

A CORPUS BASED EXPLORATION OF THE PROGRESSIVE *-KO ISS* CONSTRUCTION
IN L1, L2, AND TEXTBOOK KOREAN

By

Steven G. Gagnon

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Second Language Studies – Doctor of Philosophy

2024

ABSTRACT

Due to the typological differences between Korean's aspect system and English's aspect system in terms of progressive construction *-ko iss*, learners can no doubt have difficulty acquiring and using the *-ko iss* construction in learner Korean. This dissertation investigates two main points: (i) how is the *-ko iss* construction used in real-world Korean, including L1 Korean and L1 English and L1 Japanese learner Korean, and (ii) the way *-ko iss* is taught and used in textbooks as a main source of input for learners of Korean. To answer these questions, I use collostructional analysis to assess association strengths between verbs and the *-ko iss* (progressive) and simple (non-progressive) constructions to identify verbs that are well-attested with L1 and L2 Korean. Finally, I take an exploratory approach to using logistic regression to model L1 and L2 Korean data, the results of which can provide some insights into L1-L2 *-ko iss* usage, and insights from this initial regression analysis provide meaningful information to improve modeling of L1 and L2 Korean in future studies. The main takeaways from this study are:

- (a) Verbs co-occurring with *-ko iss* in the written Sejong corpus included a wide variety of usage cases, including many instances of stative or mental-type verbs, including *al* (*know*), *mid* (*believe*), among others.
- (b) Verbs co-occurring with *-ko iss* in the learner data showed a positive sign in that learners use and acquire the *-ko iss* construction's various semantic meanings, including its use with stative verbs. However, semantic domains used with *-ko iss* are limited when compared with the L1 data.
- (c) In textbooks, a limited number of verbs is introduced with *-ko iss* at the beginner levels. *-ko iss* is also taught in textbooks as a prototypical *action in progress* progressive

construction, without clear direction instruction on other senses of *-ko iss*. Further, across both textbook series, the frequency *-ko iss* is used at is low (maximum around 300 occurrences in a textbook series). Textbooks incidentally use *-ko iss* outside of the prototypical *action in progress* usage at later levels, however, frequencies are quite low. Findings from this dissertation can be used to inform language pedagogy. The list of verbs co-occurring with the *-ko iss* construction from the collostructional analysis provides teachers and textbook developers with a list of attested to function with *-ko iss* across a variety of usages beyond *action in progress*. Plenty of examples are also pulled from the corpus for materials developers to reference when designing textbook materials. As the aim of language teachers and materials developers is to use data-driven insights to improve teaching materials, exposing learners to a variety of verbs within contexts or lexical chunks they appear in via textbooks can aid in learning complex constructions in Korean.

TABLE OF CONTENTS

I. INTRODUCTION	1
II. LITERATURE REVIEW.....	4
III. METHOD	26
IV. RESULTS	40
V. DISCUSSION	101
REFERENCES	114
APPENDIX A: DISTINCTIVE COLLEXEME ANALYSIS I.....	120
APPENDIX B: DISTINCTIVE COLLEXEME ANALYSIS II	136
APPENDIX C: DISTINCTIVE COLLEXEME ANALYSIS III	137

I. INTRODUCTION

Corpus-driven explorations into real-world language use allow language teachers and researchers to uncover language patterns that occur frequently, which allows for development of teaching materials that mirror real-life usage rather than a speaker's language intuition. Led by John Sinclair, corpus linguistics and its potential as a proper field in applied linguistics was bolstered forward in 1987 with the publication of the *COBUILD English Language Dictionary*, a project which stemmed from the goal of building corpus-driven materials for second language learners of English. Such work helped push corpus-based and usage-based approaches forward towards recognizing that co-occurrence patterns of lexical items and syntactic structures are inseparably intertwined, thus revealing the link between form and meaning, particularly for commonly occurring collocation patterns (Sinclair, 2004).

The move towards corpus-driven, evidence-based approaches with large corpora as the backbone have influenced the development of learning materials such as textbooks, dictionaries (e.g., *COBUILD*), and so forth. As Sinclair puts it as the title of his 2004 book, we must “trust the text.” In other words, we must rely on real-world language examples in corpora to guide our understanding of language, which is a notable deviation from generative approaches to linguistics which rely heavily on linguistic intuition (Abbot & Tomasello, 2006). This is of particular importance as the use of corpora has allowed researchers to discover patterns which have otherwise gone unnoticed (e.g., Sinclair, 1997) when relying solely on a researcher's own intuition.

Since the seminal work on the *COBUILD* project, a multitude of corpus-based and corpus-driven learner materials such as dictionaries have been developed, particularly in the case of English as a second language teaching. However, despite the continued rise in corpus-based

studies and publications, there is much work to be done by corpus linguists to draw meaningful connections between research, teaching practices, and educational materials. In fact, even in the case of English language teachers, Römer (2011) argues that “the practice of English Language Teaching (ELT)... seems to be only marginally affected by the advances of corpus research” (p. 206). The sentiment was borne out in results in an earlier study by Römer (2005) which revealed that real-world English indeed differs from the English found in language teaching materials. Similarly, corpus-based explorations of Korean language learning materials (e.g., Jung, 2022) have shown deviations between language use in real-life and the language that appears in textbooks.

I agree with Römer that in the field of corpus linguistics “much work still remains to be done in bridging the gap between research and practice” (Römer, 2011, p. 206).

To that end, this dissertation outlines a corpus-driven approach to identify language patterns in L1, L2, and Textbook Korean with the goal of illuminating key differences that arise in learner language and offering suggestions for the creation of teaching materials such as language textbooks. Following in the footsteps of previous work on corpus data and textbook analysis (e.g., Jung, 2022), and Römer, 2005)), and using robust statistical methods to tease apart the differences between L1, L2, and language materials, the proposed dissertation project will be conducted with the goal of identifying verbs with co-occur with the progressive in L1 and L2 Korean, what factors predict the choice of a progressive form, and how these usage cases compare with Korean language textbook materials. This project will help move the field of Korean second language acquisition forward and help close the gap between teachers and researchers. Ultimately, this project addresses the needs of stakeholders in language education,

namely students, teachers, and language education material developers to provide learners with robust materials for language acquisition.

II. LITERATURE REVIEW

This study has two main aims which both work towards (i) understanding and modeling usage of the progressive and simple form (i.e., the “(non-)progressive”) both within and across language varieties of L1 and L2 Korean, and (ii) comparing L1 use of the progressive with language appearing in textbooks. Association strengths between verbs and the progressive and non-progressive are assessed using collocation analysis (Gries & Stefanowitsch, 2004) a statistical method to assess the attraction of a verb to a particular construction. Following that, I dig deeper to identify what predictors may impact the choice to use to use a progressive versus a non-progressive (e.g., aktionsart category, semantic domain, and L1) using logistic regression. To set the scene, studies on corpora and textbooks, L1 and L2 Korean and the progressive construction in Korean (and other languages), and descriptions of the progressive in Korean are provided.

2.1 Corpora and textbooks

In nearly all language education contexts, textbooks provide an important source of input for language learners. Textbooks provide many benefits to learners: they allow for concentrated practice on vocabulary, grammatical forms and functions, and practice activities for learners on the one hand, while simultaneously easing the burden on the language teacher by providing some materials to use in the classroom or for homework. When designed well, textbooks can provide learners with the chance to practice their target language (Lam, 2009) even if they do not have ample opportunities to communicate with L1 and expert interlocutors. As such, stakeholders in language education (including not only language teachers, but also textbook developers and publishers, and even students) should be invested in the research and development of language textbooks.

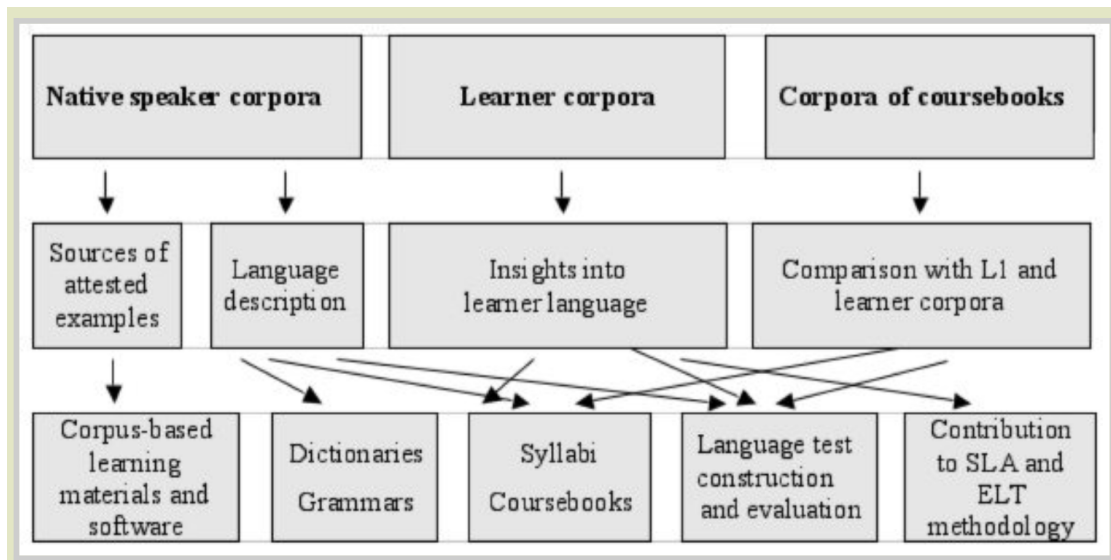
While the value of textbooks as learning resources is undeniable, there is a need for more robust explorations and comparisons of features in corpora and comparing them with how they are presented in textbooks and materials. However, insights from robust corpus studies seem to not be implemented often, and this may mean that certain criticisms of textbooks, namely that they often do not reflect the way language is used in the real-world due to contrived examples for the sake of grammar instruction but at the expense of authentic and meaningful input (e.g., Timmis, 2014). It may be for these very reasons that the usefulness of learner corpus research and its application to teaching and materials development in general has been called into question. For example, Flowerdew (1998) stated that “...the implications for pedagogy are not developed in any great detail with the consequence that the findings have had little influence on... syllabus and materials design” (p. 550). Römer (2006) also commented on this issue saying that while the value of using corpora to inform materials development has “obvious and recognized strengths... it seems that there is still a strong resistance towards corpora from the side of students, teachers, and materials writers” (p. 124). Further, when insights from corpora are used in textbook development, it seems that there are still weaknesses in their implementation. Timmis (2013) highlights this in a chapter on developing materials citing Koprowski (2005) who noted that while textbook and materials designers are open to incorporating language chunks and multi-word units in their materials, their selection is often *not* informed by corpora and still largely relies on the developers’ own sense and intuition. Koprowski also notes that the diversity of multi-word units selected for materials (in the case of English language materials) often place excessive emphasis on simple collocations rather than, say, phrasal verbs or longer multi-word expressions which could be identified in L1 corpora. Römer (2006x) also comments on mismatches between language found in corpora and materials,

stating that: “For all items investigated, researchers found considerable mismatches between naturally-occurring English and the English that is put forward as a model in pedagogical descriptions” (p. 126).

Research and development of language textbooks using corpus-based methodologies can provide stakeholders with insights to assess the language that appears in textbooks and how well that language reflects language use in real life. In an aptly titled paper, *Corpora and Language Teaching: Just a Fling or Wedding Bells?* on the state of corpus-based research and the field of language teaching where she equates corpus-based research and language teaching as either in *a fling* or on the verge of great collaboration (*wedding bells*), Gabrielatos (2005) highlights how corpora can be leveraged to create meaningful outcomes, such as developing learning and teaching materials, and examining textbooks to identify (i) what language forms learners are exposed to and (ii) facilitate the research and development of textbooks and other materials or assessments. This is illustrated in the figure borrowed from Gabrielatos (2005), which shows the potential to leverage corpora in textbook development. To summarize the main points, L1 corpora can be used to identify real-world usage of language, and this can be compared with both L2 corpora and textbook corpora. This is especially critical as textbooks have been found to not be an accurate portrayal of L1 language use (e.g., Römer 2004, 2005), sometimes due in part to language as it dynamically shifts and changes over time.

