **VVD DT NN IN** sequences were followed by the node verbs are determiners and a noun phrase (See **Figure 3.14**).

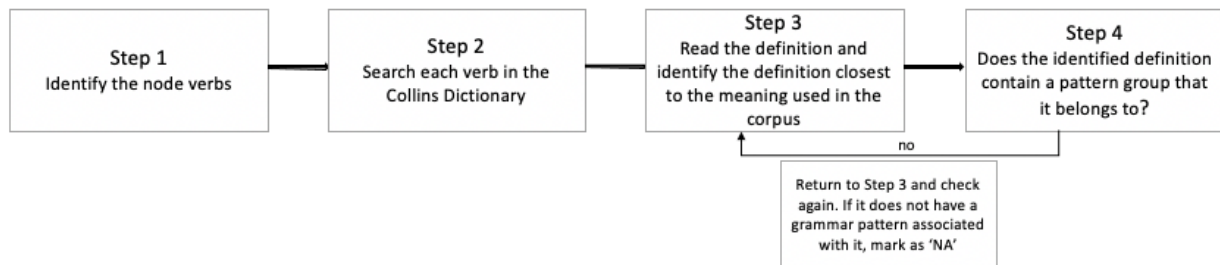
---

<sup>4</sup> The screenshot only includes the first three definitions of the word ‘arrange’



**Figure 3.14** Procedure for **V n** identification

In the case of the **VV IN DT NN** and **VVN IN DT JJ** sequences, the node verbs are followed by a preposition. The same procedure that took place for these sequences is the ‘verb + noun phrase’ sequences; however, the examples were classified by grammar patterns only, and meaning groups were not classified due to the wide range of prepositions prevalent in these groups (See **Figure 3.15**).



**Figure 3.15** Procedure for **V preposition** identification

### 3.4.3 Comparing L1 Korean Speakers with Other L1 speakers and L1 English Speakers

Part three of the experiment consisted of comparing the L1 Korean speakers’ use of verb 4-gram sequences with two groups: other L1 speakers and L1 English speakers. As mentioned, the purpose of the comparison of other L1 speakers and L1 English speakers was to understand whether the core and emerging sequences identified among L1 Korean speakers mirror that of other L1 speakers as well as L1 English speakers. Additionally, it sought to explore whether the patterns of language use in the core and emerging sequences are specific to L1 Korean speakers

or similar across other L1 speakers as well as L1 English speakers. That being said, the following two sequences were chosen for detailed analyses. Core sequence:

**VV DT NN IN**

Emerging sequence:

**VVN IN DT JJ**

Due to space constraints, it was not feasible to conduct an in-depth analysis of all of the core and emerging sequences. Despite this, the top core and emerging sequences in the L1 Korean sub-corpus was also identified to be most prominent in both the other L1 speaker and L1 English speakers, thereby being considered sufficient.

#### **3.4.3.1 Comparing with Other L1 Speaker Sub-Corpus**

As mentioned above, the other L1 speaker sub-corpus was constructed based on the CLC. Therefore, in order to ensure that the data mirrors that of the L1 Korean speaker sub-corpus, the grammar patterns and meaning groups were identified for the core sequence **VV DT NN IN** and grammar patterns for the emerging sequence **VVN IN DT JJ** among A2 to C2 level groups. The procedure was identical to that of the L1 Korean speaker sub-corpus (see **Figure 3.14** and **Figure 3.15**).

#### **3.4.3.2 Comparing with L1 English Sub-Corpus**

The comparison with the English sub-corpus only took into consideration the highest level, the C2 level. Since the purpose of the comparison is to understand whether L1 Korean speakers' use of verbs mirrored that of L1 Korean speakers at the highest proficiency level, no comparison was made for other levels. Furthermore, the top 25 verb 4-gram sequences are compared among L1 Korean speakers as well as other L1 speakers to show the comparison of the ranking of core and emerging sequences. No further comparisons are made across all three groups with regard to grammar patterns. The procedure for classifying grammar patterns and meaning groups in the L1 English sub-corpus was identical to that of L1 Korean speakers (see **Figure 3.14** and **Figure 3.15**).

### 3.6 Reliability

The subjective nature of this study required a strict protocol to control for the possibility of bias or human error, thereby leading to a faulty result. In particular, the categorization and interpretation of the verb patterns and their respective meaning groups are not entirely straightforward and, therefore, are open to debate (Hunston & Francis, 2000). To strengthen the claims made about the categorization of the patterns and to ensure that they are consistent in terms of how each of the verb sequences is assigned to a meaning group, much work was dedicated to formulating unique criteria for this study based on the examples provided in Collins Dictionary.

Firstly, when selecting random 50 verb 4-gram sequences from the chosen core and emerging sequences, it was made sure that the sequences were correctly categorized according to their tagging. For instance, “a lot of,” which should be tagged as pronoun phrases, were tagged individually in the CLC. Hence, the phrase “a lot of” was tagged as article-noun-preposition, which is incorrect. Once the data was filtered to ensure that the examples corresponded to their tagging, the next step was to store the data. In a few weeks, the same procedure was replicated by the researcher to ensure that the verb sequences were accurately recorded.

With respect to the categorization of the examples from the verb sequences and their meaning groups, the samples were first roughly divided into different meaning groups. The verbs where the categorization was straightforward and did not require further discussion were divided first. Then, in case of the examples that could belong in multiple meaning groups were drafted and set aside. These examples were categorized according to the information provided in Collins Dictionary. It should be noted that a second coder was not involved during this process. This was due to practical reasons such as lack of funds and finding a suitable coder. Instead, intra-coder reliability was achieved by initially categorizing the examples according to grammar patterns by the researcher and then re-visiting the data after one month. At this point, the researcher re-categorized the entire data again. Any discrepancies were consulted with the supervisor, and changes to the coding were implemented after discussion. Although studies call for multiple coders in order to establish reliability, within-coder reliability is also used when attaining a second coder is not feasible.

### 3.7 Data Analysis

This section presents the various methods used for analyzing the findings of this study. Firstly, the analysis of the node verbs and the core and emerging sequences are elaborated. Then, the analysis of grammar patterns is provided.

#### 3.7.1 Analysis of Node Verbs

The data analyses for node verbs consisted of calculating normalized frequencies and percentages across different proficiency levels in the L1 Korean sub-corpus. Once the core and emerging sequences were identified, the next step was to attain the raw frequencies of the node verbs. As a reminder, node verbs are the main verb in the verb 4-gram sequences. It is also the first word of the sequence. The raw frequencies were attained by calculating the number of instances of the verbs in the L1 Korean sub-corpus.

For instance, after the node verbs for the VV IN DT NN sequence in the A2 level were extracted from the L1 Korean speaker sub-corpus, the total number of words in the VV IN DT NN sequence was calculated. To clarify, these are the total number of words in the VV IN DT NN sequence in the A2 level from the L1 Korean speaker sub-corpus. If there are 30 examples in the VV IN DT NN sequence, each sequence containing four words, then the total number of words would be 120.

$$30 \text{ examples} \times 4 \text{ words per each sequence} = 120 \text{ words}$$

To calculate the percentage of the node verbs in each sequence, the raw frequency was divided by the total number of words. For example, if the word “go” appeared 20 times and the total number of words in that particular verb 4-gram sequence was 120, then the percentage would be 16.7%.

$$\frac{20 \text{ raw frequency}}{120 \text{ total \# of words}} = 0.1666667 = 16.7\%$$

Again, this calculation was conducted separately for each of the node verbs in the core and emerging sequences in the respective proficiency levels.

### 3.7.2 Analysis of the Core and Emerging Sequences

In case of the verb 4-gram sequences, the same data analysis took place for the L1 Korean sub-corpus, the other L1 speaker sub-corpus, and the L1 English sub-corpus. In each of the sub-corpus, the verb 4-gram sequences had to be normalized in order to make the comparisons between the proficiency levels and across the different sub-corpora possible. Thus, the following steps were taken to normalize the raw frequencies.

From each frequency level, the top 100 verb 4-gram sequences were downloaded, along with the raw frequency for each. In order to calculate the normalized frequency of each sequence, the raw frequency was divided by the total number of words.

$$\frac{\text{raw frequency}}{\text{total number of words}} \times 1,000,000$$

For example, if the VV IN DT NN sequence appeared 2,000 times in a corpus consisting of 300,000 words, then it would result in a normalized frequency of 6,666.66.

$$\frac{2,000 \text{ raw frequency}}{300,000 \text{ total number of words}} \times 1,000,000 = 6,666.666667 \text{ norm. frequency}$$

### 3.7.3 Analysis of the Grammar Patterns

The grammar patterns did not need to be normalized because a random sample of fifty was extracted from each of the verb 4-gram sequences. Since the sample sizes were identical across the levels, the number of instances was tallied by hand. All comparisons were made through the means of raw frequencies.

The bulk of the data analysis of the grammar patterns consisted of analyzing the examples into the categories of pattern grammar qualitatively. To do so, the concordance lines of the node verbs and the following three words were extracted from the corpus. Then, the entire sentence was consulted to appropriately categorize the sequences according to the descriptions of the different verb patterns. If any of the words were mistagged due to the wrong usage, then they were omitted. Again, this procedure took place for the selected core and emerging sequences across all proficiency levels, as well as the sub-corpora.

### 3.8 Potential Methodological Issues with the Study Design

There are several potential methodological issues with the study design which should be addressed. First and foremost, “development” in the current study is not measured purely longitudinally. Rather, the study adopts a quasi-longitudinal research design, where the proficiency levels are used to make assertions about the changes in learner language as users become more proficient in the L2. While a pure longitudinal research design would have been ideal, the practicality of collecting a large amount of data over a long period of time is difficult and not feasible. To compensate, studies have used quasi-longitudinal methods (e.g., Meunier, 2015), where proficiency is used as a proxy to benchmark development. While this method does not trace learners long-term, it can be used to make claims about language patterns at different learning trajectories.

Furthermore, the number of participants in both sub-corpora in the CLC is inconsistent throughout the proficiency levels. In the case of the L1 Korean sub-corpus, most learners fell within the range of the B1 to C1 levels (e.g., B1, B2, C1), with fewer participants in the C2, A1, and A2 levels. Thus, in order to make possible the comparison among both within (e.g., proficiency levels) and between (e.g., different sub-corpora), all frequencies have been adjusted to a common base. Taking into consideration the number of words in the CLC as well as following the standard procedure in the literature (McEnery & Hardie, 2012), the frequencies are normalized to 1 million words. Secondly, the participants’ age gap across the entire sub-corpus of L1 Korean speakers, as well as other L1 speakers, ranged from teens to 50+. While arguments can be made about the unevenness of data with regards to age, the age factor was not taken into account because the Common European Framework Reference of Languages (CEFR) provides a uniform and reliable benchmark of learners’ language ability regardless of age.

Another issue to consider is that the current study did not take into consideration the various factors such as topic (Yoon, 2017) and genre (Bi, 2020). Furthermore, some exams in the CLC were dedicated specifically to certain proficiency levels (e.g., CELSV for B2 level and CELSP for C1 level). Thus, the data could have been skewed due to the differences in the nature of these exams. However, as shown in **Table 3.16**, a wide array of styles and formats were present across all proficiency levels in the L1 Korean sub-corpus. It should also be noted that in the A1 and A2 levels, both the styles and formats were limited due to significantly smaller data compared to the other proficiency levels. A close scrutinization of these figures indicates that the

distribution is not entirely identical across the groups. However, focusing on specific formats and styles is not ideal since much of the data would be lost in the process. Despite these discrepancies, the CLC is a representative of the writing tasks in major mainstream English as a Second Language (ESL) and English as a Foreign Language (EFL) settings. That being said, the sub-corpora does not discriminate against any particular styles or format.

<b>Style</b>	<b>A1</b>	<b>A2</b>	<b>B1</b>	<b>B2</b>	<b>C1</b>	<b>C2</b>
Advice	-	13.19%	2.38%	4.18%	5.18%	5.68%
Application/ response	-	-	3.89%	3.94%	1.70%	0.87%
Argumentative/ opinion	-	20.73%	27.15%	30.78%	41.05%	47.11%
Business	-	15.18%	3.71%	1.47%	2.08%	0.45%
Complaint/ apology/ response	1.57%	1.07%	9.38%	16.29%	18.03%	5.68%
Critical	-	-	0.28%	0.59%	1.57%	2.96%
Descriptive/ creative autobiographical	-	19.40%	4.98%	7.01%	19.45%	48.64%
Informative/ news	98.43%	56.29%	22.08%	18.44%	15.25%	7.62%
N/A	-	-	32.28%	26.13%	11.11%	3.54%
<b>Format</b>	<b>A1</b>	<b>A2</b>	<b>B1</b>	<b>B2</b>	<b>C1</b>	<b>C2</b>
Article	-	-	1.74%	5.53%	8.72%	3.36%
Composition/ essay	-	-	21.69%	20.22%	24.61%	33.96%
Informative/ instructional text	16.41%	2.41%	4.56%	5.36%	8.38%	16.93%
Letter/ reference	-	55.38%	20.95%	23.73%	27.51%	14.34%
Note/ email/ memo	83.59%	30.69%	2.10%	0.53%	0.44%	-
Proposal	-	-	0.38%	-	1.68%	1.28%
Report	-	-	14.21%	16.22%	11.11%	3.92%
Review	-	-	0.35%	0.81%	1.14%	
Story	-	10.62%	1.73%	2.59%	6.65%	23.26%
Survey/ questionnaire/ form	-	-	-	0.06%	-	-
N/A	-	-	32.28%	26.13%	11.11%	3.54%

**Table 3.16** Percentage distribution of format across the six proficiency levels in the L1 Korean sub-corpus as listed in the CLC

Next, the comparison of the data from the CLC and the BNC could pose methodological concerns (Ädel, 2015; Gilquin & Granger, 2015) with regards to whether it is feasible to make a comparison between two corpora when they were compiled in drastic different settings.

However, the practicality of compiling a native-speaker corpus that is a perfect mirror of the learner corpus is not easy. Thus, the rationale for using the BNC in this study is under two main considerations: 1) the BNC has been used in a myriad of studies as a proxy for L1 English written language owing to its immense data, which could illustrate the most reliable representation of language use 2) A sub-corpus was constructed on the BNC, selecting only the academic portion in order to mirror CLC as much as possible. Therefore, the BNC was selected as a reference corpus for this study.

Lastly, the study adopts the term “tendency to Zipfian distribution” as a way of illustrating that the frequency of the verb 4-gram sequences is inversely proportional to its ranking. Zipfian distribution, also known as Zipf’s law, is a term that was initially coined by Zipf (1935). It emphasizes that the frequencies of words are inversely proportional to their ranking. Simply put, few highly frequent words account for the majority of the tokens in any given text. The majority of the words in the text are then made up of low-frequency words. The Zipfian distribution can be calculated mathematically, but it was not feasible in the case of this study due to data restrictions. Despite this limitation, the normalized frequency distribution indicates that the top verb 4-gram sequence occurred twice as frequently as the second highest ranked verb 4-gram sequence. Aside from the few that ranked among the top, there were many sequences that only appeared once or twice. While the mathematical calculation is necessary to claim that the distribution of the sequences is “Zipfian,” the current study utilized the term “tendency to” to illustrate such a shortcoming.

### **3.9 Summary**

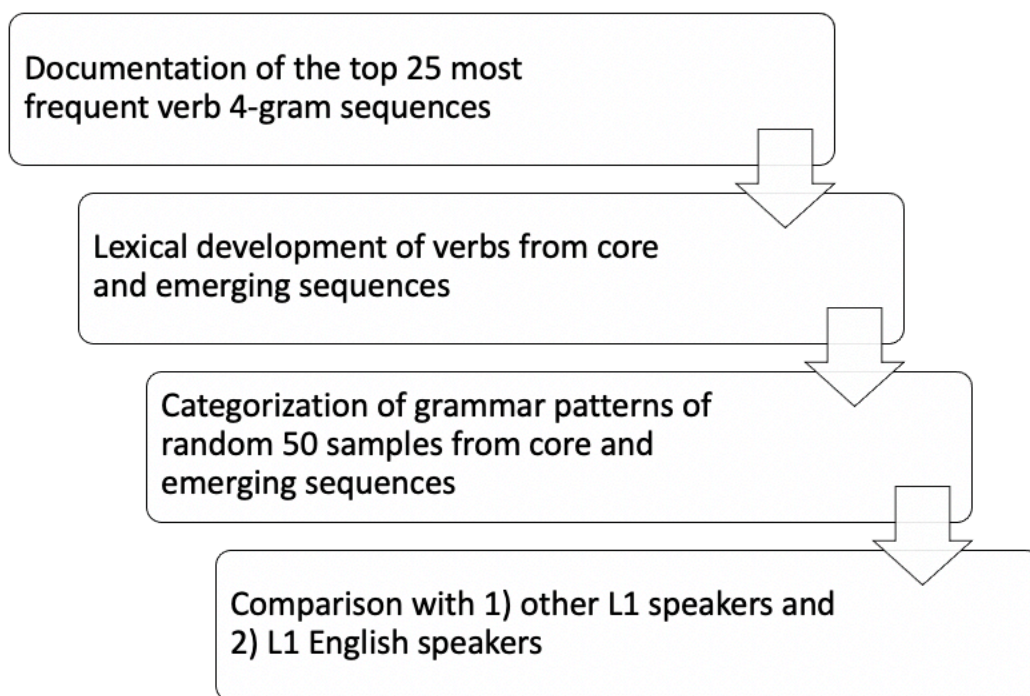
This chapter summarized the methodology that was implemented to answer the aforementioned research questions. To do so, the chapter began with research questions. Following the research questions, the methodology for investigating the sub-corpora was presented. First, the study design was introduced, followed by ethical consideration. Then, the instruments implemented in this study were elaborated. It started with a broad discussion of the CLC in terms of its data source as well as metadata. The two sub-corpora from the CLC were introduced. Next, the BNC was presented as well as the Collins WordBanks Online Corpus and the Collins English Dictionary.

Following the elaboration of the instruments, the procedures were introduced. The identification of verb 4-gram sequences as well as the node verbs was elaborated in great detail. Then, a lengthy description of identifying the grammar patterns was presented. The last section discussed the comparison of L1 Korean speakers with other L1 speakers and L1 English speakers. Next, the steps for attaining reliability were elaborated. Lastly, the methods for analyzing the findings and potential methodological issues and limitations with the study design were discussed.

## Chapter 4 Results

### 4.1 Introduction

This chapter presents the findings derived from the three sub-corpora mentioned in the methodology section: BNC academic, L1 Korean speakers in the CLC, and other L1 speakers in the CLC. **Figure 4.1** outlines the structure of this section.



**Figure 4.1** Outline of the results section

By the end of this chapter, readers will have a better understanding of the salient verb patterns in the various stages of L2 proficiency among L1 Korean speakers. Subsequently, these patterns will reveal the processes of language development and the ways in which language restructuring occurs as learners gain more experience through meaningful interaction and thus, become more proficient in the L2. Finally, the comparison with different cohorts (e.g., other L1 speakers and L1 English speakers) will reveal whether the identified patterns are specific to L1 Korean speakers or similar across all L1 and L2 English users.

This section first documents the top 25 most frequently used 4-gram verb sequences among L1 Korean speakers from the CLC. Based on the frequency distribution of the sequences,

two types of verb 4-gram sequences are identified: “core” and “emerging” sequences. As mentioned core sequences refer to 4-slot verb sequences that are consistently ranked in the top 10 across all proficiency levels from the B1 CEFR level (see 3.4.1). On the other hand, emerging sequences are verb sequences that ranked below the top 10 thresholds from the A2 to C1 levels but surfaced in the top 10 ranking in the C2 level. The distinction between these two types of sequences illustrates that some language patterns are consistent across all proficiency levels, while others become more prominent with proficiency level. This section aims to understand *how* language changes or remains the same as learners become more proficient L2 writers.

The second part of the chapter further probes into the lexical development of the core and emerging sequences. Taking apart the ‘node’ verbs, or main verbs, from each core and emerging sequence, the overused and underused verbs across the proficiency levels are compared based on their normalized frequency. Afterward, the usage of the node verbs is investigated to understand how learners in the higher proficiency level peruse more pragmatic or idiomatic expressions that go beyond the meaning of the lexis.

The following section transitions to an in-depth analysis of the random 50 examples selected from the core and emerging sequences. It first discusses the language patterns in the selected core and emerging 4-gram verb sequences. It does so by first categorizing the examples in accordance with the grammar patterns and their respective meaning groups. It then discusses the most and least common patterns in each proficiency level through means of adjusting frequency and distribution. Finally, it observes the changes in the use of grammar patterns and meaning groups across the proficiency levels.

The fourth section explores the extent to which L1 affects learners’ use of verb 4-gram sequences. To reiterate, the aim of this study was not to conduct a contrastive analysis of the structural similarities and differences between languages. Instead, it aimed to broadly identify the linguistic differences that surface among L1 Korean speakers compared to other L1 speakers in the CLC. Thus, the L1 speakers in the CLC consist of a wide range of L1s that may not share typological similarities.

The last section then discusses how the use of 4-gram verb sequences differs between L1 English speakers and L1 Korean speakers. Like the comparison with other L1 speakers, the

comparison with the academic portion of the BNC served a purpose as to whether near-native L2 speakers mimic L1 English speakers' use of English.

#### 4.2 Most Frequently Used Verb 4-gram Sequences Among L1 Korean Speakers

The first research question takes a bird's eye view by scrutinizing the most commonly used verb 4-gram sequences in each proficiency level in the L1 Korean speaker sub-corpus. The rankings of the sequences were established by normalizing raw frequencies per million words. **Table 4.1** displays the top 25 verb 4-gram sequences from each proficiency level in the CLC. Due to the insufficiency of learner data at the beginning level, however, A1 was excluded from all analyses, and A2 was included for the analysis of core sequences only.

Rank	A2	B1	B2	C1	C2
1	VBP VVG TO VV	VV DT NN IN	VV DT NN IN	VVN IN DT NN	VV DT NN IN
2	VV DT NN IN	VV IN DT NN	VVN IN DT NN	VV DT NN IN	VVN IN DT NN
3	VV IN DT NN	VVN IN DT NN	VV IN DT NN	VV IN DT NN	VV IN DT NN
4	VV PP IN PP\$	VV DT JJ NN	VV DT JJ NN	VV DT JJ NN	VV DT JJ NN
5	VV DT JJ NN	VVG IN DT NN	VVG IN DT NN	VVG IN DT NN	VVG IN DT NN
6	VBZ RB JJ SENT VV DT NN SENT	VBP VVG TO VV	VBZ DT JJ NN	VVN IN DT JJ	VBZ DT JJ NN VVN IN DT JJ
7	VV PP RB SENT	VV DT NN SENT	VV DT NN SENT	VBZ DT JJ NN	VVD DT NN IN
8	VV PP\$ NN SENT	VBZ DT JJ NN	VVP DT NN IN	VBZ RB JJ IN	VVN IN PP\$ NN
9	VVP TO VV IN VVP PP MD VV	VV IN PP\$ NN	VV PP\$ NN IN	VVD DT NN IN	VV PP\$ NN IN
10	VBZ RB JJ IN	VVP IN DT NN VV PP\$ NN IN	VBP VVG TO VV	VV DT NN SENT	VVN IN JJ NNS VVD DT JJ NN VBD DT JJ NN
11	VV PP SENT PP	VVP DT NN IN	VVP IN DT NN	VVD IN DT NN VV IN DT JJ	VVG DT JJ NN VVD IN DT NN VBZ DT NN IN
12	VVD IN DT NN	VV JJ NN IN	VVD IN DT NN	VVG DT NN IN	VVG DT NN IN
13	VVP TO VV DT	VV TO VV DT	VV TO VV DT	VB VVN IN DT VV PP\$ NN IN VBZ JJ TO VV VBP VVG TO VV VB JJ TO VV	VBD RB JJ IN VV IN DT JJ VB JJ TO VV VB DT JJ NN
14	VVG TO VV DT VV PP\$ NN IN VV IN PP\$ NN	VV PP\$ JJ NN	VVN IN JJ NN VBZ RB JJ IN	VVN IN PP\$ NN	VV DT NN SENT VBD VVN IN DT
15	VBP PP SENT PP	VBZ JJ TO VV	VVN IN DT JJ VV TO VV IN VBP RB JJ IN	VBD VVN IN DT	VV IN PP\$ NN VBD DT NN IN
16	VHP DT JJ NN	VVN IN JJ NN VVD DT NN IN	VBZ JJ TO VV	VV DT NNS IN	VBZ RB JJ IN VB VVN IN DT
17	VVP IN DT NN VVD PP IN PP\$	VVP TO VV IN	VV IN PP\$ NN VV DT NNS IN	VVP IN DT NN	VBP RB JJ IN

18	VVP TO VV PP VV TO VV DT	VVD IN DT NN VBZ RB JJ SENT	VB JJ TO VV	VVP DT NN IN	VV TO VV IN
19	VV TO VV PP VV IN NN IN	VVG DT NN IN VV TO VV IN	VVZ DT NN IN	VBD DT JJ NN	VVP IN DT NN VVN TO VV DT VBZ RB JJ TO VBZ JJ TO VV VBD RB JJ CC
20	VVG IN DT NN VV IN PP SENT VV DT NN NN	VV DT NNS IN VB JJ TO VV	VVG DT NN IN VV PP\$ NN SENT	VV PP\$ JJ NN VBZ DT NN IN	
21	VVP RB PP VVP VV PP IN DT VBZ DT JJ NN	VVP TO VV DT	VVP TO VV IN VVN IN JJ NNS VV JJ NN IN	VVN IN JJ NNS VBD RB JJ IN	VVN DT JJ NN VV PP\$ NN SENT VV PP TO VV VV DT NN , VBD JJ TO VV VVP DT NN IN VVP DT JJ NN VV PP IN DT VB DT NN IN VV TO VV DT VV PP\$ JJ NN VV IN DT NNS VV DT NN CC VBZ VVN IN DT VBP RB VVN IN VBD VVN TO VV VVN TO VV IN VVN IN NN SENT VVG IN PP\$ NN VVD TO VV IN VVD NN PP\$ NN VV DT JJ NNS VVN IN NN CC VVD PP\$ NN IN VVD IN DT JJ VV DT NNS IN VBP VVG TO VV
22	VVP TO VV PP\$ VVP IN PP\$ NN VBP VVG IN DT	VV IN NN IN	VB VVN IN DT	VB DT JJ NN	
23	VBP RB JJ IN	VBP RB JJ IN	VV JJ NNS IN VBP JJ TO VV VBD DT JJ NN	VVN IN DT NNS VVG IN DT JJ	
24	VVP RB VV TO	VBZ RB JJ NN	VVD DT JJ NN VV TO VV PP VV IN DT JJ	VV TO VV IN VV PP IN DT	
25	VVG TO VV IN VBP RB JJ SENT	VVD DT JJ NN	VVD DT NN IN	VV PP\$ NN SENT VV IN PP\$ NN	

**Table 4.1** Top 25 verb sequences for L1 Korean speakers

The verb sequences were tagged in accordance with the English TreeTagger POS Tagset with Sketch Engine modifications (see 3.3.3.1.1). Thus, taking the sequence VV NN IN DT as an example, this sequence would then be interpreted as verb-noun-preposition-determiner. One thing to note, however, is that the node verb is further distinguished by different tenses. The initial tag VV in the example stands for verbs in their base form (e.g., listen), the VVN tag indicates verbs in the past participle form and VVD for verbs in the past tense.

Based on the top 25 rankings of the verb 4-gram sequences, a few observations can be made. First and foremost, the frequency distribution of the sequences across all five proficiency levels shows a tendency for Zipfian distribution (Zipf, 1935) (see Chapter 3 for more information

regarding how the frequency distribution is Zipfian), where the frequency of the words is inversely proportional to their ranking. To briefly reiterate, the most frequently occurring sequence is likely to occur much more frequently than the following highest sequence in the ranking and even more frequently than the third highest sequence in the ranking. From **Table 4.1**, the most frequently occurring sequence in the B1 level, the VV DT NN IN, occurred 394 times (1654.25 per million words). The second most frequently occurring sequence, VV IN DT NN, occurred 258 times (1083 per million words). The third most frequently occurring sequence, VVN IN DT NN, appeared 196 times (822.93 per million words). Based on the normalized frequency per million words, the difference between the first, second, and third most frequently occurring sequences were significantly different.

Secondly, there is a pattern of developmental trajectory that surface with the increase in proficiency levels. Such a pattern is manifested through the core and emerging sequences identified in this study. In the case of the core sequences, the top five verb 4-gram sequences remained in the top across all levels from A2 to C2, although only three of the five verb 4-gram sequences made it within the top 5 in the A2 level. A further investigation into these sequences, categorized under the term “core” verb 4-gram sequences, will take place in the following section. Contrary to the core sequences, it is perhaps more interesting to note the “emerging” sequences where the sequences ranked low in lower proficiency levels (A2, B1) but ranked within the top 10 in the highest level (C2). The emerging sequences are to be further investigated in the following section, but it should be mentioned that three out of the five sequences in the core sequences are headed by verbs in the base form, whereas the node verbs in the emerging sequences were headed by either verb in the past tense or past participle form. The following two sections tap deeper into the core and emerging sequences to further investigate the developmental processes of the selected verb sequences.

#### **4.2.1 What are the “Core” Verb 4-gram Sequences?**

Based on the frequency and the ranking of the verb 4-gram sequences, the below five sequences were identified as core verb sequences from A2 to C2 levels.

1. **VV DT NN IN** (*verb-determiner-noun-preposition*)  
e.g., read a letter from

2. **VV IN DT NN** (*verb-preposition-determiner-noun*)  
e.g., go to the university
3. **VVN IN DT NN** (*verb [past participle]-preposition-determiner-noun*)  
e.g., retired from the position
4. **VV DT JJ NN** (*verb-determiner-adjective-noun*)  
e.g., learn a foreign language
5. **VVG IN DT NN** (*verb [gerund/present participle]-preposition-determiner-noun*)  
e.g., looking for a job

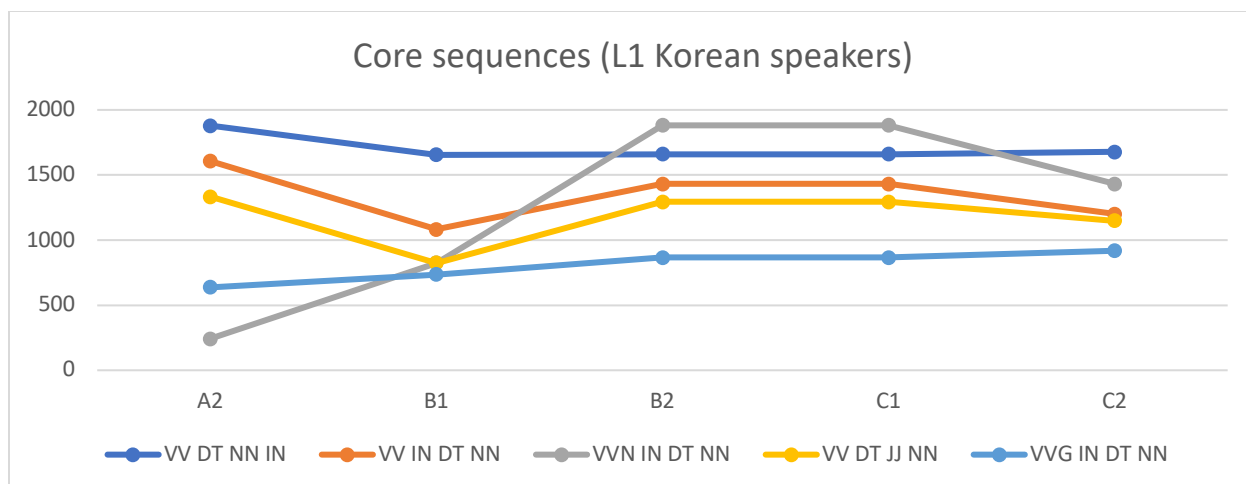
The identified core sequences reveal that three of the five sequences were headed by simple present verbs. The VVN IN DT NN and VVG IN DT NN sequences were the only two sequences where the node verb was either a past participle or present participle. **Table 4.2** displays the ranking and normalized frequency of each of the identified core sequences.

Sequences	A2	B1	B2	C1	C2
<b>VV DT NN IN</b>	2 (1879.47)*	1 (1654.25)	2 (1660.41)	2 (1660.41)	1 (1677.17)
<b>VV IN DT NN</b>	3 (1606.64)	2 (1083.24)	3 (1431.39)	3 (1431.39)	3 (1200.50)
<b>VVN IN DT NN</b>	33 (242.51)	3 (822.93)	1 (1881.25)	1 (1881.25)	2 (1430.01)
<b>VV DT JJ NN</b>	5 (1333.82)	3 (822.93)	4 (1292.34)	4 (1292.34)	4 (1147.54)
<b>VVG IN DT NN</b>	20 (636.60)	4 (734.76)	5 (867.01)	5 (867.01)	5 (918.03)

\**ranking (normalized frequency)*

**Table 4.2** Ranking and normalized frequency of “core” sequences (per million words)

All of the core sequences from B1 to C2 levels stayed within the top five in ranking, whereas two of the sequences in the A2 level were marked much lower in the ranking (20th and 33rd). The two sequences that did not make the top five thresholds in the A2 level were headed by past tense and present participle verbs. Thus, it could be that the learners in the A2 level have not solidified their knowledge beyond simple present verbs. It should also be noted, however, that the A2 data was much smaller in size and thus could have potentially caused the data to skew. Nonetheless, three of the five sequences, which all pertain to the regular verb form, were ranked high in the A2 level as well. **Figure 4.2** below illustrates the normalized frequency of the five core sequences in the L1 Korean sub-corpus.



**Figure 4.2** Distribution of normalized frequency of the core sequences from A2 to C2 level

The line graph displays the distribution of the core sequences across the five proficiency levels. As discussed earlier, the normalized frequency among the five core sequences is more dispersed at the A2 level. On the other hand, the normalized frequency of the core sequences is clustered closely together at the C2 level. Furthermore, the ranking stayed the same in the B2 and C1 levels and changed only slightly in the C2 level, where the VVN IN DT NN (e.g., lived in a city) sequence ranked second. Overall, **Figure 4.2** confirms that some sequences occur more frequently with the progression of proficiency level. Whether this is the case when compared with other speakers of L1 and L1 English speakers will be discussed in the later sections. Additionally, the following sections provide a deeper understanding of *how* language usage changes across the levels. This knowledge is crucial to discern the learners' language ability at each proficiency level and thus will be explored further later.

#### 4.2.2 What are the “Emerging” Verb 4-gram Sequences?

Based on the rankings of the frequency distribution, the following six emerging sequences were identified. The A2 level was omitted due to a lack of data where some raw frequency ranked below ten, and therefore, it was decided that the data was insufficient for further analyses. Thus, the analyses of emerging sequences ranged from B1 to C2 level only.

1. **VVN IN DT JJ** (*verb [past participle]-preposition-determiner-adjective*)  
e.g., compared to the other

2. **VVD DT NN IN** (*verb [past tense]-determiner-noun-preposition*)  
e.g., reached the age of
3. **VVN IN PP\$ NN** (*verb [past participle] -preposition-possessive pronoun-noun*)  
e.g., bought from your shop
4. **VVN IN JJ NNS** (*verb [past participle] -preposition-adjective-noun [plural]*)  
e.g., based on current conditions
5. **VVD DT JJ NN** (*verb [past tense]-determiner-adjective-noun*)  
e.g., ordered a beautiful shirt
6. **VBD DT JJ NN** (*verb [past tense]-determiner-adjective-noun*)  
e.g., was a great evening

Unlike the core sequences, the node verbs in the emerging sequences are headed by verbs in the past tense or past participle. The contrast between the core and emerging sequences manifests that learners tend to use verbs beyond the simple present as they become more proficient. **Table 4.3** illustrates the ranking and the normalized frequency of the six emerging sequences across the four proficiency levels.

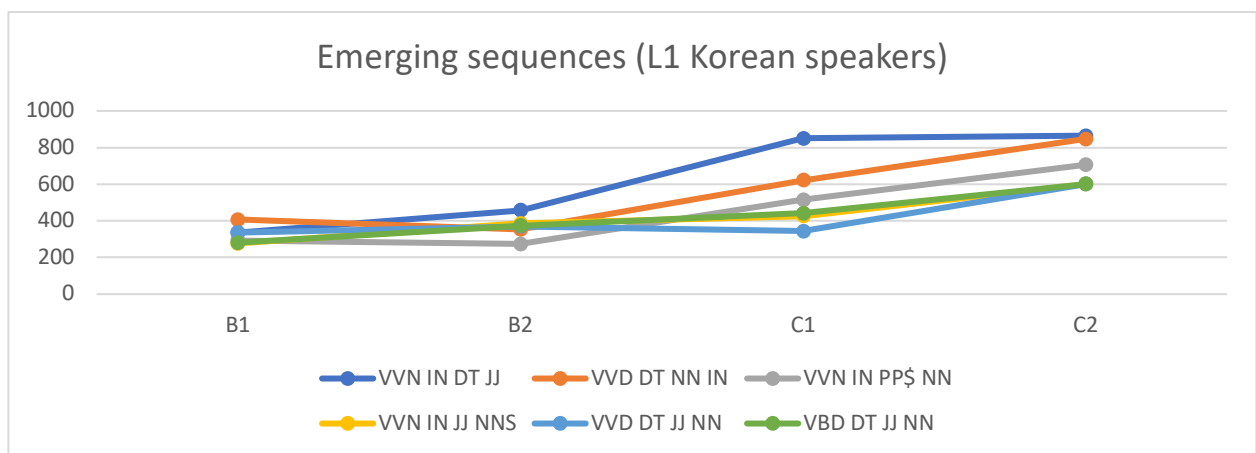
Sequences	B1	B2	C1	C2
<b>VVN IN DT JJ</b>	26 (335.89)*	15 (457.21)	6 (850.65)	6 (865.07)
<b>VVD DT NN IN</b>	16 (407.27)	25 (353.30)	9 (621.63)	7 (847.41)
<b>VVN IN PP\$ NN</b>	32 (293.90)	36 (274.32)	14 (515.30)	8 (706.18)
<b>VVN IN JJ NNS</b>	36 (277.11)	21 (386.55)	21 (425.33)	10 (600.25)
<b>VVD DT JJ NN</b>	25 (335.89)	24 (369.92)	28 (343.53)	10 (600.25)
<b>VBD DT JJ NN</b>	35 (281.31)	23 (374.08)	19 (441.69)	10 (600.25)

\**ranking (normalized frequency)*

**Table 4.3** Ranking and normalized frequency of “emerging” sequences

The table reveals that the emerging sequences ranked particularly low in the B1 and B2 levels compared to more advanced proficiency levels such as C1 and C2. For instance, only the VVD DT NN IN sequence in the B1 level ranked in the top 20 (e.g., located in the central), and the VVN IN PP\$ NN (e.g., paid for my ticket), the VVN IN JJ NNS (e.g., developed in many years), and the VBD DT JJ NN (e.g., was a lovely time) sequences all ranked below 30. In the B2 level, however, only the VVN IN PP\$ NN (e.g., practiced with her friend) sequence ranked below 30, and the rest of the sequences were either 25 or above. At the C1 level, there was a

drastic change where the VVN IN DT JJ (e.g., based on a real) and VVD DT NN IN (e.g., became a chairman of) sequences ranked above ten, and no sequences ranked below 28. Finally, at the C2 level, all six emerging sequences ranked in the top 10. The gradual increase of usage is also evident in the normalized frequency, which illustrates that there is a clear pattern of higher use of verb 4-gram sequences from B1 to C2 levels. Some sequences (e.g., VVN DT NN IN in B1 and B2) dropped in both ranking and frequency in the lower levels, but the pattern became more prominent with higher proficiency. **Figure 4.3** shows a visual representation of how the frequency changed across the four proficiency levels.