Figure 1.1.

Corpora and ELT (copied from Gabrielatos' 2005 article *Corpora and Language Teaching: Just a Fling or Wedding Bells?*).



Such analyses can also uncover trends in changes in language and how they are (or are not) represented in learning materials. A recent study by Belli (2018) investigated stative verbs in the progressive aspect in English language textbooks. This is key for textbook developers as the progressive in English has been undergoing a shift in meaning to include stative readings. Traditionally, stative verbs in English such as *love*, *want*, *feel*, etc., “have been known [as] the verbs which cannot or rarely occur in the progressive form as evidenced in a number of previously written English textbooks” (Belli, 2008, p. 2018). In fact, some previous textbooks have even claimed that stative verbs are incompatible with the progressive (e.g., Anderwald, 2012 as cited in Belli, 2008). However, such extreme statements are challenged by the fact that stative verbs do appear in the progressive form (e.g., see Granath & Wherrity, 2014). In the case of Belli’s (2008) study, the textbooks under investigation were “corpus-informed,” meaning that they “were designed by authors who made use of various native English corpuses [corpora],

which reflected the target language as it is currently written and spoken” (p. 126). Corpora used in the design of the textbooks included the *Cambridge International Corpus* and the *Cambridge Learner Corpus*, *The Corpus of Contemporary American English* (COCA), among others. The aim of the study was to identify how stative verbs in the progressive aspect were incorporated into corpus-informed textbooks. An interesting finding in this study was that stative verbs which can be associated with the progressive in English, particularly verbs expression emotion (e.g., *want, love, feel*) were in fact included in the corpus-informed textbooks with the progressive form, in contrast to previous textbooks that include descriptions of stative verbs being ungrammatical with progressive forms. This is also in-line with studies such as Freund (2016) which show that “...certain statives attract progressive aspect in particular contexts, while others remain resistant to it...” (p. 59), and that certain verbs may have increased in their usage with a progressive (in colloquial British English), a nuanced insight which can inform textbook development. In the case of the corpus-informed textbooks, Belli notes that the usage cases for stative progressives in English, such as referring to a situation as dynamic, were in fact explained in the corpus-informed textbooks.

Another textbook analysis by Jung (2022) investigated Korean language textbooks and how postpositions (specifically, particles such as *-ey* or *-eyseo*) were incorporated into the textbooks. Traditionally, certain Korean postpositions are introduced in textbooks with functions that express either static place or dynamic location (e.g., Jeong, 2011; Kim, 2011). As Jung (2022) notes, “this dichotomy of postpositions that share similar functions in textbooks may confuse language learners when they are exposed to the natural language use environment, which is not consistent with what is presented in the textbooks.” (p. 202). In Jung’s own analysis, she investigated postpositions by addressing their frequency of occurrence, checking for commonly

co-occurring verbs, and keyness analysis. The corpora employed in that study were the Sejong corpus (written and spoken) as an L1 reference, and a corpus of two series of Korean textbooks. For the textbook corpus, data were compiled from 16 volumes in total and covered four proficiency levels (typical of what might be used in a four-year Korean program at a university). The analysis revealed that as the textbooks' level (beginner through advanced) increased so too did the number and variety of verbs co-occurring with postpositions. Jung was also able to uncover that certain verbs which commonly occurred with post-positions in L1 Korean were lacking representation in the textbooks. Also of note is that the "location, position, and existence" function when a postposition is used with the predicate *-iss*, to exist, was largely lacking in the textbook corpus. Jung notes that this is in line with previous work where learners exhibited lower accuracy with location functions and higher accuracy with direction functions of constructions with the postposition *-ey* (e.g., Kim & Guo, 2016). For the purposes of this dissertation, this recent example of a study of L1 and textbook Korean shows how a comparison of (a) L1 corpora and (b) language learning materials can provide (i) insights into how real-world and textbook language differ and (ii) highlight areas where language learning materials can be modified and improved.

2.2. Textbooks and corpora – the way forward and towards incorporating robust analyses

As discussed above, corpora can be a powerful tool when creating and designing textbooks and materials for language learning. However, that is not to say that teacher-researchers and materials developers need to put intuition to the wayside when designing textbooks or preparing lesson plans. I believe Timmis (2014) aptly put it when suggesting the term *corpus-referred* materials (as opposed to *corpus-based* or *corpus-informed*) may be the way forward. Timmis states that:

“A corpus-referred approach, I would argue, explicitly allows an honorable place for intuition, experience, local need, cultural appropriacy and pedagogic convenience in determining syllabus content and the order in which items are taught.” (Timmis, 2014, p. 470).

Taking a corpus-referred approach allows stakeholders to consider the data that can be gleaned from corpora while also taking note of what intuition tells us or what has been shown to work in the classroom. For example, in this view, there is not necessarily a need to present grammatical structures based on their frequency in a corpus, especially if such features could be difficult for learners or require some scaffolding of perceivably simple features in advance (e.g., Biber & Conrad, 2010).

Additionally, there is mounting evidence that input from textbooks can have a positive impact on language acquisition. A study by Northbrook and Conklin (2019) investigated whether students were able to lexical bundles appearing in their textbooks faster than others. They found that, in fact, the input learners received from learning materials including textbooks led to students being able to process the lexical bundles they encountered in their textbooks faster. And, as the students who participated were lower in proficiency, their study also provided evidence for the effectiveness of input from textbooks even for learners at lower levels. Relating to the present study, analyzing potential disparities between L1 and textbook representation can open the door to suggest revamping the representation of certain linguistic features in textbooks.

I also point out that the field of corpus linguistics has been rapidly evolving, and since the publication of some aforementioned studies, robust statistical techniques have started to take center stage in research conducted in the corpus domain. For example, while traditional corpus enquiries into language focused on evaluating frequency counts running keyness analyses, it is becoming more common to involve advanced statistical techniques to corpus data. I specifically

refer to the use of regression analysis in the field of corpus linguistics as put forward by Gries and Deshors (2014). While a complete review of the paper is beyond the scope of this dissertation, key points will be highlighted here. In their paper, Gries and Deshors outline how historically corpus studies on interlanguage between L1 and L2s have relied on raw frequency counts from often *comparable corpora*. However, there are clear weaknesses when only considering frequency counts void of context to account for over- and underuse of linguistic features by learners when compared with L1 speakers. Thus, rather than relying on frequency counts of certain features when L1 and L2 speakers, for example, write essays about the morality of smoking, we can consider the *context* as *linguistic/contextual features* (p. 114, emphasis added). To quote Gries and Deshors:

“...we should look at NSs’ choices of *can* versus *may* when the subject is animate, singular, when the clause is interrogative,... and then compare this to NNSs’ choices of *can* versus *may* when the subject is animate, singular, when the clause is interrogative. In this view, ‘comparable situation’ is now defined much more comprehensively in terms of linguistic/contextual features... give way to what we think should be one of the fundamental questions of SLA/FLA research: ‘in a situation, *S*, characterized by features F_{1-n} that the learner is now in, what would a native speaker do (and is that what the learner did do)?’” (pp. 113-141).

Using robust statistical methods (e.g., regression modelling, generalized linear mixed effects modeling, or collocation analysis) can help teachers and researchers reveal what makes an L2 speaker’s speech sound markedly different from an L1 speaker despite target grammar and vocabulary usage being largely correct. A nuanced approach is that over/underuse of features contribute to “foreign-soundingness even in the absence of downright errors” (Granger, 2004, p. 132).

As a construction which is known to be difficult for students to acquire, the progressive has been the focus of many studies. In-line with the robust statistical methods mentioned above, recently, several scholars have begun using robust corpus-based and corpus-driven methodologies and advanced statistical techniques, such as regression modeling, generalized linear mixed effects modeling, and collocation analysis (e.g., Römer, 2005; Kranich, 2010; Hundt & Vogel, 2011; Rautioaho, 2014; Deshors & Rautioaho, 2018; Fuchs & Werner, 2018; Rautioaho, 2020) to measure the attraction of words to certain syntactic constructions or to each other within a construction (e.g., Stefanowitsch & Gries, 2003). However, to date, much of the work done on the progressive (and indeed, in much of the field of corpus linguistics) has focused heavily on learner English and world Englishes (e.g., Römer, 2005; Rautioaho, 2014, 2020, among others). While this is no doubt due to the widespread usage of English as a world language and a lingua franca which lead to a natural need for corpus based educational materials and dictionaries (e.g., COBUILD), a gap in the literature exists when it comes to other languages.

To summarize, corpus linguists have a great opportunity to serve in the role of textbook and learning materials development, particularly when evaluating what types of forms and functions should be included in materials. To that end, robust corpus-based analyses can arm us with data on constructions, their semantic meanings, and usage cases (e.g., Gries et al. 2005) as usage data pulled from corpora offer clear insights into a construction (and lexical items that appear in it as well as semantic descriptions). As has been shown in the aforementioned studies, corpus-based investigations can (i) reveal gaps between real-world and textbook language, (ii) identify changes in language over time, and (iii) aid in the improvement of learning materials that better match natural language. In the next section, I will discuss corpus-based work on the

progressive. I will discuss research on the acquisition and usage of the progressive in learner language and highlight key aspects of the progressive such as its use with stative verbs in both English and Korean. I will touch on statistical methods used in corpus linguistics to address the progressive, which will lead into the present study's methodological considerations.

2.3. Theoretical underpinnings: usage-based approaches to second language acquisition

In usage-based viewpoints of language acquisition, language learning is driven by previous experiences with language, and these repeated exposures and experiences over time result in the cumulative frequency effects necessary for uptake of linguistic constructions. Put simply, language acquisition happens after repeated exposure as language learners subconsciously tally up co-occurrence rates of forms with functions, which over time become entrenched and automatized (e.g., Bybee, 2013; Tomasello, 2003). As frequency effects and exposure to linguistic features are key to the automatization of constructions in usage-based viewpoints, studying both how learners use linguistic constructions and the input they are exposed to can provide a workable framework for teachers and materials developers in their pedagogical practices.

Thus, to analyze the choice of the progressive construction or the non-progressive construction in both L1, L2, and textbook Korean, I take a usage-based approach and consider that frequency effects drive language acquisition for both L1 and L2 speakers. In other words, the more often a speaker encounters a certain word, construction, or collocation, the more entrenched that piece of language becomes as learners are sensitive to the exposure patterns developing probabilistic knowledge (e.g., Ellis, 2002; Ellis 2008), and these repeated exposures which lead to entrenchment of language form-function mappings are as important, if not more so, than conscious noticing and becoming aware of form-function mappings (Schmidt, 1990).

Thus, it is important to identify which verbs co-occur with the progressive in textbooks and how these co-occurrence patterns differ from real-world usage, which is the aim of identifying ways to improve the textbooks, which serve as a main source of input, for learners. That is to say, while L1 influence can certainly play a role in a learner's acquisition of a linguistic form, ensuring that their input (such as textbooks) matches real-world language as much as is realistically possible can propel learners to notice and acquire constructions in their L2.