**Figure 4.3** Distribution of normalized frequency of the emerging sequences from B1 to C2 level

The normalized frequencies of the emerging sequences reveal that there is a noticeable pattern across the four proficiency levels. From B1 to C2 levels, all six verb 4-gram sequences were clustered together to around 200 to 400 times per million words and then gradually increased almost in unison. At the C1 level, however, the six emerging sequences were much more dispersed. For instance, the VVN IN DT JJ sequence occurred about 200 times per million words more than the second highest emerging sequence, VVD DT NN IN. However, the six sequences clustered again at the C2 level.

While the rankings of the verb sequences provided evidence of a changing trend in the use of verb sequences, a more granular analysis of the sequences is necessary to understand how the usage instances shift in conjunction with the proficiency levels. The following section takes a closer look at the ways in which the core and emerging sequences were used across proficiency levels.

### 4.3 Most Frequently Used Node Verbs for the ‘core’ and ‘emerging’ Verb 4-gram Sequences

This section explores the most frequently used node verbs in each proficiency level. In particular, it looks at the differences in their usage within context. **Table 4.4** shows the percentage of the most frequently used node verbs from A2 to C2 levels in the VV DT NN IN sequence.

VV DT NN IN	A2	B1	B2	C1	C2
<b>Give</b>	[1]* 9.7%**	[3] 4.6%	[4] 4.0%	[6] 2.5%	[5] 2.1%
<b>Make</b>	[1] 9.7%	[1] 7.9%	[2] 5.6%	[1] 5.9%	[6] 1.1%
<b>See</b>	[1] 9.7%	[2] 5.3%	[5] 3.5%	[6] 2.5%	[2] 6.3%
<b>Get</b>	[4] 3.2%	[2] 5.3%	[1] 6.1%	[6] 2.5%	[3] 4.2%
<b>Take</b>	[3] 4.8%	[4] 3.6%	[3] 4.9%	[2] 4.9%	[1] 9.5%

(\*: rank in the frequency; \*\*: percentage from the entire corpus)

**Table 4.4** Percentage of the top node verb from each proficiency level in the “VV DT NN IN” sequence

The percentage of the top-ranked node verbs reveals that 'make' ranked first across three of the five proficiency levels. Furthermore, 'give' appeared more frequently in the A2 and B1 levels, while 'take' occurred more frequently in higher proficiency levels (B2 to C2 levels). In the case of the verb 'get,' there were most instances in the intermediate CEFR levels, B1 and B2, but according to the concordance lines, there was a heavy repetition of "**get a job**" (e.g., "*I need to **get a job** before starting college.*", B2 level, FCE).

It is also evident that the relationship between the node verbs and the proficiency levels was not always linear. This was particularly the case for the node verbs 'see' and 'get. In the case of the verb 'see,' it ranked first in A2 and second in B1. In the B2 and C1 levels, however, the ranking dropped to fifth and sixth, respectively. Then in the C2 level, the rank jumped back up to second. There was even more fluctuation in the verb 'get' where although it was the most frequently used node verb in the B2 level, it ranked low in A2 and C1 levels but higher in B1 and C2 levels. The underuse and overuse of the node verbs alone cannot be used to make a conclusion about the learners' knowledge. However, it may be the case that the learners with

higher proficiency tend to use the 'delexical verb + noun' structure more frequently than lower proficiency levels. This will be further discussed later in the discussion section.

In order to further investigate the observations made from **Table 4.4**, the concordance lines were consulted. Taking 'make' as an example, learners in the A2 level tended to misuse the word. This can be seen in instances such as "*I will help you **make a book** at a hotel...*" (A2 level, PET) or "*I wish we can **make a meeting**...*" (A2 level, BEC 2). However, at the B1 level, phrases such as "***make a reservation***" and "***make a plan***" were frequently utilized, illustrating the entrenchment of phraseology and the learners having a more solidified understanding of the word 'make.' That is not to say, however, that learners at the B1 level did not produce any errors. Some errors in the B1 level included "***make a contract with***." At the B2 level, it was clear that the learners were capable of utilizing 'make' in a wider range of contexts such as "***make a plan***," "***make a balance***," "***make an arrangement***," and "***make a contribution***." A similar pattern was present in the C1 level, with the word 'make' being used in an even wider context, such as "***make a choice***," "***make a living***," and "***make a difference***." At the C2 level, 'make' was only used once in "***make any excuse***." This may be due to the paucity of the C2 level data.

In the case of the verb 'give,' which ranked higher in the lower levels, tended to be overused in phrases such as "***give a presentation***," "***give some advice***," "***give a chance***," and "***give a lot of***." Although there is no apparent progression between levels A2 and B2, there appeared to be a jump in the C2 level. At the C2 level, these are further developed into phrases such as "***give a guidance***" and "***give any credibility***," demonstrating a clear development in the learners' ability to use the word in broader contexts.

Another noticeable change was present in the node verb 'take.' In the lower proficiency levels, it was evident that the use of 'take' was more restricted and repetitive. For instance, "***take a step***," "***take a rest***," and "***take a picture***" were among the most common phrases in A2 to B2 proficiency levels. In the C1 and C2 levels, there was more variety in the learners' use of the verb 'take.' Some examples include "***take the position of***," "***take an advantage***," "***take this matter***," and "***take a glimpse of***." When compared with the most prevalent phrases in the lower levels, it is evident that the learners in the higher levels were able to use 'take' in a broader range of contexts.

Lastly, two apparent observations can be made for verbs 'get' and 'see.' Firstly, it should be acknowledged that the fluctuation in the ranking across the proficiency levels may be due to

the size of the data. For instance, there was much more data on the B1 and B2 levels compared to the other proficiency levels. Thus, although ‘get’ ranked third with 4.2% in the C2 level, it was only used four times, which makes it difficult to make claims about learner usage. However, ‘get’ was used significantly more in B1 and B2 levels, where “*get a job in*” was used extensively in both levels. In the advanced levels, instances of “*get a job*” existed but other instances such as “*get an access*” and “*get the benefit of*” were also prevalent as well. In the case of the verb ‘see,’ although it ranked second in the C2 level, there were only six instances throughout the entire level. Thus, it was not possible based on the evidence to claim that learners in the C2 level started to use the verb ‘see’ more extensively.

In the second core sequence, VV IN DT NN, the node verb ‘go’ was most frequently used across all five proficiency levels (see **Table 4.5**).

VV IN DT NN	A2	B1	B2	C1	C2
<b>Go</b>	[1]* 54.7%**	[1] 14.3%	[1] 12.4%	[1] 12.6%	[1] 11.8%

(\*: rank in the frequency; \*\*: percentage from the entire corpus)

**Table 4.5** Percentage of the top node verbs from each proficiency level in the “VV IN DT NN” sequence

The concordance lines associated with the verb ‘go’ showed that the majority of the examples in the A2 consisted of “*go to a meeting*,” “*go to the cinema*,” and “*go to the theatre*”. These examples illustrate the learners’ limited knowledge of the word ‘go’ where the lexeme was used to indicate the literal sense of going to a place. A similar pattern was also prevalent in the B1 level, where “*go to the cinema*” and “*go to the museum*” were among the most common verb ‘go.’ However, unlike the A2 level, a wider range of prepositions (e.g., “*go on a business trip*,” “*go over the wall*”) followed the node verb ‘go.’ A similar pattern was prevalent in B2 to C2 levels, where many of the examples were followed by the preposition ‘to.’ Additionally, a wider range of nouns (e.g., shop, school, trip, sightseeing, restaurant, town) and prepositions (e.g., go for, go on, go through, go to) were present in the higher proficiency levels.

The node verbs in the rest of the core sequences (VVN IN DT NN, VV DT JJ NN, VVG IN DT NN) are presented in **Table 4.6**. Unlike the core sequences, there were fewer data across all five proficiency levels in the other core sequences. For example, in VVN IN DT NN, where verbs are located in the past participle tense (e.g. ‘was located’), the node verb ‘located’ ranked first in the A2 level, but it only appeared twice. However, it still ranked first because the other

seven node verbs in the A2 level only occurred once. Hence, the percentage was also much higher (25%). At other proficiency levels, there was much more data. For instance, a total of 196 node verbs were found in the B1 level, 327 node verbs in the B2 level, 230 node verbs in the C1 level, and then 81 node verbs in the C2 level. Interestingly, the node verb ‘located’ also ranked first in B2 and C1 levels. In the B2 level, *“located in the middle”* occurred most frequently, whereas in the C1 level, there was a wide range of examples such as *“located in the mountain,” “located in the exhibition,” and “located in the centre,” “located between the river,” “located under the sea,”* and *“located in the middle.”* In the case of the C2 level, the node verb ‘situated’ ranked the highest with a raw frequency of three. Again, due to the scarcity of data, it may not be possible to discern whether the node verb ‘situated’ is the most frequent node verb in the VVN IN DT NN sequence.

VVN IN DT NN	A2	B1	B2	C1	C2
Located	[1]* 25.0%**	[4] 2.0%	[1] 4.0%	[1] 3.9%	[3] 1.2%
Seen	-	[1] 4.1%	[3] 3.8%	[7] 1.3%	[3] 1.2%
Situated	-	[7] 0.5%	-	[9] 0.4%	[1] 3.7%
VV DT JJ NN	A2	B1	B2	C1	C2
Buy	[1] 33.3%	[5] 3.1%	[5] 2.6%	[8] 0.6%	-
Make	[5] 2.2%	[1] 14.3%	[3] 5.6%	[2] 5.1%	[2] 6.2%
Get	[5] 2.2%	[2] 6.6%	[1] 7.8%	[1] 7.0%	[5] 1.5%
Find	[4] 4.4%	[6] 2.6%	[7] 1.7%	[3] 3.8%	[1] 9.2%
VVG IN DT NN	A2	B1	B2	C1	C2
Going	[1] 41.7%	[5] 2.9%	[4] 7.2%	[4] 5.7%	-
According	-	[1] 24.0%	[1] 20.8%	[1] 16.0%	[1] 11.5%

(\*: rank in the frequency; \*\*: percentage from the entire corpus)

**Table 4.6** Percentage of the top node verbs from each proficiency level in the “VVN IN DT NN”, “VV DT JJ NN”, “VVG IN DT NN” sequences

In the VV DT JJ NN sequence, 15 of the 45 examples (33.3%) were headed by the node verb ‘buy’. Most of them consisted of purchasing household items such as *“buy a big sofa,” “buy a new desk,”* and *“buy some good furniture,”* to name a few. In the B1 level, the node verb ‘make’ occurred 28 times out of 196 occurrences. There was a wide range of examples such as *“make a good memory,” “make a right budget,”* and *“make a wonderful trip.”* In the B2 and C1 levels, the node verb ‘get’ occurred 18 times out of 232 examples in the B2 level and 11 out of 158 times in the C1 level. In the C2 level, ‘get’ occurred only once out of 65 times. Illustrated

by the concordance lines, it is evident that there is a tendency towards more formulaic language as learners become more proficient.

Lastly, in the VVG IN DT NN sequence, ‘going’ occurred most frequently in the A2 sequence, with 10 out of 24 occurrences. In the B1 to C2 levels, ‘according’ occurred most frequently, taking into account 16 to 24% from the B1 to C1 level and then 11.5% in the C2 level.

	B1	B2	C1	C2
<b>Used</b>	[1]* 9.0%**	[2] 5.6%	[5] 1.0%	[2] 2.0%
<b>Compared</b>	[5] 1.3%	[1] 6.5%	[4] 2.9%	-
<b>Changed</b>	[5] 1.3%	[5] 1.9%	[1] 5.8%	-
<b>VVD DT NN IN</b>	<b>B1</b>	<b>B2</b>	<b>C1</b>	<b>C2</b>
<b>Saw</b>	[1] 16.5%	[3] 7.1%	[3] 5.3%	[2] 4.2%
<b>Reached</b>	[2] 5.2%	[1] 9.4%	[4] 3.9%	[3] 2.1%
<b>Enjoyed</b>	[4] 2.1%	[5] 2.4%	[1] 7.9%	-
<b>Became</b>	[5] 1.0%	[4] 4.7%	[4] 3.9%	[1] 6.3%
<b>VVN IN PP\$ NN</b>	<b>B1</b>	<b>B2</b>	<b>C1</b>	<b>C2</b>
<b>Satisfied</b>	[1] 12.9%	[1] 7.6%	[1] 6.5%	-
<b>Compared</b>	[5] 2.9%	[5] 1.5%	-	[1] 7.5%
<b>VVN IN JJ NNS</b>	<b>B1</b>	<b>B2</b>	<b>C1</b>	<b>C2</b>
<b>Used</b>	[1] 12.1%	[2] 4.3%	[3] 3.8%	[2] 2.9%
<b>Raised</b>	-	[1] 5.4%	-	-
<b>influenced</b>	[5] 1.5%	[4] 2.2%	[1] 9.6%	-
<b>VVD DT JJ NN</b>	<b>B1</b>	<b>B2</b>	<b>C1</b>	<b>C2</b>
<b>Made</b>	[1] 12.5%	[5] 5.6%	[1] 7.1%	[3] 5.9%
<b>Showed</b>	[5] 2.5%	[1] 12.4%	[1] 7.1%	-
<b>Did</b>	[6] 1.3%	[6] 2.2%	[1] 7.1%	-
<b>Found</b>	-	[7] 1.1%	-	[1] 11.8%

(\*: rank in the frequency; \*\*: percentage from the entire corpus)

**Table 4.7** Percentage of the top node verbs from each proficiency level in the emerging sequences

**Table 4.7** shows the most frequently used verbs of the emerging sequences. Unlike the core sequences, there is no overlap of the node verbs across the proficiency levels, aside from ‘satisfied’ in the VVN IN PP\$ NN sequence, where ‘satisfied’ occurred most frequently from B1 to C2 levels. However, as mentioned earlier, the lack of data in the emerging sequences makes it difficult to ascertain whether the patterns present are particular to the learners in respective learners. Regardless, the data is valuable in that as the emerging sequences surface towards the top ranking, there is also a change in the learners’ use of node verbs.

#### 4.4 Verb 4-gram Sequences to Pattern Grammar

Thus far, the results have revealed a trend of change in the verb 4-gram sequences from beginner to advanced proficiency levels. Additionally, there is a clear path of changes that are mediated by learners' CEFR levels. While these changes are interesting, there is scope to know more about *how* the verbs are used. This section explores core sequences in depth through the categorization of their respective grammar patterns. By doing so, the emerging grammar patterns are explored as well as the diversity in lexical choices as learners transition from low to high proficiency.

##### 4.4.1 Among the core sequences, what grammar patterns emerge in the L1 Korean sub-corpora?

The analyses of the core and emerging grammar patterns are divided into two sections. The first section explores the verb patterns in two core verb 4-gram sequences, VV DT NN IN and VVN IN DT NN. The rationale for choosing these two sequences is as follows: 1) The VV DT NN IN and VVN IN DT NN sequences are the top among the core verb 4-gram sequences. These sequences continued to rank first and second between B2 to C2 proficiency levels. 2) The first two verb 4-gram sequences were regarded to be representative of all the core sequences. For instance, in the remaining core sequences VV DT JJNN, VVG IN DT NN, and VBZ DT JJ NN, each of the verb are followed by either a determiner (DT) or a preposition (IN). Since the VV DT NN IN and VVN IN DT NN verb 4-gram sequences cover both combinations (e.g., VV + DT; VV + IN), they were selected. 3) The immensity of the data did not allow for all sequences to be included in the in-depth analyses.

**Table 4.8** shows the frequency distribution of grammar patterns of the VV DT NN IN sequence among learners in the A2 to C2 CEFR levels. From the sub-corpus, a random sample of 50 examples from each proficiency level was drafted and categorized by their grammar pattern.

VV DT NN IN	A2	B1	B2	C1	C2
V n	35	34	32	34	35
V n <i>about</i> n	-	-	-	1	-
V n <i>for</i> n	2	1	1	-	1
V n <i>from</i> n	-	-	1	-	1
V n <i>into</i> n	-	-	1	-	-

<b>V n on n</b>	-	1	-	1	-
<b>V n to n</b>	2	2	1	2	2
<b>V n prep/adv</b>	1	-	-	-	-
<b>Phrase</b>	-	-	-	1	2
<b>Collocation</b>	6	5	8	11	9
<b>Error</b>	4	7	6	-	-
<b>Total groups</b>	5	5	6	6	6

**Table 4.8** Distribution of grammar patterns in the “VV DT NN IN” sequence among L1 Korean speakers

Based on the table, it is evident that most of the examples from the VV DT NN IN sequence belonged to the **V n** group, which is categorized under ‘simple patterns’ according to Hunston’s grammar patterns (Francis & Hunston, 1996, 1998). The sequences under this group consist of a verb followed by a noun phrase. The rest of the patterns follow a **V n prep n** pattern, which consists of verbs followed by a noun, preposition, and a noun. This pattern is termed under the ‘complex patterns with prepositions and adverbs’ category. Examples of this pattern include **V n about n**, **V n for n**, and **V n to n**.

As evidenced by the random selection of 50 examples, collocations existed throughout all proficiency levels. However, learners in the higher proficiency level used more phrases and collocations and produced fewer errors than learners in the lower proficiency levels. **Table 4.9** provides examples of the grammar patterns across the five proficiency levels. There are a few important points to note regarding the categorization of grammar patterns. First and foremost, the patterns drafted are entire noun phrases (e.g., **V n about n**), whereas examples from the concordance lines only included a verb and three words following it (e.g., VV DT NN IN). It was cross-checked that all the examples that were randomly pulled from the concordance lines included a noun phrase, and thus, it was considered appropriate to categorize the verb sequences accordingly. In addition, it should be known that the second noun phrase in patterns such as **V n about n** is not provided in the example. Lastly, **V n** has been excluded in the table and will be further explored by meaning groups in a later section.

VV DT NN IN	A2	B1	B2	C1	C2
<b>V n about n</b>				▪ Tell the world about	
<b>V n for n</b>	▪ Buy some food for	▪ Save the money for	▪ Bring some money for		▪ Prepare a gift for

	▪ Buy a bed for				
<b>V n from n</b>			▪ Prevent the earth from		▪ Prevent the rent from
<b>V n into n</b>			▪ Make the earth into		
<b>V n on n</b>		▪ Blame the cause of		▪ Spend the money on	
<b>V n to n</b>	▪ Give the phone to ▪ Write a note to	▪ Bring the phone to ▪ Give a chance to	▪ Write a reference to	▪ Give a set of ▪ Give an influence (influence) to	▪ Send all correspondence to
<b>V n prep/adv</b>	▪ Put the bible on				
<b>Phrase</b>				▪ Shed a light on	▪ Judge a book by ▪ Strike the balance between
<b>Collocation</b>	<ul style="list-style-type: none"> <li>▪ Attend the meeting on</li> <li>▪ Do a research for</li> <li>▪ Give a speech about</li> <li>▪ Write a letter about</li> <li>▪ Write another letter to</li> <li>▪ Write the article in</li> </ul>	<ul style="list-style-type: none"> <li>▪ Attend the meeting on</li> <li>▪ Find the solution from</li> <li>▪ Improve the quality of</li> <li>▪ Receive a letter from</li> <li>▪ Write a letter to</li> </ul>	<ul style="list-style-type: none"> <li>▪ Book the flight to</li> <li>▪ Close an account without</li> <li>▪ Ensure the safety for</li> <li>▪ Examine the issue of</li> <li>▪ Increase the number of</li> <li>▪ Pay a fee for</li> <li>▪ Share some idea with</li> <li>▪ Take a trip to</li> </ul>	<ul style="list-style-type: none"> <li>▪ Attend a school in</li> <li>▪ Find any difference of</li> <li>▪ Find the difference in</li> <li>▪ Give some information about</li> <li>▪ Improve the quality of</li> <li>▪ Provide the environment for</li> <li>▪ Reduce the amount of</li> <li>▪ Send a letter to</li> <li>▪ Share the idea with</li> <li>▪ Take responsibility for</li> <li>▪ Write an article about</li> </ul>	<ul style="list-style-type: none"> <li>▪ Double the amount of</li> <li>▪ Enhance the quality of</li> <li>▪ Enjoy the feeling of</li> <li>▪ Explain the reason of</li> <li>▪ Face the problem of</li> <li>▪ Give a guidance to</li> <li>▪ Impede the progress of</li> <li>▪ Read the article with</li> <li>▪ Write a letter to</li> </ul>
<b>Error</b>	<ul style="list-style-type: none"> <li>▪ Agree the story in</li> <li>▪ Listen (listen) the information from</li> <li>▪ Live a town near</li> <li>▪ Make a meeting for</li> </ul>	<ul style="list-style-type: none"> <li>▪ Bring the lack of</li> <li>▪ Contract another business on</li> <li>▪ Do a homework with</li> <li>▪ Get a freedom from</li> <li>▪ Give an information about</li> <li>▪ Make a contract with</li> <li>▪ Make some problem in</li> </ul>	<ul style="list-style-type: none"> <li>▪ Attend a theater at</li> <li>▪ Concentrate that job because</li> <li>▪ Keep the manner in</li> <li>▪ Make some idea through</li> <li>▪ Rise a fraction to</li> <li>▪ Take this account for</li> </ul>		

**Table 4.9** Categorization of the “VV DT NN IN” sequence among L1 Korean speakers

The first pattern, **V n about n**, appeared once in the C1 level, indicating the action of communicating with someone. According to the description of pattern grammar, this example belongs to the ‘tell’ group, which consists of words such as ask, advise, and question. Notably, these words relate to communication. Another example is the **V n to n** pattern. Here, the transition from a simpler usage to a more complex usage of the target lexeme is evident. At the A2 level, for instance, the meaning of the verb ‘give’ is used to show the action of passing an object to someone. In that sense, the meaning of the node verb is simple and straightforward. At the C1 level, the verb ‘give’ is used to demonstrate having an effect on someone. It should be noted, however, that the verb ‘give’ is not a common word that corresponds to ‘an influence.’ Alternatively, ‘have an influence’ is a much more common expression in English. Nevertheless, there is an evident transition into an attempt to a more abstract use of the verbs. Finally, in the C2 level, the example “*send all correspondence to*” demonstrates that learners have a strong understanding of the collocates between words and are able to use them correctly in context.

A2: “She had to *give the phone* to him.” (PET, 2004)

B1: “Therefore she was supposed to *bring the phone to* her place.” (PET, 2008)

B2: “If you want, my last employer will *write a reference to* you.” (FCE, 1998)

C1: “And recent bad economic condition in Australia might *give an influence to* the young people of that age to...” (IELTS ac, 2005)

C2: “Should you require any further information, please do not hesitate to contact me and *send all correspondence to* the above address.” (CAE, 1999)

Another important thing to note is that learners in the higher proficiency level did not produce any errors. With regard to the learner errors, the majority of them were either 1) wrong use of node verb and/or 2) missing part of speech (e.g., preposition), thereby being incorrectly tagged. Some wrongly used node verbs such as ‘give’ in “*give an influence [influence]*” were not categorized as errors if the meaning was predictable based on concordance lines. On the

contrary, ‘make’ in “*make a meeting*” was marked as an error because it was indistinguishable as to whether the intention was to ‘call’ a meeting or to ‘start’ a meeting.

Additionally, collocations were present across all five proficiency levels, although they became more frequent in the higher proficiency levels. To reiterate, collocations were categorized in accordance with Collins Dictionary. In the case of the A2 level, the learners were able to construct sentences using collocations such as “*attend a meeting*,” “*do research*,” and “*write a letter/article*.” However, there were errors pertaining to articles (e.g., “*do a research*”). Starting from the B1 level, it was evident that there were no major errors pertaining to articles and the learners started to use a wider range of collocations. For instance, learners in the B1 level also used collocations such as “*attend a meeting*” and “*write a letter*,” which were similar to A2 level learners. At the B2 level, however, collocations such as “*ensure the safety*,” “*close an account*,” and “*book a flight*” surfaced, which are clearly more advanced than the collocations present in A2 and B1 levels. Such is also the case in the advanced levels, C1 and C2, where examples such as “*take responsibility*,” “*impede the progress*,” and “*enhance the quality*” surfaced. As can be seen, these examples require a deeper understanding of the relationship between lexemes.

In the case of phrases, they were present mostly in the C1 and C2 proficiency levels. According to the Collins Dictionary, phrases were used as a broad term to distinguish a group of words where the meaning cannot be derived from individual words but in chunks. Below are the contexts in which the phrases were used in sentences.

C1: “I would assert that this was high time for citizens to make campaigns against such a disastrous incidents and *shed a light on* that fact.” (CPE, 2000)

C2: “My English teacher often told me never to *judge a book by* its cover and in my opinion, it's the same for a person, too.” (CPE, 2001)

“... the key success of overcoming an unemployment problem is to *strike the balance between* encouraging more people to...and providing...” (CPE, 1997)

The examples above demonstrate that learners in the higher proficiency level are moving towards a more schematic use of words beyond the individual words. Although further investigation is necessary to understand the extent of learners' knowledge about phrases and phrasal verbs, it is evident that learners' use of language becomes more native-like with higher proficiency levels.

So far, this section has examined the different grammar patterns associated with the **VV DT NN IN** verb 4-gram sequence. However, many of the examples belonged to the **V n** pattern where the head verb is followed by a single noun group. In the case of the examples that belong to this pattern, the prepositional phrases following the noun phrases are not part of the noun. For instance, the preposition phrase “for (*the anniversary of our company*)” in “*arrange a party for*” is trivial in the sense that if it were moved to a different part of the clause, the meaning of the sentence would still stay the same (see Biber & Leech, 1999). Thus, it can be claimed that there were more instances of simple patterns where the verbs were followed by a single noun group. **Table 4.10** explores the different meaning groups of the verb pattern **V n**.

<b>VV DT NN IN</b>	<b>A2</b>	<b>B1</b>	<b>B2</b>	<b>C1</b>	<b>C2</b>
<b>The 'arrange' group</b>	▪ Arrange a party for		▪ Publish the news on		
<b>The 'approve' group</b>					▪ Support the idea about
<b>The 'beat an illness' group</b>					▪ Improve every aspect of
<b>The 'break a record' group</b>	▪ Find a information in ▪ Find any information in ▪ Find the information on	▪ Find a job with	▪ Find a way for ▪ Find the power of	▪ Find a job after	▪ Find the immunity of ▪ Find the story in ▪ Reach that level at
<b>The 'bring' group</b>	▪ Get a mail from ▪ Send a message through	▪ Send a book to ▪ Send some money into		▪ Remove the tour to	▪ Produce some form of
<b>The 'build' group</b>	▪ Build a factory in			▪ Build a super-market in ▪ Make a plan of ▪ Make an improvement for ▪ Make an window in ▪ Prepare another event on	▪ Leave any mark on

<b>The 'buy' group</b>	<ul style="list-style-type: none"> <li>▪ Buy an armchair to</li> <li>▪ Get a job in</li> </ul>		<ul style="list-style-type: none"> <li>▪ Get a benefit from</li> <li>▪ Get any kind of</li> <li>▪ Get some discount plus</li> </ul>	<ul style="list-style-type: none"> <li>▪ Choose the way of</li> <li>▪ Lose a sense of</li> <li>▪ Provide the education for</li> </ul>	<ul style="list-style-type: none"> <li>▪ Get a job in</li> </ul>
<b>The 'change' group</b>	<ul style="list-style-type: none"> <li>▪ Change the time of</li> <li>▪ Change the time of</li> <li>▪ Improve the product at</li> </ul>	<ul style="list-style-type: none"> <li>▪ Change the village between</li> <li>▪ Reduce the difference between</li> <li>▪ Reduce the distance between</li> <li>▪ Reduce the emission of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Improve the place for</li> <li>▪ Improve the value of</li> <li>▪ Limit the use of</li> <li>▪ Raise the fee for</li> </ul>		<ul style="list-style-type: none"> <li>▪ Open another set of</li> </ul>
<b>The 'control' group</b>		<ul style="list-style-type: none"> <li>▪ Keep this policy for</li> </ul>			
<b>The 'cover' group</b>					<ul style="list-style-type: none"> <li>▪ Hide the redness of</li> </ul>
<b>The 'describe' group</b>	<ul style="list-style-type: none"> <li>▪ Suggest a date for</li> </ul>	<ul style="list-style-type: none"> <li>▪ Describe the fear of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Maintain the peace of</li> <li>▪ Present both point of</li> <li>▪ Put the value of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Discuss the benefit of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Describe the place because</li> <li>▪ Discuss the relationship between</li> <li>▪ Express the ideal of</li> </ul>
<b>The 'develop' group</b>	<ul style="list-style-type: none"> <li>▪ Wear any coat in</li> <li>▪ Wear the dress to</li> </ul>		<ul style="list-style-type: none"> <li>▪ Wear that sort of</li> </ul>		<ul style="list-style-type: none"> <li>▪ Assume the rest of</li> </ul>
<b>The 'do' and 'take' group</b>	<ul style="list-style-type: none"> <li>▪ Keep the health (health) of</li> <li>▪ Make a book at</li> <li>▪ Make a call to</li> <li>▪ Make a reservation of</li> <li>▪ Take a picture with</li> <li>▪ Take an exam until</li> </ul>	<ul style="list-style-type: none"> <li>▪ Do all work by</li> <li>▪ Give a book for</li> <li>▪ Give some present for</li> <li>▪ Perform this kind of</li> <li>▪ Take a rest at</li> </ul>	<ul style="list-style-type: none"> <li>▪ Make a bit of</li> <li>▪ Take a photo of</li> <li>▪ Take a photograph of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Take a picture of</li> <li>▪ Keep this spot in</li> </ul>	<ul style="list-style-type: none"> <li>▪ Hold a festival because</li> <li>▪ Take a couple of</li> <li>▪ Take a glimpse of</li> <li>▪ Take another adventure in</li> <li>▪ Throw a party for</li> </ul>
<b>The 'follow' group</b>			<ul style="list-style-type: none"> <li>▪ Follow this campaign for</li> </ul>		
<b>The 'form' group</b>				<ul style="list-style-type: none"> <li>▪ Cause the failure of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Create a sense of</li> <li>▪ Make any excuse in</li> </ul>
<b>The 'give an impression' group</b>		<ul style="list-style-type: none"> <li>▪ Give a damage to</li> </ul>			
<b>The 'hear' group</b>	<ul style="list-style-type: none"> <li>▪ See a number of</li> <li>▪ See the river through</li> </ul>	<ul style="list-style-type: none"> <li>▪ Enjoy a variety of</li> <li>▪ Learn the value of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Enjoy the adventure with</li> <li>▪ Learn the value of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Learn a second-language at</li> <li>▪ Like the place for</li> </ul>	<ul style="list-style-type: none"> <li>▪ Consider the compartment of</li> <li>▪ Experience the nourishment of</li> </ul>

	<ul style="list-style-type: none"> <li>▪ See those staf (staff) in</li> </ul>	<ul style="list-style-type: none"> <li>▪ Imagine the way of</li> <li>▪ See a movie in</li> <li>▪ See a number of</li> <li>▪ See the centre of</li> <li>▪ Know the name of</li> <li>▪ Understand the actor of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Learn the value of</li> </ul>	<ul style="list-style-type: none"> <li>▪ See the development of</li> <li>▪ See the percentage of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Feel the touch of</li> <li>▪ Forget this experience since</li> <li>▪ See the correlation between</li> </ul>
<b>The ‘prepare’ group</b>			<ul style="list-style-type: none"> <li>▪ Prepare the cost by</li> </ul>		
<b>The ‘say’ group</b>	<ul style="list-style-type: none"> <li>▪ Give the presentation for</li> <li>▪ Give a presentation at</li> <li>▪ Write the success of</li> </ul>			<ul style="list-style-type: none"> <li>▪ Give a summary of</li> <li>▪ Write a report about</li> </ul>	
<b>The ‘show’ group</b>		<ul style="list-style-type: none"> <li>▪ Show the result of</li> </ul>		<ul style="list-style-type: none"> <li>▪ Tell the seriousness of</li> </ul>	
<b>The ‘start’ and ‘stop’ group</b>		<ul style="list-style-type: none"> <li>▪ Join the audience of</li> <li>▪ Join the camp in</li> <li>▪ Prevent that kind of</li> </ul>		<ul style="list-style-type: none"> <li>▪ Finish the event with</li> </ul>	
<b>The ‘turn a corner’ group</b>		<ul style="list-style-type: none"> <li>▪ Play the tennis for</li> </ul>			
<b>The ‘use’ group</b>	<ul style="list-style-type: none"> <li>▪ Use this room until</li> </ul>		<ul style="list-style-type: none"> <li>▪ Use the court during</li> </ul>	<ul style="list-style-type: none"> <li>▪ Use the space between</li> </ul>	
<b>The ‘watch’ group</b>	<ul style="list-style-type: none"> <li>▪ Read a book in</li> <li>▪ Read a letter from</li> <li>▪ Watch the film at</li> </ul>	<ul style="list-style-type: none"> <li>▪ Study a kind of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Read an advertisement in</li> <li>▪ Study the language by</li> </ul>	<ul style="list-style-type: none"> <li>▪ Read the report on</li> </ul>	
<b>The ‘watch’, ‘break a record’, and ‘approve’ groups—verbs with other meanings</b>	<ul style="list-style-type: none"> <li>▪ Attend a seminar from</li> <li>▪ Attend the semina (seminar) at</li> </ul>			<ul style="list-style-type: none"> <li>▪ Enter the university after</li> </ul>	<ul style="list-style-type: none"> <li>▪ Enter a universe before</li> <li>▪ Experience the making of</li> <li>▪ Replace the pleasure of</li> </ul>
<b>The ‘win’ group</b>				<ul style="list-style-type: none"> <li>▪ Win a ticket to</li> </ul>	
<b>None</b>	<ul style="list-style-type: none"> <li>▪ Access the database at</li> <li>▪ Increase the marketing (marketing) of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Appreciate all staff for</li> <li>▪ Decrease the percentage of</li> <li>▪ Find a mistake in</li> <li>▪ Illustrate both number of</li> <li>▪ Respect the culture of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Clean the toilet once</li> <li>▪ Decrease the number of</li> <li>▪ Discard a waste into</li> <li>▪ Quit the job for</li> <li>▪ Schedule the time for</li> <li>▪ Set an example of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Enhance the relationship between</li> <li>▪ Enclose the chart for</li> <li>▪ Grab the opportunity of</li> <li>▪ Increase the workload by</li> <li>▪ Lodge a complaint about</li> </ul>	<ul style="list-style-type: none"> <li>▪ Captivate the audience as</li> <li>▪ Endure this experience for</li> <li>▪ Execute the budget of</li> <li>▪ Govern a nation of</li> <li>▪ Pose some decrease in</li> </ul>

			▪ Unite the end of	▪ Plant the idea of ▪ Protest the plan at ▪ Regain the license after	▪ Renew the soul through
<b>Total groups</b>	<b>14</b>	<b>12</b>	<b>12</b>	<b>15</b>	<b>14</b>

**Table 4.10** Categorization of the **V n** pattern from the VV DT NN IN verb 4-gram sequence among L1 Korean speakers

As mentioned, the **V n** classification is categorized under ‘simple patterns’ in pattern grammar, indicating that the verb is followed by a noun phrase. The pattern is further categorized into different meaning groups, as shown in **Table 4.10**. Based on such classification, it did not appear that learners in the higher proficiency group used a wider range of verbs that belonged to different meaning groups. In the A2 level, for instance, 35 verbs were categorized under the **V n** pattern, of which belonged to 14 different meaning groups. On the contrary, in the C2 level, the **V n** pattern appeared 35 times, of which also belonged to 14 meaning groups.

However, there was quite a bit of repetitiveness in the lower proficiency levels. Taking the ‘break a record’ group, for example, all three examples in the A2 level were headed by the verb ‘find’ followed by ‘information’. From B1 to C2 levels, the main verbs in the ‘break a record group’ were also headed by the word ‘find’, although they were followed by different noun groups such as “*find a job*” or “*find a way for*”. A similar pattern was also prevalent in the C2 level, where the meaning group also contained the verb ‘reach’ (e.g., reach that level at). Similarly, in the ‘do’ and ‘take’ group, examples in the A2 level included “*keep the health [health]*,” “*make a book at*,” “*make a call*,” and “*take an exam*,” whereas in the C2 level, phrases such as “*take a glimpse*,” and “*hold a festival*” surfaced, which are idiomatic in nature and the meaning is depended upon the noun phrase that follows the node verb.

Overall, it appears that the general pattern in the use of the node verbs was not facilitated by the meaning groups but rather by the difficulty of the word and the level of formulaicity of the verbs. For example, in the ‘build’ group, the example in the A2 level included “*build a factory*,” where the meaning is simple and straightforward. At the C1 level, however, one of the examples was “*make a plan*,” where the meaning of the node verb ‘make’ is less straightforward and more contextual. Rather, learners would need to have an understanding that the noun ‘plan’ collocates with the verb ‘make,’ thus forming the phrase “*make a plan*.” Similarly, in the C2 level, the example in the same meaning group was “*leave any mark on*,” also indicating that the learners

are capable of constructing phrases that inform meaning that is embedded in the phrase instead of individual words.

According to Collins Dictionary, the examples belonging to the ‘none’ group denote verbs that did not belong to a meaning group. This is because the lists only account for the 400 most frequent verbs in the Collins Corpus. As such, some examples such as “lodge a complaint” could belong to a **V n** group; however, because the word “lodge” was not listed under the pattern **V n**, this example was marked as “none.” Thus, examples such as these illustrate the shortfalls of pattern grammar where the data is confined to some of the most frequently used verbs.

The **VV DT NN IN** pattern has demonstrated a unique pattern formed by both simple phrases involving a verb and a noun phrase or a complex pattern that consists of a noun phrase and a prepositional phrase. Therefore, we now explore another core verb 4-gram sequence, the **VV IN DT NN** (verb-preposition-determiner-noun) verb sequence (see **Table 4.11**).

<b>VV IN DT NN</b>	<b>A2</b>	<b>B1</b>	<b>B2</b>	<b>C1</b>	<b>C2</b>
<b>V about n</b>	3	9	4	3	2
<b>V as n</b>	-	-	-	-	1
<b>V at n</b>	-	-	-	2	-
<b>V between pl-n</b>	-	-	-	-	1
<b>V for n</b>	1	2	8	5	5
<b>V from n</b>	-	1	1	3	1
<b>V in n</b>	-	-	-	-	2
<b>V on n</b>	-	-	-	1	-
<b>V of n</b>	-	-	-	-	1
<b>V prep/adv</b>	30	6	15	18	15
<b>V to n</b>	2	7	3	3	8
<b>V with n</b>	-	2	1	2	2
<b>Phrase</b>	-	-	-	1	1
<b>Phrasal verb</b>	2	2	1	2	3
<b>None</b>	4	11	10	8	6
<b>Error</b>	8	10	6	2	2
<b>Group total</b>	6	8	8	12	13

**Table 4.11** Distribution of grammar patterns in the **VV IN DT NN** sequence among L1 Korean speakers

The node verbs in the **VV IN DT NN** sequence belong to the "simple pattern with prepositions and adverbs" in the Collins Dictionary. These are verbs followed by either a prepositional phrase such as **V about n** or **V with n**. However, some of the patterns were termed

under **V prep/adv**, which are also verbs that are followed by either a prepositional phrase or an adverb. This particular pattern could be ambiguous in the sense that an example could potentially belong to both **V prep/adv** and another group. For instance, "go to the museum" could be categorized as both **V prep/adv** as well as the **V to n** group. According to Collins dictionary, however, the verb 'go' followed by the preposition 'to' and a noun group "the museum" is actually categorized under **the V prep/adv** pattern. In order to maintain consistency throughout the data, the examples were kept separate in accordance with Collins grammar pattern.

There is an increase in the wider use of prepositional phrases with proficiency levels. For instance, among a random sample of 50, six grammar patterns were found in the A2 level, while in B1 and B2 levels, there are eight grammar patterns. At the C1 and C2 levels, there is even more of an increase to 12 and 13 patterns, respectively. This is particularly surprising, taking into consideration that the amount of data was much smaller at the C2 level. This clearly indicates that learners can utilize a more variety of prepositional phrases as they become more proficient. Furthermore, the errors also decrease with proficiency level.

Among the grammar patterns, the majority of the examples belonged to the **V prep/adv** group. This makes sense as the **V prep/adv** does not pertain to a single preposition but is inclusive of all prepositional phrases. A more granular investigation will be performed in the following section to understand the various prepositional phrases of the group.

Some prominent differences exist in the **V for n** pattern where it becomes more prevalent in B2, C1, and C2 levels, although they were used more frequently in the B2 level than any other level. Furthermore, in the **V about n** pattern, there were nine instances in the B1 level, and then gradually decreased with proficiency. Some patterns such as **V as n**, **V at n**, **V between pl-n**, **V in n**, and **V on n** did not appear in the beginning levels. **Table 4.12** shows the examples from the L1 Korean sub-corpus.

VV IN DT NN	A2	B1	B2	C1	C2
<b>V about n</b>	<ul style="list-style-type: none"> <li>▪ Enquire about the service</li> <li>▪ Talk about this product</li> <li>▪ Write about the importance</li> </ul>	<ul style="list-style-type: none"> <li>▪ Dream about this kind</li> <li>▪ Learn about an importance</li> <li>▪ Learn about the culture</li> <li>▪ Learn about the way</li> <li>▪ Think about that museum</li> <li>▪ Think about the money</li> </ul>	<ul style="list-style-type: none"> <li>▪ Enquire about the position</li> <li>▪ Know about the evening</li> <li>▪ Learn about the culture</li> <li>▪ Learn about the culture</li> </ul>	<ul style="list-style-type: none"> <li>▪ Complain about the tour</li> <li>▪ Learn about the culture</li> <li>▪ Worry about the council</li> </ul>	<ul style="list-style-type: none"> <li>▪ Complain about an article</li> <li>▪ Talk about a culture</li> </ul>

		<ul style="list-style-type: none"> <li>▪ Think about the value</li> <li>▪ Worry about the language</li> <li>▪ Write about the difference</li> </ul>	Talk about the condition		
<b>V as n</b>					▪ Work as a unit
<b>V at n</b>				<ul style="list-style-type: none"> <li>▪ Look at another fact</li> <li>▪ Look at the brochure</li> </ul>	
<b>V between pl-n</b>					▪ Choose between a girl
<b>V for n</b>	<ul style="list-style-type: none"> <li>▪ Ask for the product</li> </ul>	<ul style="list-style-type: none"> <li>▪ Apologise for the change</li> <li>▪ Study for all life-time</li> </ul>	<ul style="list-style-type: none"> <li>▪ Apply for a holiday</li> <li>▪ Apply for the position</li> <li>▪ Apply for the position</li> <li>▪ Apply for the team</li> <li>▪ Apply for the vacancy</li> <li>▪ Ask for a refund</li> <li>▪ Ask for some money</li> <li>▪ Prepare for the conference</li> </ul>	<ul style="list-style-type: none"> <li>▪ Apply for the job</li> <li>▪ Ask for a change</li> <li>▪ Look for any change</li> <li>▪ Prepare for the test</li> <li>▪ Prepare for this test</li> </ul>	<ul style="list-style-type: none"> <li>▪ Apply for the post</li> <li>▪ Apply for the travel</li> <li>▪ Cover for the league</li> <li>▪ Pay for the university</li> <li>▪ Wait for the rescue</li> </ul>
<b>V from n</b>		<ul style="list-style-type: none"> <li>▪ Suffer from a lot</li> </ul>	<ul style="list-style-type: none"> <li>▪ Suffer from the lack</li> </ul>	<ul style="list-style-type: none"> <li>▪ Choose from a variety</li> <li>▪ Depart from the appearance</li> <li>▪ Suffer from the hardship</li> </ul>	<ul style="list-style-type: none"> <li>▪ Suffer from the burden</li> </ul>
<b>V in n</b>					<ul style="list-style-type: none"> <li>▪ Participate in the university</li> <li>▪ Result in the reduction</li> </ul>
<b>V on n</b>				<ul style="list-style-type: none"> <li>▪ Comment on the result</li> </ul>	
<b>V of n</b>					<ul style="list-style-type: none"> <li>▪ Think of the homelessness</li> </ul>
<b>V prep/adv</b>	<ul style="list-style-type: none"> <li>▪ Arrive at the airport</li> <li>▪ Get on the boat</li> <li>▪ Go by a car</li> <li>▪ Go for a walk</li> <li>▪ Go to a meeting</li> <li>▪ Go to a meeting</li> <li>▪ Go to a meeting</li> <li>▪ Go to a meeting</li> <li>▪ Go to a picnic</li> <li>▪ Go to another concert</li> </ul>	<ul style="list-style-type: none"> <li>▪ Arrive at the airport</li> <li>▪ Come to the party</li> <li>▪ Go to the prison</li> <li>▪ Ride on the log</li> <li>▪ Work in the office</li> <li>▪ Work in the steel</li> </ul>	<ul style="list-style-type: none"> <li>▪ Arrive at the college</li> <li>▪ Get on the train</li> <li>▪ Go to a library</li> <li>▪ Go to a post</li> <li>▪ Go to the library</li> <li>▪ Go to the pier</li> <li>▪ Go to the supermarket</li> <li>▪ Go to the Wednesday</li> <li>▪ Live in the destitution</li> <li>▪ Live in this world</li> </ul>	<ul style="list-style-type: none"> <li>▪ Fall into the game</li> <li>▪ Go to the food</li> <li>▪ Go for the sightseeing</li> <li>▪ Go on this trip</li> <li>▪ Go to another school</li> <li>▪ Go to the bridegroom</li> <li>▪ Go to the cinema</li> <li>▪ Go to the jail</li> <li>▪ Go to the town</li> <li>▪ Live in the U.S. A</li> </ul>	<ul style="list-style-type: none"> <li>▪ Come to the port</li> <li>▪ Get on the bus</li> <li>▪ Go for a walk</li> <li>▪ Go for a walk</li> <li>▪ Go on a trip</li> <li>▪ Go through the backdoor</li> <li>▪ Go to the toilet</li> <li>▪ Go to the university</li> <li>▪ Live for the moment</li> <li>▪ Live in a seclusion</li> <li>▪ Live in that town</li> </ul>

	<ul style="list-style-type: none"> <li>▪ Go to the beach</li> <li>▪ Go to the bed</li> <li>▪ Go to the cinema</li> <li>▪ Go to the cinema</li> <li>▪ Go to the cinema</li> <li>▪ Go to the cinema</li> <li>▪ Go to the cinema</li> <li>▪ Go to the cinema</li> <li>▪ Go to the cinema</li> <li>▪ Go to the hospital</li> <li>▪ Go to the Internet</li> <li>▪ Go to the meeting</li> <li>▪ Go to the park</li> <li>▪ Go to the river</li> <li>▪ Go to the shop</li> <li>▪ Go to the shop</li> <li>▪ Go to the theatre</li> <li>▪ Go to the theatre</li> <li>▪ Go to the theatre</li> <li>▪ Lie on the bed</li> </ul>		<ul style="list-style-type: none"> <li>▪ Look at this argument</li> <li>▪ Stay in a log</li> <li>▪ Work at a business</li> <li>▪ Work for the promotion</li> <li>▪ Work in the industry</li> </ul>	<ul style="list-style-type: none"> <li>▪ Live without a cloth</li> <li>▪ Live without the internet</li> <li>▪ Look in the back</li> <li>▪ Remain in the corner</li> <li>▪ Stay at the hotel</li> <li>▪ Stay in a log</li> <li>▪ Stay in a log</li> <li>▪ Work in a team</li> </ul>	<ul style="list-style-type: none"> <li>▪ Live with a dog</li> <li>▪ Play in a playroom</li> <li>▪ Start from the hotel</li> <li>▪ Travel on a wheelchair</li> </ul>
<b>V to n</b>	<ul style="list-style-type: none"> <li>▪ Listen to the music</li> <li>▪ Listen to the music</li> </ul>	<ul style="list-style-type: none"> <li>▪ Agree to this idea</li> <li>▪ Change to another guy</li> <li>▪ Connect to this problem</li> <li>▪ Lead to a pity</li> <li>▪ Lead to the crisis</li> <li>▪ Listen to the radio</li> <li>▪ Move to another place</li> </ul>	<ul style="list-style-type: none"> <li>▪ Apply to the life</li> <li>▪ Return to a warehouse</li> <li>▪ Speak to the nature</li> </ul>	<ul style="list-style-type: none"> <li>▪ Get to a cinema</li> <li>▪ Respond to the editorial</li> <li>▪ Succumb to the urge</li> </ul>	<ul style="list-style-type: none"> <li>▪ Agree to the latter</li> <li>▪ Lead to a problem</li> <li>▪ Lead to the arrest</li> <li>▪ Leave to the town</li> <li>▪ Reach to the increase</li> <li>▪ Return to the place</li> <li>▪ Return to the world</li> <li>▪ Walk to the agency</li> </ul>
<b>V with n</b>		<ul style="list-style-type: none"> <li>▪ Agree with the solution</li> <li>▪ Agree with this statement</li> </ul>	<ul style="list-style-type: none"> <li>▪ Talk with a friend</li> </ul>	<ul style="list-style-type: none"> <li>▪ Cope with the pressure</li> <li>▪ Cope with this change</li> </ul>	<ul style="list-style-type: none"> <li>▪ Agree with the statement</li> <li>▪ Cope with the situation</li> </ul>
<b>Phrase</b>				<ul style="list-style-type: none"> <li>▪ Wait in queue</li> </ul>	<ul style="list-style-type: none"> <li>▪ Jump to a conclusion</li> </ul>
<b>Phrasal verb</b>	<ul style="list-style-type: none"> <li>▪ Join in the design</li> <li>▪ Turn on the light</li> </ul>	<ul style="list-style-type: none"> <li>▪ Deal with the problem</li> <li>▪ Deal with the problem</li> </ul>	<ul style="list-style-type: none"> <li>▪ Go through the lack</li> </ul>	<ul style="list-style-type: none"> <li>▪ Get into a university</li> <li>▪ Sail through the test</li> </ul>	<ul style="list-style-type: none"> <li>▪ Bump into any man</li> <li>▪ Come across this kind</li> <li>▪ Go through the crisis</li> </ul>

<b>None</b>	<ul style="list-style-type: none"> <li>▪ Live on the beach</li> <li>▪ Sleep in the middle</li> <li>▪ Sleep on the floor</li> <li>▪ Find on the internet</li> </ul>	<ul style="list-style-type: none"> <li>▪ Co-operate in the group</li> <li>▪ Contact vis this E-mail</li> <li>▪ Develop in any area</li> <li>▪ Help with a project</li> <li>▪ Hide in the house</li> <li>▪ Realize from the name</li> <li>▪ Return after the party</li> <li>▪ See from the pie</li> <li>▪ Take for an example</li> <li>▪ Wait for a while</li> <li>▪ Work in the morning</li> </ul>	<ul style="list-style-type: none"> <li>▪ Buy between a computer</li> <li>▪ Enjoy at the end</li> <li>▪ Face in the future</li> <li>▪ Investigate about some animal</li> <li>▪ Overspend without consideration</li> <li>▪ Meet after the school</li> <li>▪ See from the chart</li> <li>▪ Wait until the news</li> <li>▪ Work about this vacation</li> <li>▪ Work because the leader</li> </ul>	<ul style="list-style-type: none"> <li>▪ Carry as a burden</li> <li>▪ Chat in the cyber</li> <li>▪ Eat on the floor</li> <li>▪ Feel in the cinema</li> <li>▪ Infer from the food</li> <li>▪ See from the chart</li> <li>▪ See in the table</li> <li>▪ Travel as a backpacker</li> </ul>	<ul style="list-style-type: none"> <li>▪ Buy at that time</li> <li>▪ Incur for the course</li> <li>▪ Overwork at that time</li> <li>▪ Purchase at a market</li> <li>▪ Stand during the concert</li> <li>▪ Struggle for the reason</li> </ul>
<b>Error</b>	<ul style="list-style-type: none"> <li>▪ Call to a customer</li> <li>▪ Enjoy with the enviroenment (environment)</li> <li>▪ Live like the street</li> <li>▪ Meet in the burgerking</li> <li>▪ See at that day</li> <li>▪ Tell for all staff</li> <li>▪ Travel to the outside</li> <li>▪ Walk in the seaside</li> </ul>	<ul style="list-style-type: none"> <li>▪ Access in some city</li> <li>▪ Apply to a book</li> <li>▪ Begin of this paragraph</li> <li>▪ Book by the internet</li> <li>▪ Discuss about this topic</li> <li>▪ Discuss on both opinion</li> <li>▪ Invent of the car</li> <li>▪ Meet in a point</li> <li>▪ Predict in the future</li> <li>▪ Support with each family</li> </ul>	<ul style="list-style-type: none"> <li>▪ Apply for a worker</li> <li>▪ Discuss about the school</li> <li>▪ Do by the government</li> <li>▪ Go on an escalator</li> <li>▪ Lost because the text</li> <li>▪ Punish to the prisoner</li> </ul>	<ul style="list-style-type: none"> <li>▪ Promote about this project</li> <li>▪ Shower at the end</li> </ul>	<ul style="list-style-type: none"> <li>▪ Replace of a park</li> <li>▪ Spend in a football-team</li> </ul>

**Table 4.12** Categorization of the **V preposition n** patterns in the “VV IN DT NN” sequence among L1 Korean speakers

In the first grammar pattern, the examples in the **V about n** pattern are headed by a wide range of verbs across all five proficiency levels. As mentioned above, there were nine instances of the pattern in the B1 level and then gradually decreased with proficiency level. Even at the A2 level, learners were able to construct sentences correctly. Below is one example of the **V about n** pattern from each proficiency level.

A2: “I am writing to *enquire about the service* details.” (BECP, 2007)

B1: “I am going to *write about the difference* between A and B.” (FCE, 1997)

B2: “Dear sir/madam I am writing to *enquire about the position* which you advertised in the paper.” (CAE, 1999)

C1: “Dear Sir/Madam, I am writing to *complain about the tour* to Scotland which your company offered from 3rd to 6th October.” (CAE, 1998)

C2: “Dear Sir/Madam, I am writing to *complain about an article* in your newspaper regarding the charity day held in Cooper's Park on Saturday the 23rd.” (CAE, 2000)

Notably, the sentences are shorter in the A2 and B1 proficiency levels. In order to make comparisons possible across the different proficiency levels, an attempt was made to draft samples with similar topics. As a result, all of the examples above, with the exception of the B1 level, discuss writing a letter. On the other hand, the example in the B1 level discusses what the writer is going to discuss in his or her essay. All in all, the examples demonstrate that even the learners in the lower level seem to have already acquired the ‘verb + prepositional phrase’ pattern.

A similar pattern can be seen in other grammar patterns as well. For example, in the **V for n** and **V from n** pattern, the usage increases first and then decreases gradually. Regardless, because many of the sequences are topic-specific (e.g., “**apply for the position,**” “**ask for a refund,**” “**choose from a variety**”), the frequency alone cannot reveal much about the learners’ knowledge. However, the prevalence of low proficiency levels indicates that learners already have knowledge of these patterns in the earlier stage of their learning trajectory. On the contrary, it should also be noted that some patterns were only prevalent in the higher proficiency levels. These are as follows: **V as n, V at n, V between pl-n, V in n, V on n, and V of n**. It is expected, however, since the translation of the prepositions in the listed patterns is complex for L1 Korean learners because prepositions tend to translate the same in Korean. This will be further discussed later, but based on the results, it can be predicted that the difficulty of these prepositions may be one of the reasons why they were not present in the lower proficiency levels.

As evidenced by the examples, most of them belonged to the **V prep/adv** pattern. To reiterate, this pattern entails a verb followed by a prepositional phrase. Thus, this grammar

pattern consists of a wide range of prepositions. One obvious fact is that all but two examples in the A2 level were headed by “go.” Among them, all but three (e.g., on, by, for) are followed by the preposition ‘to’ and consist of many repetitions of ‘meeting,’ ‘cinema,’ and ‘theater,’ indicating potential topical effect. In the B1 level, there were significantly a smaller number of examples that belonged to this grammar pattern. Despite this, there was a wider range of verbs (e.g., arrive, come, ride, and work) and prepositions (e.g., at, to, on, in). From B2 to C2 proficiency levels, the range of verbs and prepositions broadened even further. This aligns with the observations made earlier that with the increase in proficiency level, learners are able to utilize a wider range of ‘verb + prepositional phrase’ patterns.

Lastly, phrases and phrasal verbs were also found in the VV IN DT NN sequence. Although phrases were found only in the C1 and C2 levels, phrasal verbs were found across all proficiency levels. In the A2 level, for instance, some of the examples were “**join in the design**” and “**turn on the light.**” Although these are more literal compared to those found in the C2 level (e.g., “**bump into any man,**” “**go through the crisis**”), it seems as if learners at the beginning level have already attained some understanding of these idiomatic phrases beyond the simple combination of verb and a preposition. On the contrary, errors were present across all proficiency levels, although there were significantly more in the lower level and then decreased with proficiency. As mentioned earlier, this is expected since preposition is a particularly difficult subject for Korean speakers. Also, there was quite a number of the sequences belonged to the ‘none’ group, which are either infrequent or are instances where the prepositional phrase does not modify the verb. In the lower levels, most of the examples in this category are because the preposition does not modify the verb. On the other hand, at the higher level, this may be due to infrequent verbs such as “incur,” “overwork,” and “infer.”

Based on the observations made thus far, even though the two verb 4-gram sequences ranked in the top two across the proficiency levels, there is a clear distinction between the proficiency levels. This would then lead to believe that even more difference must be prevalent in the emerging sequences where the sequences become more common with proficiency level. The following section will observe such differences.

#### 4.4.2 Among the emerging sequences, what grammar patterns emerge in the L1 Korean sub-corpus?

This section investigates the grammar patterns from two emerging sequences: VVN IN DT JJ and VVD DT NN IN. Only the top two emerging sequences will be analyzed as with the core sequences. The first sequence to be analyzed is the VVN IN DT JJ (verb [past participle]-preposition- determiner- adjective) sequence in **Table 4.13**.

VVN IN DT JJ	B1	B2	C1	C2
<b>V about n</b>	1	-	-	-
<b>V as n</b>	1	3	3	1
<b>V by n</b>	-	-	-	1
<b>V from n</b>	-	-	1	2
<b>V in n</b>	-	-	-	1
<b>V into n</b>	-	1	1	-
<b>V prep/adv</b>	1	2	2	3
<b>V to n</b>	-	4	2	5
<b>V with n</b>	1	1	1	1
<b>Phrase</b>	2	4	2	-
<b>Phrasal verb</b>	-	-	-	1
<b>None</b>	35	30	35	27
<b>Error</b>	9	5	3	-
<b>Group total</b>	6	7	8	9

**Table 4.13** Distribution of grammar patterns in the “VVN IN DT JJ” sequence among L1 Korean speakers

As with the VV IN DT NN pattern, the examples in the VVN IN DT JJ sequence also belonged to the ‘verb + prepositional phrase with a noun group. The difference, however, is that in the VVN IN DT JJ sequence, the verbs were in the past participle tense. Moreover, unlike the observations made in the VV IN DT NN sequence, where the majority of the examples belonged in the V prep/adv sequence, the majority of the examples belonged in the ‘none’ category. Two speculations can be made. First, the prepositional phrases did not modify the verbs, or the node verbs were not among the top 400 in the Collins Corpus.

Regardless, the patterns show that there is significantly less use of the ‘V preposition’ pattern in the VVN IN DT JJ sequence. For instance, **V about n** appeared only once in the B1 level, while **V as n** pattern appeared three times in the B2 and C1 levels and only once in the C2 level. Furthermore, some sequences were more prevalent in C1 and C2 levels (e.g., **V by n**, **V from n**, **V in n**, **V to n**). In contrast, others occurred more frequently in certain proficiency

levels (e.g., **V to n** pattern occurred four times in the B2 level but only twice in the B1 and C1 levels). Interestingly, phrases appeared in the B1, B2, and C1 levels, while only one phrasal verb was found in the C2 level. As expected, there were many errors in the B1 level, but they decreased significantly to none in the C2 level.

Despite the low frequency, overall, there is a gradual increase in grammar patterns with proficiency level. For instance, in the B1 level, six different groups were present, and then in the B2 level, there were seven groups in total. Then at C1 and C2 levels, the numbers increased to eight and nine, respectively. **Table 4.14** further explores the random examples selected and categorized to each of the grammar patterns.

<b>VVN IN DT JJ</b>	<b>B1</b>	<b>B2</b>	<b>C1</b>	<b>C2</b>
<b>V about n</b>	<ul style="list-style-type: none"> <li>▪ Worried about the negative</li> </ul>			
<b>V as n</b>	<ul style="list-style-type: none"> <li>▪ Regarded as an important</li> </ul>	<ul style="list-style-type: none"> <li>▪ Regarded as an important</li> <li>▪ Regarded as the high</li> <li>▪ Worked as an official</li> </ul>	<ul style="list-style-type: none"> <li>▪ Chosen as the first</li> <li>▪ Regarded as a long-term</li> <li>▪ Regarded as a same</li> </ul>	<ul style="list-style-type: none"> <li>▪ Regarded as a foolish</li> </ul>
<b>V by n</b>				<ul style="list-style-type: none"> <li>▪ Bound by the same</li> </ul>
<b>V from n</b>			<ul style="list-style-type: none"> <li>▪ Derived from this global</li> </ul>	<ul style="list-style-type: none"> <li>▪ Kept for a long</li> <li>▪ Separated from a hungry</li> </ul>
<b>V in n</b>				<ul style="list-style-type: none"> <li>▪ Caught in a bloody</li> </ul>
<b>V into n</b>		<ul style="list-style-type: none"> <li>▪ Divided into a local</li> </ul>	<ul style="list-style-type: none"> <li>▪ Changed into a big</li> </ul>	
<b>V prep/adv</b>	<ul style="list-style-type: none"> <li>▪ Moved through the high</li> </ul>	<ul style="list-style-type: none"> <li>▪ Located in the central</li> <li>▪ Located in the south</li> </ul>	<ul style="list-style-type: none"> <li>▪ Clustered in the west</li> <li>▪ Located in a small</li> </ul>	<ul style="list-style-type: none"> <li>▪ Led to a major</li> <li>▪ Lived in a big</li> <li>▪ Located on the other</li> </ul>
<b>V to n</b>		<ul style="list-style-type: none"> <li>▪ Connected to the main</li> <li>▪ Married to an elderly</li> <li>▪ Reared by the rich</li> <li>▪ Related to some crucial</li> </ul>	<ul style="list-style-type: none"> <li>▪ Given to the local</li> <li>▪ Related to the ordinary</li> </ul>	<ul style="list-style-type: none"> <li>▪ Applied to the big</li> <li>▪ Contributed to the local</li> <li>▪ Exposed to a life-long</li> <li>▪ Inclined to the second</li> <li>▪ Related to each other</li> </ul>
<b>V with n</b>	<ul style="list-style-type: none"> <li>▪ Connected with the last</li> </ul>	<ul style="list-style-type: none"> <li>▪ Coped with the financial</li> </ul>	<ul style="list-style-type: none"> <li>▪ Connected with a social</li> </ul>	<ul style="list-style-type: none"> <li>▪ Equipped with a few</li> </ul>
<b>Phrase</b>	<ul style="list-style-type: none"> <li>▪ Compared to the original</li> <li>▪ Followed by a slight</li> </ul>	<ul style="list-style-type: none"> <li>▪ Compared with the former</li> <li>▪ Compared to the other</li> <li>▪ Compared with the traditional</li> </ul>	<ul style="list-style-type: none"> <li>▪ Followed by the national</li> <li>▪ Followed by the other</li> </ul>	

Phrasal verb		▪ Followed by a slight		▪ Moved into the adjacent
<b>None</b>	<ul style="list-style-type: none"> <li>▪ Advertised in the English</li> <li>▪ Awarded as the top</li> <li>▪ Born in a rich</li> <li>▪ Considered as a serious</li> <li>▪ Declined during the same</li> <li>▪ Decreased as the mobile</li> <li>▪ Delivered to the public</li> <li>▪ Dominated than any other</li> <li>▪ Employed than the other</li> <li>▪ Failed in the first</li> <li>▪ Fallen over the last</li> <li>▪ Ignored by the young</li> <li>▪ Increased around the total</li> <li>▪ Increased as each different</li> <li>▪ Read as a different</li> <li>▪ Recorded as the second</li> <li>▪ Respected than any other</li> <li>▪ Said as an economical</li> <li>▪ Represented in these analytical</li> <li>▪ Seen as a significant</li> <li>▪ Seen from the above</li> <li>▪ Selected from the local</li> <li>▪ Showed during a whole</li> <li>▪ Sold at a half-price</li> <li>▪ Supported by a well-going</li> <li>▪ Changed by a new</li> <li>▪ Thought from the early</li> <li>▪ Trained in any special</li> </ul>	<ul style="list-style-type: none"> <li>▪ Accepted as a natural</li> <li>▪ Born in a rich</li> <li>▪ Changed by the competitive</li> <li>▪ Collapsed in the near</li> <li>▪ Constructed in the mid</li> <li>▪ Designed as an educational</li> <li>▪ Destroyed at the same</li> <li>▪ Decorated like an old</li> <li>▪ Existed in the contemporary</li> <li>▪ Explained about the Korean</li> <li>▪ Exposed in the bad</li> <li>▪ Fluctuated in the same</li> <li>▪ Followed as a second</li> <li>▪ Formed with a moderate</li> <li>▪ Handled with the utmost</li> <li>▪ Held by the local</li> <li>▪ Involved in the voluntary</li> <li>▪ Kept in a secured</li> <li>▪ Made in a recent</li> <li>▪ Made to the basic</li> <li>▪ Occurred during the past</li> <li>▪ Organised by some positive</li> <li>▪ Released by the Korean</li> <li>▪ Seen from the above</li> <li>▪ Seen from the first</li> <li>▪ Shown by the first</li> <li>▪ Sold in the global</li> <li>▪ Used in an average</li> <li>▪ Utilised in the global</li> <li>▪ Written in the extinct</li> </ul>	<ul style="list-style-type: none"> <li>▪ Admitted by a small</li> <li>▪ Adopted as an alternative</li> <li>▪ Arranged as a casual</li> <li>▪ Built for the same</li> <li>▪ Broken by the limited</li> <li>▪ Caused by the present</li> <li>▪ Composed before the sixteenth</li> <li>▪ Caused by the unhealthy</li> <li>▪ Differentiated (differentiated) from each other</li> <li>▪ Driven by the modern</li> <li>▪ Eat after the proper</li> <li>▪ Found in the big</li> <li>▪ Given by the old</li> <li>▪ Happened in the third</li> <li>▪ Held for the first</li> <li>▪ Killed by the Japanese</li> <li>▪ Labeled as a crazy</li> <li>▪ Learned in the high</li> <li>▪ Outweighed by the long-term</li> <li>▪ Placed on this modern</li> <li>▪ Posed with some applicable</li> <li>▪ Provided in a short</li> <li>▪ Rearranged for the next</li> <li>▪ Ruined by a sudden</li> <li>▪ Seen as a good</li> <li>▪ Seen as a miniature (miniature)</li> <li>▪ Shown in the recent</li> <li>▪ Stopped by a thick</li> <li>▪ Supported by the aforementioned</li> <li>▪ Surrounded by a beautiful</li> <li>▪ Thought as an upper</li> <li>▪ Tired of the lonely</li> <li>▪ Widened to the great</li> <li>▪ Worked as a part-time</li> <li>▪ Written as the top</li> </ul>	<ul style="list-style-type: none"> <li>▪ Attend for the past</li> <li>▪ Blocked by a low</li> <li>▪ Born into a wealthy</li> <li>▪ Built in a short</li> <li>▪ Caused by the contaminated</li> <li>▪ Continued for a few</li> <li>▪ Developed for some dangerous</li> <li>▪ Escaped through the other</li> <li>▪ Faced with any other</li> <li>▪ Forced to an early</li> <li>▪ Held in the first</li> <li>▪ Influenced by each other</li> <li>▪ Intensified by an unexpected</li> <li>▪ Known for the great</li> <li>▪ Lost during the Korean</li> <li>▪ Taught in a Christian</li> <li>▪ Shocked with the empty</li> <li>▪ Situated on the top</li> <li>▪ Spoken in the real</li> <li>▪ Used as a sole</li> <li>▪ Portrayed as the ultimate</li> <li>▪ Protected in those myriad</li> <li>▪ Visit since a young</li> <li>▪ Reminded of this embarrassing</li> <li>▪ Seen as a bad</li> <li>▪ Seen in the green</li> <li>▪ Selected for the first</li> </ul>

	<ul style="list-style-type: none"> <li>▪ Used by the underground</li> <li>▪ Used in an Australian</li> <li>▪ Used in an average</li> <li>▪ Used in an average</li> <li>▪ Used in an average</li> <li>▪ Used in an average</li> <li>▪ Worn by the royal</li> </ul>			
<b>Error</b>	<ul style="list-style-type: none"> <li>▪ Developed to a modern</li> <li>▪ Developed to a straight</li> <li>▪ Grown in a poor</li> <li>▪ Grown in a wealthy</li> <li>▪ Grown in a wealthy</li> <li>▪ Grown in a wealthy</li> <li>▪ Made by a great</li> <li>▪ Prepared to the future</li> <li>▪ Put in a high</li> </ul>	<ul style="list-style-type: none"> <li>▪ Impressed for the delicious</li> <li>▪ Changed to a different</li> <li>▪ Experienced to the real</li> <li>▪ Moved in the other</li> <li>▪ Used to the small</li> </ul>	<ul style="list-style-type: none"> <li>▪ Absorbed in the English</li> <li>▪ Inherited to the next</li> <li>▪ Appreciated about this friendly</li> </ul>	

**Table 4.14** Categorization of the *V preposition n* pattern in the “VVN IN DT JJ” sequence among L1 Korean speakers

Firstly, a large portion of the data belonged to the ‘none’ category across all four proficiency levels, although there were 27 instances in the C2 level compared to 35 cases in the B1 level. As can be seen in the examples, the prepositional phrase does not modify the node verb. For instance, in the example “**eat after the proper...**” in the C1 level, the preposition ‘after’ does not modify the verb, ‘eat.’ In other words, the meaning of the sentence would not be affected if the prepositional phrase was to be moved to a different part of the clause, and thus, the preposition ‘after’ in this case is not important for overall comprehension of the sentence. There were also examples such as “**delivered to the public**” in the B1 level, where ‘delivered’ followed by a noun phrase and a prepositional phrase (e.g., *V n to n*) would be considered a pattern, but ‘delivered’ as *V to n* was not recognized as a pattern according to Collins Dictionary.

Concerning the grammar pattern associated with the VVN IN DT JJ sequence, it is clear that learners in the lower level (e.g., B1) have a fair understanding of grammar patterns and their usage. For instance, “**regarded as...**” appeared across all four proficiency levels.

B1: “...that is old people's gradual interest for cinema with showing a little fluctuation (fluctuation) but they are *regarded as an important* group for...” (IELTS ac, 2005)

B2: “Over centuries the transport had been *regarded as an important* factor in life socially and economically.” (FCE, 2001)

C1: “However, as changing social value doesn't happen immediately, it should be *regarded as a long-term* plan.” (CPE, 1999)

C2: “You'll be *regarded as a foolish* or stupid person.” (CAE, 1999)

In the above four examples, it is evident that even at the lower proficiency level, learners can correctly use “regarded as” in the sentence. In fact, at the B1 level, learners can construct longer sentences with multiple clauses. The difference, however, is that ‘regarded’ is used in simple present form (e.g., “**are regarded as**”), whereas in the B2 level, it was used in the past perfect continuous tense. In the C1 level, ‘regarded’ was preceded by the auxiliary verb ‘should be.’ At the C2 level, ‘regarded’ is used in the future passive voice.

The patterns that occurred only in the C2 level were **V by n**, **V in n**, and phrasal verbs. The examples include “... *we were bound by* the same tie” (CPE, 1998), “*Caught in a bloody* traffic jam...” (CPE, 1998), and “... who had just *moved into the adjacent* house.” (CPE, 1997). It is evident that these examples are more idiomatic in nature and that learners would need to have a good understanding of not only the verbs but also the prepositional phrase that follows the verb in order to construct these sentences. On the contrary, some prevalent phrases in the B1 level were “Some people are *worried about the negative* effect of...” (IELTS gt, 2012), “The produced electric power is *moved through the high* voltage cable.” (IELTS ac, 2008), and “... the news article which is *connected with the last* music festival” (FCE, 1997). Unlike the C2 level, the meaning embedded by the verb and the prepositional phrase is more straightforward and less idiomatic.

The observation of ‘in-between’ levels, such as the B2 and C1 levels, demonstrates the continuum of so-called ‘less-idiomatic’ to ‘more-idiomatic’ phrases that the learners transition into as they become more proficient in the L2. This is evidenced by grammar patterns such as **V as n**, **V prep/adv**, **V to n**, and **V with n**. Below are samples were taken from the **V prep/adv** and **V with n**.

B1: “The produced electric power is *moved through the high* voltage cable.” (IELTS ac, 2008)

B2: “Situation "A" restaurant is *located in the central* area and it has a large parking spaces.” (FCE, 1997)

C1: “All the facilities are *clustered in the west* part of the island.” (IELTS ac, 2011)

C2: “I am a student of the Fordham college, which is *located on the other* side of the park.” (CAE 2001)

The examples demonstrate the learners' transition into less frequent words such as 'located' and 'clustered' in the B2 and C1 levels. Furthermore, in the C2 level, the verb 'located' is followed by the prepositional phrase 'on the other side,' whereas in the B2 level, the verb 'located' was followed by the prepositional phrase 'in the central area.' It is evident that as learners become more proficient, they are also becoming more cognizant of the various ways in which prepositions are used in context.

B1: “..., I would like to complaint about your News article which is *connected with the last* music festival” (FCE, 1997)

B2: “Therefore I firmly believe that children who have *coped with the financial* problems have more benefits in their future lives.” (IELTS ac, 2005)

C1: “Furthermore, self-respect seems to be *connected with a social* term, democracy.” (CPE, 1999)

C2: “You should provide a bus which is *equipped with a few* facilities for the disabled.” (CPE, 1997)

The **V with n** is another example that shows a more stable use of the verbs and prepositional phrases. Taking the B1 level example, the context in which “**connected with the last**” does not seem as natural as the example in C1, “**connected with a few.**” In fact, in the B1 example, the preposition ‘on’ seems more suitable to describe what an article was about. Typically, ‘connect with’ indicates a relationship, and this is better demonstrated in the example in the C1 level, where the learner indicates that self-respect is related to democracy. In this sense, it could be claimed that learners have a better understanding of prepositions at the C1 level than at the B1 level. On the contrary, in the example in B2 level, the learner is able to use correctly the verb ‘cope’ and the prepositional phrase ‘with the financial problem’ to indicate the overcoming of a difficult situation. Similarly, in the C2 level, the verb ‘equipped’ and the prepositional phrase ‘with a. few facilities’ are used correctly as well. Again, a careful judgment should be made in assuming learners’ knowledge at each proficiency level due to the reliance on a small number of examples. However, based on these examples, as well as examples in the previous sections, it is evident that prepositions continue to be a triggering point for L1 Korean speakers, but they do become more idiomatic and formulaic with proficiency level. The following section explores another emerging sequence, the VVD DT NN IN, as shown in **Table 4.15**.

VVD DT NN IN	B1	B2	C1	C2
<b>V n</b>	34	34	35	33
<b>V n as n</b>	1	-	-	-
<b>V n for n</b>	-	-	1	-
<b>V n prep/adv</b>	1	-	-	1
<b>V n to n</b>	-	3	-	1
<b>V n with n</b>	1	-	1	-
<b>Phrase collocation</b>	-	1	1	1
	8	8	9	10
<b>Error</b>	5	4	3	1
<b>Total groups</b>	5	4	5	5

**Table 4.15** Distribution of grammar patterns in the “VVD DT NN IN” sequence among L1 Korean speakers

The verb 4-gram sequence VVD DT NN IN is identical to VV DT NN IN. The only difference is that the VVD DT NN IN sequence is headed by verbs in the past tense, whereas the VV DT NN IN sequence is headed by verbs in the base form. Thus, the grammar patterns in the

VVD DT NN IN sequence also belong to either **V n** or a more complex pattern of **V n preposition n**.

The table indicates that, as with the VV DT NN IN sequence, the majority of the sequences in the VVD DT NN IN sequence also belonged to the simple pattern, verbs followed by a noun phrase. Aside from the **V n** pattern, there was only a tiny distribution of samples across complex patterns with prepositional phrases. Furthermore, it did not appear that the learners in the higher proficiency level were using a wider range of verb patterns. In fact, four different grammar patterns were used in the B1 level, whereas only three were found in the C2 level. Despite these small differences, two clear conclusions can be made: 1) The number of errors decreases with proficiency, and 2) there is a gradual increase of collocations with proficiency level. **Table 4.16** further explores how these patterns were used differently in context.

VVD DT NN IN	B1	B2	C1	C2
<b>V n as n</b>	<ul style="list-style-type: none"> <li>▪ Described this festival as</li> </ul>			
<b>V n for n</b>			<ul style="list-style-type: none"> <li>▪ Asked the driver for</li> </ul>	
<b>V n prep/adv</b>	<ul style="list-style-type: none"> <li>▪ Left some money in</li> </ul>			<ul style="list-style-type: none"> <li>▪ Led the crisis to</li> </ul>
<b>V n to n</b>		<ul style="list-style-type: none"> <li>▪ Gave a key to</li> <li>▪ Gave the suitcase to</li> <li>▪ Paid some money to</li> </ul>		<ul style="list-style-type: none"> <li>▪ Murmured a question to</li> </ul>
<b>V n with n</b>	<ul style="list-style-type: none"> <li>▪ Provided the public with</li> </ul>		<ul style="list-style-type: none"> <li>▪ Shared a plenty of</li> </ul>	
<b>Phrase collocation</b>	<ul style="list-style-type: none"> <li>▪ Reached the age of</li> <li>▪ Received a call from</li> <li>▪ Recived (received) a complaint from</li> <li>▪ Studyed (studied) another subject in</li> <li>▪ Take a trip on</li> <li>▪ Took the drug for</li> <li>▪ Won the prize in</li> <li>▪ Wrote a letter to</li> </ul>	<ul style="list-style-type: none"> <li>▪ Became a father of</li> <li>▪ Entered the house with</li> <li>▪ Offered a variety of</li> <li>▪ Reached the age of</li> <li>▪ Resolved the problem with</li> <li>▪ Took a trip in</li> <li>▪ Took this kind of</li> <li>▪ Won a prize in</li> </ul>	<ul style="list-style-type: none"> <li>▪ Made a joke about</li> <li>▪ Attracted the attention of</li> <li>▪ Became a chairman of</li> <li>▪ Caused every problem in</li> <li>▪ Gave some advice to</li> <li>▪ Owned the house in</li> <li>▪ Reached the age of</li> <li>▪ Received a phone [call] about</li> <li>▪ Took an interest in</li> <li>▪ Won the prize in</li> </ul>	<ul style="list-style-type: none"> <li>▪ Casted a spell on</li> <li>▪ Added the excitement of</li> <li>▪ Called the police for</li> <li>▪ Found an answer to</li> <li>▪ Grabbed the chance of</li> <li>▪ Heard a rumour of</li> <li>▪ Heard some noise from</li> <li>▪ Liked the idea of</li> <li>▪ Sealed the envelope with</li> <li>▪ Sealed the envelope with</li> <li>▪ Wrote a message to</li> </ul>
<b>Error</b>	<ul style="list-style-type: none"> <li>▪ Fell a sleep as</li> <li>▪ Gave a labour to</li> <li>▪ Looked the dictionary for</li> </ul>	<ul style="list-style-type: none"> <li>▪ Declined the rate of</li> <li>▪ Felt the program of</li> <li>▪ Heard the incident from</li> <li>▪ Made an unfortune in</li> </ul>	<ul style="list-style-type: none"> <li>▪ Disgraced the festival by</li> <li>▪ Experienced any row between</li> </ul>	<ul style="list-style-type: none"> <li>▪ Pointed the office with</li> </ul>

	<ul style="list-style-type: none"> <li>▪ Mixed another purpose for</li> <li>▪ Remained the stable between</li> </ul>		<ul style="list-style-type: none"> <li>▪ Made an appointment for</li> </ul>	
<b>Total groups</b>	5	4	5	4

**Table 4.16** Categorization of the **V n preposition n** pattern in the “VVD DT NN IN” sequence among L1 Korean speakers

As with the VV DT NN IN pattern, the table excludes the verb pattern **V n** as they will be discussed in a separate table.