In addition to frequency effects, a learner's L1 can also influence the uptake and acquisition of a linguistic feature. For example, existing literature on progressive and continuous aspect constructions shows evidence for variation between L1 and L2 usage, and that L2 usage may be influenced by interlanguage effects from the L1. In a corpus-based study of argumentative essays from the ICLE (International Corpus of Learner English), Virtanen (1996) found that differences in a learner's L1 lead learners to use the progressive construction in different amounts. A comparison of essay data from the L1 Finnish, L1 Finland-Swedish (a dialect of Swedish spoken in Finland), and L1 Swedish revealed statistically significant differences in the usage rates of the progressive. Virtanen noted that the rate at which a learner used a progressive in their writing differed depending on their L1. Notably, L1 Finnish learners used the progressive significantly less in their writing than the other two learner groups (L1 Finnish-Swedish and L1 Swedish). Virtanen attributed this difference in usage to L1 influence, stating that students' usage of the progressive "seems to vary according to their mother tongue background" (p. 301).

When considering frequency effects on the uptake of form-function mappings, it is possible to tease apart more detail than simply a verb's association with a particular construction. Within the domain of usage-based explorations on interlanguage, variationist approaches are

useful when analyzing what predictors, such as the lexical aspect or semantic domain, of the verb in question will gear a speaker towards the choice of one variant over another variant. For example, Deshors (2011) and later Gries and Deshors (2014) exhibited how the variation between the choice to use *may* or *can* in English interlanguage can be explained with several predictors such as *speaker* (e.g., L1/L2), form (*may/can*), and subject animacy, among others. The analysis showed that certain grammatical features (e.g., aspect and negation) can lead L2 speakers of English to use *may* and *can* in different ways.

For the progressive specifically, recent studies have taken usage-based and variationist approaches to explore patterns in a speaker's choice of the progressive and non-progressive (e.g., Deshors & Rautionaho, 2018; Fuchs & Werner, 2018; Hundt & Vogel, 2011; Kranich, 2010; Rautionaho, 2014; Rautionaho, 2020; Römer, 2005). In the case of English, corpus-based variationist explorations have shown that the progressive construction may most often be chosen with verbs in the present tense, verbs which are dynamic, and when the subject is animate, as revealed by multifactorial analyses by Hundt, Rautionaho, and Strobl, 2020, and Rautionaho et al. 2018. Specifically, Hundt and colleagues identified that tense, modality, verb type, and animacy of the subject were all important predictors in the choice of a progressive or non-progressive. More specifically, they revealed that dynamic verbs and the present tense were significant predictors of the choice to use the progressive aspect in the corpus data. Rautionaho and Hundt (2021) considered the context in which the progressive occurred in their data (the International Corpus of English) and found that in addition to durative situations calling for the progressive (e.g.: *consider*, *dance*, and *stay*, p. 616), having a progressive appear in the preceding context also lead to an increased usage of the progressive through syntactic priming. A preceding study conducted by Deshors and Rautionaho (2018) explored the variation between

the choice of the progressive and non-progressive construction based on semantic domain and lexical aspect category (aktionsart) of the verb (among other categories) and found that “more often than not, writers’ constructional choices are not influenced by a single linguistic factor... but rather by the combined influence of (or the interaction between) two factors...” (p. 238). Their multivariate analysis showed how semantic domain plays a role in the choice to use, or not use, a progressive construction, as it was the only annotated feature which did not have an interaction effect with other features. Thus, we can conclude that multivariate usage-based and variationist explorations of patterns in the progressive and non-progressive should include the semantic domain of the verb as it appears to be a significant factor regardless of variety of the speaker (e.g., L1/L2), lexical category of the verb, or genre.

In particular, when considering data from a usage-based perspective with a focus on variation between a choice of construction *a* or construction *b*, it is important to consider constructions which are as functionally and semantically similar as possible. That is to say, we can investigate alternation when the choice of either construction is possible. Rautionaho and Hundt (2021) also point this out and note that “what allows us to treat progressives as part of an alternation is that we carefully limit our dataset to instances where both variants are a potential choice” (p. 602). In short, to ensure that variation between the choice of progressive and non-progressive is represented, studies such as Rautionaho and Deshors (2018), Hundt et al. (2020), Rautionaho (2020), Rautionaho and Hundt (2021), extract set amounts of exemplars appearing in the progressive and the non-progressive. Post-extraction, exemplars are randomized and manually checked to be included or excluded based on certain criteria. By following strict criteria, this allows researchers to identify what factors can impact the choice of a progressive or non-progressive in a variationist approach. Namely, all verbs in the aforementioned studies only

included verbs which could appear in both the progressive and the non-progressive. To borrow Hundt et al.'s (2020) example, it is a difference that can be seen in sentence variations like *he was driving along the road* and *he drove along the road* (p. 82). As Rautionaho & Deshors (2018) put it: "To strengthen our analysis, we further limited the data to only include such lexical verbs that both occur with progressive and non-progressive constructions in our data set" (p. 232). Thus, through careful extraction and selective data cleaning, a variationist approach can be used to assess what linguistic and contextual factors can influence the choice of a progressive or non-progressive construction in different speaker varieties.

Bringing this discussion back to influence of frequency effects from the input from a usage-based perspective, this dissertation also incorporates textbook data to determine potential effects of input (with a focus on the verbs appearing with *-ko iss*) on the usage of the Korean progressive in learner language, and, to identify potential gaps in the textbook language which may need to be addressed considering the verbs appearing in the progressive in L1 data. This is of particular note as input and the frequency at which input occurs plays a major role in the uptake of form-function associations. For learners, textbooks are one of the major sources of input (e.g.: Römer, 2004) and so textbook language merits investigation as it is one source of input that can be controlled, to some extent, to provide learners with useful input for language learning. Of course, textbooks can never be considered as holistic or exhaustive representations of language for learners, but working towards a more comprehensive representation of language in textbooks as it is used in modern Korean is one goal of this study.

2.4. The *-ko iss* construction

In this section, I outline the progressive *-ko iss* construction in Korean which can express a continuous or progressive meaning. Functionally, *-Ko iss* is comparable to the *be... ing*

progressive construction in English. Key typological differences between Korean and the target L2s, English and Japanese, are discussed in terms of potential interlanguage transfer effects as well. This will set the stage for discussing the selection of corpora and the statistical methods for exploring the use of the continuous and progressive constructions in L1 and L2 varieties of Korean. As the main focus of this dissertation is *-ko iss*, that construction is discussed most in-depth.

An in-depth account of Korean grammar is Yeon and Brown's (2011) book *Korean: A Comprehensive Grammar*. Thus, I use descriptions provided by Yeon and Brown for clarity and consistency. Furthermore, their descriptions of Korean are of "the standard Seoul speech in the Central dialectal zone" (p. 1) which is most often the target for second language learners of Korean.

To form the progressive, *-ko*, a suffix, is attached to the base form of a verb, and then *iss* is added after *-ko*. In writing, there is a space between *-ko* and *iss*. While *-ko* does not change form, verb endings (to denote past/present/future tense) honorifics, or conjunctions may be added to *-iss*. This construction is most similar to the English *be... ing* or the Japanese *-te iru* for denoting an action in progress as can be seen in example (1):

(1) 미나가 지금 저녁 식사를 준비하고 있다.

Mina-ka jigeum jeonyeog sigsa-reul junbiha-ko iss-ta.

Mina-NOM now dinner-ACC prepare-PROG-DECL

Mina is preparing dinner now.

However, while the *-ko iss* form seems similar to the English progressive in example (1), there are a few key factors which set it apart from the English progressive. Perhaps the most surprising difference between the English and Korean progressive forms is that, unlike in English, the

Korean progressive is “usually optional and used only for emphasis” (Yeon & Brown, 2011, p. 214) and “unlike the English progressive or the Japanese *-te i...* the Korean *-ko iss-* is not obligatory for an ongoing event interpretation (Lee & Kim, 2007, p. 656). A common example of this is when someone asks what you are doing, and in Korean, a pragmatically appropriate response could take the simple present tense, whereas in English, the progressive is preferred, such as in the conversational example (2):

(2) A: 지금 뭐 해?

Jigeum mwo hae?

Now what do-PRES.

What are you doing now?

B: 지금 공부해.

Now study-PRES.

Now, I am studying.

Another key difference is that the progressive in Korean cannot usually hold a futurate meaning. For example, while in English the progressive can be used in the futurate to denote an action you are about to do. For example, an English speaker can say *I am going now* right before they depart their location. The equivalent in Korean, in example (3), can only be used to describe the action in progress. In other words, one must have already departed their location in order to use the progressive *-ko iss* alongside the verb *go*:

(3) 지금 가고 있어요.

Jigeum ga-ko iss-eo-yo.

Now go-PROG-PRES.

I am going now.

The Korean progressive construction in question for this study, *-ko iss*, is interesting because it differs in its usage from other languages like English, and even more typologically similar languages such as Japanese. For example, the Korean progressive can often be used with stative and mental verbs which do not often take the progressive in other languages such as English (some verbs that fall into this category include *believe, desire, feel, have, know, realize, and remember*). In particular, a common verb co-occurring with *-ko iss* is *al* (*know*), as can be seen in example (4):

(4) 원경 누나가 이제 그 사실을 알고 있다.

Wonkyung nuna-NOM ijae geu sasil-eul al-ko iss-ta.

Older sister Wonkyung-NOM now that fact-ACC know-PROG-PRES-DECL.

Older sister Wonkyung knows that fact now.

In terms of what may influence the choice of a progressive or non-progressive with such verbs, Lee (2006) says that such verbs “belong to a class of inchoative eventualities which describes an instantaneous inception event that starts a continuous state” (p. 697). Thus, a choice to use the Korean progressive *-ko iss* construction with a verb such as *know* can be because progressive aspect in Korean is used with not only actions in progress, but also to states that come about due to some event, such as coming to know new information, which would then lead a Korean speaker to choose to use the progressive *-ko iss* construction with the verb *know*. So, another possible translation of (4) above, depending on the context, can include *older sister Wonkyung is now aware of that fact*.

In the case of psychological or cognitive verbs, such as *believe* or *know*, one reason Lee points out that may allow the verbs to take a progressive marking is that psychological verbs in Korean “...do not have to occur every moment afterwards to maintain their effect” (p. 715). In

other words, once one becomes aware of some fact or situation, for example, the choice to use *-ko iss* with *know* is appropriate as they have entered a continuous state being aware of the fact from that point onwards. Given this, it is noted by not only Lee (2006) but other scholars as well that the categorization of verbs co-occurring with *-ko iss* is still up for debate. While in this dissertation I explore verbs, such as *know*, as stative and mental verbs, some scholars categorize verbs such as *know* as accomplishments (e.g., Hong, 1991) or resultative achievements (Ahn, 1995) since they come about as inchoative events. This issue, to my knowledge, is yet to be settled as how best to classify such verbs in Korean. So, I do my best to account for such verbs by considering them as stative verbs and mental verbs (as appropriate) but acknowledge a future study could account for other categorizations of stative or mental verbs in Korean.

Continuing with the discussion of *-ko iss* and typological differences between English and Japanese, the Korean progressive construction can be used with imperatives commonly, which is a unique feature of the progressive in Korean. While the progressive construction in English can be used to give a command or instruction in some situations, this usage may not be as common as it is in Korean. For example, as an example of a progressive used with an imperative in English, one might say: *you are not to be driving late at night*. However, the semantic difference is that English example indicates an action the speaker intends for the listener to do in the future (such as instructing someone to avoid driving late at night going forward). In Korean, however, you may use the progressive alongside the verb *wait* to express to your listener you want them to stay at the place they are currently at and to instruct they wait for you at that location, and this statement has the same nuance as *wait here*, directing the listener to do the action in the present moment, not in the future. A similar meaning in English could be expressed using *keep* or *stay*. In real-world usage, imperatives or commands with *-ko iss* in

Korean are generally used in plain or intimate speech styles rather than formal or honorific speech styles. Take note of example (4):

(5) 여기서 기다리고 있어.

Yeogiseo gidari-ko iss-eo.

Here wait-PROG-PRES.

Wait here (*stay* waiting here).