Although the examples from the grammar patterns are few, the same observations can be made about the grammar patterns for the VVD DT NN IN sequence. First of all, the learners in the lower proficiency levels (e.g., B1 level) are able to use complex sentence patterns. Therefore, below are two examples taken from each of the proficiency levels, which are to be used to illustrate the prevailing differences in the context in which the grammar patterns are used.

B1: “However, you *described this festival as* a disaster.” (FCE, 1997)

“I *left some money in* your chest draw because I made some long distance call home.” (FCE, 1999)

B2: “Staff in hotel *gave a key to* me and I have never seen a staff who wanted to give me services after that time.” (CAE, 1998)

“Actually I *paid some money to* her but I bought it very cheaply.” (FCE, 1997)

C1: “Worst of all, Mr. Brown *made a joke about* disabled person and my friend took offence at his remarks.” (CPE, 1998)

“For example, my father is government officer..., he *shared a plenty of* time with us, and I had to think over...” (IELTS ac 2005)

C2: “Not only has it *led the crisis to* our social welfare system but it has also brought about unhappiness and endless depression for some people.” (CPE, 1997)

“I wondered where she might have come from and *murmured a question to* her.” (CPE, 1998)

The randomly selected examples did not show any apparent differences between the levels. However, the noun groups in the B1 level consisted of either a noun or a possessive pronoun, followed by a noun. Similarly, in the B2 level, the noun group followed by the prepositions was also either a noun or possessive determiner. In the second example in the C1 level, however, the noun phrase consists of the preposition ‘with’ followed by the pronoun ‘us.’ This may be an indication that with the increase in the proficiency level, the sentences also become longer and more complex. However, further investigation needs to take place in order to ascertain to what extent the sentences become more complex.

However, based on the concordance lines, there is evidence that complex patterns are prevalent across all proficiency levels. The difference, however, might be in the choice of verbs. For instance, at the B1 level, learners tended to use higher frequency words such as ‘described,’ ‘led,’ and provided.’ In the C2 level, however, the node verbs included ‘led’ and ‘murmured,’ which are of lower frequency.

Another difference between the proficiency levels is the rise in the collocations and phrases usage. Predictably, more collocations were present in higher proficiency levels. Despite this difference, the learners in the lower proficiency levels also demonstrated their ability to utilize a wide range of collocations. On the other hand, there were more errors in the B1 level than in the C2 level. These errors include spelling errors such as ‘a sleep’, which was tagged as ‘determiner + noun.’ Other errors were ‘gave a labour to,’ which does not make much sense in the meaning, and ‘labour’ is typically used without an article. The errors gradually declined but still existed in the near-native proficiency level, C2, indicating that even at the highest CEFR level, errors do exist, although they may be less noticeable.

The following section explores the meaning groups belonging to the **V n** pattern across the four proficiency levels (see **Table 4.17**).

<b>VVD DT NN IN</b>	<b>B1</b>	<b>B2</b>	<b>C1</b>	<b>C2</b>
<b>The ‘allow’ group</b>	<ul style="list-style-type: none"> <li>▪ Needed some information of</li> </ul>			
<b>The ‘be’ group</b>		<ul style="list-style-type: none"> <li>▪ Remained a plateau from</li> </ul>		<ul style="list-style-type: none"> <li>▪ Stayed an hour</li> </ul>
<b>The ‘become’ group</b>	<ul style="list-style-type: none"> <li>▪ Become a part of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Became a reason for</li> <li>▪ Became a topic of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Became an activist for</li> </ul>	<ul style="list-style-type: none"> <li>▪ Became a mother of</li> <li>▪ Became a stranger to</li> <li>▪ Became a tradition of</li> </ul>
<b>The ‘break a record’ group</b>	<ul style="list-style-type: none"> <li>▪ Reached a peak in</li> <li>▪ Reached a peak with</li> </ul>	<ul style="list-style-type: none"> <li>▪ Reached a peak of</li> <li>▪ Reached a peak of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Found a hotel after</li> <li>▪ Reached a peak of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Reached a climax with</li> </ul>

		<ul style="list-style-type: none"> <li>▪ Reached a plateau over</li> <li>▪ Reached the plateau of</li> </ul>		
<b>The 'bring' group</b>		<ul style="list-style-type: none"> <li>▪ Got a prize in</li> </ul>		<ul style="list-style-type: none"> <li>▪ Led the party down</li> </ul>
<b>The 'build' group</b>	<ul style="list-style-type: none"> <li>▪ Created the money for</li> <li>▪ Made a dam to</li> <li>▪ Made a relationship with</li> </ul>			<ul style="list-style-type: none"> <li>▪ Made the relationship between</li> </ul>
<b>The 'buy' group</b>	<ul style="list-style-type: none"> <li>▪ Bought a present for</li> <li>▪ Bought the computer at</li> <li>▪ Received a present (present) from</li> </ul>	<ul style="list-style-type: none"> <li>▪ Spent the money to</li> </ul>	<ul style="list-style-type: none"> <li>▪ Got a job at</li> <li>▪ Got the permission from</li> <li>▪ Received a job in</li> <li>▪ Received the prize of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Chose a tunnel in</li> <li>▪ Chose the basement as</li> <li>▪ Got a job in</li> <li>▪ Got a job in</li> </ul>
<b>The 'change' group</b>				<ul style="list-style-type: none"> <li>▪ Opened the lid of</li> </ul>
<b>The 'concern' group</b>				<ul style="list-style-type: none"> <li>▪ Affected the relationship in</li> </ul>
<b>The 'cover' group</b>				<ul style="list-style-type: none"> <li>▪ Hid the jar in</li> </ul>
<b>The 'describe' group</b>	<ul style="list-style-type: none"> <li>▪ Reported the information of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Answered the value of</li> <li>▪ Described the beginning of</li> </ul>		
<b>The 'develop' group</b>		<ul style="list-style-type: none"> <li>▪ Got a function of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Caught a cold with</li> </ul>	
<b>The 'do' and 'take' group</b>	<ul style="list-style-type: none"> <li>▪ Held a party in</li> <li>▪ Lost a friend of</li> <li>▪ Took a lot of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Took an hour in</li> <li>▪ Took the airplane in</li> </ul>	<ul style="list-style-type: none"> <li>▪ Made a booking for</li> <li>▪ Paid a visit to</li> <li>▪ Took a tour to</li> <li>▪ Took an outing to</li> </ul>	<ul style="list-style-type: none"> <li>▪ Did a make-up like</li> </ul>
<b>The 'eat' group</b>	<ul style="list-style-type: none"> <li>▪ Ate a meal at</li> </ul>			
<b>The 'end' group</b>				<ul style="list-style-type: none"> <li>▪ Ended this bit of</li> </ul>
<b>The 'form' group</b>			<ul style="list-style-type: none"> <li>▪ Established an organization for</li> </ul>	
<b>The 'follow' group</b>			<ul style="list-style-type: none"> <li>▪ Left the house without</li> <li>▪ Visited every house of</li> </ul>	
<b>The 'hear' group</b>	<ul style="list-style-type: none"> <li>▪ Enjoyed the beginning of</li> <li>▪ Enjoyed the festival because</li> <li>▪ Preferred the Education (education) at</li> <li>▪ Preferred the water than</li> <li>▪ Saw a article of</li> <li>▪ Saw an advertisement on</li> <li>▪ Saw an article in</li> <li>▪ Saw the advertisement from</li> </ul>	<ul style="list-style-type: none"> <li>▪ Felt the competence of</li> <li>▪ Enjoyed the festival because</li> <li>▪ Heard the siren of</li> <li>▪ Saw an advertisement in</li> <li>▪ Needed some money for</li> <li>▪ Saw the advertisement at</li> </ul>	<ul style="list-style-type: none"> <li>▪ Enjoyed the day as</li> <li>▪ Enjoyed the dip in</li> <li>▪ Recalled the memory of</li> <li>▪ Saw a group of</li> <li>▪ Saw the plan of</li> <li>▪ Saw this offer from</li> </ul>	<ul style="list-style-type: none"> <li>▪ Experienced the feeling of</li> <li>▪ Felt the perfection of</li> <li>▪ Felt the surge of</li> <li>▪ Learnt the mechanism of</li> <li>▪ Remembered the expression on</li> <li>▪ Saw the banner across</li> <li>▪ Saw the power of</li> </ul>

	<ul style="list-style-type: none"> <li>▪ Saw the advertisement in</li> <li>▪ Saw the advertisement at</li> <li>▪ Saw the advertisement in</li> <li>▪ Saw the advertisement of</li> <li>▪ Saw the show on</li> <li>▪ Saw the ticket for</li> <li>▪ Saw the view of</li> </ul>			
<b>The ‘include’ group</b>			▪ Involved the use of	
<b>The ‘kill’ group</b>			▪ Killed the leader of	
<b>The ‘meet’ group</b>			▪ Met the murderer without	▪ Met a girl in
<b>The ‘move’ group</b>		▪ Turned the sight to		
<b>The ‘operate’ group</b>	▪ Rang the bell on			
<b>The ‘protect’ group</b>		▪ Helped the organisation during		
<b>The ‘record’ group</b>		▪ Indicated the rate of	▪ Recorded a peak with	▪ Recorded every step in
<b>The ‘say’ group</b>				▪ Sang the song of
<b>The ‘seem’ group</b>		▪ Seemed no doubt about		
<b>The ‘show’ group</b>	▪ Showed the satisfaction about		▪ Showed an increase in	▪ Meant the world to
<b>The ‘start’ and ‘finish’ group</b>	▪ Finish a course in	<ul style="list-style-type: none"> <li>▪ Began the rate with</li> <li>▪ Finished the life in</li> </ul>	▪ Started the business with	
<b>The ‘top’ group</b>		▪ Passed the examination so	<ul style="list-style-type: none"> <li>▪ Dominated the amount of</li> <li>▪ Passed the examination so</li> <li>▪ Passed the test by</li> </ul>	
<b>The ‘watch’ group</b>				▪ Studied a degree in
<b>The ‘watch’, ‘break a record’, and ‘approve’ groups-verbs with other meanings</b>		▪ Joined a campsite like	▪ Attended this kind of	▪ Introduced the word with
<b>The ‘win’ group</b>		▪ Lost every battle on		
<b>None</b>	▪ Cleaned the stage for	<ul style="list-style-type: none"> <li>▪ Degraded the percentage of</li> <li>▪ Celebrate all night by</li> <li>▪ Drew the house with</li> <li>▪ Found some error in</li> <li>▪ Increased the attendance to</li> </ul>	<ul style="list-style-type: none"> <li>▪ Modernized the way of</li> <li>▪ Organized a schedule of</li> <li>▪ Rearranged the order of</li> <li>▪ Surveyed a group of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Found some vacuum in</li> <li>▪ Searched every place in</li> <li>▪ Searched some information about</li> <li>▪ Woke the guide at</li> </ul>
<b>Total groups</b>	<b>12</b>	<b>17</b>	<b>16</b>	<b>19</b>

**Table 4.17** Categorization of the **V n** pattern in the “VVD DT NN IN” sequence among L1 Korean speakers

Unlike the **V n** patterns in the VV DT NN IN sequence, it appeared that in the VVD DT NN IN sequence, the number of meaning groups rose with the proficiency level. In the A2 level, for instance, a total of 12 groups were found, whereas, in the C2 level, 19 groups were present. There was also fluctuation in the B2 level, where 17 meaning groups were found, and 16 meaning groups were found in the C1 level. Regardless, it is evident that more meaning groups were found with higher proficiency levels.

The meaning groups found in the intermediate proficiency levels, B1 and B2, are as follows: The ‘allow’ group, the ‘describe’ group, the ‘eat’ group, the ‘move’ group, the ‘operate’ group, the ‘protect’ group, the ‘seem’ group, and the ‘win’ group. On the contrary, these groups were found only in the advanced proficiency levels, C1 and C2:

- The ‘change’ group.
- The ‘concern’ group.
- The ‘cover’ group.
- The ‘end’ group, the ‘form’ group, and the ‘follow’ group.
- The ‘include’ group.
- The ‘kill’ group, the ‘meet’ group, the ‘say’ group, and the ‘watch’ group.

The ‘become’ group, the ‘break a record’ group, the ‘buy’ group, the ‘do’ and ‘take’ group and the ‘hear’ group were found consistently across all four groups. Some groups, such as the ‘build’ group, were only found in B1 and C2 levels.

Below are samples taken from the ‘become’ group:

B1: “Fashion *became a part of* our life and some people think that what clothes you wear is most important thing.” (FCE, 2000)

B2: “Over the past few decades, the issue of competition and cooperation *became a topic of* discussion in the world. (IELTS ac, 2008)

C1: “...in other words, he *became an activist for* the black people of South Africa.” (FCE, 2000)

C2: "... some choose to dress in simple and dark style when they are unhappy and this *became a tradition of* dressing in black in meaning." (CPE, 2001)

Based on the examples, there is evidence that learners across all proficiency levels have a good understanding of the verb + noun phrase + prepositional phrase sentence structure. However, it should be noted that the meaning groups only pertain to the **V n** structure since none of the examples belong to the complex sentence pattern according to the grammar pattern in the Collins Dictionary. The reason for this is that the prepositional phrases in the examples do not modify the node verbs.

In the example from the 'break a record' group, one example from each proficiency level was pulled out for the node verb 'reached':

B1: "It *reached a peak in* 1996 afterwards it plautaued" (IELTS ac, 2005)

B2: "In 1990, around 90% of this group went to the movie, at least once a year, and it *reached a peak of* 95% in 1996." (IELTS ac, 2005)

C1: "In 1970, the numbers employed *reached a peak of* 200 thousands" (IELTS ac, 2008)

C2: "... which gradually had gone up as the concert *reached a climax with* all the audience screaming around me as if I were there again." (CPE 2001)

First of all, these examples show that learners have acquired the structure "**reached a peak**" at the low intermediate level (e.g., B1 level), although it could be argued that this is due to the writing prompt. Regardless, there is a slight difference in the structure "**reached a peak/climax**" across the four levels. In B1, "**reached a peak**" is followed by the preposition 'in' and the year. In the B2 level, the sentence is extended to "**reached a peak of**," followed by the percentage and the preposition 'in' and the year. This could be an indication that learners are able to construct longer and more complex sentences. In the C1 level, the preposition 'of' is

followed by a noun phrase, and finally, in the C2 level, the preposition ‘with’ is followed by a longer noun phrase.

In addition to the differences in the **V n** noun meaning groups, the ‘hear’ group had a particularly large number of examples in the B1 level, although it was the largest group across all four proficiency levels. However, in the B1 level, there were many repetitions of “saw the advertisement,” which indicates that perhaps this is due to the topic. There were also repetitions of ‘enjoyed’ and ‘preferred’ as well. In the B2 level, however, there is a wider range of node verbs such as ‘felt,’ ‘enjoyed,’ ‘heard,’ ‘saw,’ and ‘need.’ Although there are only three different node verbs in the C1 level, in the C2 level, there are five different node verbs. Based on the observation made from the ‘hear’ group, it can be predicted that learners in the lower proficiency level are more reliant on vocabulary in the prompt or that they tend to overuse a repertoire of their limited vocabulary.

Overall, the analyses of the sample core and emerging verb 4-gram sequences illustrated that there is indeed a difference at the level of both patterns and meaning groups. In the case where the patterns did not show a clear difference, the possibility could be due to the number of samples. However, a more microscopic observation at the level of meaning groups has shown that there is a difference in terms of verb uses. While the result of L1 Korean speakers has shown a pattern, what we are still unaware of is whether this pattern can be generalized across all L2 learners. Furthermore, the other L2 speakers’ subcorpus is much larger in size; thus, the findings can be more robust. The purpose of this study was not to contrast the two cohorts but rather to understand the extent to which the result is specific to L1 Korean speakers or among all L2 learners. The following section compares L1 Korean speakers and other L2 speakers in the CLC.

#### **4.5 How do verb 4-gram sequences differ between L1 Korean speakers and other speakers of L1?**

This section explores the top 25 ranking of the verb 4-gram sequences across other L1 speakers. The similarities and differences are compared with the rankings of the L1 Korean subcorpus to contrast the changes between the two groups of learners. **Table 4.18** displays the top 25 ranking of A2 level across other speakers of L1.

Rank other L2	A2	L1 KO
1	VV IN DT NN	3
2	VV DT NN IN	2
3	VBP VVG TO VV	1
4	VV PP IN PP\$	4
5	VVP PP MD VV	9
6	VVD IN DT NN	12
7	VBZ DT JJ NN	21
8	VV DT JJ NN	5
9	VV PP IN DT	21
10	VVP TO VV IN	9
11	VV IN PP\$ NN	14
12	VVG IN DT NN	-
13	VVP IN DT NN	17
14	VV TO VV PP	19
15	VBP RB JJ IN	23
16	VVD DT JJ NN	26
17	VVP TO VV PP	18
18	VV PP DT NN	27
19	VV DT NN NN	20
20	VV TO VV IN	32
21	VVD DT NN IN	27
22	VV IN DT JJ	27
23	VVP TO VV DT	13
24	VV PP\$ NN IN	14
25	VVP DT NN IN	34

**Table 4.18** Top 25 rankings of A2 levels in the other L1 speaker sub-corpus

Four out of five of the core sequences (VV DT NN IN, VV IN DT NN, VV DT JJ NN, VVG IN DT NN) and two out of six emerging sequences (VVD DT JJ NN, VVD DT NN IN) appeared within top 25 among other L2 speakers. This is contrary to the L1 Korean speaker sub-corpus, where only three of the five core sequences (VV DT NN IN, VV IN DT NN, VV DT JJ NN) and one emerging sequence (VVG IN DT NN) were ranked in the top 25. Although careful judgment should be made due to the small data size among L1 Korean speakers, a stronger argument can be made in that even at the beginning proficiency level, learners are starting to grasp the core sequences that become more prominent as their proficiency level also increases.

With respect to the rankings, the top four sequences in the other L1 speaker were like that of L1 Korean speakers. There were fluctuations in the ranking where the VV IN DT NN sequence ranked first in the other L1 speaker sub-corpus, whereas it ranked third among L1 Korean speakers. On the contrary, there was also quite a bit of disparity between the sub-corpora (12 of the 25 sequences differed by five or more rankings between the two sub-corpora; six of them differed by more than ten rankings). For example, the core sequence, VVG IN DT NN,

ranked 12 among other L2 speakers, while it did not appear among the L1 Korean sub-corpus. Furthermore, aside from a handful of sequences (e.g., VVP PP MD VV, VV DT JJ NN, VVP TO VV IN, VVP IN DT NN, VVP TO VVPP, VV DT NN NN), the difference in the ranking among the majority of the sentences were more than five. For instance, the VVP DT NN IN sequence ranked 25 among other L1 speakers, while it ranked 34 among L1 Koreans. These disparities require further investigation into the differences between L1 Koreans and other L1 speakers; however, the scope of this research was to broadly explore whether or not L1 Korean speakers and other L1 speakers in the same proficiency level use similar verb sequences. In that sense, it is evident that although the most frequently used verb sequences mirrored one another, the differences seem much larger in the lower-ranked sequences. **Table 4.19** explores the top 25 ranking across B1 levels.

Rank other L2	B1	L1 KO
1	VV DT NN IN	1
2	VV IN DT NN	2
3	VV DT JJ NN	3
4	VBZ DT JJ NN	7
5	VBP VVG TO VV	5
6	VVG IN DT NN	4
7	VVN IN DT NN	3
8	VVP PP MD VV	36
9	VV TO VV PP	27
10	VV PP IN PP\$	41
11	VV PP IN DT	67
12	VVD DT NN IN	16
13	VBZ RB JJ NN	24
14	VV DT NN NN	28
15	VVD IN DT NN	18
16	VVP TO VV DT	21
17	VV IN PP\$ NN	9
18	VVP IN DT NN	10
19	VV TO VV DT	13
20	VVP TO VV IN	17
21	VV DT NNS IN	20
22	VBP RB JJ IN	23
23	VV IN DT JJ	33
24	VV PP\$ NN IN	10
25	VVD IN PP\$ NN	61

**Table 4.19** Top 25 rankings of B1 levels in the other L1 speaker sub-corpus

In the case of the B1 level, all five core sequences identified in the L1 Korean sub-corpus also ranked in the top seven sequences among other L1 speakers. In fact, the first three core

sequences (VV DT NN IN, VV IN DT NN, VV DT JJ NN) were identical in both sub-corpora. Furthermore, the VBP VVG TO VV sequence, which ranked fifth among other L1 speakers, also ranked fifth among L1 Korean speakers. On the other hand, only one emerging sequence (VVD DT NN IN) appeared in the top 25 among other L1 speakers.

While the top rankings stayed either identical or differed by less than five rankings (VBZ DT JJ, VVD DT NN IN, VVD IN DT NN, VVP TO VV DT, VVP TO VV NN, VV DT NNS IN), there were also drastic changes. In fact, 13 out of 25 sequences differed by more than five rankings between other speakers of L1 and L1 Korean speakers; 9 of those differed by more than ten rankings. For instance, the VVP PP MD VV sequence ranked eighth among other L1 speakers while ranking 36th among L1 Korean speakers. There were also sequences such as VV PP NN PP\$, which ranked 11th among other L1 speakers and then 67th in the L1 Korean sub-corpus. **Table 4.20** explores the top 25 ranking across B2 levels.

Rank other L1	B2	L1 KO
1	VV DT NN IN	1
2	VV IN DT NN	3
3	VVN IN DT NN	2
4	VV DT JJ NN	4
5	VBZ DT JJ NN	6
6	VVG IN DT NN	5
7	VVD IN DT NN	12
8	VVN IN DT JJ	15
9	VVG DT NN IN	20
10	VVD DT NN IN	25
11	VBZ DT NN IN	28
12	VV PP\$ NN IN	9
13	VV DT NNS IN	17
14	VBZ RB JJ IN	14
15	VBP VVG TO VV	10
16	VV TO VV PP	24
17	VV IN DT JJ	24
18	VV PP IN DT	54
19	VV DT NN NN	35
20	VV IN PP\$ NN	17
21	VBP RB JJ IN	15
22	VVG DT JJ NN	33
23	VVP IN DT NN	11
24	VB JJ TO VV	18
25	VVD DT JJ NN	24

**Table 4.20** Top 25 rankings of B2 levels in the other L1 speaker sub-corpus

As with the B1 level, all five of the core sequences found among L1 Korean speakers also ranked within the top six among other L1 speakers. In fact, the ranking of the top six between the two groups stayed within the top six, with the fluctuation being one or two rankings. In the case of the core sequences, three of the five sequences (VVN IN DT JJ, VVD DT NN IN, VVD DT JJ NN) ranked within the top 25. With respect to differences, 12 of the 25 rankings differed by five rankings or more between other L1 learners and L1 Korean speakers. Among them, seven of the sequences (VVG DT NN IN, VVD DT NN IN, VBZ DT NN IN, VV PP IN DT, VV DT NN NN, VVG DT JJ NN, VVP IN DT NN) differed by more than ten rankings. **Table 4.21** explores the top 25 ranking across C1 levels.

Rank other L1	C1	L1 KO
1	VV DT NN IN	2
2	VVN IN DT NN	1
3	VV DT JJ NN	4
4	VV IN DT NN	3
5	VBZ DT JJ NN	7
6	VVG IN DT NN	5
7	VVN IN DT JJ	6
8	VBZ DT NN IN	20
9	VVG DT NN IN	12
10	VB JJ TO VV	13
11	VV PP\$ NN IN	13
12	VB VVN IN DT	12
13	VV IN DT JJ	11
14	VV PP IN DT	24
15	VVG DT JJ NN	26
16	VV TO VV PP	33
17	VVD IN DT NN	11
18	VBZ RB JJ IN	8
19	VV DT NNS IN	16
20	VVD DT NN IN	9
21	VV TO VV DT	28
22	VVN IN PP\$ NN	14
23	VB DT JJ NN	22
24	VVP IN DT NN	17
25	VV DT NN NN	33

**Table 4.21** Top 25 rankings of C1 levels in the other L1 speaker sub-corpus

Again, all five of the core verb sequences as identified among L1 Korean speakers were found in the top 10 among other L1 speakers. In particular, the VVN IN DT NN sequence, which ranked in the top two from B1 to C2 levels among L1 Korean speakers, also started to rank in the top two at the C1 level. In the A2 level, the VVN IN DT NN sequence did not appear in the top

25, whereas in the B1 level, it ranked seventh and then third in the B2 level. In the case of the emerging sequences, only two (VVN IN DT JJ, VVD DT NN IN) of the five that were found among L1 Korean speakers ranked in the top 25 among other L1 speakers.

In terms of differences, 11 of the 25 sequences (VBZ DT NN IN, VVPP IN DT, VVG DT JJ NN, VV TO VV PP, VVD IN DT NN, VBZ RB JJ IN, VVD DT NN IN, VV TO VV DT, VVN IN PP\$ NN, VVP IN DT NN, VV DT NN NN) differed by a ranking of five or more. Among them, six of them (VBZ DT NN IN, VVPP IN DT, VVG DT JJ NN, VV TO VVPP, VBZ RB JJ NN, VVD DT NN IN) differed by more than ten rankings. Finally, **Table 4.22** shows the comparison of the rankings of C2 levels between L1 Korean speakers and other L1 speakers in the CLC.

Rank other L1	C2	L1 KO
1	VV DT NN IN	1
2	VVN IN DT NN	2
3	VV DT JJ NN	4
4	VV IN DT NN	3
5	VBZ DT JJ NN	5
6	VVG IN DT NN	5
7	VVN IN DT JJ	6
8	VBZ DT NN IN	11
9	VVG DT NN IN	12
10	VV PP\$ NN IN	9
11	VB VVN IN DT	16
12	VB JJ TO VV	13
13	VVD IN DT NN	11
14	VV IN DT JJ	13
15	VVG DT JJ NN	11
16	VVD DT NN IN	7
17	VVN IN PP\$ NN	8
18	VV DT NNS IN	25
19	VVP IN DT NN	19
20	VBZ RB JJ IN	16
21	VVN IN JJ NNS	10
22	VB DT JJ NN	13
23	VBZ JJ TO VV	19
24	VV DT NN CC	23
25	VV PP IN DT	22

**Table 4.22** Top 25 rankings of C2 levels in the other L1 speaker sub-corpus

In the C2 level, the ranking of the top nine sequences in the other L1 speaker sub-corpus was identical to that of the C1 level. When compared to L1 Korean speakers, the top five sequences among other L1 speakers were also among the top five in the L1 Korean sub-corpus.

Similar to the C1 level, all five core sequences were also found in the other L1 speaker sub-corpus, whereas three of the emerging sequences were found (VVN IN DT JJ, VVD DT NN IN, VVN IN JJ NNS). Unlike other proficiency levels, the rankings in the C2 level did not differ as drastically across the two groups. Six of the 25 sequences differed by five or more rankings (VB VVN IN DT, VVD DT NN IN, VVN IN PP\$ NN, VV DT NNS IN, VVN IN JJ NNS, VB DT JJ NN), and among them, only one differed by more than 10 (VVN IN JJ NNS). **Table 4.23** zooms closer to the difference in the rankings of core and emerging sequences across the two learner groups.

Sequences	A2		B1		B2		C1		C2	
	L1 KO	Other L1	L1 KO	Other L1	L1 KO	Other L1	L1 KO	Other L1	L1 KO	Other L1
VV DT NN IN	2	2	1	1	2	1	2	1	1	1
VVN IN DT NN	33	45	3	7	1	3	1	2	2	2
VV IN DT NN	3	1	2	2	3	2	3	4	3	4
VV DT JJ NN	5	9	3	3	4	4	4	3	4	3
VVG IN DT NN	20	16	4	6	5	6	5	6	5	6
VVN IN DT JJ	-	-	26	38	15	8	6	7	6	7
VVD DT NN IN	-	-	16	12	25	10	9	20	7	16
VVN IN PP\$ NN	-	-	32	60	36	40	14	22	8	17
VVN IN JJ NNS	-	-	36	144	21	58	21	43	10	21
VVD DT JJ NN	-	-	25	40	24	25	28	35	10	32
VBD DT JJ NN	-	-	35	59	23	41	19	30	10	29

**Table 4.23** Comparison of core and emerging sequence rankings in the L1 Korean and other L1 speaker sub-corpora

Based on the ranking of the sequences, there did not appear to be any considerable fluctuation in the core sequences across all proficiency levels. The only exception is in the A2 proficiency level, where the VVN IN DT NN sequence ranked 33rd among L1 Korean speakers, whereas it ranked 45th among other L1 speakers. Still, in both groups, this sequence ranked significantly lower than in other proficiency levels. Furthermore, in the A2 proficiency level, the VV DT NN IN, VV IN DT NN, and VV DT JJ NN sequences ranked in the top 10 in both groups. The VVG IN DT NN sequence, on the other hand, ranked 20th and 16th in among L1 Korean speakers and other L1 speakers, respectively. Since the emerging sequences were not taken into consideration in the A2 level among L1 Korean speakers due to a lack of data, no such comparison was made in the analysis.

In the B1 level, the VV DT NN IN, VV IN DT NN, and VV DT JJ NN sequences ranked in the top five among other L1 speakers. The VVG IN DT NN and VVN IN DT NN sequences

ranked sixth and seventh, respectively. On the other hand, the fluctuation in the ranking was much greater across emerging sequences. For instance, the VVD DT NN IN sequence ranked 16th among L1 Korean speakers and 12th among other L1 speakers. As illustrated in the table, the other five sequences differed by more than ten rankings. In particular, the rankings for VVN IN JJ NNS differed by 108 rankings.

From B2 to C2 levels, all of the core sequences, with the exception of one (VVG IN DT NN), ranked in the top five among other L1 speakers. As mentioned above, the ranking of the core sequences did not differ among C1 and C2 proficiency levels, although there were slight differences in the B2 level. On the other hand, the emerging sequences fluctuated greatly across all proficiency levels. In the B2 level, with the exception of two sequences (VVN IN PP\$ NN, VVD DT JJ NN), the other four sequences differed by more than five rankings. The ranking of the VVN IN JJ NNS sequence differed greatly between the two groups, although the difference was much lower than in the B1 level (differed by 37 rankings). In the C1 and C2 levels, all of the sequences, with the exception of VVN IN DT JJ, differed by at least five rankings. However, the difference in the ranking seemed to be getting smaller. For instance, the VVN IN JJ NNS sequence decreased by 108 rankings in the B1 level to 37 in the B2 level, and then 22 in C1, and then the difference of 11 rankings places in the C2 level. Thus, it is evident that the differences in the rankings were getting closer with proficiency level.

Based on the observation made thus far, there are both similarities and differences in the rankings among L1 Korean speakers and other L1 speakers. However, further analysis of the core and emerging grammar patterns should provide a thorough understanding of how L1 Korean speakers differ in their use of verb 4-gram sequences from that of the other L1 speaker counterpart. The following section compares the grammar pattern of one core (VV DT NN IN) and emerging (VVN IN DT JJ) sequence. The comparison is made across end levels of each ‘beginner,’ ‘intermediate,’ and ‘advanced’ level. For instance, in the core sequence VV DT NN IN, the comparison is made across A2, B2, and C2 levels. In the case of the emerging sequence VVN IN DT JJ, comparisons are made across B2 and C2 levels since the A2 level was omitted in the L1 Korean sub-corpus due to a lack of data.

#### **4.5.1 A Comparison of the Core Sequences Between L1 Korean Speakers and Other L1 speaker**

In order to observe the differences in the use of the core sequences, the core sequence, VV DT NN IN, was selected. A random sample of 50 examples was extracted from both the L1 Korean speakers and other L1 speaker sub-corpora, which were then categorized according to pattern grammar. **Table 4.24** illustrates the grammar categorization across A2 levels.

VV DT NN IN	A2 (other L1)	A2 (L1 Korean)
V n <i>for</i> n		<ul style="list-style-type: none"> <li>▪ Buy some food for</li> <li>▪ Buy a bed for</li> </ul>
V n <i>to</i> n	<ul style="list-style-type: none"> <li>▪ Tell the truth to</li> </ul>	<ul style="list-style-type: none"> <li>▪ Give the phone to</li> <li>▪ Write a note to</li> </ul>
V n <i>with</i> n	<ul style="list-style-type: none"> <li>▪ Play this game with</li> </ul>	
V prep/adv	<ul style="list-style-type: none"> <li>▪ Go to the theatre because</li> <li>▪ Put the table in</li> </ul>	<ul style="list-style-type: none"> <li>▪ Put the bible on</li> </ul>
Phrase	<ul style="list-style-type: none"> <li>▪ Keep an eye on</li> <li>▪ Take a note of</li> </ul>	
Collocation	<ul style="list-style-type: none"> <li>▪ Attend a meeting at</li> <li>▪ Cancel the appoint with</li> <li>▪ Drink a cup of</li> <li>▪ Give a talk at</li> <li>▪ Give some advice about</li> <li>▪ Write a letter to</li> <li>▪ Write a letter to</li> <li>▪ Write an article for</li> <li>▪ Write an article about</li> <li>▪ Write an article about</li> <li>▪ Write an article for</li> <li>▪ Write the article for</li> <li>▪ Write this letter because</li> </ul>	<ul style="list-style-type: none"> <li>▪ Attend the meeting on</li> <li>▪ Do a research for</li> <li>▪ Give a speech about</li> <li>▪ Write a letter about</li> <li>▪ Write another letter to</li> <li>▪ Write the article in</li> </ul>
Error	<ul style="list-style-type: none"> <li>▪ Make a tour in</li> <li>▪ Research the internet of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Agree the story in</li> <li>▪ Lesten (Listen) the information from</li> <li>▪ Live a town near</li> <li>▪ Make a meeting for</li> </ul>
<b>Total groups</b>	<b>5</b>	<b>4</b>

**Table 4.24** Categorization of the “VV DT NN IN” sequence in A2 level (L1 Korean vs. other L1 speaker)

A side-by-side comparison of the grammar patterns showed that a total of five groups of grammar patterns were identified among other L1 speakers, whereas four groups were found among L1 Korean speakers. However, there was a discrepancy between the number of examples since those that belonged in the V n pattern were omitted for further analysis. That being said, 15 examples were found to belong to four different grammar pattern groups, and 20 sequences belonged to five different grammar patterns. The most prominent difference between the two groups was found in the collocations. In the other L1 speaker sub-corpus, seven different

collocations were found (e.g., “attend a meeting,” “cancel the appointment,” “drink a cup of \_\_,” “give a talk,” “give advice,” “write letter,” “write article”) with several repetitions for “write letter” and “write article.” On the contrary, five different collocations were found (e.g., “attend meeting,” “do research,” “give speech,” “write letter,” “write article”), and two instances of “write letter” were found among L1 Korean speakers.

Aside from collocations, among the other L1 speaker, **V n to n**, **V n with n**, and **V prep/adv** patterns were found. Among these, the **V n with n** pattern was only found in the other L1 speaker sub-corpus. On the contrary, **V n for n**, **V n to n**, and **V prep/adv** patterns were found among A2 level learners in the L1 Korean sub-corpus, of which the **V n for n** pattern was only found in the L1 Korean sub-corpus. In the **V n to n** pattern, one example was found in the other L1 speaker sub-corpus, “tell the truth to,” while two examples were found among L1 Korean speakers (e.g., “give the phone to,” “write a note to”). Below are the concordance lines:

Other L1 speakers:

\* “To be honest if you *tell the truth to* her will be better.” (L1 Turkish, PET, 2011)

L1 Korean speakers:

\* “So she had to *give the phone to* him.” (PET, 2008)

\* “Well, Vitros bye bye I have to *write a note to* another friends.” (KET, 2000)

Both examples in the L1 Korean sub-corpus consist of either ‘have to’ or ‘had to’ + infinitive, showing an obligation for something. On the contrary, the example in the L1 speaker sub-corpus denotes an ‘if-statement,’ following the structure of the first conditional if+ present simple, will + infinitive. Evidently, even within the same grammar pattern, the concordance lines show that the use of verbs differed.

Another example is taken out from the **V prep/adv** pattern where the node verb ‘put’ appeared in both sub-corpora. Below are the examples:

Other L1 speakers:

\* “I *put the table in* front of my bed.” (L1 Spanish-European, PET, 2009)

L1 Korean speakers:

\* “...it was broken by my young brother, so I can’t *put the bible on* the table.” (PET, 2001).

In the other L1 speaker sub-corpus, the node verb ‘put’ is used in the past tense form, whereas in the L1 Korean speaker sub-corpus, the verb ‘put’ is in the based form followed by the negative marker ‘can’t.’

Lastly, there were twice as many errors among L1 Korean speakers than in other L1 speaker sub-corpus. In the L1 Korean sub-corpus, three of the four errors were due to missing prepositions (e.g., missing ‘with’ in “**agree the story in**”), and one pertained to lexical choice (e.g., “**make a meeting**”). On the other hand, in the other L1 sub-corpus, the error pertained to either the lexical choice (e.g., “**make a tour in**”) or article (e.g., “**research the internet**”).

The following section discusses the meaning groups in the **V n** verb pattern. As can be seen in **Table 4.25**, 29 of the 50 examples in the other speakers of the L1 sub-corpus belonged to the **V n** group. Similarly, 34 of the 50 examples among L1 Korean speakers were found in this group.

<b>VV DT NN IN</b>	<b>A2 (other L1)</b>	<b>A2 (L1 Korean)</b>
<b>The ‘arrange’ group</b>	▪ Arrange a hotel for	▪ Arrange a party for
<b>The ‘break a record’ group</b>		▪ Find a information in ▪ Find any information in ▪ Find the information on
<b>The ‘bring’ group</b>	▪ Bring a friend with	▪ Get a mail from ▪ Send a message through
<b>The ‘build’ group</b>		▪ Build a factory in
<b>The ‘buy’ group</b>	▪ Buy a present for ▪ Buy all kind of ▪ Find the owner of	▪ Buy an armchair to ▪ Get a job in
<b>The ‘change’ group</b>	▪ Change the time because ▪ Change the time from ▪ Change the time of	▪ Change the time of ▪ Change the time of ▪ Improve the product at
<b>The ‘control’ group</b>	▪ Control the time in	
<b>The ‘describe’ group</b>	▪ Confirm the date before ▪ Suggest a date for	▪ Suggest a date for
<b>The ‘develop’ group</b>		▪ Wear any coat in

		<ul style="list-style-type: none"> <li>▪ Wear the dress to</li> </ul>
<b>The ‘do’ and ‘take’ group</b>	<ul style="list-style-type: none"> <li>▪ Give a discount of</li> <li>▪ Make an arrangement for</li> <li>▪ Make a change to</li> <li>▪ Make a change to</li> <li>▪ Make this change because</li> </ul>	<ul style="list-style-type: none"> <li>▪ Keep the helth (health) of</li> <li>▪ Make a book at</li> <li>▪ Make a call to</li> <li>▪ Make a reservation of</li> <li>▪ Take a picture with</li> <li>▪ Take an exam until</li> </ul>
<b>The ‘follow’ group</b>	<ul style="list-style-type: none"> <li>▪ Visit the castle of</li> </ul>	
<b>The ‘hear’ group</b>	<ul style="list-style-type: none"> <li>▪ Know the deadline of</li> <li>▪ Know the play so</li> <li>▪ Need a course in</li> <li>▪ Remember the name of</li> </ul>	<ul style="list-style-type: none"> <li>▪ See a number of</li> <li>▪ See the river through</li> <li>▪ See those staf (staff) in</li> </ul>
<b>The ‘prepare’ group</b>	<ul style="list-style-type: none"> <li>▪ Prepare a reading from</li> </ul>	
<b>The ‘say’ group</b>	<ul style="list-style-type: none"> <li>▪ Give a presentation at</li> <li>▪ Give a presentation at</li> <li>▪ Give some brief of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Give the presentation for</li> <li>▪ Give a presentation at</li> <li>▪ Write the success of</li> </ul>
<b>The ‘use’ group</b>		<ul style="list-style-type: none"> <li>▪ Use this room until</li> </ul>
<b>The ‘watch’ group</b>	<ul style="list-style-type: none"> <li>▪ Watch another kind of</li> <li>▪ Watch the competition as</li> </ul>	<ul style="list-style-type: none"> <li>▪ Read a book in</li> <li>▪ Read a letter from</li> <li>▪ Watch the film at</li> </ul>
<b>The ‘watch’, ‘break a record’, and ‘approve’ groups—verbs with other meanings</b>	<ul style="list-style-type: none"> <li>▪ Attend this conference about</li> <li>▪ Spend some time with</li> </ul>	<ul style="list-style-type: none"> <li>▪ Attend a seminar from</li> <li>▪ Attend the semina (seminar) at</li> </ul>
<b>None</b>		<ul style="list-style-type: none"> <li>▪ Access the database at</li> <li>▪ Increase the marketting (marketing) of</li> </ul>
<b>Group total</b>	13	14

**Table 4.25** Meaning groups of the “VV DT NN IN” sequence in A2 level (L1 Korean vs. other L1 speakers)

Among other L1 speakers, 13 different meaning groups were found, whereas, in the L1 Korean sub-corpus, there were 14 different meaning groups along with a 'none' group. Those in the 'none' group denote examples that were categorized under the **V n** pattern in the Collins Dictionary but were not found in the meaning groups, potentially due to the node verb being outside the top 400 most frequently used verbs according to the Collins Dictionary.

As shown in the table, there are a number of similarities between the two groups of learners in terms of the node verbs. Most prominently, both examples followed a 'arrange + determiner + noun + preposition' pattern in the 'arrange' group. Below is an example of the usage:

Other L1 speakers:

\* “I’d be grateful if you could *arrange a hotel for* two days.” (L1 Chinese, BECP, 2002)

L1 Korean speakers:

\* “I have to *arrange a party for* the anniversary of our company, sorry it’s really important for our company.” (PET, 2008)

Interestingly, these two examples show that in the other L1 speaker sub-corpus, “**arrange a hotel for**” was followed by the modal verb ‘could.’ In the L1 Korean speaker sub-corpus, however, the phrase ‘arrange a party for’ is followed by ‘have + infinitive’ to talk about an obligation.

Other similarities in the two groups can be seen in the ‘change’ group, where the node verb ‘change’ is followed by the noun phrase ‘time.’ Below is an example of “**change the time of,**” which was present in both groups:

Other L1 speakers:

\* “Dear Sally, Sorry, but I have to *change the time of* our meet on Tuesday...” (L1 Spanish, PET, 2008)

L1 Korean speakers:

\* “Dear Sally, I need to *change the time of* our meeting because...” (PET, 2008)

In the two examples, the context in which “**change the time of**” was used was almost identical in both groups, where the node verb ‘change’ is followed by either “have + infinitive” or “need + infinitive.” Aside from the examples mentioned above, further similarities can be found in meaning groups such as the ‘say’ group, where “**give a/the presentation**” was prevalent in both other L1 speakers and L1 Korean speakers.

Some of the prominent differences between the two groups are that meaning groups such as the ‘break a record,’ ‘build,’ ‘develop,’ and ‘use’ groups were found only among L1 Korean speakers. On the contrary, the ‘control’ group, the ‘follow’ group, and the ‘prepare’ group were only prevalent among other L1 speakers. There are also differences in the node verbs within the examples as well. For instance, in the ‘hear’ group, all of the three examples in the L1 Korean speaker group consisted of the node verb ‘see’ followed by a noun phrase to indicate what they are ‘looking at.’ On the other hand, a wide range of verbs was prevalent among other L1 speakers, such as ‘know,’ ‘need,’ and ‘remember’.

**Table 4.26** illustrates the grammar pattern categorization of the VV DT NN IN sequence in intermediate levels, the B2 level.

VV DT NN IN	B2 (other L1)	B2 (L1 Korean)
V n <i>for</i> n		▪ Bring some money for
V n <i>from</i> n	▪ Prevent this situation from	▪ Prevent the earth from
V n <i>into</i> n	▪ Turn the canteen into	▪ Make the earth into
V n <i>to</i> n		▪ Write a reference to
V n <i>with</i> n	▪ Fill the atmosphere with	
Phrase	▪ Join the club	
Collocation	<ul style="list-style-type: none"> <li>▪ Arrange a meeting because</li> <li>▪ Make a film about</li> <li>▪ Make the difference between</li> <li>▪ Realise the importance of</li> <li>▪ Receive the information from</li> <li>▪ Share the bathroom with</li> <li>▪ Solve the problem because</li> <li>▪ Spend some time in</li> <li>▪ Surf the internet on</li> </ul>	<ul style="list-style-type: none"> <li>▪ Book the flight to</li> <li>▪ Close an account without</li> <li>▪ Ensure the safety for</li> <li>▪ Examine the issue of</li> <li>▪ Increase the number of</li> <li>▪ Pay a fee for</li> <li>▪ Share some idea with</li> <li>▪ Take a trip</li> </ul>
Error		<ul style="list-style-type: none"> <li>▪ Attend a theater at</li> <li>▪ Concentrate that job because</li> <li>▪ Keep the manner in</li> <li>▪ Make some idea through</li> <li>▪ Rise a fraction to</li> <li>▪ Take this account for</li> </ul>
<b>Total groups</b>	<b>5</b>	<b>5</b>

**Table 4.26** Categorization of the “VV DT NN IN” sequence in B2 level (L1 Korean vs. other L1 speakers)

As mentioned previously, the table excludes the V n pattern, where most of the examples were categorized in. Instead, the remaining examples were categorized into their respective grammar patterns. Unlike the A2 level, a much smaller sample remained for both other L1 and L1 Korean sub-corpora. Among other L1 speakers, 13 examples were dispersed among five

groups, whereas in the L1 Korean sub-corpus, 18 examples were found among five groups. In both groups, a wide range of collocations was found. However, a more comprehensive range of collocations was present in both groups. In the other L1 speaker sub-corpus, nine different collocations and eight different collocations were present among L1 Korean speakers.

The **V n with n** pattern and the phrase “**join the club**” was only found among other L1 speakers, whereas the **V n for n** and the **V n to n** patterns were found only in the L1 Korean sub-corpus. In both groups, **V n from n** and **V n into n** patterns were found. In the **V n from n** pattern, both examples were headed by the node verb ‘prevent.’ Below are the concordance lines:

Other L1 speakers:

\* “The overcrowding problem of our company will become an important issue soon but we still have time to **prevent this situation from** getting worst.” (L1 Spanish, BECH, 2010)

L1 Korean speakers:

\* “In fact, a group of people like a “Green peace” are policing strongly to **prevent the earth from** getting ruined.” (FCE, 1998)

In both examples, the node verb ‘prevent’ is followed by the preposition ‘to’ + infinitive (prevent). Furthermore, both examples follow the ‘verb (prevent) + preposition (from) + verb-ing (getting). Thus, the meaning of stopping something from happening is manifested identical in both examples.

Lastly, unlike the other L1 speakers, six errors were found among L1 Korean speakers. Some errors pertain to wrong word choice (e.g., “**attend a theater**”), but the majority of the errors are due to prepositions (e.g., “**concentrate (on) that job,**” “**take this (into) account for**”). **Table 4.27** below displays the meaning groups of the **V n** pattern.

VV DT NN IN	B2 (other L1)	B2 (L1 Korean)
The ‘arrange’ group		▪ Publish the news on
The ‘beat an illness’ group	▪ Improve the situation of	

<b>The 'become' group</b>	<ul style="list-style-type: none"> <li>▪ Become a profesional (professional) at</li> </ul>	
<b>The 'break a record' group</b>		<ul style="list-style-type: none"> <li>▪ Find a way for</li> <li>▪ Find the power of</li> </ul>
<b>The 'build' group</b>	<ul style="list-style-type: none"> <li>▪ Create a park in</li> </ul>	
<b>The 'buy' group</b>	<ul style="list-style-type: none"> <li>▪ Buy a computer from</li> <li>▪ Get all kind of</li> <li>▪ Offer the food in</li> </ul>	<ul style="list-style-type: none"> <li>▪ Get a benefit from</li> <li>▪ Get any kind of</li> <li>▪ Get some discount plus</li> </ul>
<b>The 'change' group</b>	<ul style="list-style-type: none"> <li>▪ Close the city-centre to</li> <li>▪ Improve the look of</li> <li>▪ Reduce the harm of</li> <li>▪ Reduce the number of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Improve the place for</li> <li>▪ Improve the value of</li> <li>▪ Limit the use of</li> <li>▪ Raise the fee for</li> </ul>
<b>The 'count against' group</b>	<ul style="list-style-type: none"> <li>▪ Recommend the proposal of</li> </ul>	
<b>The 'describe' group</b>	<ul style="list-style-type: none"> <li>▪ Report the weather on</li> </ul>	<ul style="list-style-type: none"> <li>▪ Maintain the peace of</li> <li>▪ Present both point of</li> <li>▪ Put the value of</li> </ul>
<b>The 'develop' group</b>		<ul style="list-style-type: none"> <li>▪ Wear that sort of</li> </ul>
<b>The 'do' and 'take' group</b>	<ul style="list-style-type: none"> <li>▪ Make a reservation for</li> </ul>	<ul style="list-style-type: none"> <li>▪ Make a bit of</li> <li>▪ Take a photo of</li> <li>▪ Take a photograph of</li> </ul>
<b>The 'follow' group</b>		<ul style="list-style-type: none"> <li>▪ Follow this campaign for</li> </ul>
<b>The 'hear' group</b>	<ul style="list-style-type: none"> <li>▪ Consider some kind of</li> <li>▪ Feel any kind of</li> <li>▪ Forget the meaning of</li> <li>▪ Imagine a world without</li> <li>▪ Know the news from</li> <li>▪ Know the place at</li> <li>▪ See a decrease in</li> <li>▪ See a house with</li> </ul>	<ul style="list-style-type: none"> <li>▪ Enjoy the adventure with</li> <li>▪ Learn the value of</li> <li>▪ Learn the value of</li> </ul>
<b>The 'prepare' group</b>		<ul style="list-style-type: none"> <li>▪ Prepare the cost by</li> </ul>
<b>The 'protect' group</b>	<ul style="list-style-type: none"> <li>▪ Help the rest of</li> </ul>	
<b>The 'say' group</b>	<ul style="list-style-type: none"> <li>▪ Give a description of</li> <li>▪ Give an overview of</li> <li>▪ Write a report about</li> </ul>	
<b>The 'start' and 'stop' group</b>	<ul style="list-style-type: none"> <li>▪ Continue the project of</li> <li>▪ Start another language (language) in</li> </ul>	
<b>The 'use' group</b>		<ul style="list-style-type: none"> <li>▪ Use the court during</li> </ul>
<b>The 'watch' group</b>	<ul style="list-style-type: none"> <li>▪ Examine the turnover as</li> </ul>	<ul style="list-style-type: none"> <li>▪ Read an advertisement in</li> <li>▪ Study the language by</li> </ul>
<b>None</b>	<ul style="list-style-type: none"> <li>▪ Clarify the issue through</li> <li>▪ Compare the turnover of</li> <li>▪ Discipline the behavioure (behaviour) of</li> <li>▪ Drop a email on</li> <li>▪ Misunderstand this conduct as</li> <li>▪ Monitor the quality of</li> <li>▪ Pose a threat to</li> <li>▪ Print a correction in</li> <li>▪ Refund the difference of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Clean the toilet once</li> <li>▪ Decrease the number of</li> <li>▪ Discard a waste into</li> <li>▪ Quit the job for</li> <li>▪ Schedule the time for</li> <li>▪ Set an example of</li> <li>▪ Unite the end of</li> </ul>

<b>Group total</b>	14	13
--------------------	----	----

**Table 4.27** Meaning groups of the “VV DT NN IN” sequence in B2 level (L1 Korean vs. other L1 speakers)

Compared with the A2 level proficiency group, it did not seem as though the learners in the B2 level were able to implement a broader range of meaning groups. In fact, 13 and 14 meaning groups were found in the other L1 speaker and L1 Korean speakers, respectively. That being said, there were no drastic changes in the number of meaning groups in the B2 level. However, compared to the A2 proficiency level, more examples fell in the ‘none’ category for both other L1 speakers and L1 Korean speakers. This is an indication that learners are starting to utilize lower frequency verbs.

The main overlaps between the two cohorts are in the ‘buy’ group, the ‘change’ group, the ‘describe’ group, the ‘do’ and ‘take’ group, the ‘hear’ group, and the ‘watch’ group. Interestingly, there were not as many overlaps in the examples. Some of the overlaps include ‘get \_\_\_ kind of’ in the ‘buy’ group and ‘improve the \_\_\_ of’ in the ‘change’ group. The concordance lines were extracted for these examples:

The ‘buy’ group

Other L1 speakers:

\* “You can *get all kind of* exercise if you do this once a day.” (L1 Swedish, FCE, 1999)

L1 Korean speakers:

\* “Modern shopping centres are very convenient to *get any kind of* goods.” (IELTS gt, 2012)

As can be seen from the examples, in the other L1 speaker group, the node verb ‘get’ is followed by the auxiliary verb ‘can’ to illustrate what one *can* get by exercising daily. On the other hand, in the example in the L1 Korean sub-corpus, the node verb ‘get’ is part of the infinitive phrase ‘to get any kind of goods’ to give more detail about the subject ‘modern shopping centres’.

The ‘change’ group

Other L1 speakers:

\* “I am here in order to explain my ideas about how to *improve the look of* our village.”  
(L1 Italian, FCE, 1993)

L1 Korean speakers:

\* “I believe we have to educate young musicians and *improve the value of* music.” (FCE, 1997)

In the case of the examples in the ‘change’ group, both examples contain ‘verb + infinitive’ and a noun phrase after the preposition ‘of.’

With respect to differences, seven of the meaning groups were found only in the other L1 speaker group (e.g., the ‘beat an illness’ group, the ‘become’ group, the ‘build group, the ‘count against’ group, the ‘protect’ group, the ‘say’ group, and the ‘start’ and ‘stop’ group) whereas six groups were found only among L1 Korean speakers (e.g., the ‘arrange’ group, the ‘break a record’ group, the ‘develop’ group, the ‘follow’ group, the ‘prepare’ group, and the ‘use’ group). Unlike the A2 level, both the L1 Korean speakers and other L1 speakers utilized a wider range of verbs and verb sequences, and fewer repetitions of the identical phrases can be found.

The following section compares the grammar patterns and meaning groups found in the VV DT NN IN verb sequence among C2 levels (see **Table 4.28**).

VV DT NN IN	C2 (other L1)	C2 (L1 Korean)
V n <i>for</i> n		▪ Prepare a gift for
V n <i>from</i> n		▪ Prevent the rent from
V n <i>to</i> n	▪ Connect the PC to	▪ Send all correspondence to
Phrase	▪ Make a reference to ▪ Make the impression from ▪ Run the risk of ▪ Run the risk of	▪ Judge a book by ▪ Strike the balance between
Collocation	▪ Acquire some knowledge of ▪ Improve the quality of ▪ Improve the quality of ▪ Increase the popularity of ▪ Lay the foundation for ▪ Pursue the competition of ▪ Submit a proposal about	▪ Double the amount of ▪ Enhance the quality of ▪ Enjoy the feeling of ▪ Explain the reason of ▪ Face the problem of ▪ Give a guidance to ▪ Impede the progress of ▪ Read the article with ▪ Write a letter to
Error	▪ Forget the factor of	

<b>Total groups</b>	<b>3</b>	<b>5</b>
---------------------	----------	----------

**Table 4.28** Categorization of the “VV DT NN IN” sequence in C2 level (L1 Korean vs. other L1 speaker)

Interestingly, in the C2 level, fewer complex grammar patterns (**V n prep n**) were found, especially among other L1 speakers. Only one grammar pattern, **V n to n**, was found among other L1 speakers, whereas three patterns were found in the L1 Korean sub-corpus (**V n for n**, **V n from n**, **V n to n**). Similar to the B2 level, a wide range of collocations were found in both other L1 speakers and L1 Korean speaker sub-corpora. Aside from one instance of repetition in the collocation “improve the quality of” in other L1 speaker sub-corpus, there were no other overlaps. Lastly, fewer errors were present in the C2 level, although one error appeared in the other L1 speaker corpus. Below in **Table 4.29** is the meaning groups found in the **V n** pattern.

<b>VV DT NN IN</b>	<b>C2 (other L1)</b>	<b>C2 (L1 Korean)</b>
<b>The ‘approve’ group</b>	▪ Push the issue of	▪ Support the idea about
<b>The ‘be’ group</b>	▪ Make a living if	
<b>The ‘beat an illness’ group</b>		▪ Improve every aspect of
<b>The ‘break a record’ group</b>	▪ Find a bunch (bunch) of ▪ Find a job in ▪ Find a solution for ▪ Find no meaning at ▪ Find the site as	▪ Find the immunity of ▪ Find the story in ▪ Reach that level at
<b>The ‘bring’ group</b>		▪ Produce some form of
<b>The ‘build’ group</b>	▪ Build a place for ▪ Perform some sort of	▪ Leave any mark on
<b>The ‘buy’ group</b>	▪ Buy the product because ▪ Get an accommodation for	▪ Get a job in
<b>The ‘call out’ group</b>	▪ Attract a plethora of	
<b>The ‘change’ group</b>		▪ Open another set of
<b>The ‘cover’ group</b>		▪ Hide the redness of
<b>The ‘describe’ group</b>	▪ Describe the number of	▪ Describe the place because ▪ Discuss the relationship between ▪ Express the ideal of
<b>The ‘develop’ group</b>		▪ Assume the rest of
<b>The ‘do’ and ‘take’ group</b>	▪ Take a look at ▪ Take a walk around ▪ Take the aeroplane to ▪ Throw a joke at	▪ Hold a festival because ▪ Take a couple of ▪ Take a glimpse of ▪ Take another adventure in ▪ Throw a party for
<b>The ‘form’ group</b>	▪ Show any kind of ▪ Show any sign of	▪ Create a sense of ▪ Make any excuse in
<b>The ‘hear’ group</b>	▪ Consider the number of ▪ Imagine a world without	▪ Consider the compart of

	<ul style="list-style-type: none"> <li>▪ Notice the difference among</li> <li>▪ Understand the world around</li> <li>▪ Want a repetition of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Experience the nourishment of</li> <li>▪ Feel the touch of</li> <li>▪ Forget this experience since</li> <li>▪ See the correlation between</li> </ul>
<b>The ‘meet’ group</b>	<ul style="list-style-type: none"> <li>▪ See the owner of</li> </ul>	
<b>The ‘start’ and ‘stop’ group</b>	<ul style="list-style-type: none"> <li>▪ Escape the danger of</li> </ul>	
<b>The ‘use’ group</b>	<ul style="list-style-type: none"> <li>▪ Use some improvement on</li> <li>▪ Use the power of</li> </ul>	
<b>The ‘watch’ group</b>	<ul style="list-style-type: none"> <li>▪ Check the size of</li> </ul>	
<b>The ‘watch’, ‘break a record’, and ‘approve’ groups—verbs with other meanings</b>	<ul style="list-style-type: none"> <li>▪ Introduce the topic of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Enter a universe before</li> <li>▪ Experience the making of</li> <li>▪ Replace the pleasure of</li> </ul>
<b>None</b>	<ul style="list-style-type: none"> <li>▪ Dedicate a day on</li> <li>▪ Discontinue the course from</li> <li>▪ Enhance the income of</li> <li>▪ Expand the range of</li> <li>▪ Guarantee the excitement of</li> <li>▪ Minimise the number of</li> <li>▪ Obstruct the movement of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Captivate the audience as</li> <li>▪ Endure this experience for</li> <li>▪ Execute the budget of</li> <li>▪ Govern a nation of</li> <li>▪ Pose some decrease in</li> <li>▪ Renew the soul through</li> </ul>
<b>Group total</b>	<b>16</b>	<b>15</b>

**Table 4.29** Meaning groups of the “VV DT NN IN” sequence in C2 level (L1 Korean vs. other L1 speakers)

As mentioned earlier, in the random sample of 50 examples, most of them belonged to the simple pattern **V n**. Among them, 37 of the 50 examples were found in the other L1 speaker sub-corpus, while 35 examples were found in the L1 Korean speaker sub-corpus. These examples were categorized into 16 meaning groups in the other L1 speaker and 15 L1 Korean speakers, sub-corpus. Although the number of meaning groups was similar between the two groups, there were also differences in the distribution of the meaning groups. For instance, the ‘be’ group, the ‘call out’ group, the ‘meet’ group, the ‘start’ and ‘stop’ group, the ‘use’ group, and the ‘watch’ group were found only in the other L1 speaker group. On the contrary, the ‘beat an illness’ group, the ‘bring’ group, the ‘change’ group, the ‘cover’ group, and the ‘develop’ group were groups that only existed in the L1 Korean speaker sub-corpus.

Furthermore, what is evident from both groups when compared to the lower levels A2 and B2 is the wider range of verbs and their usage even in the same meaning group. For instance,

in the ‘do’ and ‘take’ group, three of the four examples were headed by the node verb ‘take’ in other L1 speaker’ sub-corpus:

- Take a look
- Take a walk
- Take the aeroplane

Similarly, three of the five examples in the L1 Korean sub-corpus were also headed by the verb ‘take’:

- Take a couple of [days off]
- Take a glimpse of
- Take another adventure

As illustrated by these examples, all six of the examples display different use of the verb ‘take.’ This is also evident in other meaning groups, such as the ‘break a record’ group, where the numerous appearances of the verb ‘find’ all have a different meaning that is not stored on the node verb itself but in the sequence that follows the verb.

Thus far, this section has explored the transition of the core verb sequence VV DT NN IN and how it differed between the two groups. From a holistic perspective, there is a general trend of more complex language use and a stronger form-meaning connection between the verbs and their meaning with higher proficiency. The differences were found in areas such as errors in the use of articles, although at the C2 level, no errors were found in the L1 Korean speaker subcorpus. The next section explores the change in the use of the VVN IN DT JJ sequence, which appeared to be an emerging sequence in both the L1 Korean speaker sub-corpus and the other L1 speaker sub-corpus.

#### **4.5.2 A Comparison of the Emerging Sequence VVN IN DT JJ between L1 Korean Speakers and Other L1 speakers**

This section compares the VVN IN DT JJ sequence, an emerging sequence found in both the other L1 speaker’ and the L1 Korean speaker sub-corpora. Since the A2 level was omitted in

the L1 Korean speaker sub-corpus, the analyses in this section only pertain to B2 and C2. Below is the grammar pattern for the B2 level (see **Table. 4.30**).

<b>VVN IN DT JJ</b>	<b>B2 (other L1)</b>	<b>B2 (L1 Korean)</b>
<b>V as n</b>		<ul style="list-style-type: none"> <li>▪ Regarded as an important</li> <li>▪ Regarded as the high</li> <li>▪ Worked as an official</li> </ul>
<b>V in n</b>	<ul style="list-style-type: none"> <li>▪ Built in the same</li> </ul>	
<b>V into n</b>	<ul style="list-style-type: none"> <li>▪ Devided (divided) into an old</li> </ul>	<ul style="list-style-type: none"> <li>▪ Divided into a local</li> </ul>
<b>V to n</b>		<ul style="list-style-type: none"> <li>▪ Connected to the main</li> <li>▪ Married to an elderly</li> <li>▪ Reared by the rich</li> <li>▪ Related to some crucial</li> </ul>
<b>V with n</b>	<ul style="list-style-type: none"> <li>▪ Treated with the anticrime</li> </ul>	<ul style="list-style-type: none"> <li>▪ Coped with the financial</li> </ul>
<b>V prep/adv</b>	<ul style="list-style-type: none"> <li>▪ Placed in a cetric</li> <li>▪ Placed in a quiet</li> <li>▪ Reached at the second</li> </ul>	<ul style="list-style-type: none"> <li>▪ Located in the central</li> <li>▪ Located in the south</li> </ul>
<b>Phrase</b>	<ul style="list-style-type: none"> <li>▪ Compared to a top</li> <li>▪ Compared to the other</li> <li>▪ Compared to the other</li> <li>▪ Compared to the previous</li> <li>▪ Compared to the previous</li> <li>▪ Followed by a general</li> <li>▪ Followed by a steady</li> <li>▪ Followed by a subtle</li> </ul>	<ul style="list-style-type: none"> <li>▪ Compared with the former</li> <li>▪ Compared to the other</li> <li>▪ Compared with the traditional</li> <li>▪ Followed by a slight</li> </ul>
<b>Phrasal verb</b>	<ul style="list-style-type: none"> <li>▪ Added in the second</li> </ul>	
<b>None</b>	<ul style="list-style-type: none"> <li>▪ Changed in the last</li> <li>▪ Bitten by a big</li> <li>▪ Built around the old</li> <li>▪ Built during the last</li> <li>▪ Caused by a new</li> <li>▪ Demonstrated by a real</li> <li>▪ Doubled over the past</li> <li>▪ Filmed into a short</li> <li>▪ Finished within the next</li> <li>▪ Held at the central</li> <li>▪ Hired by the Spanish</li> <li>▪ Increased in the last</li> <li>▪ Involved in the different</li> <li>▪ Located near the main</li> <li>▪ Made at the same</li> <li>▪ Maintained at the same</li> <li>▪ Mentioned in the first</li> <li>▪ Offered by the second</li> <li>▪ Offered in the local</li> <li>▪ Provided with a wide</li> <li>▪ Recognized in the third</li> <li>▪ Separate in a small</li> <li>▪ Secured by a special</li> </ul>	<ul style="list-style-type: none"> <li>▪ Accepted as a natural</li> <li>▪ Born in a rich</li> <li>▪ Changed by the competitive</li> <li>▪ Collapsed in the near</li> <li>▪ Constructed in the mid</li> <li>▪ Designed as an educational</li> <li>▪ Destroyed at the same</li> <li>▪ Decorated like an old</li> <li>▪ Existed in the contemporary</li> <li>▪ Explained about the Korean</li> <li>▪ Exposed in the bad</li> <li>▪ Fluctuated in the same</li> <li>▪ Followed as a second</li> <li>▪ Formed with a. moderate</li> <li>▪ Handled with the utmost</li> <li>▪ Held by the local</li> <li>▪ Involved in the voluntary</li> <li>▪ Kept in a secured</li> <li>▪ Made in a recent</li> <li>▪ Made to the basic</li> </ul>

	<ul style="list-style-type: none"> <li>▪ Shown in the first</li> <li>▪ Sold in the first</li> <li>▪ Surprised about the free</li> <li>▪ Taken with a slight</li> <li>▪ Used as a botanic</li> <li>▪ Used as a main</li> <li>▪ Used in an average</li> <li>▪ Written in the past</li> </ul>	<ul style="list-style-type: none"> <li>▪ Occurred during the past</li> <li>▪ Organised by some positive</li> <li>▪ Released by the Korean</li> <li>▪ Seen from the above</li> <li>▪ Seen from the first</li> <li>▪ Shown by the first</li> <li>▪ Sold in the global</li> <li>▪ Used in an average</li> <li>▪ Utilised in the global</li> <li>▪ Written in the extinct</li> </ul>
<b>Error</b>	<ul style="list-style-type: none"> <li>▪ Advertised at the international</li> <li>▪ Found in the different</li> <li>▪ Impressed from the big</li> <li>▪ Reached to a completely</li> </ul>	<ul style="list-style-type: none"> <li>▪ Impressed for the delicious</li> <li>▪ Changed to a different</li> <li>▪ Experienced to the real</li> <li>▪ Moved in the other</li> <li>▪ Used to the small</li> </ul>
<b>Group total</b>	<b>7</b>	<b>7</b>

**Table 4.30** Categorization of the “VVN IN DT JJ” sequence in B2 level (L1 Korean vs. other L1 speakers)

From a holistic view, seven different groups are found in the other L1 speaker and the L1 Korean speaker sub-corpora. Similar to that of the L1 Korean speakers, the majority of the examples in the other L1 speaker sub-corpus were in the ‘none’ group. Again, the examples in this category belong to a grammar pattern according to the Collins Dictionary; however, the specific node verbs were not listed under the grammar pattern.

Most notably, there were several repetitions for both the other L1 speaker’ and L1 Korean speakers’ sub-corpora. However, although there were repetitions within specific grammar patterns in each group, they did not overlap between the two learner groups. For instance, in the **V prep/adv** pattern, three examples were found in the other L1 speaker sub-corpus, while two were found in the L1 Korean speaker group. Among them, two of the three examples in the other L1 speaker group were headed by the verb ‘place,’ whereas, in the L1 Korean sub-corpus, the examples contained the node verb ‘located.’ Below are the examples from concordance lines:

Other L1 speakers:

\* “From my point of view this big store is very useful it is very big, there are a lot of products and it is *placed in a cetric* area of the town.” (L1 Spanish, CELSV, 2003)

\* “This (my grandpa’s private garden) is *placed in a quiet* place, in the countryside of north Italy.” (L1 Italian, CELSV, 2003)

L1 Korean speakers:

\* “Otherwise, in the middle of the island, reception building is *located in the central* of land between Eastern accommodation and Western accommodation...” (IELTS ac, 2011)

\* “For tourist’s exciting pier is *located in the south* part of island.” (IELTS ac, 2011)

Firstly, the two instances of “**placed in a...**” in the other L1 speaker sub-corpus came from the CELSV exam, while the two instances of “**located in the...**” in the L1 Korean speaker sub-corpus were found in the IELTS. Thus, the learners’ choice of these node verbs may have been mediated by the task. Nevertheless, in all four examples, the node verbs were used in the past tense to indicate either the placement or the location of something.

Two instances were found in the case of phrases: “**compared to/with**” and “**followed by.**” In the other L1 speaker sub-corpus, there were five instances of “**compared to**” and three instances of “**followed by.**” On the contrary, in the L1 Korean speaker corpus, “**compared with**” appeared two times, and “**compared to**” and “**followed by**” appeared once each. With respect to errors, there were four and five errors in the other L1 speaker sub-corpus and L1 Korean sub-corpus, respectively. Errors in the other L1 speaker sub-corpus pertained to prepositions (e.g. “**advertised at the student magazine**”), article (e.g. “**found in the different places**”). In the L1 Korean sub-corpus, however, all errors pertained to prepositions (e.g., “**impressed for the delicious**”).

Lastly, as mentioned above, most of the examples fell in the **V n** pattern (30 in the L1 Korean sub-corpus and 31 in the other L1 speaker sub-corpus). In both groups, many of the patterns fell in either the **V in n** pattern or the **V by n** pattern (note that the **V by n** pattern is labeled under either **V by amount** or **V by -ing**). In the L1 Korean speaker sub-corpus, 13 instances of ‘in’ were found, whereas, in the other L1 speaker sub-corpus, 11 examples were found. Among them, “sold in,” “used in,” and “written in” appeared in both sub-corpora. In the case of the preposition ‘by,’ it occurred five times in the L1 Korean speaker sub-corpus while appearing six times in the other L1 speaker sub-corpus. There were no overlaps between the two

groups. Aside from the prepositions ‘in’ and ‘by,’ other prepositions include ‘about,’ ‘as,’ ‘at,’ ‘during,’ ‘from,’ ‘like,’ ‘to,’ and ‘with’ in the L1 Korean speakers’ sub-corpus. On the other hand, in the other L1 speaker’ sub-corpus, prepositions included ‘about,’ ‘around,’ ‘as,’ ‘at,’ ‘during,’ ‘into,’ ‘near,’ ‘over,’ ‘with,’ and ‘within.’

The following section in **Table 4.31** discusses the VVN IN DT JJ sequence found among C2 learners.

VVN IN DT JJ	C2 (other L1)	C2 (L1 Korean)
V <i>about</i> n	▪ Asked about the main	
V <i>as</i> n	▪ Regarded as a cost-effective	▪ Regarded as a foolish
V <i>by</i> n		▪ Bounded by the same
V <i>for</i> n	▪ Designed for the modern	
V <i>from</i> n		▪ Kept for a long ▪ Separated from a hungry
V <i>in</i> n	▪ Embedded in the final ▪ Included in the next	▪ Caught in a bloody
V <i>to</i> n	▪ Adapated (adapted) to the modern ▪ Related to a non-profit ▪ Related to the personal	▪ Applied to the big ▪ Contributed to the local ▪ Exposed to a life-long ▪ Inclined to the second ▪ Related to each other
V <i>with</i> n	▪ Linked with a certain	▪ Equipped with a few
V prep/adv	▪ Led to the bizarre ▪ Led by an old ▪ Led to a culinarc ▪ Led to the above ▪ Located at an isolated ▪ Located in a friendly	▪ Led to a major ▪ Lived in a big ▪ Located on the other
Phrase	▪ Compared to the good ▪ Followed by a slight	▪ Mode into the adjacent
Phrasal verb	▪ Settled in the English	
None	▪ Built by some gregarious ▪ Caused by the aggressive ▪ Caused by these low-flying ▪ Changed in the past ▪ Considered as a good ▪ Defined as the scientific ▪ Defined until a certain ▪ Doubled in the last ▪ Done by the foreign-invested ▪ Doubled in the last ▪ Erradicated (eradicated) from the whole ▪ Expanded to the whole	▪ Attend for the past ▪ Blocked by a low ▪ Born into a wealthy ▪ Built in a short ▪ Caused by the contaminated ▪ Continued for a few ▪ Developed for some dangerous ▪ Escaped through the other ▪ Faced with any other ▪ Forced to an early ▪ Held in the first ▪ Influenced by each other

	<ul style="list-style-type: none"> <li>▪ Faced with the same</li> <li>▪ Fathered by an Indian</li> <li>▪ Found in an old</li> <li>▪ Given at an early</li> <li>▪ Heard on the local</li> <li>▪ Preserved in a good</li> <li>▪ Published in a forthcoming</li> <li>▪ Published in a forthcoming</li> <li>▪ Published in the next</li> <li>▪ Sent to an exclusive</li> <li>▪ Set in the post-war</li> <li>▪ Situated near the main</li> <li>▪ Sold by the local</li> <li>▪ Solved in a different</li> <li>▪ Stressed by the frenetic</li> <li>▪ Tracked to the excessive</li> <li>▪ Tracked to the excessive</li> <li>▪ Written in the early</li> <li>▪ Written as a personal</li> </ul>	<ul style="list-style-type: none"> <li>▪ Intensified by an unexpected</li> <li>▪ Known for the great</li> <li>▪ Lost during the Korean</li> <li>▪ Taught in a Christian</li> <li>▪ Shocked with the empty</li> <li>▪ Situated on the top</li> <li>▪ Spoken in the real</li> <li>▪ Used as a sole</li> <li>▪ Portrayed as the ultimate</li> <li>▪ Protected in those myriad</li> <li>▪ Visit since a young</li> <li>▪ Reminded of this embarrassing</li> <li>▪ Seen as a bad</li> <li>▪ Seen in the green</li> <li>▪ Selected for the first</li> </ul>
<b>Error</b>	▪ Build at that old	
<b>Group total</b>	10	9

**Table 4.31** Categorization of the “VVN IN DT JJ” sequence in C2 level (L1 Korean vs. other L1 speakers)

Among a random sample of 50 examples from the VVN IN DT JJ sequence, ten grammar pattern groups were found in the other L1 speaker sub-corpus, while nine grammar pattern groups were found among L1 Korean speakers. In both groups, slightly more meaning groups were identified than those from the B2 proficiency level.

Grammar patterns such as **V *about* n** and **V *for* n** were found only in the other L1 speaker sub-corpus. On the other hand, groups such as **V *by* n** and **V *from* n** were found only among L1 Korea speaker sub-corpus. In the case of the **V *prep/adv*** pattern, six examples were found among other L1 speakers, while four were found in the L1 Korean speaker sub-corpus. Interestingly, two different node verbs were found among other L1 speakers (e.g., ‘led,’ ‘located’) and four in the L1 Korean speaker sub-corpus (e.g., ‘led,’ ‘lived,’ ‘located,’ ‘related’). With respect to the other patterns, examples from patterns such as **V *as* n** and **V *prep/adv*** contained node verbs that were present in both sub-corpora.

Similar to the B2 levels, the majority of the examples were categorized in the 'none' category (31 from other L1 speaker sub-corpus and 27 from the L1 Korean speaker sub-corpus). Among them, ten different prepositions (as, at, by, from, in, near, on, to, until, with) and 26 node verbs were found in the other L1 speaker sub-corpus. On the other hand, a wider range of 12

prepositions (as, by, during, for, in, into, of, on, since, through, to, with) and 27 node verbs were found in the L1 Korean sub-corpus. There were a few overlaps between the two sub-corpora, such as 'caused by,' 'situated near/on,' and 'faced with.' Most of the node verbs in both groups were topic-specific (e.g., "**fathered by an Indian,**" "**lost during the Korean**").

#### 4.6 How do verb 4-gram sequences differ between L1 Korean speakers and L1 English speakers?

So far, the analyses have revealed that there is a clear difference in the use of both the verb sequences and the grammar patterns among the top two core and emerging sequences in both the L1 Korean speakers and other L1 speakers. This section explores the top 25 verb 4-gram sequences that are present in the British National Corpus (BNC). As a reminder, the academic subset of the BNC was chosen as a suitable reference for this study, following the myriad of studies that have also used the corpus as a proxy for L1 English, owing to its immense data and robustness. The main purpose of comparing L1 Korean speakers and L1 English speakers is to explore whether there is any resemblance between near-native L1 Korean speakers with L1 English speakers. **Table 4.32** displays the top 25 verb 4-gram sequences in the BNC corpus and compares the frequency of the sequences (per mil. words) with the C2 level from the L1 Korean sub-corpus and other L1 speaker sub-corpus.