Finally, and most importantly for the analysis to follow, the Korean simple present and the present progressive *-ko iss* can often be used interchangeably (e.g., *I eat my lunch* and *I am eating my lunch* are both possible to express an action in progress in Korean), and thus there is a need to understand how the Korean progressive is used across L1, L2, and textbook Korean.

I have provided an overview of the *-ko iss* construction in Korean. In this study, my analysis will focus on *-ko iss*'s prototypical usage, that is, when it appears sentence final without any other connectors or tenses attached. Future studies can explore *-ko iss* in future/past tenses, as well as in conjunctions or negations.

2.5. Research on continuous aspect constructions in Korean

There have been several notable studies on continuous aspect constructions in Korean. In this section, I highlight some studies on L1 and L2 Korean. As is the case with many learners, regardless of the L2, acquiring the progressive construction can be tricky due to differences in semantic usage as well as form and function mappings; Korean is no exception.

Crosslinguistic variation in the usage of the Korean aspect construction has been explored by several scholars through various lenses. One study by Lee and Kim (2007) was an empirical approach to the *-ko iss* action in progress and *a/eo iss* continuous state construction, analyzed through the lens of the Aspect Hypothesis. When studying constructions related to temporality,

for example, imperfective aspect or continuous aspect constructions, one lens linguists have relied on is the Aspect Hypothesis (see Andersen, 1990, 1991; Andersen & Shirai, 1994; Bardovi Harlig & Comajoan-Colomé, 2020). To briefly summarize, the Aspect Hypothesis attempts to describe the acquisition order and usage of aspect constructions in L1 and L2 language. While results have been largely mixed, generally, we observe that in languages with progressive aspect, progressive marking is used first with activity verbs (e.g., verbs expressing activities that happen over a period of time, but where the endpoint is arbitrary as *run* in *they ran around the park*), and later with accomplishment (verbs where the action has a duration and a definitive endpoint, such as *run* in *run a mile*) and achievement (verbs expressing an event which takes place in an instant or a moment, such as *recognize*, *die*, or *reach* in the context of *reach the top*) verbs (Andersen & Shirai, 1996), based on Vendler's (1957) four-way classification of a verb's inherent lexical aspect.

Lee and Kim (2007) thus explored the acquisition of the progressive continuous (imperfect aspect) constructions *-ko iss* (action in progress) and *-a/eo iss* (continuous state) using cross-sectional data from over 100 learners of Korean. Data were collected through sentence interpretation and guided picture description tasks. The results from their findings confirmed, largely as expected, that among the continuous aspect constructions the action in progress *-ko iss* develops before resultative *-ko iss* and *-a/eo iss* constructions, which is in-line with the Aspect Hypothesis. Further, learners exhibited less frequent usage of the continuous state *-a/eo iss* constructions than the progressive *-ko iss*. Given typological differences between Korean and English such results may be expected. Further analysis using corpora can help glean which verbs and verb-types may be more or less associated with each aspect construction in both L1 and L2

Korean. As *-ko iss* has a much higher rate of usage in learner language, it was selected as the main focus for extraction from the corpus data in the present dissertation.

As mentioned previously, a major source of input for learners comes from the textbooks and materials used in class. There have been a few studies to date on the progressive in Korean in textbooks, and so far the general trend appears to be that continuous aspect constructions are not appearing in the abundance or variety that they have the potential for in L1 speech. For example, the resultative use of the *-ko iss* construction is not featured in all textbooks (Brown & Yeon, 2010), but when it was, it was often used without explanation and in the context of *wear* verbs regarding clothing, as the *-ko iss* construction can be used to express both (i) the act of putting on clothes and also (ii) the act of currently wearing the clothes. In some literature, this “resultative” meaning of *-ko iss* is discussed as a separate construction (e.g., Chae, 2018). In the present study, I am exploring the form-function of *-ko iss* but will only address this distinction if distinctive collexemes including verbs with resultative (such as *to wear* verbs) meanings are identified.

From a usage-based perspective, including *wear* verbs with the *-ko iss* construction is important as it is one common use and the form significantly differs from English. However, it leaves teachers wanting when other instances of the resultative meaning are not included. In that vein, a study by Jang (2005) tallied the number of textbooks which included a discussion of the resultative *-ko iss* and found that only a fraction of the textbooks introduced the various meanings and semantic uses of *-ko iss*, and most books only touched on the standard progressive form across textbooks prepared for general learners of Korean and learners with specific L1s (e.g., Japanese). Kim (2014) carried out a comprehensive study on the change of the usage of *-ko iss* historically and diachronically, as well as using the spoken section of the Sejong Corpus to

identify frequency of the progressive in the corpus based on the Vendlerian (1957) categories (accomplishment, activity, achievement, state). Through the analysis of the distributions of the types of *-ko iss*, they found that, as perhaps expected given the Korean *-ko iss* construction's wide variety of usage cases (e.g., action in progress, iterative progressive, narrative present, stative progressive, resultative, habitual, etc.), that the distribution between *-ko iss* functioning semantically as a prototypical progressive (e.g., action in progressive) or otherwise (such as stative or resultative meanings) were quite similar (around 40% and 45%, respectively) (p. 48). Thus, Kim states that due to the fact that the progressive *-ko iss* construction actually conveys not only progressive meanings, but also stative and resultative meanings, that the construction itself may need to be reassessed as just expressing "the general imperfective, encompassing the habitual and the non-Progressive use(es)..." (p. 49).

I agree with Kim's assessment of the Korean *-ko iss* construction and argue that due to its complex semantics, a robust analysis of what factors predict the choice of a progressive is necessary for not only describing the Korean language but also informing the development of textbooks. As mentioned, the way the progressive is introduced in textbooks is often lacking or incomplete, with some texts including only the purely progressive usage, others including some variety but without clear explanation as the function of the construction with various semantic meanings. In this study, I hope to build on the existing literature by marrying the findings from studies on the Korean imperfective and continuous aspect constructions outlined above and build on them using approaches that are becoming more common in the field of corpus linguistics. In the methodology section below, I outline the choice of corpora, predictor variables, statistical tools, and collostructional analysis.

III. METHOD

In this section I outline the methodology for this study. First, details about the L1 and L2 corpora are provided, and a description of the textbook corpora compiled for the study is also provided. Then, the data extraction methods used to identify progressive and non-progressive forms for analyzing variation patterns within the corpora is discussed. The factors and levels for annotation of the data are also discussed in detail, for example, aspect (progressive versus non-progressive), variety (L1 Korean, L2 Korean, Textbook language), semantic domain (based on Biber et al., 1999, Biber et al. 2021) and Aktionsart (following Deshors & Rautionaho, 2018; Rautionaho, 2020; who borrowed Vendler's 1957 model of Aktionsart). Finally, data analysis methods are discussed, including collostructional analysis (distinctive collexeme analysis), and regression modeling.

3.1 Corpora description

3.1.1 Sejong corpus

The Sejong Corpus was a corpus of L1 Korean written and spoken language made publicly available by the National Institute for Korean Language in South Korea¹. The Sejong Corpus provides L1 Korean language corpora in both spoken and written formats (and has recently expanded to include other modalities such as a *text message corpus*). As outlined by Lee (2022) in the *Routledge Handbook of Korean as a Second Language*, the development of the corpus was funded by the Korean government over the span of roughly ten years (1998 to 2007). The total size of the corpus is about 200,320,000 *ejeols*². I use the Written section of the corpus in this study, which in total is about 36,879,143 *ejeols*. The contents of this include news articles, books

¹ I used part of the Sejong Corpus for this study, however, this corpus is no longer available due to copyright issues. An updated corpus is now available, titled *Modu-eui Malmoongchi* (Korean: 모두의 말뭉치; English: Everyone's Corpus) and is available online at the following URL: <https://kli.korean.go.kr/>

² An *ejeol* is a word and any grammatical suffixes attached in Korean.

and novels, and essays on a wide range of topics. However, due to the size of the raw corpus and needing to manually convert UTF-16 files to UTF-8 to make the files readable by R and AntConc, I randomly selected 100 files from this written corpus for analysis in this dissertation. Data were annotated using UDPipe (Straka et al., 2017) on a personal computer to ensure that the POS-tagger applied to all data is the same. The collection of the 100 files is coined as the KOR100. The KOR100 is 4,784,997 *ejeols* in size.

3.1.2 Learner corpus

The National Institute of Korean Language Korean Learner Corpus was selected as the learner corpus for the present study. The corpus is compiled by the National Institute of Korean Language (NIKL) and is freely available to download from the NIKL website (<https://kcorpus.korean.go.kr>). The Korean learner corpus is a large corpus which also includes error annotation. The corpus includes data from learners from over 100 countries and 90 different L1 backgrounds (Lee, 2022) and is roughly 3.78 million *ejeol* in size. Data samples were provided by learners at university-level language institutions in Korea, Korean immigrant educational institutions, as well as universities and King Sejong institutes outside of Korea over from 2015 to 2021. The data were collected through collaboration at these language learning institutions, and an Excel file provided by the NIKL gives an overview of sample topics learners were prompted with when writing their essays. Students were tasked with writing about various prompts, and topics ranged from writing about one's daily schedule to describing wedding customs, writing about their future in 10 years, and the need to install CCTV cameras in daycares, among others (full list available from the NIKL 2015~2021 Learner Corpus Sampling Information Spreadsheet). The corpus consists of both spoken and written data, though only the written subsections for the L1 English L2 Korean and L1 Japanese L2 Korean groups are

included for analysis here. Version 4.1 (released in 2021) of the learner corpus was analyzed for this dissertation. In total, 1,639 essays (184,181 *ejeols*) written by L1 speakers of English were included for this analysis, and 4,090 essays (495,391 *ejeols*) written by L1 speakers of Japanese were included for this analysis.

3.1.3 Textbook corpora

To examine the nature of language that Korean language learners are exposed to, two series of textbooks were selected for corpus compilation. The two textbook series selected for the present study are *New Sogang Korean* published by Sogang University Press, the recently updated edition of *KLEAR Integrated Korean* published by University of Hawaii Press. Each textbook was selected as they are currently used in Korean language programs in South Korea and the North American higher education context and an analysis of the texts within each set of learning materials can provide a view of the input learners are exposed to in Korean classes.

Table 3.1 shows the number of tokens present in the textbook data stratified by level. Token number was identified using AntConc (Anthony, 2023). Files were manually and semi-automatically converted into machine readable formats. Files were then converted to UTF-8 to allow for token counting and extraction. The textbook analysis is a frequency analysis using raw and relative frequencies to quantify usage of lemmas across textbook volumes.

Table 3.1

Summary of data in New Sogang Series and KLEAR Integrated Korean textbook series (number of tokens)

	Level 1	Level 2	Level 3	Level 4	Total
New Sogang Korean	5802	8730	13679	25519	53730
KLEAR Integrated Korean	7386	13704	18836	24298	64224

3.2. Part of speech tagging and extraction

The data used in this study come from two main sources: The National Institute of Korean Language (NIKL) corpora, and a textbook corpus. To identify and extract all instances of the target constructions the data was Part-of-Speech (POS) tagged. Tagging of corpus data is a key step as it allows for the extraction of both the progressive and non-progressive forms of target verbs. While some parts of the L1 Corpus include POS annotation, running the raw data through POS-tagging myself has some advantages, namely that all the corpora can be POS-tagged using the same POS-tagging models. Following Jung (2022), I used *UDPipe* (Straka et al., 2016), a package available for R which allows for POS-tagging, tokenization, lemmatization, among other Natural Language Processing tasks. However, the main justification for the use of *UDPipe* at data cleaning, preparation, and extraction steps is that it includes pre-trained models of Korean that can be used during the POS-tagging and annotation process. I used the function ‘`udpipe_annotate`’ and called for the pre-trained model for Korean (korean-gsd-ud-2.5-191206.udpipe) to tag the data.