Tag	Rank (BNC)	Per million words	C2 level rank (L1 Korean)	Per million words	C2 level rank (other L1)	Per million words
VVN IN DT NN	1	1318.33	2	1430.01	2	9152.82
VV DT NN IN	2	807.00	1	1677.17	1	10542
VVN IN DT JJ	3	668.75	6	865.07	7	4632.68
VVD DT NN IN	4	513.09	7	847.41	16	2532.28
VV DT JJ NN	5	511.92	4	1147.54	3	7172.42
VV IN DT NN	6	432.40	3	1200.50	4	6446.3
VVD IN DT NN	7	408.72	11	547.29	13	2960.09
VB VVN IN DT	8	398.90	23	406.05	11	3099.08
VVG IN DT NN	9	383.57	5	918.03	6	4728.96
VVG DT NN IN	10	383.19	12	529.63	9	3582.48
VBD VVN IN DT	11	373.06	14	441.36	45	1564.79
VVD DT JJ NN	12	354.02	10	600.25	32	1882.77
VBZ DT JJ NN	13	320.35	5	918.03	5	5651.70
VVN IN DT NP	14	306.73	N/A	N/A	192	580.36
VBZ DT NN IN	15	279.66	11	547.29	8	3924.19

VVG DT JJ NN	16	265.56	11	547.29	15	2708.56
VVN IN JJ NNS	17	243.54	10	600.25	21	2156.00
VBD DT JJ NN	18	224.65	5	918.03	29	1958.70
VVZ DT NN IN	19	215.93	29	176.54	28	2011.59
VBZ VVN IN DT	20	214.20	23	282.47	39	1662.42
VVN TO VV DT	21	213.96	20	353.09	51	1457.67
VVN DT NN IN	22	207.58	26	229.51	1662	97.63
VVN IN DT NNS	23	205.31	27	211.85	37	1690.90
VVN IN JJ NN	24	200.82	28	194.20	98	956.64
VVN IN NN IN	25	197.11	32	123.58	115	837.99
...						
VVN IN PPS NN	40	138.54	8	706.18	17	2457.02

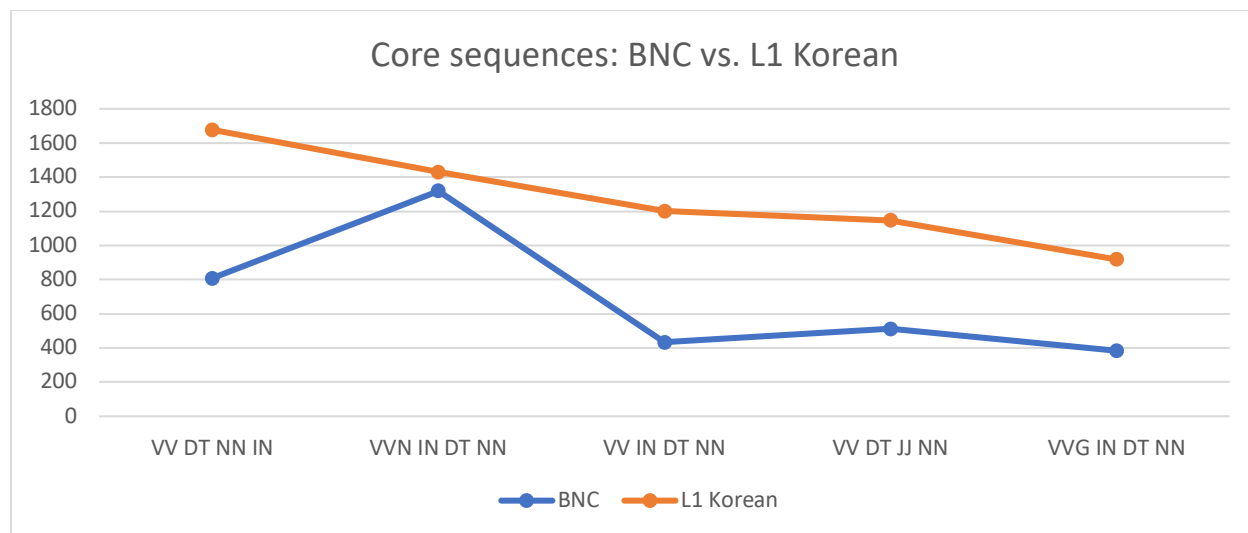
**Table 4.32** Comparison of the top 25 verb 4-gram sequences in the BNC and C2 level in the L1 Korean sub-corpus and other L1 speakers

So far, the analyses have revealed that there is a clear difference in the use of both the verb sequences and the grammar patterns among the top two core and emerging sequences in both the L1 Korean speakers and other L1 speakers. This section explores the top 25 verb 4-gram sequences that are present in the British National Corpus (BNC). As a reminder, the academic subset of the BNC was chosen as a suitable reference for this study, following the myriad of studies that have also used the corpus as a proxy for L1 English, owing to its immense data and robustness. The main purpose of comparing L1 Korean speakers and L1 English speakers is to explore whether there is any resemblance between near-native L1 Korean speakers with L1 English speakers. **Table 4.32** displays the top 25 verb 4-gram sequences in the BNC corpus and compares the frequency of the sequences (per mil. words) with the C2 level from the L1 Korean sub-corpus and the other L1 speaker sub-corpus.

Interestingly, the top five sequences in the L1 Korean sub-corpus, also known as the core sequences, ranked within the top 10 in the BNC corpus as well. In fact, three of the sequences in the BNC (VVN IN DT NN, VV DT NN IN, VV DT JJ NN) also ranked within the top five among near-native L1 Korean speakers. On the other hand, the VV IN DT NN and VVG IN DT NN sequences ranked sixth and ninth, respectively. Even still, the difference in ranking between the BNC sub-corpus and the L1 Korean sub-corpus was three and four ranks. Additionally, the VVN IN DT NN and VV DT NN IN sequences ranked first and second in the BNC sub-corpus, while they ranked second and first in both the L1 Korean and other L1 speaker sub-corpus respectively. Although the sequences between the L1 Korean speakers and other L1 speakers were more similar than that of the BNC, this could be due to topic differences and the nature of

the data. Despite these differences, there is evidence that the core sequences in the L1 sub-corpus were also ranked high in the BNC corpus.

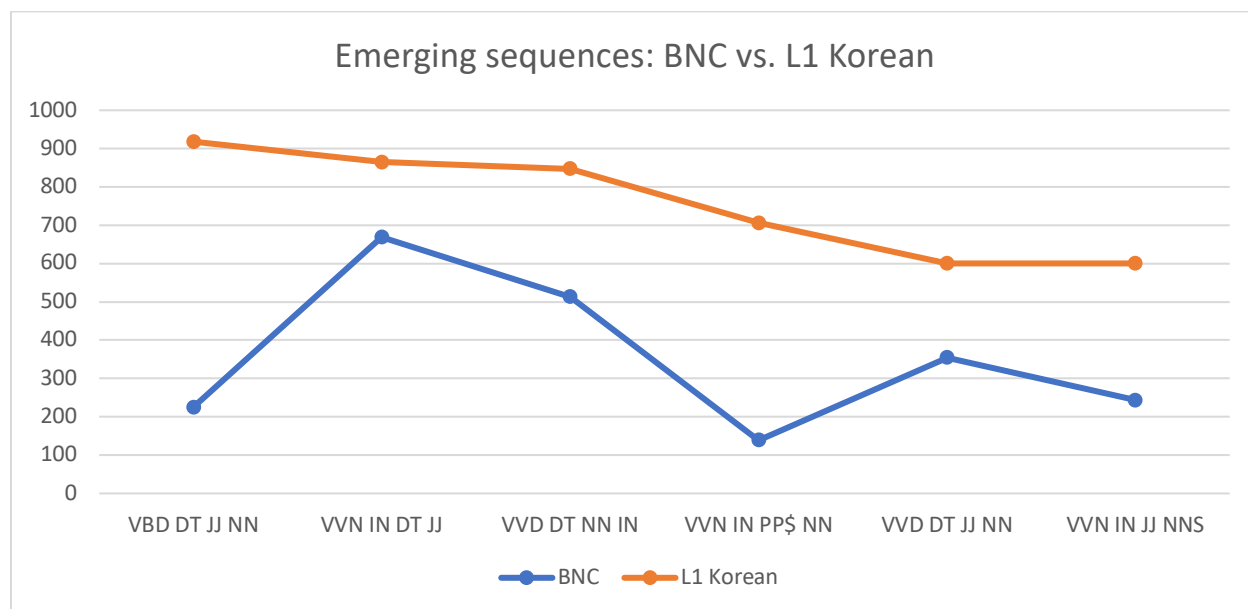
On the contrary, there was quite a bit of dispersion among emerging sequences. The top two sequences, VVN IN DT JJ and VVD DT NN IN, stayed within the top ten in the BNC and the L1 Korean sub-corpus, although the VVD DT NN IN sequence ranked 16 in the other L1 speaker sub-corpus. In the case of the VVD DT JJ NN sequence, it ranked 12th in the BNC academic sub-corpus while ranking 10th in the L1 Korean speaker sub-corpus. However, in the VBD DT JJ NN, VVN IN PP\$ NN, VVD DT JJ NN, and VVN IN JJ NNS sequences, the gap in the ranking between the two sub-corpora was even more significant. For instance, the VVN IN JJ NNS sequence ranked 10th in the L1 Korean sub-corpus while ranking 17th in the BNC sub-corpus. **Figure 4.4** displays the graph that compare the normalized frequency of the core sequences between the BNC sub-corpus and the L1 Korean sub-corpus.



**Figure 4.4** Core sequences: BNC vs. L1 Korean speakers

As mentioned above, there were differences in the first two sequences between L1 English speakers and L1 Korean speakers, where in the L1 Korean sub-corpus, the VV DT NN IN sequence ranked first, while in the BNC sub-corpus, the VVN IN DT NN sequence ranked first. However, other than the fact that the normalized frequency was generally higher among L1 Korean speakers, the general trend in the ranking of the five sequences was identical for core sequences.

On the contrary, the dispersion was even more significant in the VBD DT JJ NN sequence, which ranked 19th in the BNC while ranking fifth among L1 Korean speakers. Interestingly, the same sequence ranked 29th among other L1 speakers. Lastly, the VVN IN PP\$ NN sequence ranked 40th in the BNC sub-corpus while ranking eighth in the L1 Korean sub-corpus and 17th in the other L1 speaker sub-corpus. **Figure 4.5** illustrates the normalized frequency of the emerging sequences in the BNC and L1 Korean sub-corpus.



**Figure 4.5** Emerging sequences: BNC vs. L1 Korean speakers

Unlike the core sequences, there is much more dispersion among the emerging sequences. In particular, the VBD DT JJ NN and VVN IN PP\$ NN sequences in the BNC ranked far below L1 Korean speakers. Since these sequences were not among the top two emerging sequences, they are not analyzed further in this study. However, it should be noted that although the top most frequently occurring sequences were similar among both L1 English speakers and L1 Korean speakers, as well as among other L1 speakers, the difference dispersed greatly further down in the ranking.

#### 4.6.1 A Comparison of the Core Sequences Between L1 Korean Speakers and L1 English Speakers

The comparison of the top 25 sequences among L1 English, L2 Korean, and other L1 speaker sub-corpora was similar (with the exception of the VVD DT NN IN sequence in the other L1 speaker sub-corpus), particularly in the core sequences and the top emerging sequences. This section takes a closer look at the core sequence, VV DT NN IN, which ranked second in the BNC and first among L1 Korean speakers. **Table 4.33** illustrates the comparison of the grammar patterns between the BNC and the near native-speaker proficiency level of L1 Korean speakers, the C2 level.

VV DT NN IN	BNC	C2 (L1 Korean)
V n <i>for</i> n		▪ Prepare a gift for
V n <i>from</i> n	▪ Protect the body from	▪ Prevent the rent from
V n <i>in</i> n	▪ Place the incident in	
V n <i>to</i> n		▪ Put the blame on ▪ Send all correspondence to
V n prep/adv	▪ Turn the lens on	
Phrase	▪ Make a habit of ▪ Stand the test of	▪ Judge a book by ▪ Strike the balance between
Collocation	▪ Enlist the help of ▪ Raise the question of ▪ Watch the video of	▪ Double the amount of ▪ Enhance the quality of ▪ Enjoy the feeling of ▪ Explain the reason of ▪ Face the problem of ▪ Give a guidance to ▪ Impede the progress of ▪ Read the article with ▪ Write a letter to
<b>Group total</b>	<b>5</b>	<b>5</b>

**Table 4.33** Categorization of the “VV DT NN IN” sequence in C2 level (BNC vs. L1 Korean speakers)

The categorization of the grammar patterns excludes the V n pattern, which will be discussed below. As can be seen from the table, eight sequences in the BNC sub-corpus belonged in complex patterns (e.g., V n *from* n) or phrase/collocation. On the contrary, 15 sequences in the C2 levels in the L1 Korean speak sub-corpus were categorized under either complex patterns or phrase/collocation. In both groups, five pattern groups were identified. The majority of the examples in the L1 Korea speaker sub-corpus were found in collocations (9 examples), whereas only three were found in the BNC sub-corpus. There were two phrases in

each of the two sub-corpora. With respect to the sequences, **V n in n** and **V n prep/adv** was found only in the BNC sub-corpus, whereas **V n for n** and **V n to n** were found only among the L1 Korean sub-corpus. The **V n from n** pattern was found in both groups. Below are the concordance lines:

L1 English speakers:

\* “Eventually there are not enough T4 cells to *protect the body from* infection.”  
(selection of leaflets from T Higgins Trust, 1985-1993)

L1 Korean speakers:

\* “It is also important to *prevent the rent from* rising annually regardless of economic growth.” (CPE, 1999)

In the case of the first example, the node verb ‘protect’ is followed by an object (‘body’) + preposition (‘from’) + noun. With respect to the example from L1 Korean speakers, the node verb ‘prevent’ is followed by an object (‘rent’) + preposition (‘from’) + gerund. The following table (Table 4.34) scrutinizes the meaning groups of the examples categorized under the **V n** pattern.

VV DT NN IN	BNC	C2 (L1 Korean)
The ‘approve’ group	<ul style="list-style-type: none"> <li>▪ Accept a variety of</li> <li>▪ Promote a sense of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Support the idea about</li> </ul>
The ‘beat an illness’ group		<ul style="list-style-type: none"> <li>▪ Improve every aspect of</li> </ul>
The ‘become’ group	<ul style="list-style-type: none"> <li>▪ Become a trainer in</li> <li>▪ Become the basis for</li> </ul>	
The ‘break a record’ group		<ul style="list-style-type: none"> <li>▪ Find the immunity of</li> <li>▪ Find the story in</li> <li>▪ Reach that level at</li> </ul>
The ‘bring’ group	<ul style="list-style-type: none"> <li>▪ Fit the batten above</li> </ul>	<ul style="list-style-type: none"> <li>▪ Produce some form of</li> </ul>
The ‘build’ group	<ul style="list-style-type: none"> <li>▪ Place an order on</li> <li>▪ Prepare a pamphlet on</li> </ul>	<ul style="list-style-type: none"> <li>▪ Leave any mark on</li> </ul>
The ‘buy’ group	<ul style="list-style-type: none"> <li>▪ Collect the man of</li> <li>▪ Get a sleep if</li> <li>▪ Share the gleaning of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Get a job in</li> </ul>
The ‘change’ group	<ul style="list-style-type: none"> <li>▪ Improve the setting of</li> <li>▪ Reduce the length of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Open another set of</li> </ul>

	<ul style="list-style-type: none"> <li>▪ Reduce the power of</li> <li>▪ Train a student in</li> </ul>	
<b>The ‘cover’ group</b>		<ul style="list-style-type: none"> <li>▪ Hide the redness of</li> </ul>
<b>The ‘describe’ group</b>		<ul style="list-style-type: none"> <li>▪ Describe the place because</li> <li>▪ Discuss the relationship between</li> <li>▪ Express the ideal of</li> </ul>
<b>The ‘develop’ group</b>		<ul style="list-style-type: none"> <li>▪ Assume the rest of</li> </ul>
<b>The ‘do’ and ‘take’ group</b>	<ul style="list-style-type: none"> <li>▪ Do that movement because</li> <li>▪ Find the time for</li> </ul>	<ul style="list-style-type: none"> <li>▪ Hold a festival because</li> <li>▪ Take a couple of</li> <li>▪ Take a glimpse of</li> <li>▪ Take another adventure in</li> <li>▪ Throw a party for</li> </ul>
<b>The ‘end’ group</b>	<ul style="list-style-type: none"> <li>▪ Complete the job of</li> </ul>	
<b>The ‘form’ group</b>	<ul style="list-style-type: none"> <li>▪ Order the creation of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Create a sense of</li> <li>▪ Make any excuse in</li> </ul>
<b>The ‘hear’ group</b>	<ul style="list-style-type: none"> <li>▪ Feel the pressure of</li> <li>▪ Hear the tick of</li> </ul>	<ul style="list-style-type: none"> <li>▪ Consider the compart of</li> <li>▪ Experience the nourishment of</li> <li>▪ Feel the touch of</li> <li>▪ Forget this experience since</li> <li>▪ See the correlation between</li> </ul>
<b>The ‘interest’ group</b>	<ul style="list-style-type: none"> <li>▪ Attract the interest of</li> </ul>	
<b>The ‘kill’ group</b>	<ul style="list-style-type: none"> <li>▪ Remove any trace of</li> </ul>	
<b>The ‘start’ and ‘stop’ group</b>	<ul style="list-style-type: none"> <li>▪ Avoid the opprobrium of</li> <li>▪ Join the plan before</li> <li>▪ Start the programme with</li> </ul>	
<b>The ‘watch’, ‘break a record’, and ‘approve’ groups- verbs with other meanings</b>		<ul style="list-style-type: none"> <li>▪ Enter a universe before</li> <li>▪ Experience the making of</li> <li>▪ Replace the pleasure of</li> </ul>
<b>The ‘turn a corner’ group</b>	<ul style="list-style-type: none"> <li>▪ Score a century on</li> </ul>	
<b>The ‘use’ group</b>	<ul style="list-style-type: none"> <li>▪ Use the toilet during</li> </ul>	
<b>None</b>	<ul style="list-style-type: none"> <li>▪ Assess the extent to</li> <li>▪ Complement the showing of</li> <li>▪ Connect the pattern of</li> <li>▪ Cure the sickness of</li> <li>▪ Finance a range of</li> <li>▪ Grab a couple of</li> <li>▪ Implement the directive by</li> <li>▪ Induce a goalkeeper by</li> <li>▪ Obviate the need for</li> <li>▪ Recapture the feeling of</li> <li>▪ Secure a share of</li> <li>▪ Shut the window between</li> <li>▪ Specify the end of</li> <li>▪ Specify the number of</li> <li>▪ Tarnish the gift with</li> </ul>	<ul style="list-style-type: none"> <li>▪ Captivate the audience as</li> <li>▪ Endure this experience for</li> <li>▪ Execute the budget of</li> <li>▪ Govern a nation of</li> <li>▪ Pose some decrease in</li> <li>▪ Renew the soul through</li> </ul>
<b>Group total</b>	16	15

**Table 4.34** Meaning groups of the “VV DT NN IN” sequence in C2 level (BNC vs. L1 Korean speakers)

In the BNC sub-corpus, 42 of the 50 grammar patterns belonged to the **V n** pattern, whereas 35 of the 50 patterns in the L1 Korean speaker sub-corpus were categorized under the **V n** pattern. With respect to the meaning groups, 16 meaning groups were found in the BNC sub-corpus, and 15 meaning groups were in the L1 Korean speaker sub-corpus. Among these patterns, some were found only in the BNC sub-corpus (e.g., the ‘become,’ ‘end,’ ‘interest,’ ‘kill,’ ‘start’ and ‘stop,’ ‘turn a corner,’ and ‘use’ group) while others were found only in the L1 Korean speaker sub-corpus (e.g., the ‘beat an illness,’ ‘break a record,’ ‘cover,’ ‘describe,’ ‘develop,’ and ‘watch,’ ‘break a record,’ and ‘approve’ groups- verbs with other means’ group). Some meaning groups were found in both groups (e.g., the ‘approve,’ ‘bring,’ ‘build,’ ‘buy,’ ‘change,’ ‘do’ and ‘take,’ ‘form’ and ‘hear’ group). The side-by-side exploration of the two groups demonstrates that not only are there fewer repetitions, but the language is more formulaic in nature. For instance, in the BNC sub-corpus, examples such as “**promote a sense of,**” “**fit the batten above,**” and “**score a century**” illustrate that the meaning of the verbs is embedded within the phrase rather than on the node verb. Similarly, in the L1 Korean sub-corpus, phrases such as “**hold a festival**” and “**take a glimpse**” illustrate that the meaning of the node verbs cannot be explained without the surrounding words.

Lastly, 15 examples in the BNC sub-corpus were categorized under ‘none’ as opposed to six examples in the L1 Korean sub-corpus. This indicates that L1 English speakers used less frequent node verbs compared to L1 Korean speakers.

#### 4.6.2 A Comparison of the Emerging Sequences Between L1 Korean Speakers and L1 English Speakers

This section discusses the VVN IN DT JJ sequence, an emerging sequence that ranked third in the BNC sub-corpus while it ranked sixth in the L1 Korean sub-corpus. **Table 4.35** illustrates the grammar pattern categorization of random 50 samples pulled from the BNC and the C2 level L1 Korean sub-corpus.

VVN IN DT JJ	BNC	C2 (L1 Korean)
V about n	▪ learnt about the new	
V as n		▪ Regarded as a foolish

<b>V by n</b>		▪ Bounded by the same
<b>V from n</b>		▪ Kept for a long ▪ Separated from a hungry
<b>V in n</b>	▪ Included in a protective ▪ Indulged in the ages-old	▪ Caught in a bloody
<b>V into n</b>	▪ Etched in a week	
<b>V to n</b>	▪ Adjusted to a 24-hour ▪ Added to each eligible ▪ Left to the private ▪ Married to a successful ▪ Transferred to the south	▪ Applied to the big ▪ Contributed to the local ▪ Exposed to a life-long ▪ Inclined to the second ▪ Lived in a big ▪ Located on the other ▪ Related to each other
<b>V with n</b>	▪ Faced with this clever ▪ Punctuated with these abrupt	▪ Equipped with a few
<b>V prep/adv</b>	▪ Led to the gradual ▪ Passed by the armed ▪ Passed into a pleasant ▪ Stumbled through the shallow	▪ Led to a major
<b>Phrasal verb</b>	▪ Fitted into that pre-ordained	▪ Moved into the adjacent
<b>None</b>	▪ Accumulated in the late ▪ Blocked by a downward ▪ Built by the Athenian ▪ Caused by the extra ▪ Chosen by a postal ▪ Described as a priceless ▪ Described as a significant ▪ Elected for a one-year ▪ Expected in the Permo-Triassic ▪ Frozen into an eternal ▪ Implemented in the near ▪ Initiated in the psychiatric ▪ Led by the deputy ▪ Located for all major ▪ Looked in those last ▪ Made for the same ▪ Needed in the armed ▪ Occupied in the social ▪ Produced by each domestic ▪ Published in the Philosophical ▪ Sealed behind a transparent ▪ Seen as the main ▪ Specified by a maximum ▪ Spoken by the old ▪ Spread in the latter ▪ Stimulated by a blank ▪ Succeeded by a thin ▪ Taken at an early ▪ Taken by the administrative ▪ Used as a liquid ▪ Used as a mild ▪ Used in the initial	▪ Attend for the past ▪ Blocked by a low ▪ Born into a wealthy ▪ Built in a short ▪ Caused by the contaminated ▪ Continued for a few ▪ Developed for some dangerous ▪ Escaped through the other ▪ Faced with any other ▪ Forced to an early ▪ Held in the first ▪ Influenced by each other ▪ Intensified by an unexpected ▪ Known for the great ▪ Lost during the Korean ▪ Taught in a Christian ▪ Shocked with the empty ▪ Situated on the top ▪ Spoken in the real ▪ Used as a sole ▪ Portrayed as the ultimate ▪ Protected in those myriad ▪ Visit since a young ▪ Reminded of this embarrassing ▪ Seen as a bad ▪ Seen in the green ▪ Selected for the first

	<ul style="list-style-type: none"> <li>▪ Won by the nationalist</li> <li>▪ Written by the 30-year-old</li> </ul>	
<b>Group total</b>	<b>8</b>	<b>9</b>

**Table 4.35** Categorization of the “VVN IN DT JJ” sequence in C2 level (BNC vs. L1 Korean speakers)

In the VVN IN DT JJ sequence, eight grammar pattern groups were found in the BNC and nine groups in the L1 Korean sub-corpus. Patterns such as **V in n**, **V to n**, **V prep/adv**, and **V with n** were found in both groups, whereas **V about n** and **V into n** was only found in the BNC group. Furthermore, patterns such as **V as n**, **V by n**, and **V from n** were found only in the L1 Korean sub-corpus.

One phrasal verb was found in each of the two groups, but the majority of the examples belonged to the ‘none’ group (34 examples in the BNC, 27 instances in the L1 Korean sub-corpus). Among them, examples such as “**blocked by**,” “**caused by**,” and “**seen as**” were found in both groups. In the BNC sub-corpus, seven different prepositions were present (e.g., as, at, behind, by, for, in, into), and 12 prepositions were found in the L1 Korean sub-corpus (e.g., as, by, during, for, in, into, of, on, since, though, to, with). Furthermore, a wide range of node verbs was found in both sub-corpora with no repetition found in the L1 Korean sub-corpus and a few found in the BNC (e.g., ‘described,’ ‘used’).

Based on the observation of the grammar patterns in the BNC and the C2 level L1 Korean sub-corpus, it is evident that C2 level L1 Korean learners’ use of verb sequences does not differ drastically from that of L1 English speakers. Below are two examples taken from the concordance lines to identify whether there are differences in other usages.

Other L1 speakers:

\* “These twin processes **led to the gradual** displacement of people from rural areas.”  
(Written books and periodicals, 1986)

\* “Because of this, work was **transferred to the south** coast and then to Malvern.”  
(Written books and periodicals, 1992)

L1 Korean speakers:

\* “This *led to a major* debate whether the government should provide free university education.” (IELTS ac, 2010)

\* “Therefore, I can proudly say that 85per cent of the income raised by the Charity Day was *contributed to the local* hospital.” (CAE, 2000)

The first example from both groups was headed by the node verb ‘led’ followed by a prepositional phrase (‘to’) and a noun group. In the second example, both node verbs are in the passive form, also followed by a prepositional phrase. The concordance lines display that the L1 Korean speakers’ use of verb sequences at the highest level does indeed resemble that of L1 English speakers.

#### 4.7 Summary

This chapter explored the usage of verb 4-gram sequences among L1 Korean speakers and compared them to that of both other L1 speakers and L1 English speakers. Among L1 Korean speakers, some main differences between beginner (A2) and advanced learners (C1 and C2) were found. Firstly, learners in the higher proficiency levels demonstrated their ability to utilize more idiomatic and formulaic expressions such as “go for a walk.” In contrast, in the lower level (e.g., A2), the words were used more repetitively and directly (e.g., “go to a meeting”). Furthermore, the variety in the language in the higher proficiency levels illustrates that as learners become more proficient, so does their use of language as they have access to more words and expressions. In addition, it is also clear that language does not evolve in a linear fashion. Based on the categorization of the samples into grammar patterns, learners, even at the lowest level, have some command of phrasal verbs and collocations. The difference, however, could like in the degree of schematicity of the phrases; however, this did not lie within the scope of this study.

The second part of the study scrutinized the L1 Korean data in conjunction with other L1 speaker sub-corpus in the CLC. The result highlighted that L1 transfer was visible among L1 Korean speakers, particularly with regard to learners’ use of prepositions and articles, the two most difficult concepts to acquire for L1 Korean speakers even at the most advanced level. In addition, the side-by-side comparison also revealed that the core evidence as coined based on the

L1 Korean sub-corpus was also evident in the other L1 speaker sub-corpus. Thus, it is clear that the most commonly used 4-gram verb sequences were not specific to L1 Korean speakers but most frequent across all L2 learners. On the other hand, the same did not appear for emerging sequences, thereby illustrating that the emergence of some of the verb sequences may indeed be specific to L1 Korean speakers.

Finally, the last part of the study compared L1 Korean speaker data with L1 English speakers. The comparison was conducted by consulting the academic written portion of the BNC. Interestingly, similar to findings from the comparison of L1 Korean speakers and other L1 speaker sub-corpora, the core sequences were found among L1 English speakers as well. Thus, the most commonly used verb sequences may not be specific to L2 learners but common across all English users. Similarly, however, the same did not hold for emerging sequences. The following section discusses the main findings with respect to empirical research.

## Chapter 5 Discussion

### 5.1 Introduction

Underpinned by the usage-based approaches and claims that language restructures and develops as the learners' experience with the language expands (Ellis, Römer, O'Donnell, 2016; Pérez-Paredes et al., 2020; Tomasello, 2003), the current study looked at the changes in the frequency of language use and attempted to make sense of the process of language restructuring and development. In this study, the L2 writing development of L1 Korean speakers was explored using a bottom-up approach.

Many of the studies to date have scrutinized specific types of linguistic features such as collocations (Chang, 2018; Reppen, 2009), lexical bundles (Bychkovska & Lee, 2017; Chen & Baker, 2010), and constructions (Mostafa & Crossley, 2020; Römer, 2019) to name a few. While zoning into them individually has the merit of providing a clear pattern of the overuse, underuse, and misuse of the linguistic features, these findings are often limited to the specific linguistic feature being investigated. This is one of the limitations of using a top-down approach as it concentrates on a particular aspect of language. Instead, an investigation of all surfacing linguistic features would make it possible to draw a broad picture of the process of language development, which would also bring to the surface the learners' use of various linguistic features at the level of both individual lexical items as well as chunks. That being said, this study adopted a bottom-up approach to provide a holistic perspective of language use in L2 writing. More specifically, it sought to illustrate the developmental process as well as features of interlanguage as L1 Korean speakers become more proficient L2 writers. This study examines the developmental pattern by looking at POS n-grams. More specifically, it examines the most frequently occurring verb 4-gram sequences as well as emerging sequences. The 4-gram sequences encompass a wide range of linguistic features and could include a random sequence of words with punctuation in between (e.g., VV PP SENT PP -> “**invite you. I**”; “**show you. I**”). Since the purpose of this study was to gain a bird's eye view of language use and restructuring, it was expected to include such sequences. Regardless, these sequences did not appear to be part of either the 'core' or 'emerging' sequences.

The main goals of this study are twofold. Firstly, it seeks to identify the verb 4-gram patterns across the proficiency levels. This was achieved by drafting and analyzing the most

commonly used verbs among the core and emerging sequences in each proficiency level. The second goal is to explore the processes and the restructuring of learner language. These claims are built around the myriad of studies that have shown that language is sensitive to frequency, where the learners' experience with the language is highly relevant when it comes to language acquisition (Bybee, 2010; Ellis, 2012; Tomasello, 2003). The current study contributes to the ongoing quest for understanding the linguistic patterns of interlanguage at different trajectories of development. Perhaps one of the most significant contributions of this study to learner corpus research literature is the use of the CEFR to benchmark learners' proficiency, an international standard for measuring language ability.

The following sections discuss the findings with respect to empirical research. The findings are discussed in the order they were presented in the results section. It first examines the findings from L1 Korean speakers with respect to the core and emerging sequences across the six proficiency levels, the node verbs, and their correlation with proficiency level. It also discusses the encoding of the sequences of patterns. The second section then shifts the attention to the differences in the verb sequences between L1 Korean speakers and other speakers of L1, where the top 24 verb 4-grams are compared. Next, the top core and emerging sequences are compared between L1 Korean speakers and other L1 speakers based on the categorization of grammar patterns. Finally, the last section discusses the findings from the comparison between L1 Korean speakers at the C2 level and L1 English speakers.

## **5.2 L1 Korean Speakers**

This section discusses the findings pertaining to the L1 Korean speakers. It begins by revisiting the top 25 verb 4-gram sequences across the six proficiency levels. The discussion then shifts to the node verbs and strives to understand the results from the current study based on empirical research. Finally, the section concludes by pinpointing the main findings of the grammar patterns and illustrating the prevalent language patterns at different language trajectories.

### **5.2.1 Ranking of 4-gram Verb Sequences**

The first part of the discussion takes place around the top verb 4-gram sequences across the proficiency levels. The normalized frequency has a tendency to Zipfian distribution (Zipf,

1935) in that the sequences among the top occurred much more frequently than the subsequent sequences. **Table 5.1** illustrates the frequency distribution of the top 10 sequences across the A2 to C2 levels.

Rank	A2 Tag	Norm. Freq.	B1 Tag	Norm. Freq.	B2 Tag	Norm. Freq.	C1 Tag	Norm. Freq.	C2 Tag	Norm. Freq.
1	VBP VVG TO VV	2394.81	VV DT NN IN	1654.25	VV DT NN IN	1766.48	VVN IN DT NN	1881.25	VV DT NN IN	1677.17
2	VV DT NN IN	1879.47	VV IN DT NN	1083.24	VVN IN DT NN	1359.15	VV DT NN IN	1660.41	VVN IN DT NN	1430.01
3	VV IN DT NN	1606.64	VVN IN DT NN VV DT JJ NN	822.93	VV IN DT NN	1138.86	VV IN DT NN	1431.39	VV IN DT NN	1200.50
4	VV PP IN PP\$	1394.45	VVG IN DT NN	734.76	VV DT JJ NN	964.29	VV DT JJ NN	1292.34	VV DT JJ NN	1147.54
5	VV DT JJ NN VBZ RB JJ SENT	1333.82	VBP VVG TO VV	671.78	VVG IN DT NN	860.38	VVG IN DT NN	867.01	VVG IN DT NN VBZ DT JJ NN	918.03
6	VV DT NN SENT	1303.50	VV DT NN SENT VBZ RB JJ IN	638.19	VBZ DT JJ NN	710.75	VVN IN DT JJ	850.65	VVN IN DT JJ	865.07
7	VV PP RB SENT	1212.56	VBZ DT JJ NN	554.22	VV DT NN SENT	598.52	VBZ DT JJ NN	744.32	VVD DT NN IN	847.41
8	VV PP\$ NN SENT	1182.25	VV PP\$ NN SENT	524.83	VVP DT NN IN	532.02	VBZ RB JJ IN	629.81	VVN IN PP\$ NN	706.18
9	VVP TO VV IN VVP PP MD VV	1060.99	VV IN PP\$ NN	520.63	VV PP\$ NN IN	507.08	VVD DT NN IN	621.63	VV PP\$ NN IN	653.21
10	VBZ RB JJ IN	1000.36	VVP IN DT NN VV PP\$ NN IN	508.63	VBP VVG TO VV	502.93	VV DT NN SENT	580.73	VVN IN JJ NNS VVD DT JJ NN VBD DT JJ NN	600.25

**Table 5.1** Normalized frequency distribution of the top 10 sequences among L1 Korean speakers

The verb sequence in the top rank occurred much more frequently than in the second-highest rank in all five proficiency levels. Furthermore, the frequency dropped significantly in proportion to the ranking. In addition to the distribution showing Zipfian tendency, such a claim is in line with various studies, which have illustrated that, in usage, the most frequent words and phrases account for the majority of the tokens (Ninio, 2006; O'Donnell & Ellis, 2010). In a similar vein, the core sequences, which occurred most frequently across all of the proficiency levels, provide support to the current usage-based understanding of language learning in that the most noticeable or commonly used language items are acquired first (Barlow & Kemmer, 2000; Bybee, 2008; Ellis, 2008; Goldberg, 2006; Langacker, 2000; Tomasello, 2003). As will be

discussed later, these patterns are prevalent even among other L1 speakers, as well as L1 English speakers, which further illustrates that the most commonly occurring sequences do not pertain to L1 Korean speakers only but that they are prevalent across all English speakers.

One point should be made clear, however, regarding beginner learners. The current study illustrated that the core and emerging sequences did not surface the way they did so for intermediate (B1 and B2 levels) and advanced (C1 and C2 levels) learners. In fact, the written production rubric of the CEFR states that at the A1 level, learners “can write simple isolated phrases and sentences” (Council of Europe, 2001). At the A2 level, learners “can write a series of simple phrases and sentences linked with simple connectors like ‘and,’ ‘but,’ and ‘because’” (Council of Europe, 2001), which contrast drastically with the B1 level where learners “can write straightforward connected text on topics, which are familiar, or of personal interest” (Council of Europe, 2001). Thus, it may be the case that the beginner learners simply did not have the linguistic capability to produce tangible sentences. Alternatively, it could be due to the lack of data at the A2 level, although it will be illustrated later why this is not likely the case.

Overall, the patterns (e.g., core and emerging sequences) of the verb sequences identified in this study illustrate that language use is not random but subject to patterns where there must be some association between the words (Hunston & Francis, 2000) and syntax (Sinclair, 1991). For instance, the frequent use of the VV IN DT NN sequence across the proficiency level indicates that some verbs are used more frequently used with a preposition rather than others (e.g., VV DT NN IN). What is unknown about these patterns based on the sequences alone is the extent to which these patterns are formulaic and whether the same level of formulaicity is prevalent across all levels. The following sections will further disentangle the ways in which the sequences identified in this study can tell us about usage and pattern.

The discussion will now shift more specifically towards the core and emerging sequences. To reiterate, the core sequences in this study consisted of the following:

VV DT NN IN (verb-determiner-noun-preposition)

VV IN DT NN (verb-preposition-determiner-noun)

VV DT JJ NN (verb-determiner-adjective-noun)

VVN IN DT NN (verb [past participle]- preposition-determiner-noun)

VVG IN DT NN (verb [gerund/present participle]-preposition-determiner-noun)

As discussed in Chapter 4, three of the five core sequences (VV DT NN IN, VV IN DT NN, VV DT JJ NN) were headed by simple present verbs, followed by either a determiner or a preposition and a noun phrase. Interestingly, these three sequences were the only core sequences found among L2 learners in the L1 Korean sub-corpus. This finding is in line with empirical research, which has shown that past tense verbs are particularly challenging to L2 learners (Hinkel, 1997), and that they appear later in the L2 development of verb morphemes (Dulay, Burt, & Krashen, 1982; Ellis, 1994). Furthermore, in the developmental trajectory, verbs initially surface as uninflected forms (Housen, 1995, 2002). Similarly, past tense verbs in this study didn't surface until the B1 level, which lends support to the notion that past tense verbs may be difficult at the beginning levels. This is not to claim, however, that learners at the beginner level were 'incapable' of producing sentences using simple past verbs. Rather, they were ranked much lower in frequency compared to verbs in the simple present. That being said, learners at lower levels are also capable of using different verb forms. For example, O'Keeffe and Mark (2017) illustrated through their exploration of the CLC that learners at the beginner level (L1) were able to utilize the past simple affirmative form competently, albeit with limited usage due to their vocabulary.

In addition to the core sequences, the findings have also illustrated an upward trend in a group of sequences. These sequences, which were labeled as the 'emerging' sequences, were ranked below the top 10 at the B1 level but gradually surfaced into the top 10 at the C2 level. The following are the emerging sequences identified in this study:

VVD DT NN IN (verb [past tense]-determiner-noun-preposition)

VVD DT JJ NN (verb [past tense]-determiner-adjective-noun)

VBD DT JJ NN (verb [past tense]-determiner-adjective-noun)

VVN IN PP\$ NN (verb [past participle]-preposition-possessive pronoun-noun)

VVN IN DT JJ (verb [past participle]- preposition-determiner-adjective)

VVN IN JJ NNS (verb [past participle]-preposition-adjective-noun [plural])

Based on these emerging sequences, several conclusions can be made. First and foremost, all six sequences were headed by either a past tense verb or a past participle verb. This is

different from the core sequences, where only two of the five sequences were headed by a past or present verb. Again, the surfacing of the past tense verbs provides further support that they emerge much later in the L2 developmental stage (Dulay, Burt, & Krashen, 1982; Ellis, 1994). Next, the emerging sequences provide robust evidence for usage-based approaches to language learning. One of the main tenets of usage-based theories is that language develops over the course of learners' history of language use through their engagement in meaningful interactions (Bybee, 2010; Ellis, O'Donnell, & Römer, 2013; Tomasello, 2003). The gradual surfacing of the emerging sequences can be seen as a manifestation of learners' language competency in that as they become more proficient users of the language, they are able to utilize more complex language sequences (e.g., past tenses). Previously, syntactic complexity in L2 writing has been most commonly measured by indices such as mean length of T-unit, clause, clause per t-unit, and dependent clauses per clause (Wolfe-Quintero, Inagaki, & Kim, 1998), but these indices are restricted to the length of clauses and do not take into consideration the different forms of the verb. While conceptualizing complexity according to these measures has gained scholarship, what could also be beneficial is to probe further by exploring the link between both forms and the length of clauses. Ortega (2003) points out that syntactic complexity matrices alone are insufficient in measuring the pure 'complexity' nature of language as "more complex may mean more developed in many different ways, and the nature of L2 development cannot be sufficiently investigated by means of these global measures alone" (p. 494). In that sense, the surfacing of the emerging sequences provides opportunities for further investigation. One disclaimer to be made, however, is that this result did not take into account the various factors that contribute to learners' language use (e.g., genre, topic). Regardless, considering that the CLC consisted of writing samples from various topics and registers across all proficiency levels, it would be fair to claim that there is a clear link between proficiency and usage in the context of EFL.

### **5.2.2 Node Verbs**

The second part of the study explores the use of 'node' verbs from the core and emerging sequences. Before delving into further discussion about the node verbs, it should be stated that the current study fully complies with the notion that lexis and grammar are interdependent, as shown in many usage-based studies (Sinclair, 1987a; Sinclair, 1987b; Römer, 2009). The purpose of taking apart the node verbs from the sequences was to identify the most commonly

used verbs in the core and emerging sequences and compare how they differ across both the proficiency levels and sequences. Furthermore, it also strives to bridge the gap between learner corpus research and pattern grammar by illustrating that language usage is not a random collection of words but rather consists of patterns.

Before delving into further analysis, it might be worthwhile to see the ways in which the top node verbs identified in this study differ from that of another similar study. Römer (2019) conducted a study where the author used a learner corpus, English First-Cambridge Open Language Database (EFCAMDAT)<sup>5</sup>, to identify the use of verb argument constructions in L2 writing at different trajectories of development. **Table 5.2** illustrates the 50 most frequent verbs found in the EFCAMDAT corpus. Interestingly, all of the top-ranked node verbs, which were headed by a simple present verb, were present among the top 50, as identified in **Table 5.2**.

Taking the node verbs in the VV DT NN IN sequence, for example, the verbs “give,” “make,” and “see,” ranked first in the A2 level among L1 Korean speakers, which are ranked 26, 9, and 6, respectively. In the B1 level, “make” (rank 6) was the most frequent, whereas in B2, “get” (rank 10) occurred most frequently. At the C1 level, “make” (rank 6) occurred most frequently, and at the C2 level, “take” (rank 12) was the top-ranked node verb. However, the same did not hold for node verbs that were headed by past tense. One interpretation could be that the changes in the ranking of the emerging sequences in proportion to the proficiency level could be an indication that the learners are starting to utilize more infrequent words.

1. Be	11. Play	21. Wear	31. Learn	41. Talk
2. Have	12. Take	22. Know	32. Tell	42. Read
3. Do	13. Think	23. Love	33. Let	43. Call
4. Go	14. Hope	24. Buy	34. Write	44. Listen
5. Like	15. Want	25. Help	35. Speak	45. Say
6. See	16. Look	26. Give	36. Use	46. Bring
7. Live	17. Come	27. Find	37. Thank	47. Apply
8. Work	18. Eat	28. Feed	38. Walk	48. Ask
9. Make	19. Need	29. Start	39. Study	49. Feel
10. Get	20. Meet	30. Watch	40. Wash	50. Visit

**Table 5.2** 50 most frequently used verbs in the EFCAMDAT learner corpus (Römer, 2019)

<sup>5</sup> The EFCAMDAT corpus consists of texts written by students upon completing coursework for an online school. The data is composed of CEFR levels from A1 to C1 and consists of the following data: German A1, A2, B1, B2, C1; Mexican A1, A2, B1, B2, C1

Based on the percentage of the top node verbs from the proficiency levels, it is evident that even among the same verb sequences, the use of verbs differed between the proficiency levels. However, as elaborated extensively in the literature review, usage-based approaches to language learning highlight that meanings don't derive from individual words but from constructions and phrases (Goldberg, 1995; Hunston, 2008; Langacker, 1987; Tomasello, 2003). In other words, form and meaning are inseparable, and the meaning of the words does not lie in the individual words but at various levels of complexity. Such was particularly the case among learners with higher proficiency. For instance, the verb "go" in the A2 level was used literally and more commonly with the preposition "to" (e.g., "**go to a meeting**"), whereas, in the more advanced levels such as B2 to C2, learners used the verbs more idiomatically (e.g., "**go over the wall**") and with various prepositions (e.g., "**go on a business trip**"). There were also cases where learners in the A2 level had made attempts to use verbs in a non-literal sense, albeit with errors. In the case of the verb "make," examples such as "**make a book**" and "**make a meeting**" show that the learners have some understanding of the ways in which "make" can be used to mean "arrange." With the progression of proficiency, however, the errors diminish, and other uses of "make" arises (e.g., "**make a choice**," "**make a living**"). Such an example illustrates linguistic forms are not stored as formulas but rather in the continuum of entrenchment (Ellis et al., 2015). If the verb "make" was stored as a formula for how the learner would have been exposed to it (e.g., "**make a reservation**"), then it should not be possible that the learners are able to generalize the meaning of the word "make" by experimenting with the various ways in which "make" can collocate with other nouns. As learners become more proficient in the L2, they are able to draw upon words that tend to go together, as illustrated in the examples above. Thus, these findings lend support to the myriad of studies that claim that there is a strong correlation between lexical richness and L2 competence (e.g., Crossley, Cobb, & McNamara, 2013, Jarvis & Daller, 2013), as well as multi-word units (Bestgen & Granger 2014; Kyle & Crossley, 2015).

As a way of summary, a few main claims can be made from the analyses of the node verbs. Firstly, the percentage of the individual node verbs in this study cannot be used to arrive at a conclusion about learners' vocabulary knowledge. Instead, they provide a further probing point of investigation to tap into the most frequently used node verbs. Taking the VV DT NN IN

sequence as an example, the word “give,” which made up 9.7% of VV DT NN IN in the A2 level, was also among the most frequently used verbs from B1 to C2 levels. Judging by the frequency of the node verb “give,” it may seem reasonable to think that learners of all proficiency have equal knowledge. However, further ad-hoc analyses are needed to look at how the verbs are used differently in context. Secondly, learner errors were much greater at the lower level than at the higher level. This is expected as learners at the beginning level strive to internalize and experiment with different linguistic features. The above example, “**make a book at a hotel**,” also lends some support to the current understanding of interlanguage and how language learning doesn’t necessarily entail echoing what has been heard by the learner. Instead, through exposure and various individual cognitive processes, the learners continue to experiment with different inputs until they have finally mastered their use (Ellis, Römer, & O’Donnell, 2016). Next, it can also be seen that the use of the individual verbs was driven by the topic across the tasks. In the examples provided in the VVN IN DT NN sequence, the verbs “located,” “seen,” and “situated” were among the most frequently used node verbs across the proficiency levels. At the B2 level, the phrase “**located in the middle**” was repeated multiple times throughout, whereas, at the C1 level, it started to vary (e.g., “**located in the mountain**,” “**located between the river**”). While it can be presumed that the learners at the high level have a better command of the wide range of ways in which the verb “located” can be used, it is also evident that the concordance lines are highly specific to the topic. This finding has also been shown in previous studies that the topic effect, in conjunction with the learners’ unique background knowledge, can affect their use of modal verbs (Hinkel, 2006). Lastly, there is no overlap of the most frequently used node verbs in the emerging sequences. Such finding could be due to the lack of data found in lower levels but may also be due to the verbs being more topic-specific (e.g., “buy,” “find”).

### 5.2.3 Categorization of Sequences to Patterns

Based on the derived core and emerging sequences, two candidate sequences were chosen for the analysis in Chapter 4. (Core: VV DT NN IN, VVN IN DT NN; Emerging: VVN IN DT JJ, VVD DT NN IN). These verb 4-gram sequences were selected since they were the top two ranked core and emerging sequences in the C2 level. Due to constraints in space, a random sample of 50 examples was selected. From a holistic perspective, by categorizing the sequences

according to pattern grammar (Francis, Hunston & Manning, 1996, 1998), evidence was found that the language used by learners does indeed evolve with their proficiency levels in their use of more varied words, phrases, and collocations. The findings lend support partly to empirical research on L2 writing measurements such as complexity, accuracy, and fluency (Wolfe-Quintero, Inagaki, & Kim, 1998) in that language use becomes more complex at both the syntactic and semantic levels. However, rather than using these measures as a benchmark, the current study used pattern grammar to illustrate the ways in which the form-meaning mapping changes as learners become more competent.

In addition to the observable changes in language use, the findings also indicated that language use evolves in a dynamic manner as learners make their transition from beginner to intermediate, and finally, to advanced level (Verspoor, Lowie, & Van Dijk, 2008). Simply put, language doesn't evolve in a linear fashion where a clear path of learners' development can be drawn (Verspoor, Lowie, & Van Dijk). Rather, it progresses and digresses in ways in which complex phrases and chunks may not always surface. Such a claim is in line with dynamic systems theory (Bot, Lowie, & Verspoor, 2007; Verspoor, Schmid, & Xu, 2012) and other research which claims that language uses may differ drastically depending on a wide range of factors such as topic and the learner. One of the major shortfalls of the studies to date is the skewness of data where the learner corpus is compiled among a small cohort of learners in a single institution. However, the CLC encompasses not only a wide range of learners but also various topics and academic registers, which makes the generalization made in this study much more robust and reliable. The discussion now shifts more specifically to the grammar patterns of the selected core and emerging sequences.

First and foremost, it should be mentioned that the current study did not attempt to align patterns and constructions (e.g., Hunston & Su, 2019). However, by categorizing the sequences using a systematic approach, it sought to understand the changes taking place around the proficiency levels and bring awareness to the changes in the form-meaning mapping of Korean learners. Although some of the patterns found in the study could potentially be identified as verb constructions (e.g., V about n, V for n), no such reference was made.

In the two core sequences, VV DT NN IN and VV IN DT NN, several remarks can be made. First, in the VV DT NN IN sequence, the most prominent pattern was the V n pattern, consisting of a simple present verb followed by a noun phrase. It accounted for the majority of

the samples across all of the proficiency levels. There were also more complex patterns, such as V n about n and V n for n. It was not the case that the syntactic structures were complex only as the learners' language was becoming more proficient. This is supported by Biber et al. (1999), where complex noun phrases can be defined by factors such as non-compulsory elements in the noun phrase (e.g., a study of x) and recursive embedding. One aspect to keep in mind, however, is that the broad categorization V n pattern could encompass longer, more complex structures. Following previous research, it could potentially be that the length of the noun phrases increases with proficiency level (Bulté & Housen, 2014), although such was not in the scope of this study. Since the study only accounted for the verb 4-gram sequences, it may be more appropriate to consider that the use of complex patterns V n into n, V n prep/adv was surfacing at both the beginner level as well as at the most advanced level.

In the case of the V n, even after each of the identified examples was categorized into meaning groups (see **Table 4.10**), no major differences between the proficiency groups were evident based on the frequencies alone. It is not until delving into the examples of the concordance lines that there are clear differences across the proficiency levels. Firstly, there was quite a bit of repetition at the beginning. Taking the 'hear' group as an example, all three occurrences in the A2 level were headed by the verb "see," followed by a noun phrase. At the B1 level, in addition to the verb "see," other verbs also surface, such as "enjoy," "learn," and "imagine." At the C2 level, the verbs become even more various (e.g., "consider," "experience," "feel"), and there is a clear transition of the expansion of the learners' knowledge of the verb as manifested in use. This finding can be understood from the perspective of usage-based models in that language is shaped largely by usage events (Kemmer & Barlow, 2000) and that language expands and becomes more complex as learners gain more experience with it. Clearly, as demonstrated by the examples pulled from the concordance lines, the learners at the most advanced level have attained the ability to not only use the words correctly but also incorporate them in a way where the meaning isn't embedded in the single lexeme itself but in the phrase as a chunk. As the learners become more proficient, they would have a larger repertoire of language to choose from, thereby providing them with ways to express themselves using a wider range of words and expressions.

Next, as can be predicted, the language at the beginning level contained errors (e.g., spelling, misuse) or language uses that were awkward or unnatural. Such is typically normal at

the beginning level owing to the understanding that language is based on the knowledge gained from the usage of the language and the generalizations that the learners make through their exposure (Langacker, 1987, 1991; Croft, 1991; Tomasello, 2003; Goldberg, 2006, Bybee, 2010). Thus, the learner errors manifested at the beginning levels can be seen as the processes of language learning as the learners continue to internalize the exposures and try to make sense of their usage. Taking the “see” examples mentioned above, in the beginning, the examples contained phrases such as “see those staf (staff) in,” indicating the literal sense of meeting someone. Following the concordance line for context, the more appropriate word would have been “meet those staff” or “talk to those staff.” At the most advanced levels, the word “see” is used in examples such as “see the development of” and “see the correlation between,” illustrating not only error-free uses of the word “see” but also utilizing them in a much more natural way. Thus, the learner errors manifested by beginning learners serve as a crucial stepstone in the process in which learners take on the words and phrases they pick up through exposures and attempt to make sense of them by experimenting with them in a wide range of contexts.

Further observation can be made with regard to the categorization of ‘none’ in the VV DT NN IN sequence. As elaborated in the results section, examples categorized under ‘none’ are words that belonged to the V n group according to pattern grammar but were not labeled under the groups. Again, this could be due to the pattern grammar accommodating for the top 400 most frequently used verbs. That being said, the increase in the number of examples in the ‘none’ category across the proficiency levels can be used as an indication that with proficiency, the learners are indeed able to utilize a wider range of verbs.

With respect to the core sequences, VV IN DT NN, a similar pattern can be observed with respect to the learners’ overuse of verbs at the beginner level. Taking the V prep/adv as an example, the majority of the examples were headed by the verb “go” and then followed by the preposition “to” and a noun phrase. There are also frequent repetitions of “go to a meeting” and “go to the cinema,” which illustrates that the learners’ use of the given grammar structure may have been mediated by the topic. Interestingly, the word “go” is also seen at the advanced levels; however, different prepositions also appear, such as “go for a walk,” “go through the back door,” and “go to the toilet.” It is evident that at the highest level, the learners have acquired that one can “go to” a specific place but can also “go on a trip” or “go for a walk.” The learners’ ability to

utilize verbs with various prepositions is an indication that they have a better grasp of the wide range of meaning that can be encompassed by the lexeme. In the B2 level, the error “go on an escalator” illustrates the ways in which the learner may have learned that the verb “go” can be used in a wide range of contexts but haven’t yet fully acquired the breadth. Errors are an important process in language learning, and through these trials and errors, the learners finally arrive at the correct usage. On that note, there is a clear indication of a decrease in errors with proficiency. Such indication is in line with studies that have shown that prepositions are particularly difficult for L2 learners (Geist, 2017; Larsen-Freeman & Celce-Murcia, 2016) and that oftentimes, learners are stuck because they are unable to discern which preposition to use. Such is particularly the case among L1 Korean speakers, mainly because prepositions aren’t as distinct as they are in English. For instance, the preposition “to,” “in,” and “on” can all be translated the same, and thus, L1 Korean speakers do not have the intuition of which prepositions are most appropriate as do L1 English speakers. Back (2011) found that many prepositions errors in L1 Korean speakers can occur due to either L1 transfer or confusion in the lexemes. Other studies have also investigated preposition errors in various contexts, such as prepositional verbs (Yu & Yoo, 2010); the preposition of (Lah & Yoo, 2015) and agreed that there were many preposition-related errors among L1 Korean speakers. Whether the learner produced errors in this study is due to lexical confusion is impossible to discern; however, it is sensible to think that L1 transfer is a contributing factor to the errors. On the other hand, although some errors persisted even at the highest level, it can be claimed that the knowledge of prepositions improves with proficiency and that the learners eventually develop an intuition about them.

Another interesting observation can be made with respect to phrases and phrasal verbs. It is a well-known fact that much of the English language is made up of chunks or prefabricated expressions and that they are an important component in acquiring a native-like proficiency (Pawley and Syder, 1983; Lewis, 1997). However, it is widely known that phrasal verbs or lexical chunks are particularly difficult points for L2 learners because meaning is parsed in the phrases, not in the individual words. The evidence from this study also lends support to the ways in which phrases and phrasal verbs are more prevalent in the advanced levels. Furthermore, phrasal verbs were also found across all proficiency levels, even at the beginning level. One disclaimer to make is that the phrasal verbs were used rarer in the beginning levels with just one occurrence) or were awkward in context (e.g., “go through the lack”). This finding provides

strong support to the previous work on phrasal verbs (Garnier & Schmitt, 2016; Nassaji & Tien, 2010) but also reveals that even at the beginning level, the learners are striving toward making sense of the lexical chunks.

The discussion now shifts to the two emerging sequences, VVN IN DT JJ and VVD DT NN IN. The first sequence, VVN IN DT JJ, is headed by past participle verbs, and the second sequence VVD DT NN IN, is headed by past tense verbs. In the VVN IN DT JJ sequence, the majority of the examples belonged in the ‘none’ group, similar to the findings from the VV DT NN IN. Interestingly, the number was higher at the B1 and C1 levels but was lower in B2 and even lower at the C2 level. One interpretation of this finding could be that the frequency level of words does not necessarily encompass the learner’s linguistic ability. Another interpretation can be that the use of words in writing can differ depending on factors such as the topic. Since the CLC consisted of a wide range of exam topics, this could have potentially led the learners to use certain verbs more than others. On the other hand, there was a gradual decrease in errors, which illustrates that the learners at the highest level have a better grasp of prepositions, thereby producing fewer errors.

A final observation can be made with respect to learners’ use of phrases and phrasal verbs. Interestingly, no phrases and one phrasal verb were found at the C2 level, whereas a few phrases were found from B1 to C1 level, but no phrasal verbs were found. Since the categorization only includes a random sample of 50 sequences, it would be difficult to make a concrete conclusion about learners’ ability to use phrasal verbs. However, the bottom-up approach has shed potential to further investigate learners’ knowledge of lexical chunks at various developmental trajectories. The results in this study could also lend support to the study by Granger and Bestgen (2014), which claimed that learners at the intermediate level tended to overuse higher frequency collocations and underused low-frequency collocations.

In the VVD DT NN IN sequence, a similar pattern was found. The majority of the examples belonged to the V n meaning group, errors decreased with proficiency, and there was a gradual increase in collocations. The less dramatic difference in the emerging sequences, as opposed to the core sequences, is the omission of the A2 level in the emerging sequences due to a lack of data. Thus, all analyses were conducted among B1 to C2 levels, which are low-intermediate to near-native proficiency.

A close observation of the examples has revealed that there is not a clear distinction between intermediate and advanced levels. This complies with the findings mentioned above, along with the claims of usage-based models that language does not evolve gradually. Based on the findings in this study, it could be true that the learners at the intermediate level have already acquired as much knowledge as the learners at the C2 level as far as word order in phrases goes. What could potentially set the two levels apart could be in other areas such as syntactic complexity and the content.

### **5.3 L1 Korean Speakers vs. Other L1 speakers**

This section discusses the findings from the comparison of the sequences between L1 Korean speakers and other L1 speakers. It first explores the difference by comparing the top 25 verb 4-gram sequences as well as the core and emerging sequences in order to provide a clear view of the interlanguage differences. Then, the selected samples of the patterns are discussed to further probe into the differences between L1 Koreans and other L2 speakers in the corpus.

Before further discussion, however, it should be revisited that the main concerns of this lie in the realm of L1 Korean speakers. Therefore, the comparison was not made with any particular L1 speakers but included all other L1 speakers in the CLC.

#### **5.3.1 Comparison of Verb 4-gram Sequences**

The top 25 verb sequences were compared across each of the proficiency levels from A2 to C2. The rankings of the normalized frequency revealed that the sequences among the top ranking were similar across all levels. For instance, the top four in the A2 level, VV IN DT NN, VV DT NN IN, VBP VVG TO VV, and VV PP IN PP\$, were the top four in both other L1 speakers and L1 Korean speakers. Among them, two of the core sequences, VV IN DT NN and VV DT NN IN, ranked first and second in the other L1 speaker's sub-corpus and third and second in the L1 Korean speakers' sub-corpus. The dispersed rankings of the A2 levels among L1 Korean speakers (see 5.2) may not be reliable due to the small sample. However, given that the A2 sub-corpus of other L1 speakers consisted of 9 million words, it can then be claimed that at the A2 level, the learners have yet to internalize the core emerging sequences that become more prevalent at the B1 level. In fact, from rankings 1 to 5 among other L1 speakers, all of the sequences are headed by a present verb, whereas past verbs do not surface until rank 12. The

conclusion drawn from this finding is not necessarily that L1 Korean speakers use past tenses later than other L1 speakers but that regardless of the learners' L1, the language use at the beginning level is unstable.

This claim lends support to Wolfersberger (2003), whose study illustrated that in the case of writing tasks that demand an L2 writing proficiency that is much higher than that of the learners' actual proficiency, L1 strategies do not tend to transfer to the L2 writing process. Thus, this could explain why none of the emerging sequences surfaced in the top 25 in both L1 Korean and other L1 speakers and that even among the core sequences, those that ranked high in one group also tended to rank high in the other.

From B1 to C2 levels, it is evident that the core sequences were among the top in both the L1 Korean and other L1 speakers. Although the rankings did not align perfectly, the difference in ranking did not exceed more than four rankings. In fact, at the C1 and C2 levels, the core sequences were very similar, and the rankings only differed by one. The rankings in both sub-corpora lend support to the aforementioned idea that frequencies are Zipfian (Zipf, 1935) in that the top most frequent sequences accounted for the majority of the sequences. Given that the two sub-corpora mirror each other in terms of potential differential factors, such as topic, it can be claimed that the learners most frequently used sequences, otherwise known as core sequences, are common regardless of the learners' L1.

On the other hand, there seems to be a different story associated with the emerging sequences. The findings based on L1 Korean speakers illustrated that certain patterns became more and more frequent with the proficiency level. However, this finding was not validated by other L1 speakers. In fact, most of the emerging sequences identified among L1 Korean speakers differed in rankings of more than 5. The difference was much bigger at the lower levels, such as B1 and B2. For instance, some sequences such as VVN IN JJ NNS differed by more than 100 rankings. Whether or not emerging sequences can be identified among other L1 speakers did not fall into the scope of this study. However, given that the clear pattern of more frequent usage of past tense verbs and past participle verbs surfaced in the L1 Korean sub-corpus only, this lends to the possibility that this may be due to learners L1. Alternatively, other factors could be at play, such as that L1 Korean speakers utilize more passive voice with proficiency level.

### 5.3.2 Comparison of Patterns

To probe further into the differences between L1 Korean speakers and other L1 speakers, candidate groups were selected. Firstly, the end levels of ‘beginning,’ ‘intermediate,’ and ‘advanced’ levels were chosen (e.g., A2, B2, C2) for the selected core (VV IN DT NN, VV DT NN IN) and emerging sequences (e.g., VVN IN DT JJ, VVD DT NN IN) used earlier.

The comparison of the A2 and B2 levels in the core sequences revealed that there were no distinguishing features of L1 transfer. In fact, in the VV IN DT NN sequence, “go to” was overused by both groups of learners and their errors pertained mostly to prepositions. Again, this is clear evidence that prepositions are particularly for L2 learners (Geist, 2017). Even at the C2 level, prepositions were revealed to be a difficult concept, although the errors are much more subtle.

As for the VV DT NN IN sequence, the most prominent error in the L1 Korean sub-corpus seems to be the learners' use of articles. This is predictable since articles do not exist in Korean and, thus, are particularly difficult for L1 Korean speakers (Choi, 2017). Another interesting point to consider is the learners' use of collocations. Despite that there must have been some overlap in the prompt as each of the sub-corpus is a sub-set of the CLC, there isn't overlap in the learners' use of collocations. One potential explanation could be aligned with the usage-based view where learners' L1 plays an active role when acquiring a second language (Bybee, 2008). In a similar vein, research has also shed light on the ways in which L1 influence is prevalent in the learners' use of lexical bundles, collocations, and lexico-grammatical patterns (Paquot, 2013). Although the results from this study would require further investigation, it may be the case that the L1 difference may have affected the different usage of the collocations as manifested by the selected samples in this study.

The data from the emerging sequences were taken from B2 and C2 levels only. In the VVN IN DT JJ sequence, a similar pattern can be seen with respect to the learners' use of patterns. In both groups, “compared to” and “followed by” were dominant, which are both transitional phrases. One explanation could be that the similar use of language could be due to the topic. It could also be an indication that the learners tended to overuse these phrases in their writing. A similar pattern can be found in the C2 level as well, where even though the sample was randomly selected, there were overlaps in the use of node verbs in both groups. The findings can potentially contradict previous studies that claim that the L1 transfer effect is prevalent

among L2 learners, especially as learners become more and more proficient in the L2. It could also be that at the most advanced level, learners have acquired a native-like proficiency to discern the most appropriate use of language both at the individual vocabulary level and also at the lexical bundle level. The following section investigates the extent to which the most advanced learners mirror the language of L1 English speakers.

## **5.4 L1 Korean Speakers vs. L1 English Speakers**

This section discusses the differences in the verb 4-gram sequences among L1 Korean speakers and L1 English speakers. The native-speaker reference corpus consulted in this study was the BNC. As illustrated in research, it is a widely known issue that comparing a learner corpus with a native-speaker corpus could be problematic (Barlow, 2005). However, finding a reference corpus that is identical to that of the learner corpus is not feasible. Thus, in order to match the L1 English corpus to the L1 Korean sub-corpus as much as possible, the academic BNC corpus was chosen as a candidate for this study.

### **5.4.1 Comparison of Verb 4-gram Sequences**

The ranking of the top 25 sequences revealed a similar pattern to both the L1 Korean speaker sub-corpus as well as the comparison of L1 Korean and other L1 speaker sub-corpus. In fact, the top two rankings were swapped, but the top six rankings were also identical for the most part. By now, it could be assumed that the most frequently used verb sequences are common among all L1 and L2 users. Although it isn't in the scope of the study, it might be worth mentioning that when compared with the C2 level in other L1 speaker sub-corpus, the sequence VVD DT NN IN ranked 16th, whereas in the BNC and the L1 Korean sub-corpus, it ranked 4th and 7th, respectively. As the other L1 sub-corpus consists of learners of various L1, further investigation should be carried out.

With respect to the core sequences, all five of them were ranked within the top nine. However, emerging sequences did not surface the same way in the BNC as they did in the L1 Korean sub-corpus, further lending to the support that although the most frequently used sequences are prevalent across all speakers, the less frequent sequences do not align as well due to various factors which may include the learners' L1, topic, and experience with the language to

name a few. The selected core and emerging sequences are probed further to better understand the similarities and the differences between the two groups.

#### **5.4.2 Comparison of Patterns**

In the core sequence VV DT NN IN, both the L1 English and the L1 Korean sub-corpus contained verbs in five groups. Interestingly, there were more collocations among L1 Korean speakers, which could indicate their tendency to overuse collocations more often than native speakers. However, as supported in empirical research (Li & Schmitt, 2010), the majority of the collocations are unidiomatic and have an equivalent L1 translation. For instance, some of the collocations found in the L1 Korean sub-corpus were “double the amount of,” “enjoy the feeling of,” whereas in the BNC, collocations such as “enlist the help of” and “raise the question of” were present. While the study did not investigate erroneous use of collocations, it is a widely known fact that L1 transfer plays a critical role in misusing collocations in the L2 (Granger, 1998; Laufer & Waldman, 2011; Nesselhauf, 2003, 2005). This assertion is also claimed to be true among L1 Korean speakers. For instance, Lee (2016) investigated the L1 influence on L2 collocations among L1 Korean speakers and found that even at the advanced level, learners tended to make more errors, especially with respect to incongruent collocations and that learners relied heavily on L1 knowledge.

With respect to the meaning groups in the VV DT NN IN sequence, both sub-corpora consisted of similar numbers, although more examples from the BNC corpus belonged to the ‘none’ group. To reiterate, the “none” group consists of examples that did not belong in one of the meaning groups due to infrequent use. Thus, an inference can be made that L1 English speakers tended to use more low frequency verbs.

The findings from the emerging sequences also did not differ drastically from the core sequences. Although there weren’t any noticeable differences, however, significantly more examples fell into the “none” group, indicating that the L1 English speakers contained less frequent words. Several interpretations can be made based on this finding. Undermining that although the academic portion of the BNC was chosen as it best exemplified the data from the CLC, it cannot be claimed that the two corpora are identical in terms of the topics, genre, and text lengths. Moreover, plenty of studies have investigated that linguistic features and L2 writing performance can vary depending on a wide range of factors such as topic (Cho, 2019) and genre

(Bi, 2020). With that respect, the findings in this portion of the study should be carefully interpreted. Thus, further investigation is necessary in order to understand L1 Korean speakers' lexical knowledge at the C2 level.

## 5.5 Summary

This chapter provided an overview of the L2 writing developmental pattern among L1 Korean speakers and triangulated the findings with two different sub-corpora in hopes of shedding light on the extent to which the process of development is specific to L1 Korean speakers. From a holistic perspective, the findings of this study provide robust evidence to the growing body of research in the field of usage-based theories in that diverse structural patterns emerge as learners gain more exposure and thus, become more proficient in the language. It is through these meaningful interactions that the learners are able to use language at their discretion as they gain proficiency in the L2. As illustrated in this study, the language usage may differ in multiple levels throughout various proficiency levels. Language may differ in the use of difficulty of individual lexical items, although this was not measured extensively in this study. Language may also differ at the level of chunks or sequences. One certainty is that there is a clear pattern in learners' use of sequences that emerge with proficiency level. Thus, drastically different use of language both within and among levels explain why there is not a clear trajectory of language development.

Despite such differences, there is still much to be said about learners' language learning in that the development is not random. That is, there are aspects of usage that can be anticipated with respect to the learners' proficiency. This was evident through the identification of the core and emerging sequences. While plenty of language research has proven that the frequency of discrete language items and structures is Zipfian, only a few studies to date have explored the emergence of language use, more specifically, the kind of language that emerges with frequency in attested data. Furthermore, the current study has shown that some emerging sequences are specific to L1 Korean users. If the same analysis was conducted among other L1 speakers, a different type of trend could potentially emerge.

Although the usage-based approaches to language learning stand strongly on the notion that the role of frequency experience with the language plays a crucial role in language development, research should make a clear judgment when making conclusions. For instance,

learners' use of language can differ drastically based on a wide range of personal and group characteristics as well as the learners' language background (Gablasova, Brezina, & McEney, 2017). While it would be ideal to control for all potential interfering factors, doing so is very difficult to achieve. Despite such shortfall, corpus-based frequency information is still useful in illustrating the changes in the use of language from a holistic perspective.

## **Chapter 6 Conclusions**

### **6.1 Introduction**

This chapter concludes the thesis by first addressing a summary of the main findings. It then discusses the implications of the present study by discussing the ways in which this study has expanded our understanding of L2 usage. Then the limitations of the study are addressed. Finally, some suggestions for future research are presented.

### **6.2 Summary of Major Findings**

From a holistic perspective, the findings of this study provide robust evidence to the growing body of research in usage-based approaches to language learning and pattern grammar. Revisiting the conceptual framework of this study as introduced in Chapter 3, the current study provides evidence that language usage is largely based on patterns and the different patterns emerge as learners become more proficient in the L2.

Situating the study in the realm of usage-based approaches comes with the underlying assumption that language is learned through meaningful interaction with it. In other words, as learners gain more exposure to the language, they use language in a way that is more complex and abstract. Throughout the study, the term “complex” is used somewhat liberally, but it is used to illustrate that language becomes “more developed” in ways where it cannot be measured with a few indices (Ortega, 2003). One thing to bear in mind, however, is that as Ortega (2003) emphasized, “complex” use of linguistic features does not necessarily dictate better. As this study did not observe syntactic complexity to coin the term “complex,” it should be understood that what is meant by “complex” is different. Rather, the term “complex” is used in this study to encompass the wide range of ways in which language is more developed at the holistic level.

Taking the view of usage-based approaches, frequency plays a major role when discussing the entrenchment of language. Two major claims can be made here. Firstly, the most frequently used words and language structures are learned first. Next, there should be a change in the use of language structures as learners become more proficient L2 writers. These claims were made apparent through the identification of the core and emerging sequences in these studies among L1 Korean speakers. The core sequences were evident across all proficiency levels, whereas the emerging sequences became more apparent with proficiency.

Another major claim to make about usage-based approaches is the idea that language learning is meaning-based. Usage-based theories reject the traditional view that meaning is based on lexical items. Rather, it draws on the notion of form-meaning mapping in that the meaning of the word should be viewed at the syntactic level. Thus, while the current study is situated around verb usage, it is viewed at the level of sequences. In particular, it regroups the “node” verbs according to pattern grammar as a way to make further claims about L2 writing development. The findings not only support the ways in which patterns can be used to make claims about the developmental processes but also that the use of patterns changes and varies. This illustrates that while there are similar patterns that can be identified across the proficiency levels, language learning is also subject to much variability.

Lastly, the findings of the patterns and sequences identified among L1 Korean speakers were compared with that of two other cohorts: other L1 speakers and L1 English speakers. The comparison of the two sub-corpora with that of L1 Korean speakers revealed that the core sequences identified were among the top more frequently occurring sequences in all three groups. These findings revealed that just as most frequent words tend to be acquired and automatized more quickly than less frequent words (Verspoor & Schmitt, 2013), the most frequent sequences are acquired first among all L2 learners regardless of L2. However, the emerging sequences identified among L1 Korean speakers did not surface in the same way in both the L1 English sub-corpus and the other L1 speaker sub-corpora. Thus, it is evident that the emerging sequences found among L1 Korean speakers may be an indication of interlanguage that is not prevalent among other L1 speakers. These sequences may reveal aspects and processes of language learning that fall outside the scope of this research but need to be further explored in future studies that examine, perhaps longitudinally, the emergence of the unique patterns of variation in the L1 Korean community.

### **6.3 Critical Reflection on the Overall Study**

This study aimed to provide a bird’s eye view of the processes of L2 writing development among L1 Korean speakers using verb 4-gram sequences and pattern grammar. This aim was fulfilled through the identification of the core and emerging sequences as well as the categorization of the examples into grammar patterns. The study showed evidence of usage-based approaches to language learning by illustrating that the simple and most frequent verb 4-

gram sequences were present across all proficiency levels. Furthermore, this study highlighted that language learning is a process. For instance, the emerging sequences became more frequent as learners gained more experience with the language and developed a wider understanding of different usage. The samples from the core and emerging sequences also demonstrated that learners at the higher proficiency level have greater access to a wider range of words and expressions. Their language use also started to become more abstract and error-free. Additionally, this research illustrated an extensive scope of learners' linguistic capabilities at various proficiency levels. While there were more erroneous words and sentences in the lowest proficiency level, learners started to utilize more complex linguistic structures and verb sequences as they gained more experience with the language. However, with regards to the approach taken, a few points need to be revisited.

Firstly, this study used a bottom-up approach to investigate linguistic features in order to provide a holistic understanding of language use. Furthermore, it did not intend to offer rich insight into any particular feature. Studies that have taken a more granular approach by looking at collocations (Reppen, 2009) and constructions (Römer, 2019) provide a focus on a single linguistic feature, providing a richer understanding of how the particular linguistic feature changes across different groups of learners. While the approach taken in this study did not zone in on any specific linguistic feature, it provided clear evidence of the dynamic nature of L2 development. Language is indeed multifaceted and does not develop in a linear fashion (Verspoor, Lowie, & Van Dijk, 2008). For instance, phrasal verbs and idiomatic expressions, which are often recognized to be used by more advanced language users, were prevalent across all levels. Such usage may be due to L1 transfer or unintentional usage, but multiple instances provide robust support to the notion that learners at the lowest proficiency level can also use a wide range of linguistic features, albeit at varying degrees of complexity. However, the frequencies alone cannot provide an answer to learners' linguistic knowledge. Rather, the learners start to develop a greater repository of language and thus, can learn to utilize the more appropriate linguistic structure depending on the style and the format of the task.

Secondly, this study utilized pattern grammar as a tool to systematically categorize the verb 4-gram sequences. As mentioned in the methodology section (see 3.6), the verb patterns and the meaning groups are largely open to debate (Hunston & Francis, 2000) and thus can differ depending on the interpretation. Additionally, as language is largely dependent on the context of

its usage (Goldberg, 1995), the nuance of sentences can differ greatly. The unclarity of intended meaning is particularly the case in learner language since the learners may misuse words, leaving the researcher to make speculations about the intended use. This perhaps is one of the caveats of learner data, where sometimes it is not clear as to what the learners intend to communicate. However, language usage tends to be subjective in nature, and the intended meaning may not be clear, particularly in written language. Thus, no single measurement can be used as absolute indices for pinpointing the learning trajectory. That is not to say, however, that all measurements are faulty or useless. Rather, categorizing indices such as pattern grammar can still offer valuable insight into language development broadly, but it should not be used to make executive decisions about learners' abilities.

The last point sums up the two issues discussed. Can L2 writing development be predicted? Empirical research claims that L2 learning is subjective to various cognitive, social, and environmental factors (De Bot, Lowie, & Verspoor, 2007). While L2 development cannot be measured with a concrete developmental benchmark in the ways L1 development can, this study has illustrated that L2 development may not be entirely random. For instance, the emerging sequences from this study showed that some patterns of language become more frequent with proficiency level. This may depend on the topic and the type of writing, but the current study stands on the idea that some aspects of L2 writing development can be predicted.

#### **6.4 Implications of the Present Study**

The current study traces the learners' language use in different stages of their developmental trajectory using a large learner corpus. Although other studies have observed language patterns from a top-down approach, where they start with a confined set of constructions or language patterns as a starting point (e.g., Römer, Skalicky, Ellis, 2020), only a few have investigated such emerging development from a bottom-up approach using reliable data regarding language proficiency. In this sense, this study is original and unique in bringing together a data-driven approach and the largest corpus of written learner language available. One of the biggest merits of such an approach is that it provides a comprehensive analysis of patterning beyond the level of the lexeme by aligning form and meaning (Hunston & Su, 2019). In particular, the approach used in this study encompasses a wide range of linguistic patterns

such as formulaic sequences, constructions, and idiomatic expressions, to name a few. That being said, the implications can be found in broad sectors.

The first implication is geared towards researchers. The current study can serve as a starting point to further investigate various linguistic patterns in more depth. For instance, what sort of grammar patterns surface at each proficiency level? When do learners start to use phrasal verbs? What type of noun phrases emerge across the board? This approach could potentially lead to the analysis of emerging constructions in L2 learner data. Based on the findings from this study, each of the identified linguistic patterns can be further investigated through other methods such as experiments to probe into whether L1 Korean speakers have attained knowledge of the linguistic patterns. The sequences identified in this can, therefore, provide extremely useful information about language development that can lead researchers to embrace different theories and methodologies (e.g., formulaic sequences, lexical bundles, VACs, construction grammar, and pattern grammar) to approach emergent properties of the L2 using a myriad of approaches in the theoretical claims.

In a similar vein, the current study strived to overcome the limitations of analyzing language use based on single lexical items. As shown through the node verb analyses in this study, the meaning of words is often embedded at the phrase and sentence levels rather than in individual words. This was particularly the case among more advanced learners, where the meaning of the words became more abstract in nature. The study sheds light on how and when learners are able to use verbs to construct phrases and sentences where the meaning becomes more abstract (e.g. “face the problem of” -> C2). Thus, it would be misleading to make assumptions about learners’ proficiency level based on their use of individual words only. This study is in line with recent calls (Gries & Deshors, 2020) to explore language use using multi-factorial approaches rather than analyses of discrete, individual lexical features.

The current study also holds implications for pedagogy. It has long been known that L2 writing is a particularly challenging skill for learners due to its multi-faceted nature. It is often the case that teachers are unable to pinpoint what needs to be taught, resulting in them relying on curriculum and standards rather than on students’ needs. While error analysis and other methodological approaches have shed a tangible understanding of some of the students’ problematic areas, it would also be helpful to know what the students are capable of producing at different proficiency levels as a way to inform both teachers and curriculum designers to help

students attain higher proficiency. In that sense, the current study, which is based on a large dataset encompassing learners from a wide range background, provides an open view of language development that can be of valuable use in designing appropriate language curricula for learners.

## **6.5 Contribution of the Study**

This section describes the contributions of the study. First, the methodological contributions are elaborated, and then the educational contributions are discussed.

### **6.5.1 Methodological Contribution of the Study**

The methodology behind combining usage-based approaches with pattern grammar provides an innovative way of understanding learner language using learner corpora. Many studies in the field of learner corpus research have been deterministic; that is, the research was conducted with specific linguistic features as the starting point. While this method has enriched our understanding of the underuse, misuse, and overuse of specific features (Granger, 2015), there is a limited association with such a method. First of all, a conclusion can only be drawn with regard to the features under investigation. Next, although learner errors serve as a useful tool for understanding the processes of SLA, focusing on errors in isolation is restricted to what the learners cannot do (Dagneaux, Denness & Granger, 1998). The current study allowed a clearer understanding of the various linguistic features that surface in academic writing across different proficiency levels. On that note, the current study is based on a large learner corpus with the CEFR proficiency levels as a proxy for different stages of language development. Although pure longitudinal data is difficult to attain, particularly at the scale of the CLC, a quasi-longitudinal study serves as a reliable source in providing robust claims about language development and the processes.

### **6.5.2 Educational Contribution of the Study**

From a pedagogical standpoint, instructors' ability to identify and forecast some of the most common learner errors at different stages of language development is beneficial. Such information can inform instructors with regard to some of the challenging points for learners. The instructor can curate lessons to focus on these errors as a way to provide learners with extra

support. However, understanding what learners can produce at different developmental trajectories brings a new paradigm to both teaching and research.

For teachers, they will have a better idea to help students at different proficiency levels by anticipating the general path of development and thereby providing learners with appropriate resources and guidance. Furthermore, they will be able to provide learners with a bite-sized amount of information at a time instead of overloading them with too much information, which may be detrimental to their overall learning. Having a comprehensive understanding of learners' capabilities at different learning trajectories can help in deciding how much input the learner may be able to process at a time.