After tagging the data in R using *UDPipe*, I used the freeware tool AntConc version 4.2.1 (Anthony, 2023) to extract progressives and non-progressives. This was a multi-step process. First, I extracted the examples of the progressive *-ko iss* and compiled a list of lemmas appearing

in the progressive (for this study, I focus on present progressive for the collostructional analysis and regression modeling, so those were extracted). Then, following previous corpus studies on the progressive, I extracted corresponding non-progressive examples in the simple present based on the list of lemmas that co-occur with *-ko iss*. As Korean is agglutinative, regular expressions had to be written to be able to call for verbs which had a grammatical morpheme both attached and unattached. As an example, the verb *to party* in Korean is *patihada* (파티하다). *Pati* corresponds to *party*, and *hada* corresponds to the English equivalent of *do* (so the singular verb means *to party*). However, sometimes, the choice to include a grammatical morpheme on the lexical part of the verb (here, *pati*) is also possible, resulting in the form *patireul hada*, with the accusative case marker *reul* 를 attached directly to the lexical word *pati*, and causing a space between the two parts of the verb (파티하다 → 파티를 하다). This is a unique feature of *hada* verbs (verbs which include *hada* 하다), and so, to ensure accurate extraction, two versions of the regular expressions were submitted to AntConc to call for the target verb forms in the POS-tagged data.

To illustrate data extraction, I will borrow an example from Rautionaho (2020) who extracted key verbs appearing in stative progressive constructions. Take the verb *want* as an example. Rautionaho illustrates how a regular express used in POS tagged data can extract first the forms in the non-progressive using the following expression first:

`\bwant\S*(VBD|VB|VBZ|VBP|VBN)\b` (p. 188), as the expression includes tags for verb forms such as present or past tense. The second step, then, is to swap the tags for the present participle form (VBG) which aids in identifying instances of the progressive. These extractions can then be organized in a spreadsheet for further cleaning and annotation. Of note is that Rautionaho

employs a two-step process in her extraction method as it allows for keeping data organized (namely, keeping the target constructions and the rest of the data separate from each other). In the present study, I also employ a two-step extraction process for each target construction. For stative progressives in Korean, I refer to Yeon and Brown (2011) who identify certain verbs in Korean which appear with the progressive. Namely, those verbs are: *know*, *not know*, *love*, *believe*, *want*, *remember*, and *feel* (p. 215). Instances of stative progressives in L1, L2, textbook corpora will thus be quantitatively and qualitatively explored. It is possible that stative progressives not listed above may appear in the data.

Examples of the regular expressions used include:

1. \b 받 VV EF\b (to call for non-*hada* verbs; here, the verb stem is 받 *bad*, and it can be swapped out for another form or left blank to call for all verbs in the written simple present form).
2. \b 생각 NNG |한다| XSF EF\b (to call for intact *hada* verbs; here, 생각 *to think* is used as an example)
3. \b 생각 NNG JKO |한다| VV EF\b (to call for *hada* verbs with accusative case marking attached to the lexical part of the verb; the tag JKO calls for accusative case marking)
4. \b NNG JKS 된 VV EF\b (to call for the verb 되다 *to become* in the data; in the simple present, when this verb means *to become*, it corresponds to the nominative case marker which the addition of JKS calls for).

3.3. Annotation of explanatory variables

Annotation and coding of explanatory variables (predictors) are discussed in this section. An overview of all explanatory variables is listed in Table 3.2. The explanatory variables in this

study include: aktionsart category, animacy of the subject, aspect (progressive or non-progressive), semantic domain of the verb, and variety (L1/L2 Korean).

Table 3.2.

Proposed predictors to annotate for the logistic regression

Predictor	Levels (adapted from Rautionaho et al. (2018), Rautionaho (2020).
Aktionsart (Vendler, 1957)	Accomplishment, achievement, activity, stative (e.g., Vendler, 1957)
Animacy	Animate, human, inanimate
Aspect (dependent factor)	Progressive <i>-ko iss</i> , Non-progressive
Semantic domain (Biber et al., 1999)	Activity, aspectual, causative, communication, existence, mental, occurrence
Variety	(i) L1 (NIKL Sejong), (ii) L2 (NIKL Learner Corpus: L1 English and L1 Japanese subsections), (iii) Textbook

Aktionsart categories are based on Vendler’s (1957) classification of verbs is a “four-way classification of the inherent semantics of verbs” (Andersen & Shirai, 1996, pp. 531-532) based on a verb’s inherent lexical aspect. which includes four distinct semantic types: states, activities, accomplishments, and achievements. These categories are determined by three elements, namely, a verb’s dynamism, durativity, and telicity and have been used for lexical verb classifications in numerous studies (e.g., Rautionaho & Deshors, 2020; Salaberry & Shirai, 2000). It is important to note that aktionsart annotation must consider the context in which the verb occurs. For example, *run* can be either an activity (e.g., *she is running in the park*) or an achievement (e.g., *he ran a mile*). That is to highlight that the context in which the verb occurs, and the semantics of the verb phrases, must be considered.

The aktionsart classification, then, falls to the verb's inherent lexical aspect and temporality, the duration the action the verb describes takes place, and the endpoint. Telicity refers to whether a verb has an endpoint: telic verbs have endpoints and fall into the achievement or accomplishment aktionsart categories. Atelic verbs do not have a clear endpoint and are thus categorized as activity verbs with aktionsart categories. Breaking telic verbs down further, whether they are achievements or accomplishments depends on the durativity of the verb, where punctual verbs with abrupt endpoints are classified as achievements, and verbs where the endpoint takes some time to culminate are categorized as accomplishments (accomplishment verbs are also often said to be verbs with "goals"). States are those verbs which are durative and describe the state of something, such as *to know*. Aktionsart has been useful in corpus-based studies on the progressive in interlanguage and World Englishes. Generally, it has been found that the progressive is predicted (or "triggered") by verbs whose subjects are animate and where the lexical category is activity (e.g., Biber et al. 1999). However, as the Korean language allows for statives to occur in the progressive there may be some variation between L1, L2, and textbook language that is worth investigating with aktionsart categories as a predictor. Table 3.3 provides a list of the aktionsart categories with example sentences based on Andersen and Shirai (1996). Of note is how depending on the context, a verb can be categorized in different aktionsart categories, such as *run*. This demonstrates the importance of considering the entire verb phrase, not just the verb itself, when annotating for aktionsart.

Table 3.3.*Aktionsart categories with examples*

Aktionsart category	Example
Accomplishment: telic, time span of action has a clear terminal endpoint.	Verb: Read. Sentence: I read the magazine in an hour. Verb: Run. Sentence: Brittany ran a mile.
Achievement: telic, endpoint is punctual, and the event takes place instantaneously at a single point in time.	Verb: Recognize. Sentence: I suddenly recognized his voice on the phone. Verb: Die. Sentence: She died in her home last Tuesday.
Activity: atelic, duration of a period without a terminal endpoint, or an endpoint which is arbitrary.	Verb: Run. Sentence: He is running in the park. Verb: Play. Sentence: Boram is playing with her doll.
State: durative, describe a state. Verbs lacking a habitual reading in simple present are states.	Verb: Love. Sentence: Romeo loves Juliet. Verb: Want. Sentence: Serena wants to go back to college.

The predictor *animacy* refers to the verb's main subject and whether it is alive and sentient, though in linguistic research animacy falls across a spectrum rather than being binary animate/inanimate. The progressive construction was first explored in terms of animacy of the subject by Strang (1982). Strang coded animacy across a continuum, first with "subjects [that] which are human or otherwise viewed as capable of activity," (p. 443): (a) human, (b) quasi-human and/or animal, and finally (c) inanimate subjects. Other studies have included more factors within the animacy category, such as Zaenen et al. (2004) who discussed annotating for subject animacy distinctions including "collectives of humans when displaying some degree of group identity," computers as "intelligent machines," and even vehicles (pp. 3-5). For the purposes of the present study, animacy will be coding as human, animate, or inanimate.

The predictor *semantic domain* pertains to the semantic meaning of a verb in context. First discussed in 1999, and again, in 2021 by Biber, Johansson, Leech, Conrad, and Finegan, the seven-level classification of verbs is based on a verb’s core meanings, or “the meaning that speakers tend to think of first” (p. 359). It is important to consider a verb not only in isolation but in the context in which it appears. For example, a verb such as *get* in English can mean *obtain*, but it can also mean *become* (consider *I got the money from him yesterday* versus *I got so scared when I thought he didn’t have the money*). Thus, when annotating for semantic domain, verbs are considered in the context of the sentence in which they appear.

Table 3.4.

Semantic domains and descriptions based on Biber et al., 1999 and 2021

Semantic domain category	Descriptions
Activity verbs	Denote actions/events associated with a choice. Buy, carry, go, leave, run, work... Transitive/intransitive
Communication verbs	Special subcategory of activity verbs that involve speaking and writing. Ask, announce, call, discuss, explain, say, shout, speak, suggest, yell, tell, write...
Mental verbs	Denote activities and states experienced by humans, but do not involve physical action (and not always volition). Subject is usually the recipient. Cognitive and emotional meanings included. Think, know, love, want, see, taste, read, hear
Causative verbs	Indicate that the person or inanimate object brings about a new state of affairs. Allow, cause, enable, force, help, let, require, permit

Table 3.4 (cont'd).

Occurrence verbs	Also called <i>verbs of simple occurrence</i> in Biber et al. (1999, 2021). Report events that occur apart from any volitional activity. Become, change, happen, develop, grow, increase, occur
Existence verbs	Also called as existence and relationship verbs. Existence: Be, seem, appear Relationship: contain, include, involve, represent
Aspectual verbs	Characters the stage something is at, or the progress of an event or activity Kept, stopped, started, began, continue

3.4. Collostructional analysis

Collostructional analysis, broadly, is a family of statistical methods which allow for the measurement of the degrees of attraction between words and grammatical constructions (see: Gries & Stefanowitsch, 2003; Stefanowitsch & Gries, 2004). The name comes from the combination of *construction* and *collocation* (Gries & Stefanowitsch, 2003, p. 100) as the aim of the method is to assess collocation patterns between words and constructions (distinctive collexeme analysis), or between words within constructions (co-varying collexeme analysis). This is useful when assessing a verb which may appear in multiple constructions with a similar meaning as “[the verb] may ‘alternate’ between two constructions if (or to the degree that) the verb’s meaning is compatible with the meanings of both constructions” (Gries & Stefanowitsch, 2004). This is true in the case of Korean where oftentimes the simple present can be used interchangeably with the progressive *-ko iss* construction in many cases. Verbs which exhibit a preference for a construction based on their calculated association strengths are referred to as distinctive collexemes of that construction.

Such assessments of variation have been undertaken in corpus studies on L1/L2 English. In order to assess the progressive versus non-progressive alternation, for example, Rautioaho (2020) employed distinctive collexeme analysis (DCA). Specifically, she targeted co-occurrence

patterns of stative verbs in the two grammatical constructions (progressive versus non-progressive) to assess which stative verbs are attracted to the progressive construction in different varieties of English. To assess the collostructional strength of a word to constructions using DCA, the absolute frequencies (of words in the construction) are assessed alongside the observed and expected frequencies in each construction (Hilpert, 2006). To run a DCA, Gries (2022) provides an R Script, *Coll.Analysis 4.0*, which allows the analyst to submit the data tables including the words extracted from each construction. It is important to make sure that prior to this the data has been adequately cleaned so that all target words are lemmatized in the same way to accurately account for their frequencies in each construction and that all exemplars are included (for raw frequency counts). Once the data tables are loaded into R using the script, the data undergoes a Fisher-Yates test, which provides the analyst with collostructional strength scores. Higher collostructional strength scores correspond to stronger associations between the words and the constructions, and likewise suggest higher entrenchment of the syntax-lexis links between said words and constructions in the speaker's mind. In this way, as collostructional analysis is able to account for more than just raw frequencies:

“...it [collostructional analysis] identifies not only the expressions which are frequent in particular constructions' slots; rather, it computes the degrees of association between the collexeme and the collostruction, determining what psychological research has become known as one of the strongest determinants of prototype formation, namely the cue validity of, in this case, a particular collexeme for a particular construction.” (Stefanowitsch & Gries, 2003, p. 237).