In case of curriculum designers and educational researchers, it will inspire them to further probe into what learners are able to produce by contributing to our ongoing search for what exactly is entailed by L2 development. In particular, the study will help curriculum designers to make an informed decision on which learning materials are most appropriate for learners at different proficiency levels and how these linguistic features can be introduced in a systematic way.

## **6.6 Limitations**

There are several limitations to this study that should be addressed. First and foremost, while studies have illustrated the appropriateness of 4-gram sequences in capturing learners' language use, by no means does it capture the full scope of language development. For instance, shorter patterns may have been neglected, and longer patterns may have been cut short. In the process of analyses, it was evident that learners utilized longer sequences that were difficult to make sense of based on the 4-gram sequences. However, the bigger problem might be in shorter phrases (e.g., 2- and 3- gram sequences), which may have been ignored since the study only sought out 4-gram sequences. Regardless, there are also shortfalls to shorter n-gram sequences. For instance, it would be hard to make sense of 2-gram sequences if they consisted of a be-verb and a determiner (e.g., "is a"). Thus, an investigation into both short and long sequences is necessary to make a full conclusion about learners' use of verb sequences.

Another caveat to this study is the lack of access to the raw data. While syntactic complexity and patterns were not in the scope of this study, having access to the raw data would

allow for triangulation of observing learner development from various angles, thereby providing a more comprehensive understanding of L2 writing development.

Lastly, perhaps the biggest limitation of this study may have to do with the corpus itself. The CLC has gained its merit as one of the biggest learner corpora to date. Furthermore, unlike other learner corpora, the CLC is not only systematic but the proficiency levels are distinguished according to the CEFR. While the corpus and its size account for learners from a wide range of backgrounds, it doesn't account for potential predictors of variation such as aptitude and task. Additionally, the data is inconsistent across the proficiency levels, posing a few difficulties. First and foremost, when comparing across the different groups, it was impossible to draft the most frequently used examples from the core and emerging sequences due to there being little data, particularly at A2 and C2 levels. While it makes sense that most of the learners would fall in the 'average' (e.g., B1 and B2 levels), the small data in these marginal proficiency levels made it difficult to grasp what the most frequently used examples may be. Therefore, the analyses were conducted by drafting a random sample of 50 examples from each of the frequency levels. This may, however, be a caveat to taking a granular approach.

## **6.7 Recommendations for Future Research**

The current study is by no means exhaustive and is not intended to capture the breadth of language development in its entirety. Rather, the study attempted to illustrate a bird's eye view of the changes and restructuring of the languages observed as L1 Korean speakers become more proficient in L2 writing. Thus, the findings in this study offer a myriad of recommendations for future research.

First, future research could probe further into some of the linguistic features that surfaced in this study, such as collocations and phrasal verbs. While studies such as Siyanova-Chanturia (2015) offer insight into the use of collocations longitudinally, using a large set of data with a solid indication of the proficiency levels would provide a deeper understanding of what the learners can produce at different learning trajectories. Second, the current study measured development regarding learners' use of verbs only. However, conducting the same study using different POS tags such as nouns and adjectives could provide different findings from the current study.

Additionally, future studies should consider various writing genres and tasks. The current study did not take into consideration the various writing genre and tasks despite many studies that have shown genre effects and task effects on writing. The main reason for rejecting these claims was the ground that the goal of this study was to use a large dataset under the broad term “academic writing,” which are writing sample written under exam conditions. Whether genre and tasks had a great impact on learners’ use of sequences and patterns can be further investigated.

Lastly, as mentioned above, one of the limitations of this study is the lack of raw data. Having access to the data would make it possible to measure some of the findings in this study more statistically. For instance, the syntactic complexity could be measured if the raw data was available. Although complexity, accuracy, and fluency measures were not within the scope of this study, these findings could help to supplement the evidence about L2 writing development found in this study.

## REFERENCE

- Aarts, J., & Granger, S. (1998). Tag sequences in learner corpora: A key to interlanguage grammar and discourse. In S. Granger (Ed.) *Learner English on computer* (pp. 132-141). London/New York: Addison Wesley Longman.
- Ädel, A. (2008). Involvement features in writing: Do time and interaction trump register awareness? In G. Gilquin, S. Papp, & B. Díez-Bedmar (Eds.), *Linking up contrastive and learner corpus research* (pp. 33-53). Amsterdam: Rodopi.
- Ädel, A. (2015). Variability in learner corpora. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *Cambridge handbook of learner corpus research* (pp. 379-400). Cambridge: Cambridge University Press.
- Atkins, S., Clear, J., & Ostler, N. (1992). Corpus design criteria. *Literary & Linguistic Computing*, 7(1), 1-16.
- Babanoglu, M. (2014). A corpus-based study on the use of pragmatic markers as speech-like features in Turkish EFL learners' argumentative essays. *Social and Behavioral Sciences*, 136, 186-193.
- Back, J. (2011). Preposition errors in writing and speaking by Korean EFL learners: A corpus-based approach. *Studies in British and American Language and Literature*, 99, 227-247.
- Barlow, M. (2000). Usage, Blends and Grammar. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language* (pp. 287-313). Stanford: CSLI.
- Barlow, M., & Kemmer, S. (2000). *Usage-based models of language*. Stanford, CA: CSLI Publications.
- Barlow, M. (2005). Computer-based analyses of learner language. In R. Ellis & G. Barkhuizen (Eds.), *Analysing learner language* (pp. 335-357). Oxford: OUP.
- Behrens, H. (2009). Usage-based and emergentist approaches to language acquisition. *Linguistics*, 47, 383-411.
- 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.
- Bi, P. (2020). Revisiting genre effects on linguistic features of L2 writing: A usage-based perspective. *International Journal of Applied Linguistics*, 30(3), 429-444.

- Biber, D. (2000). Investigating language use through corpus-based analyses of association patterns. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language* (pp. 287-313). Stanford: CSLI.
- Biber, D. (2009). Corpus-based and corpus-driven analyses of language variation and use. In B. Heine & H. Narrog (Eds.), *The Oxford handbook of linguistic analysis* (pp. 1-22). Oxford: The Oxford University Press.
- Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, 45(1), 5-35.
- Biber, D., & Leech, G. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education.
- Bolton, K., Nelson, G., & Hung, J. (2002). A corpus-based study of connectors in student writing; Research from the International Corpus of English in Hong Kong (ICE-HK). *International Journal of Corpus Linguistics*, 7(2), 165-182.
- Bot, K, Lowie, W., & Verspoor, M. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and Cognition*, 10(1), 7-21.
- Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348-393.
- Bowerman, M. (1976). Semantic factors in the acquisition of rules for word use and sentence construction. In D. Morehead, and A. M. Morehead (Eds.), *Normal and deficient child language* (pp. 99-179). Baltimore: University Park Press.
- Braine, M. (1976). Children's first word combinations. *Monographs of the Society for Research in Children Development* 41(1), 1-104.
- Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing*, 26, 42-65.
- Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, 10, 425-455.

- Bybee, J. (2008). Usage-based grammar and second language acquisition. In P. Robinson & N. Ellis (Eds.), *Handbook of cognitive linguistics in second language acquisition* (pp. 216-236). London: Routledge.
- Bybee, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Bychkovska, T., & Lee, J. (2017). At the same time: Lexical bundles in L1 and L2 university student argumentative writing, *Journal of English for Academic Purposes*, 30, 38-52.
- Cambridge University Press (2013). *Introductory Guide to the Common European Framework of Reference (CEFR) for English Language Teachers*. [Booklet].
- Cancino, H., Rosansky, E., & Schumann, J. (1978). The acquisition of English negatives and interrogatives by native Spanish speakers. In E. M. Hatch (Ed.), *Second language acquisition: A book of readings* (pp.207-230). Rowley, MA: Newbury House.
- Carter, R. & McCarthy, M. (2004). *Cambridge Grammar of English: A Comprehensive Guide*. Cambridge: Cambridge University Press.
- Chang, Y. (2018). Features of lexical collocations in L2 writing: A case study of Korean adult learners of English. *English Teaching*, 73(2), 3-36.
- Chen, Y., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14(2), 30-49.
- Cho, M. (2019). The effect of prompts on L2 writing performance and engagement. *Foreign Language Annals*, 52, 576-594.
- Cho, Y. (2014). Perception of second language writing: Comparing two ESL learners. *English Language Teaching*, 26(2), 23-42.
- Choi, M. (2017). *Use of English Articles by Korean Students*. Unpublished master's thesis. St. Cloud State University. Minnesota, U.S.
- Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.
- Chomsky, N. (1959). A review of B.F. Skinner's *Verbal behavior*. *Language* 35, 26-58.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. (1978). *Topics in the theory of generative grammar*. Berlin: de Gruyter.
- Chomsky, N. (1980). *Rules and representations*. Oxford: Blackwell.
- Chomsky, N. (1988). *Language and Problems of Knowledge: The Managua Lectures*. Cambridge, MA: MIT Press.

- Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics*, 5, 161-169.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 213-238.
- Croft, W. (1991). *Syntactic categories and grammatical relations: The cognitive organization of information*. Chicago: University of Chicago Press.
- Croft, W. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Croft, W., & Cruse, D. A. (2004). *Cognitive linguistics*. New York: Cambridge University Press.
- Crossley, S. A., Cobb, T., & McNamara, D. S. (2013). Comparing count-based and band-based indices of word frequency: Implications for active vocabulary research and pedagogical applications. *System*, 41, 965-981.
- Dagneaux, E., Denness, S., & Granger, S. (1998). Computer-aided error analysis. *System*, 26, 163-174.
- Daskalovska, N. 2014. Corpus-based versus traditional learning of collocations. *Computer Assisted Language Learning* 28(2), 130-144.
- 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.
- Díez-Bedmar, M. B. (2018). Fine-tuning descriptors for CEFR B1 level: Insights from learner corpora. *ELT Journal*, 72(2), 199-209.
- Domaradzki, M. (2007). Cognitive critique of generative grammar. *Lingua* 17, 38-58.
- Dulay, H., Burt, M., & Krashen, S. (1982). *Language Two*. Oxford: Oxford University Press.
- Ellis, N. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, 24(2), 143-188.
- Ellis, N. (2008). Usage-based and form-focused language acquisition: The associative learning of constructions, learned attention, and the limited L2 endstate. In P. Robinson & N. Ellis (Eds.) *Handbook of cognitive linguistics and second language acquisition* (pp. 372-405). London/New York: Routledge.

- Ellis, N. (2009). Optimizing the input: Frequency and sampling in usage-based and form-focused learning. In M. Long & C. Doughty (Eds.) *The handbook of language teaching* (pp. 139-158). New Jersey: Wiley-Blackwell.
- Ellis, N. (2012). Second language acquisition. In G. Trousdale & T. Hoffmann (Eds.), *Oxford handbook of construction grammar*. Oxford, UK: Oxford University Press.
- Ellis, N., & Cadierno, T. (2009). Constructing a second language: Introduction to the special section. *Annual Review of Cognitive Linguistics* 7, 111-139.
- Ellis, N., & Robinson, P. (2008). An introduction to cognitive linguistics, second language acquisition, and language instruction. In P. Robinson & N. Ellis (Eds.) *Handbook of cognitive linguistics and second language acquisition* (pp. 3-24). London/New York: Routledge.
- Ellis, N., & Larsen-Freeman, D. (2009). Constructing a second language: Analyses and computational simulations of the emergence of linguistic constructions from usage. *Language Learning*, 52(1), 90-125.
- Ellis, N., Simpson-Vlach, R., Römer, U., O'Donnell, M., & Wulff, S. (2015). Learner corpora and formulaic language in second language acquisition research. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research (Cambridge handbooks in language and linguistics)* (pp. 357-378). Cambridge: Cambridge University Press.
- Ellis, N. C., O'Donnell, M. B., & Römer, U. (2013). Usage-based language: Investigating the latent structures that underpin acquisition. *Currents in Language Learning*, 1, 25-51.
- Ellis, N., Römer, U., & O'Donnell, M. (2016). *Usage-based approaches to language acquisition and processing: Cognitive and corpus investigations of construction grammar*. Oxford: Wiley.
- Ellis, R. (1994). *The Study of Second Language Acquisition*. Oxford: Oxford University Press.
- Eskildsen, S. (2008). Constructing another language- Usage-based linguistics in second language acquisition. *Applied Linguistics* 30(3), 335-357.
- Eskildsen, S., & Cadierno, T. (2007). Are recurring multi-word expressions really syntactic freezes? Second language acquisition from the perspective of usage-based linguistics. In M. Nenonen and S. Niemi (Eds.). *Collocations and Idioms 1: Papers from the First*

- Nordic Conference on Syntactic Freezes, Joensuu, May 19-20, 2007*. Studies in Languages, University of Joensuu, Vol. 41. Joensuu: Joensuu University Press.
- Field, Y., & Yip, L. M. O. (1992). A comparison of internal conjunctive cohesion in the English essay writing of Cantonese speakers and native speakers of English. *RELC Journal*, 23(1), 15-28.
- Fillmore, C. (1977). The case for case reopened. In P. Cole & J. Sadock, (Eds.), *Syntax and semantics 8: Grammatical Relations* (pp. 59-82). New York: Academic Press.
- Fillmore, C. (1989). Grammatical construction theory and the familiar dichotomies. *North-Holland Linguistic Series: Linguistic Variations 54*, 17-38.
- Fillmore, C. J., Kay, P., & O'Connor, C. (1988). Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language* 64, 501-538.
- Flowerdew, L. (2001). 'The exploitation of small learner corpora in EAP materials design'. In M. Ghadessy and R. Roseberry (Eds.) *Small Corpus Studies and ELT* (pp. 363-379). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Fogal, G., & Verspoor, M. (2020). *Complex Dynamic Systems Theory and L2 Writing Development*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Francis, G. (1993). A corpus-driven approach to grammar- principles, methods and examples. In M. Baker, G. Francis & E. Tognini-Bonelli (Eds.) *Text and Technology: In Honour of John Sinclair* (pp. 137-156). Amsterdam/Philadelphia, PA: Benjamins.
- Francis, G., Hunston, S. & Manning, E. (1996). *Collins Cobuild Grammar Patterns 1: Verbs*. London/New York: HarperCollins Publisher.
- Francis, G., Hunston, S. & Manning. (1998). *Collins Cobuild Grammar Pattern 2: Nouns and Adjectives*. London/New York: HarperCollins Publisher.
- Gablasova, D., Brezina, V., & McEnery, T. (2017). Exploring learner language through corpora: Comparing and interpreting corpus frequency information. *Language Learning*, 67(1), 130-154.
- Garnier, M., & Schmitt, N. (2016). Picking up polysemous phrasal verbs: How many do learners know and what facilitates this knowledge?, *System*, 59, 29-44.
- Geist, M. (2017). Noticing grammar in L2 writing and problem-solving strategies. *Studies in Second Language Learning and Teaching*, 7(3), 471-487.

- Gilquin, G., & Granger, S. (2015). Learner language. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 418-436). Cambridge: Cambridge University Press.
- Gilquin, G., & Paquot, M. (2008). Too chatty: Learner academic writing and register variation. *English Text Construction* 1(1), 41-61.
- Gilquin, G., Granger, S., & Paquot, M. (2007). Learner corpora: The missing link in EAP pedagogy. *Journal of English for Academic Purposes*, 6, 319-335.
- Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago, IL: The Chicago University Press.
- Goldberg, A. E. (2006). *Construction at work*. New York: Oxford University Press.
- Goldberg, A. E., & Casenhiser, D. (2008). Construction learning and second language acquisition. In P. Robinson and N. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 197-215). New York: Routledge.
- Goldberg, A. E., Casenhiser, D. M., & Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive Linguistics*, 15, 289-316.
- Granger, S. (1996). From CA to CIA and back: An integrated contrastive approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg, & M. Johansson (Eds.), *Languages in contrast: Text-based cross-linguistic studies* (pp. 37-51). Lund: Lund University Press.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and application* (pp. 145-160). Oxford: Clarendon Press.
- Granger, S. (2015). Contrastive interlanguage analysis: A reappraisal. *International Journal of Learner Corpus Research*, 1(1), 7-24.
- Granger, S. (2008). Learner corpora. In A. Lüdeling, & M. Kytö, (Eds.), *Corpus Linguistics. An International Handbook. Volume 1*. 259-275. Berlin & New York: Walter de Gruyter.
- Granger, S., & Bestgen, Y. (2014). The use of collocations by intermediate vs. advanced non-native writers: A bi-gram based study. *International Review of Applied Linguistics in Language Teaching (IRAL)*, 52, 229-252.
- Green, C. (2019). Enriching the academic wordlist and secondary vocabulary lists with lexicogrammar: Toward a pattern grammar of academic vocabulary. *System* 87, 1-10.

- Green, C., & Lambert, J. (2018). Advancing disciplinary literacy through English for academic purposes: Discipline-specific wordlists, collocations and word families for eight secondary subjects. *Journal of English for Academic Purposes*, 35, 105-115.
- Gries, S. T., & Deshors, S. C. (2020). Statistical analyses of learner corpus data. In *Routledge handbook of second language acquisition and corpora* (pp. 119-132). New York/London: Routledge.
- Halliday, M. A. K. (2003). *On Language and Linguistics* (vol.3). London: Continuum.
- Hinkel, E. (1997). The past tense and temporal verb meanings in a contextual frame. *TESOL Quarterly*, 31(2), 289-313.
- Hinkel, E. (2009). The effects of essay topics on modal verb uses in L1 and L2 academic writing. *Journal of Pragmatics*, 41(4), 667-683.
- Hoey, M. & O'Donnell, M. (2008). Lexicography, grammar, and textual position. *International Journal of Lexicography*, 21(3), 209-309.
- Hopper, P. J. (1987). Emergent grammar. *Berkeley Linguistics Society*, 13, 139-157.
- Hopper, P. J. (1998). Emergent grammar. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (pp. 155- 176). Mahwah, NJ: Erlbaum.
- Hornby, A. (1974). *Guide to patterns and usage in English*. Second edition. Oxford: Oxford University Press.
- Housen A. (1995). It's about time- The acquisition of Temporality in English as a second language in a multilingual educational context. Free University of Brussels: Unpublished PhD Dissertation.
- Housen, A. (2002). A corpus-based study of the L2 acquisition of the English verb system. In S. Granger, J. Hung, and S. Petch-Tyson (Eds.), *Computer learner corpus, second language acquisition and foreign language teaching* (pp. 77-116). Amsterdam: John Benjamins.
- Hunston, S. (2008). Patterns, lexis and semantic sequences. *International Journal of Corpus Linguistics*, 13(3), 271-295.
- Hunston, S. (2014). Pattern grammar in context. In T. Herbst, H. J. Schmid & S. Faulhaber (Eds.) *Constructions, collocations, patterns* (pp. 99-120). Berlin: de Gruyter.
- Hunston, S. (2019). Patterns, constructions, and applied linguistics. *International Journal of Corpus Linguistics*, 24(3), 324-353.

- Hunston S. & Francis, G. (2000). *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English*. Amsterdam/Philadelphia, PA: John Benjamins.
- Hunston, S., & Su, H. (2019). Patterns, constructions, and local grammar: A case study of 'evaluation'. *Applied Linguistics*, 40(4), 567-593.
- Jarvis, S., & Daller, M. (2013). *Vocabulary knowledge: Human rating and automated measures*. Amsterdam: John Benjamins.
- Joffe, H., & Yardley, L. (2003). Content and thematic analysis. In D.F Marks & L, Yardley (Eds.), *Research methods for clinical and health psychology*. (pp. 56-68). Thousand Oaks: Sage.
- Johnson, K., & Johnson, H. (1999). *Encyclopedic Dictionary of Applied Linguistics*. New Jersey: Wiley-Blackwell Publishing.
- Kemmer, S., & Barlow, M. (2000). Introduction: A usage-based conception of language. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language* (pp. vii-xxviii). Stanford: CSLI.
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, 1, 7-36.
- Kilmova, B. F. (2014). Constraints and difficulties in the process of writing acquisition, *Procedia-Social and Behavioral Science*, 122, 433-437.
- Kim, T. Y. (2008). Korean L2 writers' previous writing experience: L1 literacy development in school. *Second Language Studies*, 27(1), 103-154.
- Kim, S. (2016). *L1 influence on L2 learning of English resultative constructions: The syntactic and semantic structure of Korean students' interlanguage*. Unpublished doctoral dissertation. Seoul National University. Seoul, Korea.
- Kim, H., & Sung, M. (2019). A usage-based analysis of L2 production of English resultative constructions. *Language Research*, 55(1), 151-178.
- Kim, J. Y. (2014). Predicting L2 writing proficiency using linguistic complexity measures: A corpus-based study. *English Teaching*, 69(4) 27-51.
- Kim, Y., & Emel'yanova, L. (2021). The effects of written corrective feedback on the accuracy of L2 writing: Comparing collaborative and individual revision behavior. *Language Teaching Research*, 25(2), 234-255.

- Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, 49(4), 757-786.
- Kyle, K., & Crossley, S. (2017). Assessing syntactic sophistication in L2 writing: A usage-based approach. *Language Testing*, 34(4), 513-535.
- Kyle, K., Crossley, S., & Verspoor, M. (2021). Measuring longitudinal writing development using indices of syntactic complexity and sophistication. *Studies in Second Language Acquisition*, 43(4), 781-812.
- Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago, IL: University of Chicago Press.
- Langacker, R. W. (1987). *Foundations of cognitive grammar: Theoretical Prerequisites*. Volume I. Stanford: Stanford University Press.
- Langacker, R. W. (1991). *Foundations of Cognitive Grammar*. Volume II. Stanford: Stanford University Press.
- Langacker, R. W. (1988). A usage-based model. In Brygida Rudzka-Ostyn (Ed.), *Topics in Cognitive Linguistics (Current Issues in Linguistic Theory 50)* (pp. 127-61). Amsterdam: Benjamins.
- Langacker, R. W. (2000). A dynamic usage-based model. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language*, 1-63. Stanford: CSLI.
- Lah, J., & Yoo, I. W. (2015). A corpus analysis of the preposition of *in* Korean college matriculants' writing. *English Teaching*, 70(3), 99-115.
- Lan, G., Liu, Q., & Staples, S. (2019). Grammatical complexity: 'What does it mean' and 'so what' for L2 writing classrooms? *Journal of Second Language Writing*, 46, 100673.
- Larsen-Freeman, D. (2013). Chaos/Complexity theory for second language acquisition. In C. Chapelle (Ed.) *The nyclopedia of Applied Linguistics* (pp. 590-597). Oxford, UK: Wiley-Blackwell.
- Larsen-Freeman, D., & Celce-Murcia, M. (2016). *The grammar book: form, meaning and use for English language teachers (3<sup>rd</sup> Ed.)*. Boston: National Geographic Learning.
- Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning* 61, 647-672.
- Leech, G. (1998). Preface. In S. Granger & G. Leech (Eds.), *Learner English on computer* (pp. xiv-xx). Abingdon: Routledge.

- Lewis, M. (1997). *Implementing the lexical approach: Putting theory in practice*. Hove: Language Teaching Publications.
- Li, J., & Schmitt, N. (2010). The development of collocation use in academic texts by advanced L2 learners: A multiple case study approach. *Perspectives on formulaic language: Acquisition and Communication*, 22-46.
- Lieven, E. (2014). First language learning from a usage-based approach. In T. Herbst, H. Schmid, & S. Faulhaber (Eds.) *Constructions collocations patterns* (pp. 9-32). Berlin/Munich/Boston: De Gruyter.
- Lieven, E., & Tomasello, M. (2008). Children's first language acquisition from a usage-based perspective. In P. Robinson & N. Ellis (Eds.) *Handbook of cognitive linguistics and second language acquisition* (pp. 168-196). London/New York: Routledge.
- Lin, M., & Liou, H. (2009). Expansion of EFL academic vocabulary for writing via web-enhanced lexical instruction. *English Teaching & Learning* 33(2), 95-146.
- Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures and indices of college-level ESL writers' language development. *TESOL Quarterly*, 45(1), 36-62.
- McEnery, T., & Hardie, A. (2012). *Corpus Linguistics*. Cambridge: Cambridge University Press.
- McEnery, T., & Wilson, A. (1996). *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- MacWhinney, B. (2001). Emergence from what? *Journal of Child Language*, 28, 726-732.
- Matsuda, P. (2003). Second language writing in the twentieth century: A situated historical perspective. In B. Kroll (Ed.), *Exploring the dynamics of second language writing (Cambridge Applied Linguistics)* (pp. 15-34). Cambridge: Cambridge University Press.
- Mostafa, T., & Crossley, S. (2020). Verb argument construction complexity indices and L2 writing quality. Effects of writing tasks and prompts. *Journal of Second Language Writing*, 49, 100730.
- Nassaji, H., & Tian, J. (2010). Collaborative and individual output tasks and their effects on learning English phrasal verbs. *Language Teaching Research*, 14(4), 397-419.
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24, 233-242.
- Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam: John Benjamins.
- Ninio, A. (2006). *Language and the learning curve: A new theory of syntactic development*. Oxford, UK: Oxford University Press.

- O'Donnell, M. B., & Ellis, N. C. (2010). Towards an inventory of English verb argument constructions. *Proceedings of the 11<sup>th</sup> Annual Conference of the North American Chapter of the Association for Computational Linguistics*, Los Angeles.
- O'Keeffe, A., & Mark, G. (2017). The English grammar profile of learner competence. *International Journal of Corpus Linguistics*, 22(4), 457-489.
- Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level writing. *Applied Linguistics*, 24(4), 492-518.
- Osborne, J. (2004). Top-down and bottom-up approaches to corpora in language teaching. In U. Connor & T. A. Upton (Eds.), *Applied Corpus Linguistics: A Multidimensional Perspective* (pp. 251- 265). Amsterdam: Brill.
- Palviainen, A., Kalaja, P., & Mäntylä, K. (2012). Development of L2 writing: Fluency and proficiency. *AFinLA-E: Soveltavan kielitieteen tutkimuksia*, 4, 47-59.
- Paquot, M. (2010). *Academic vocabulary in learner writing: From extraction to analysis*. London: Continuum International Publishing Group.
- Paquot, M. (2013). Lexical bundles and L1 transfer effects. *International Journal of Corpus Linguistics*, 18(3), 391-417.
- 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. 191-226). London: Longman.
- Perek, F. (2015). *Argument structure in usage-based construction grammar*. Amsterdam: John Benjamins.
- Pérez-Paredes, P., Mark, G., and O'Keeffe, A. (2020). *The impact of usage-based approaches on second language learning and teaching*. Cambridge Education Research Reports. Cambridge University Press.
- Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychon Bull Rev*, 21, 1112-1130.
- Polio, C. & Park, J. H. (2016). Language development in second language writing. In R. Manchón & P. Matsuda (Eds.), *Handbook of L2 writing*. New York: Routledge.

- Reid, J. (2002). Writing. In R. Carter & D. Nunan (Eds.) *The Cambridge guide to teaching English to speakers of other languages*, (pp. 28-33). Cambridge, UK: Cambridge University Press.
- Reppen, R. (2009). Exploring L1 and L2 writing development through collocations: A corpus-based look. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language* (pp. 49-59). London: Palgrave Macmillan.
- Robinson, P., & Ellis, N. (2008a). An introduction to cognitive linguistics, second language acquisition, and language instruction. In P. Robinson & N. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 3-24). New York: Routledge.
- Robinson, P., & Ellis, N. C. (2008b). *Handbook of cognitive linguistics and second language acquisition*. London: Routledge.
- Römer, U. (2009). English in academia: Does nativeness matter? *International Journal of English Studies*, 20(2), 89-100.
- Römer, U. (2009). The inseparability of lexis and grammar: Corpus linguistic perspectives. *Annual Review of Cognitive Linguistics* 7, 141-163.
- Römer, U. (2019). A corpus perspective on the development of verb constructions in second language learners. *International Journal of Corpus Linguistics*, 24(3), 268-290.
- Römer, U., Skalicky, S. C., & Ellis, N. C. (2020). Verb argument constructions in advanced L2 English learner production: Insights from corpora and verbal fluency tasks. *Corpus Linguistics and Linguistics Theory*, 16(2), 303-331.
- Schwartz, B. D. & Sprouse, R. (2013). Generative approaches and the Poverty of the Stimulus. In J. Herschensohn and M. Young-Scholten (Eds.), *The Cambridge handbook of second language acquisition* (pp. 137-158). Cambridge: Cambridge University Press.
- Sethuraman, N. (2002). The acquisition of verbs and argument structure constructions. Unpublished doctoral dissertation. University of California, San Diego: La Jolla.
- Shin, G. (2017). Developmental aspects of English argument structure constructions for Korean-speaking second language learners: Usage-based constructional approaches to language development. *Ampersand*, 4, 10-20.
- Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing. *TESOL Quarterly*, 27(4), 657-677.

- Sinclair, J. (1987a). Collocation: A progress report. In R. Steele & T. Threadgold (Eds.), *Language Topics. Essays in honour of Michael Halliday* (pp. 319-331). Amsterdam: John Benjamins.
- Sinclair, J. (1987b). The nature of the evidence. In J. Sinclair (Ed.), *Looking Up: An Account of the COBUILD Project in Lexical Computing* (pp. 150-159). London: HarperCollins.
- Sinclair, J. (1991). *Corpus Concordance Collocation*. Oxford: Oxford University Press.
- Sinclair, J. (1995). *Collins COBUILD English Dictionary*. London: HarperCollins.
- Sinclair, J. (1996). The search for units of meaning. *Textus*, (9)1, 75-106.
- Siyanova-Chanturia, A. (2015). Collocation in beginner learner writing: A longitudinal study. *System*, 53, 148-160.
- Stefanowitsch, A., & Gries, S. T. (2003). Collocations: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8(2), 209-243.
- Sun, Y., & Wang, L. (2003). Concordancers in the EFL classroom: Cognitive approaches and collocation difficulty. *Computer Assisted Language Learning* 16(1), 83-94.
- Thewissen, J. (2013). Capturing L2 accuracy developmental patterns: Insights from an error-tagged EFL learner corpus. *The Modern Language Journal*, 97(1), 77-101.
- Tomasello, M. (1992). *First verbs: A case study of early grammatical development*. Cambridge, MA: Cambridge University Press.
- Tomasello, M. (2000). First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics*, 11(1), 61-82.
- Tomasello, M. (2003). *Constructing a language*. MA: Harvard University Press.
- Tyler, A., & Ortega, L. (2016). Usage-based approaches to language and language learning: An introduction to the special issue. *Language and Cognition*, 8, 335-345.
- Verspoor, M., & Schmitt, N. (2013). Language and the lexicon in SLA. In P. Robinson (Ed.) *The Routledge encyclopedia of SLA* (353-360). New York: Routledge/Taylor & Francis.
- Verspoor, M., Lowie, W., & Van Dijk, M. (2008). Variability in second language development from a dynamic systems perspective. *The Modern Language Journal*, 92(2), 214-231.
- Verspoor, M., Schmid, M., & Xu, X. (2012). A dynamic usage-based perspective on L2 writing. *Journal of Second Language Writing*, 21(3), 239-263.
- Wolfe-Quintero, K., Inagaki, S., & Kim, H. (1998). *Second language development in writing: Measures of fluency, accuracy & complexity*. Honolulu: University of Hawai'i Press.

- Wolfersberger, M. (2003). L1 to L2 writing process and strategy transfer: A look at lower proficiency writers. *Teaching English as a Second or Foreign Language (TESL-EJ)*, 7(2), 1-12.
- Xu, J., & Xu, Z. (2007). Discourse management chunks in Chinese college learners' English speech: A spoken corpus-based study. *Foreign Language Teaching and Research*, 39(6), 437-443.
- Year, J., & Gordon, P. (2009). Korean speakers' acquisition of the English ditransitive construction: The role of verb prototype, input distribution, and frequency. *The Modern Language Journal*, 93(3), 399-417.
- Yoon, H. (2017). Linguistic complexity in L2 writing revisited: Issues of topic, proficiency, and construct multidimensionality. *System*, 66, 130-141.
- Yu, T., Yoo, I. W. (2010). Korean university students' use of prepositional verbs: A corpus-based study. *English Teaching*, 65(4), 403-424.
- Yunxia, S., Min, Y., & Zhuo, S. (2009). An empirical study on a computer-based corpus approach to English vocabulary teaching and learning. In W. Li and J. Zhou (Eds): *Proceedings of the 2nd International Conference on Computer Science and Information Technology*. IEEE Xplore, pp. 218–21.
- Zipf, G. K. (1935). *The Psycho-Biology of Language: An introduction to dynamic philology*. Cambridge, MA: MIT Press.

## Appendices

### Appendix 1

CEFR Global Scale (Cambridge University Press, 2013)

C2	<ul style="list-style-type: none"><li>• Can understand with ease virtually everything heard or read.</li><li>• Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation</li><li>• Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.</li></ul>
C1	<ul style="list-style-type: none"><li>• Can understand a wide range of demanding, longer texts, and recognise implicit meaning.</li><li>• Can express him/herself fluently and spontaneously without much obvious searching for expressions.</li><li>• Can use language flexibly and effectively for social, academic and professional purposes.</li><li>• Can produce clear, well-structured, detailed text on complex subject, showing controlled use of organisational patterns, connectors and cohesive devices.</li></ul>
B2	<ul style="list-style-type: none"><li>• Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation.</li><li>• Can interact with a degree of fluency and spontaneity that makes regular interactions with native speakers quite possible without strain for either party.</li><li>• Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.</li></ul>
B1	<ul style="list-style-type: none"><li>• Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc.</li><li>• Can deal with most situations likely to arise whilst travelling in an area where the language is spoken.</li><li>• Can produce simple connected text on topics, which are familiar, or of personal interest.</li><li>• Can describe experiences and events, dreams, hopes &amp; ambitions and briefly give reasons and explanations on opinions and plans.</li></ul>
A2	<ul style="list-style-type: none"><li>• Can understand sentences and frequently used expressions related to areas most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment).</li><li>• Can communicate in simple and routine tasks requiring simple and direct exchange of information on familiar and routine matters.</li></ul>

	<ul style="list-style-type: none"><li>• Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.</li></ul>
A1	<ul style="list-style-type: none"><li>• Can understand and use familiar everyday expressions and very basic phrases aimed at satisfaction of needs of a concrete type.</li><li>• Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has.</li><li>• Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.</li></ul>

## Appendix 2

### List of L1 in the CLC

Afrikaans	Haitian Creole	Rajasthani
Akan	Hausa	Riff
Albanian	Hebrew	Romanian
Amharic	Hindi	Romansch
Arabic- Egyptian	Hungarian	Russian
Arabic- Gulf	Ibo/Igbo	Samoan
Arabic- Levant	Icelandic	Serbian
Arabic- Meghreb	Igala	Shona
Arabid- Other	Indonesian	Sindhi
Armenian	Italian	Singhalese
Assamese	Japanese	Slovak
Aymara	Javanese	Slovene
Azerbaijani	Kannada	Somali
Baluchi	Kashmiri	Spanish- European
Bambara	Kazakh	Spanish- Latin American
Basque	Khmer	Spanish- Other
Bemba	Korean	Swahili
Bengali	Kurdish	Swazi
Bihari	Lao	Swedish
Breton	Latvian	Swiss German
Bulgarian	Lingala	Tagalog
Burmese	Lithuanian	Tahitian
Byelorussian	Luba	Tamil
Cantonese	Luganda	Tatar
Catalan	Luo	Telugu
Chinese	Luxemburgish	Tetum
Creole	Malagasy	Thai
Croatian	Malay	Tibetan
Czech	Malayalam	Tigrinya
Danish	Malinka	Tongan
Dari	Maltese	Trukese
Dutch	Mandarin	Tulu
Efik	Maori	Tupi/Gurani
English	Marathi	Turkish
Estonian	Marshallese	Uighur
Ewe	Masai	Ukrainian
Faeroese	Mende	Ulithian
Farsi	Mongolian	Urdu
Fijian	Nepali	Uzbek
Finnish	Norwegian	Vietnamese
Flemish	Oriya	Wolof
French	Palauan	Yao
Fulani	Panjabi	Yapese
Ga	Pashto	Yiddish
Georgian	Polish	Yoruba
German	Ponapean	Zulu
Gilbertese	Portuguese	
Greek	Pushtu	
Gujarati	Quechua	

## Appendix 3

### BNC written genres

Description
academic prose: humanities
academic prose: medicine
academic prose: natural sciences
academic prose: politics law education
academic prose: social & behavioural sciences
academic prose: technology computing engineering
administrative and regulatory texts, in-house use
print advertisements
biographies/autobiographies
commerce & finance, economics
e-mail sports discussion list
school essays
university essays
excerpts from two modern drama scripts
single- and multiple-author collections of poems
novels & short stories
Hansard/parliamentary proceedings
official/governmental documents/leaflets company annual reports etc.; excludes Hansard
instructional texts/DIY
personal letters
professional/business letters
miscellaneous texts
TV autocue data
broadsheet national newspapers: arts/cultural material
broadsheet national newspapers: commerce & finance
broadsheet national newspapers: personal & institutional editorials & letters-to-the-editor
broadsheet national newspapers: miscellaneous material
broadsheet national newspapers: home & foreign news reportage
broadsheet national newspapers: science material
broadsheet national newspapers: material on lifestyle leisure belief & thought
broadsheet national newspapers: sports material
regional and local newspapers: arts
regional and local newspapers: commerce & finance
regional and local newspapers: home & foreign news reportage
regional and local newspapers: science material
regional and local newspapers: material on lifestyle, leisure, belief & thought
regional and local newspapers: sports material
tabloid newspapers
non-academic/non-fiction: humanities

non-academic: medical/health matters  
non-academic: natural sciences  
non-academic: politics law education  
non-academic: social & behavioural sciences  
non-academic: technology, computing, engineering  
popular magazines  
religious texts, excluding philosophy

#### Appendix 4 Penn Treebank Tagset

POS Tag	Description	Example
CC	Coordinating conjunction	And
CD	Cardinal number	1, third
DT	Determiner	The
EX	Existential there	There is
FW	Foreign word	Les
IN	Preposition, subordinating conjunction	In, of, like
IN/that	That as subordinator	That
JJ	Adjective	Green
JJR	Adjective, comparative	Greener
JJS	Adjective, superlative	Greenest
LS	List marker	1)
MD	Modal	Could, will
NN	Noun, singular or mass	Table
NNS	Noun plural	Tables
NP	Proper noun, singular	John
NPS	Proper noun, plural	Vikings
PDT	Predeterminer	Both the boys
POS	Possessive ending	Friend's
PP	Personal pronoun	I, he, it
PPZ	Possessive pronoun	My, his
RB	Adverb	However, usually, naturally, here, good
RBR	Adverb, comparative	Better
RBS	Adverb, superlative	Best
RP	Particle	Give up
SENT	Sentence-break punctuation	.!?
SYM	Symbol	/[=*
TO	Infinitive 'to'	Togo
UH	Interjection	Uhhuhhhh
VB	Verb be, base form	Be
VBD	Verb be, past tense	Was, were
VBG	Verb be, gerund/present participle	Being
VBN	Verb be, past participle	Been
VBP	Verb be, sing. present, non-ed	Am, are
VBZ	Verb be, 3 <sup>rd</sup> person sing. present	Is
VH	Verb have, base form	Have
VHD	Verb have, past tense	Had
VHG	Verb have, gerund/present participle	Having
VHN	Verb have, past participle	Had
VHP	Verb have, sing. present, non-ed	Have

VHZ	Verb have, 3 <sup>rd</sup> person sing.present	Has
VV	Verb, base form	Take
VVD	Verb, past tense	Took
VVG	Verb, gerund/present participle	Taking
VVN	Verb, past participle	Taken
VVP	Verb, sing.present, non-3d	Take
VVZ	Verb, 3 <sup>rd</sup> person sing.present	Takes
WDT	Wh-determiner	Which
WP	Wh-pronoun	Who, what
WP\$	Possessive wh-pronoun	Whose
WRB	Wh-adverb	Where, when
#	#	#
\$	\$	\$
“	Quotation marks	‘ ‘
“	Opening quotation marks	‘ ‘
(	Opening brackets	( {
)	Closing brackets	) }
,	Comma	,
:	Punctuation	- ; : ...