In short, this means that collostructional analysis when used with L1 and L2 corpora, allows language researchers to assess which combinations of words and constructions are “highly

characteristic” (p. 237), and thus can aid in the development of teaching materials and lesson planning for language teachers.

Finally, when creating tables to summarize the results of the collostructional analysis and provide English definitions for all distinctive collexemes, English definitions were checked by referring to the Naver Korean Dictionary (available online at <https://dict.naver.com/>) and by manually inspecting the data to ensure polysemous words were separated (for example, multiple entries of the verb form *sseu* are in the table due to the verb’s polysemous nature).

3.4. Logistic Regression

When exploring a dependent variable with two outcomes, corpus linguists can employ (binary) logistic regression modeling to explore what explanatory variables may influence the choice of construction A or construction B. In this case, I follow statistical design from previous studies which use logistic regression to explore what factors may influence the choice of a learner or a first language speaker of a language to use the progressive or the non-progressive in their writing. I follow guidelines for planning, preparing, and interpreting the model as they are presented in Brezina’s (2018) *Statistics in Corpus Linguistics: A Practical Guide*. The binary dependent variable is the choice of a progressive or a non-progressive, and the explanatory variables include animacy, aktionsart, semantic domain, and variety.

This study addresses the following research questions:

Research questions

1. What are the distinctive collexemes of the progressive and non-progressive in L1 and L2 Korean?

2. Do any explanatory variables related to the verb (aktionsart, semantic domain, animacy) or speaker (variety) predict the use of the progressive *-ko iss* construction in L1 and L2 Korean?
3. What verbs are most commonly used and taught in the progressive in Korean language textbooks?

IV. RESULTS

4.1. Statistical approach

I closely follow Rautionaho et al. (2018), Deshors and Rautionaho (2018), Rautionaho (2020), and Jung (2022) when road mapping the statistical design for this study. Rautionaho and collaborators' studies serve as a mentor text for assessing the (non)progressive alternation between L1/L2 varieties as they employ both collostructional analysis (in particular, distinctive collexeme analysis) in tandem with robust statistical methods such as regression modeling.

4.2. Research question 1

Addressing research question 1, I discuss the results of the distinctive collexeme analysis for L1 and L2 Korean data. Each group (L1 written Korean, L1 English L2 written Korean, and L1 Japanese L2 written Korean) is discussed in its corresponding section. After discussing each group separately, comparisons between L1 and L2 results are made as appropriate. The tables with the list of distinctive collexemes (verbs that exhibited a preference for either the progressive or the non-progressive construction) are included in the appendix at the end of this dissertation due to their length. The prose in the sections below describe key highlights from the collostructional analysis. In particular, I go into detail stating (i) which verbs had an attraction or preference for the progressive and the non-progressive, and (ii) explore the semantic domains the verbs were categorized in upon qualitative analysis of the distinctive collexemes. I then (iii) provide key examples from the corpora to illustrate how verbs are used in each construction, placing emphasis on stative and mental verbs such as *al* (*know*) in particular for their potential usefulness in Korean language teaching and materials development.

4.2.1. Analysis of L1 Corpus: Distinctive Collexemes for the (non)progressive in L1 Korean Written Data

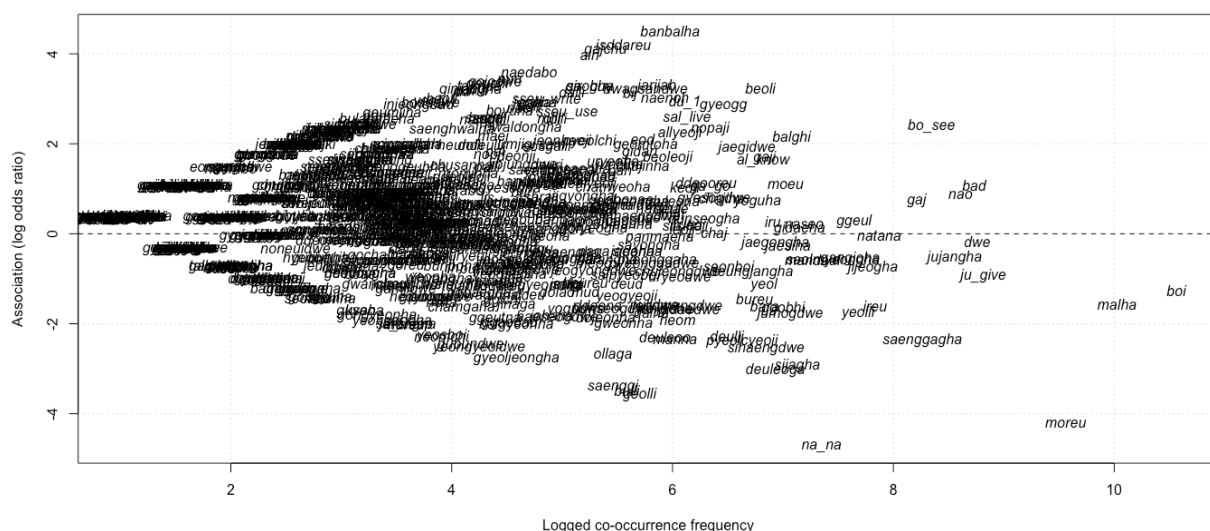
Table A-1 in the appendix shows the distinctive collexemes for the progressive on the left and the non-progressive on the right. In total, 256 distinctive collexemes were identified for the progressive, and 131 distinctive collexemes were identified for the non-progressive. The ranking is calculated by comparing a lemma's observed frequency with its expected frequency in each construction, as well as the total number of lemmas in each construction. A collostructional strength of 1.3 or greater is considered significant (Hilpert, 2006), and such lemmas are called distinctive collexemes of the construction.

Verbs and their preferences for the (non)progressive construction are visually displayed in Figure 4.1 which can be interpreted in the following way: The x-axis labeled *logged co-occurrence frequency* exhibits frequency of the lemma, and the farther to the right a lemma falls indicates its higher frequency. The y-axis labeled *association (log odds ratio)* is a visual representation of a lemma's preference for the (non)progressive. To interpret a lemma's preference for either construction on the figure, start from the dashed line in the middle (0 on the y-axis). Lemmas appearing above the dashed line were attracted to the progressive, and lemmas appearing below the dashed line were attracted to the non-progressive. As an example, take the verb *moreu* (*to not know*), which falls towards the bottom right of the figure. *Moreu* has a preference for the non-progressive construction (it falls below the dashed line), with a coll.strength score of 727. As *moreu* has a high coll.strength score and preference for the non-progressive, it is a distinctive collexeme of the non-progressive (i.e., the verb *moreu* is attracted to the non-progressive. Likewise, looking above the dashed line the stative verb *gaj* (*have*) is clearly visible. Falling above the dashed line indicates its preference for the progressive

construction (coll.strength of 39.21). Using Table A-1 (located in the appendix, tables include English translations of distinctive collexemes) and Figure 4.1 in tandem provides a comprehensive view of the verbs appearing across the progressive and non-progressive constructions and their preferences for either construction in L1 Korean written corpus data can be ascertained. Of note is that there were more distinctive collexemes found for the progressive than the non-progressive.

Figure 4.1.

Visual representation of lemmas, their frequencies, and preference for the (non)progressive



Exploring the verbs attracted to the progressive and non-progressive in L1 Korean writing, we see a variety of verbs which fall into various semantic domains (based on Biber et al., 1991 and 2021) well represented in both constructions.

Starting with the non-progressive construction, verbs fell into the following semantic domains:

- activity verbs (e.g., *deuleoo* – *come in*; *manna* – *meet*),

- communication verbs (e.g., *malha* – *speak*; *haeseogdwe* – *be interpreted*; *seolmeyongha* – *explain*; *gangjoha* – *emphasize*; *seoneonha* – *declare*),
- mental verbs (e.g., *moreu* – *to not know*; *bara* – *hope*; *johaha* – *like*; *jeulgi* – *enjoy*, *sarangha* – *love*; *weonha* – *want*),
- causative verbs (e.g., *heoyongha* – *permit*),
- occurrence verbs (e.g., *na* – *happen*; *pyeolcyeji* – *spread*; *dalha* – *reach* (e.g., *a level of something*), *dwe* – *become*),
- aspectual verbs (e.g., *sijagha* – *start*; *ggeutna* – *end*).

Of note is that in the current analysis, the existence/relationship semantic domain (e.g., represent, include, or contain) did not yield any distinctive collexemes in the L1 written data.

Moving on to verbs that were attracted to the progressive, a variety of semantic domains are also well represented by verbs appearing in the progressive. Distinctive collexemes fell into the following semantic domains:

- activity verbs (e.g., *sa* – *buy*; *dalli* – *run*; *moeu* – *gather*; *pal* – *sell*; *mojibha* – *recruit*),
- communication verbs (e.g., *jeonha* – *to tell/convey or pass on information*; *dabbyeonha* – *reply*; *nonha* – *discuss*),
- mental verbs (e.g., *bo* – *see*; *al* – *know*; *neuggi* – *feel*; *nuri* – *enjoy*; *uryeoha* – *be concerned or fearful*; *insigha* – *be aware*; *gominha* – *worry*; *mid* – *believe*),
- causative verbs (e.g., *chujinha* – *push ahead with or promote something to happen*),
- occurrence verbs (e.g., *byeonhwaha* and *baggu* – *change*; *jeunggaha* – *increase*),
- existence verbs (e.g., *daebyeonha* – *represent*; *mangraha* – *include or contain*),
- aspectual verbs (e.g., *beoli* – *start/begin*; *geuchi* – *stop*; *gyesogdwe* – *be continued*; *gyesogha* – *continue*).

Qualitatively comparing the distinctive collexemes found in the progressive and non-progressive in the L1 Korean written data shows that each semantic domain has unique verbs associated with them in each construction. For example, activity verbs in the non-progressive are largely verbs which can happen in a moment, for example *come in* or *meet*, whereas verbs found to be distinctive collexemes of the progressive inherently allow for a longer period of time, such as *gather/collect* and *recruit*. Notably, the verb *come in* (*deuleoo*) in the non-progressive was often used in the phrase *it comes into my eye* (눈에 확 들어온다), which can be translated idiomatically as *it catches my eye* in English.

(5) 4BH0004.txt

Korean: 그런데 펼쳐진 일기장의 왼쪽 페이지가 갑자기 내 눈에 확 들어온다.

English: However, in the open diary, the left page suddenly **caught my eye/attention** (**literally: entered in my eye**).

The activity verbs found in the progressive were, as expected used to express an action occurring over a larger period of time as opposed to a moment, and with an inanimate subject (showing variety in animacy of subject):

(6) 6BA02D33.txt

Korean: 지구기온 상승과 기상 이변을 일으키는 온실가스인 이산화탄소 농도의 국내 증가속도가 일본, 중국 등 주변국가를 크게 앞지르고 있는 것으로 나타나 비상한 관심을 모으고 있다.

English: The rate of increase in the concentration of carbon dioxide, a greenhouse gas that causes rises in temperatures and extreme weather events, is far exceeding that of

neighboring countries such as Japan and China, and thus is **gathering/drawing** extreme attention.

Distinctive collexemes in the communication verb category in the (non)progressive also exhibited unique trends in their usage. First, in the case of the non-progressive, the verbs appearing in the communication semantic domain were largely based around disseminating information (e.g., *haeseogdwe* – *interpret*, *seolmyeongha* – *explain*, *seoneonha* – *declare*). Notable is that these verbs imply a one-way transfer of information from the speaker to the listener(s):

(7) 4BJ01001.txt

Korean: 지은이는 우리가 당연히 여기면서 살아온 근대 자본주의 세계 자체, 그리고 그것을 지탱해온 자유주의라는 거대한 이데올로기, 그리고 이에 맞서온 저항의 지배적 형태 모두에 심각한 위기가 발상하여 더 이상 그 생명을 지속하기 어려워졌다고 선언한다.

English: The author **declares** that a serious crisis has arisen in both the modern capitalist world itself, the great ideology of liberalism that has sustained it, and the dominant form of resistance against it, making it difficult to sustain its life any longer.

As can be seen in the above example, much of the usage of communication verbs in the non-progressive tended towards conveying information, without necessarily requiring an interaction or reaction from the intended listener(s). On the other hand, in the progressive, communication verbs were largely interactional and used to describe exchanges between parties, passing information along, debating, and giving responses. As an example, take (8) which shows the verb

jeonha (to pass along or convey information) being used with the *-ko iss* construction to express conveying new facts and ideas by the author on various artistic mediums.

(8) 4BJ01001.txt

Korean: 또 하나, 저자가 갖고 있는 건축을 비롯한 미술, 사진, 음악, 오페라에 대한
저자의 식견은 풍부한 교양을 제공해 줄뿐만 아니라, 새로운 흥미로운 사실들을
전하고 있다.

English: In addition, the author's insight into art, photography, music, and opera,
including architecture, not only provides a rich culture, but also **conveys** new interesting
facts.

Mental verbs, or those verbs describing activities or states experienced by humans,
follows with several distinctive collexemes in the non-progressive. In this category, the pair of
verbs *al* (to know) and *moreu* (to not know), both stative verbs, were found to have a preference
for different constructions. *Al* (to know), was largely associated with the progressive
(coll.strength of 64.84), and *moreu* (to not know) was associated with the non-progressive
(coll.strength of 727 – *moreu* was also the distinctive collexeme with the highest coll.strength
score in the non-progressive distinctive collexeme list).

(9) 2BA90A35.txt

Korean: 그녀가 무슨 말을 하고 싶은지 다 알고 있다.

English: Everybody **knows** what she wants to say.

**al* (know) marked with progressive *-ko iss*

(10) 5BA01B07.txt

Korean: 많은 사람들이 아직 에이즈를 동성 연애자나 극소수 문제있는
사람들의 병으로만 알고 있다.

English: Many people still only **know** of AIDS as a disease affecting gay people
or a very small number of people.

**al* (know) marked with progressive *-ko iss*

(11) 4BH0004.txt

Korean: 앞으로는 게임이나 애니메이션 같은 멀티미디어 쪽 예술에서 보다
중요한 예술적 성과가 나올지 모른다.

English: Going forward, it is **not known** if more artistic achievements may come
about in multimedia fields such as gaming or animation.

As can be seen in the examples, *al* (*to know*) is widely used with the progressive in the L1 Korean written corpus, and *moreu* (*to not know*) is used at a high rate with the non-progressive. This distinction is notable as both verbs have been said to be stative verbs which can be used in the progressive *-ko iss* construction in Korean. However, according to the collostructional analysis, I find that there is a clear preference for *al* to be used with the progressive *-ko iss*, and for *moreu* to be used in the non-progressive, at least in written data.

Stative verbs beyond *al* (*know*) and *moreu* (*not know*) were present in the data. Other stative verbs that were distinctive collexemes for the progressive included *neuggi* (*feel*), *insigha* (*be aware*), *gominha* (*worry*), and *mid* (*believe*). Given the fact that Korean allows for stative progressives at a higher rate than other languages, having so few stative progressives appear in the mental verb category was surprising, though it must be said that these are only those stative

progressives which were attracted to the progressive. It is possible that other stative progressives appeared at lower frequencies and were therefore not found to be distinctive collexemes. Nonetheless, the stative progressives found in the L1 written data help illuminate how the progressive *-ko iss* can combine with stative verbs in Korean.

(12) 2BA93A22.txt

Korean: 나는 비행기보다는 철도여행을 좋아하므로 매번 불편을 느끼고 있다.

English: I prefer trips by train over plane, so I **feel** uncomfortable every time.

**neuggi* (feel) marked with progressive *-ko iss*

Example (12) illuminates how the progressive *-ko iss* may be used with a mental stative verb when the experience continues over a period of time, such as feeling uncomfortable each time one does a certain activity. This is also realized with the verb *insigha* (*to be aware, perceive, recognize*), which is used with the progressive *-ko iss* when a speaker is discussing a point which they are aware of and recognize as important.

(13) 5BA01B10.txt

Korean: 나는 인간 생명의 존엄성에 대한 윤리적 철학적 배경이 배제된

생명공학이 인류의 재앙이 될 수 있다는 점을 깊이 인식하고 있다.

English: I am deeply **aware** that biotechnology, which excludes ethical and philosophical backgrounds on the dignity of human life, can become a disaster for humanity.

**insigha* (aware) marked with progressive *-ko iss*

Beliefs can also be expressed by combining the verb *mid* (*believe*) with the progressive in Korean. It can also be used to express a belief one holds about an event they presume will happen at a future time.

(14) 5BA01A09.txt

Korean: 하지만 LG 선수들은 김태환 감독이 그에 대한 대비책을 내놓을 것으로
믿고 있다.

English: However, LG athletes **believe** that director Taehwan Kim will come up with a countermeasure.

**mid* (believe) marked with progressive *-ko iss*

Finally, when considering the mental verbs, and in particular the aforementioned stative verbs which are distinctive collexemes of the (non)progressive, an unexpected trend can be observed in regard to the supposed emotional sentiment expressed with the verbs in each construction. The collostructional analysis revealed that mental verbs which are distinctive collexemes of the non-progressive are largely associated with positive mental experiences or emotions: *bara* (*hope*), *johaha* (*like*), *jeulgi* (*enjoy*), *sarangha* (*love*), *weonha* (*want/desire*), are all verbs which can be categorized as largely positive emotional or mental experiences.

However, zooming in on the distinctive collexemes for the progressive reveals a different pattern: *uryeoha* (*be concerned, fearful*), and *gominha* (*worry*) are mental verbs which are associated with negative emotions. While beyond the scope of this dissertation, it is an interesting finding that semantic meanings associated with the (non)progressive constructions seem to follow different trends in L1 Korean writing.

For causative verbs, in the non-progressive the verb *heoyongha* (*permit*) appeared as a distinctive collexeme. In the progressive, *chujinha* (*to promote something to happen*) is the

distinctive collexeme identified. *Chujinhada* was often used with objects such as *plan* (*계획을 추진하고 있다*, *pushing ahead/promoting a plan*) and other similar objects.

(15) 5BA01B06.txt

Korean: 정부도 의료전달체계 확립을 위해 가벼운 질병으로 3 차병원을

이용하는 환자에게는 무거운 진료비를 물게 하는 방안을 추진하고 있다.

English: To establish a medical delivery system, the government is also **pushing for/promoting** a plan to impose heavy medical expenses on patients who use tertiary hospitals due to treat mild diseases.

Several verbs falling into the occurrence verb category were distinctive collexemes. Starting with the non-progressive, one of the most common verbs was *na* which, while difficult to translate into English, is a verb which means *to happen/come up*, and in some cases *break out* or *occur* depending on the context. Unique to Korean, this verb is usually preceded by a noun marked with a nominative case to denote what is happening or occurring. Qualitative analysis of the data containing *na* shows the verb appearing with a wide array of nouns, including *dust*, *disease*, *problem*, and *smell*, among others. In each case, *na* is used to denote the occurrence of the noun. While it is beyond the focus of the present dissertation, a follow-up study could explore *na* and the lexical items it associates with through co-varying collexeme analysis (another type of collocation analysis) to identify typical usage cases of *na*.

Another verb which follows a similar pattern in the data is *dwe* (*become*). In fact, due to the form *dwe* appearing in a multitude of grammatical constructions in the Korean language, to extract only those occurrences of *dwe* expressing the *become* meaning, a separate regular expression had to be written to ensure that *dwe* was being extracted alongside nouns with

nominative case, followed by manual inspection of the extractions. Qualitative inspection of the data reveals *dwe* appearing with a multitude of nouns marked with nominative case, including *reason* (...의/ㅁ/가 된/다 – *something becomes the reason for...*; ...도움이 된/다 – *something is/becomes helpful*).

(16) 5BA01B07.txt

Korean: 사춘기에 다리 안쪽이나 등에 생기는 튼살은 조기에 치료하면 회복에 도움이 된다.

English: Treatment of stretch marks during puberty on the inner or back of the leg can **be helpful**.

Finally, distinctive collexemes were identified in the non-progressive for the aspectual verb category, which includes verbs which denote the time or status something is occurring at (such as *start*, *end*, *continue*, etc.). In the L1 Korean written data, *sijagha* (*start*) and *ggeutna* (*end*) were distinctive collexemes of the non-progressive.

(17) 4BJ01001.txt

Korean: 대부분의 인간적 삶이란, 자신이 스스로의 삶을 거리를 두고 볼 수 있는 여유와 창조적 해결을 위한 판타지를 갖지 못해 종종 영원한 미궁으로 흘러 들어가거나 비극으로 끝난다.

English: As for most human lives, they are unable to put space between themselves and their lives to have the freedom to come up with creative solutions, and sometimes this causes them to fall into an endless labyrinth, which **ends** in tragedy.

Notably, there are no other aspectual verbs in the non-progressive apart from those which indicate the start or end of an event. In contrast, the progressive had *start/begin* (*beoli*), *stop* (*geuchi*), *gyesogha* (*continue*), and *gyesogdwe* (*be continued*) as distinctive collexemes. Perhaps unsurprisingly, the progressive *-ko iss* construction is used to express events as they are in the process of starting/stopping, as well as to describe their status in continuation.

(18) 5BA01A02.txt

Korean: 일부 시위대들이 정유소 봉쇄를 풀었으나, 여전히 대다수는
과도한 유류세 인하를 요구하며 연일 시위를 **계속하고 있다**.

English: Some protestors have lifted the oil refinery blockade, however others,
demanding excessive oil tax cuts, **are continuing** to protest.

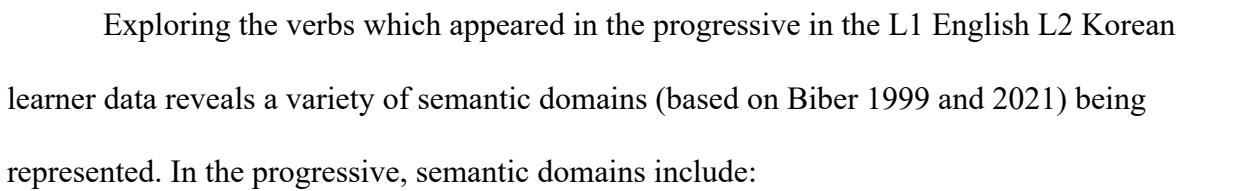
Interim summary: In this section, I have shared the results of the distinctive collexeme analysis for the L1 Korean written data. I provided a table which allows the reader to view which verbs were exhibited a preference for either the progressive or the non-progressive. Discussing the results in prose, I outlined verbs which appeared as distinctive collexemes in the progressive and the non-progressive and discussed them in terms of semantic domains verbs are categorized while providing examples of verbs in context (taken from the corpus). As this section covers L1 written data, there are more distinctive collexemes than there are for the learner data to follow. While it is natural to expect L1 data to include a wider variety of verbs, it must be noted that the amount of data is also substantially larger in the L1 corpus. In what follows, I will discuss the results of the distinctive collexeme analysis for the learner data, providing the list of the distinctive collexemes for the progressive and the non-progressive.

4.2.2. Analysis of L2 Corpus: Distinctive Collexemes for the (non)progressive in L1 English L2 Korean Writing

Table A-2 in the appendix shows the distinctive collexemes for the progressive on the left and the non-progressive on the right for L1 English L2 Korean speakers. In total, there were twenty-three (23) distinctive collexemes attracted to the progressive *-ko iss* construction, and nine (9) distinctive collexemes showing a preference for the non-progressive.

This can also be seen in Figure 4.2, which visually shows the relationship between each lemma, its frequency, and its preference for the (non)progressive. Figure 4.2 can be interpreted in the following way: The x-axis labeled *logged co-occurrence frequency* exhibits frequency of the lemma, and the farther to the right a lemma falls denotes its higher frequency. The y-axis labeled *association (log odds ratio)* is a visual representation of a lemma's preference for the (non)progressive. To interpret a lemma's preference for either construction on the figure, start from the dashed line in the middle (0 on the y-axis). Lemmas appearing above the dashed line are associated with the progressive, and lemmas appearing below the dashed line correspond with the non-progressive. Thus, as a simple example, looking to the very bottom right of the figure, the lemma *saenggagha (to think)* is the most frequent among lemmas appearing in the non-progressive as it is farthest to the right, and it exhibits a strong preference for the non-progressive, as it is far below the dashed line.

Visual representation of lemmas, their frequencies, and preference for the (non)progressive



- There was a lack of the following semantic domains in the progressive data: communication verbs, causative verbs, existence verbs, and aspectual verbs.

- activity verbs (e.g., *meog – eat; ju – give*),
- mental verbs (e.g., *saenggagha – think; bo – see; deud – listen*).

As the distinctive collexeme analysis allows an analyst to identify verbs which have a high association strength and preference for a particular construction, this suggests that when verbs appear in both constructions, their usage in the learner data will tend towards the progressive as opposed to the non-progressive. Further, it seems that in the case of L1 English L2 Korean writers, the associations between certain semantic domains and a variety of usages are stronger with verbs in the progressive than the non-progressive.

Digging deeper into the verbs associated with the (non)progressive in the L1 English L2 Korean learner data, unsurprisingly, many of the verbs are used to indicate a change occurring over time. For example, the verb most highly associated with the progressive in the learner data is *manhaji* (*increase, grow*), which was often used to express changes on a societal level:

(19) sample_9841.txt

Korean: 하지만 인터넷을 무조건 믿는 사람이 많아서 인터넷 발전으로
인해 사람들이 사기꾼 같은 사람들은 신뢰하는 경우가 **많아지고 있다**.

English translation: However, as there are many people who just believe in the internet, through the development of the internet the amount of people who believe scammers is **increasing**.

Further qualitative analysis also revealed some errors in learners' usage of the verb *manhaji* with the progressive. An intransitive, the verb *manhaji* describes an increase or growth, and thus the argument of the verb should take the nominative case marker in Korean. However, some learners in the L1 English category exhibited errors in their usage of grammatical markers with verbs in the progressive:

(20) sample_30998.txt

Korean: 명품 소비자들은 대부분 여자 있었는데 요즘 남자들도 명품에 대한 관심을 **많아지고 있다**.

English translation: Consumers of brand-name/designer products were mostly women, but these days, interest in brand-name products by men is also **increasing**.

Interestingly, the lemma directly following *manhaji* came out to be *jeunggaha*, which also means *increase*. However, its overall lower association strength and frequency in the learner data could be due to it being taught at a more advanced level, and thus learners may have been exposed to the word less.

Continuing in the investigation of distinctive collexemes in the learner data reveals an interesting trend with mental and stative verbs exhibiting a preference for and a strong association strength with the progressive construction in Korean. Several mental and some stative verbs, including *gominha* (*worry*), *al* (*know*), *neuggi* (*feel*), and a physical stative verb *sal* (*live*) exhibited a preference for the progressive. This is an interesting finding as Korean is known to allow for stative verbs at frequencies higher than English, so learners exhibiting usage of stative progressives in their Korean writing is a positive sign for acquisition of the form-meaning association between states and the progressive construction in learner Korean.

(21) sample_29420.txt

Korean: 현대 부모님들도 외국어 학습을 시작할까 **고민하고 있다**.

English: Parents in modern times are **worried** (*worrying*) about starting foreign language acquisition.

(22) sample_31034.txt

Korean: 내가 중국하고 영국에서도 살아* 본 적이 있어서 서양 문화와 아시아
문화를 잘 알고 있다.

English: Because I have experience living in both China and England, I **know**
both western and Asian cultures very well.

Note: * denotes a correction in spelling.

As can be seen in the examples above, L1 English learners of Korean exhibit use of stative progressives with both stative verbs that can take the progressive in English (e.g., it is not unheard of for a sentence such as *parents are worrying about X* in English), as well as verbs that typically do not co-occur with the progressive, such as *al (know)*.

The exploration of stative verbs appearing in the learner data is more notable when compared with the stative verbs that are distinctive collexemes in the non-progressive. Among the verbs which appeared in both constructions, the only stative verb that was found to be a distinctive collexeme in the non-progressive was *saenggagha (think)*, which exhibited an extremely strong association strength and preference for the non-progressive that was higher than any other singular verb's preference for either construction in the L1 English data.

In terms of form-meaning mappings, results show that L1 English writers of Korean associate the progressive with prolonged duration of an event or action (e.g., *increasing, preparing, changing, attending, becoming, disappearing*) and mental states (e.g., *worry, feel, know, expect*). On the other hand, verbs found to be significant collexemes of the non-progressive, with the exception of *saenggagha (think)*, are by and large (physical) actions (e.g., *see, eat, give, oppose, send, drink, use, listen*). This distinct difference between the form-meaning mapping of the (non)progressive in Korean suggests that learners may associate the

progressive form with usages of prolonged duration and mental states, and the non-progressive form with physical action verbs. Diving into the usage cases of such action verbs, we can see that learners use the non-progressive for habitual actions:

(23) sample_3890.txt

Korean: 버스나 지하철을 탈 때 헤드폰으로 항상 음악을 듣는다.

English: When I ride the bus or subway I always listen to music with headphones.

Further, learners are also correct in associating the verb *bo* (*to see*) in the non-progressive with its usage of expressing how one views a certain state or situation. In other words, for the verb *bo* (*to see*), learners have acquired its usage which extends beyond the simple *to see/to watch* function and are able to use it to express their views.

(24) sample_6714.txt

Korean: 사회가 빠르게 변화하면서 사람들의 관심 분야도 변하기 때문에

전통문화를 조금씩 현대화 시키는 것은 모두에게 좋다고 본다.

English: As society changes rapidly so too do people's interests, and so I **view** the change in traditional culture to something more modern as a good thing.

Interim summary: Overall, the distinctive collexeme analysis of verbs in the progressive and non-progressive in L1 English L2 Korean writing show interesting trends. Most notable is the trend of the progressive construction *-ko iss* being associated with stative and mental verbs when compared with the non-progressive form. Results suggest that L1 English learners of L2 Korean are able to overcome the obstacles that may be presented by typological differences between English and Korean, namely that Korean allows for more stative verbs than English. While it was hypothesized that learners from the L1 English background would demonstrate and overall lack of stative verbs in their writing, that fact that several stative verbs appeared in their writing

is a positive sign for the accurate development of the form-meaning mappings of the progressive construction in learner language. Perhaps most notable is the learner usage of *al (to know)*, a stative verb in Korean which almost never takes the progressive in English. Overall, results suggest a positive trend towards felicitous usage and promising acquisition patterns of the progressive *-ko iss* construction in L1 English L2 Korean. Finally, notable in this set of learner essays is that when it comes to verbs which appear in both constructions, more verbs appearing in those constructions are found to be distinctive collexemes of the progressive as opposed to the non-progressive. From a usage-based perspective, it could be the case that the input learners receive, either through textbooks or interactions with native speakers, is that these verbs simply appear more with the progressive than the non-progressive, and thus when focusing on verbs that can appear in both, it is natural that more distinctive collexemes are found in the progressive.

4.2.3. Analysis of L2 Corpus: Distinctive Collexemes for the (non)progressive in L1 Japanese L2 Korean Writing

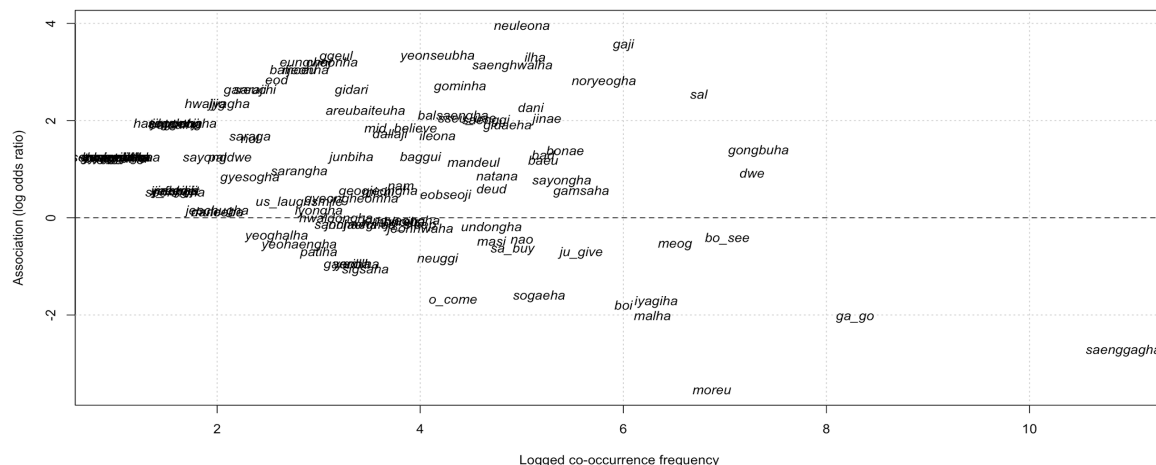
In this section, the findings from the distinctive collexeme analysis of the L1 Japanese L2 Korean data are discussed.

As can be seen in Table A-3, there is an uneven distribution of the number of distinctive collexemes in the progressive and non-progressive. As a distinctive collexeme analysis only explores those verbs which appear in both constructions, it appears that when a verb can appear in either construction, learners tend to associate verbs with the progressive more often and use the verbs in the progressive at higher frequencies. This is borne out in higher collostructional strength scores, yielding more distinctive collexemes for the progressive than the non-progressive. In total, there were fifty distinctive collexemes for the progressive construction and fourteen distinctive collexemes associated with the non-progressive construction.

This uneven distribution is visualized in Figure 4.3, which visually represents each verb and its preference for the progressive or the non-progressive. All verbs that were submitted to the collostructional analysis are featured in Figure 4.3, hence why there appears to be more lemmas listed than can be seen in Table A-3 (Table A-3 only includes those which yielded collostructional/association strengths of greater than 1.3; Figure 4.3 represents all lemmas which were extracted and submitted to the distinctive collexeme analysis). Figure 4.3 can be interpreted in the following way: The x-axis labeled *logged co-occurrence frequency* exhibits frequency of the lemma, and the farther to the right a lemma falls denotes its higher frequency. The y-axis labeled *association (log odds ratio)* is a visual representation of a lemma's preference for the (non)progressive. To interpret a lemma's preference for either construction on the figure, start from the dashed line in the middle (0 on the y-axis). Lemmas appearing above the dashed line are associated with the progressive, and lemmas appearing below the dashed line correspond with the non-progressive. Thus, as a simple example, looking to the very bottom right of the figure, the lemma *saenggagha* (*to think*) is the most frequent lemmas as it is farthest to the right, and it exhibits a strong preference for the non-progressive, as it is far below the dashed line.

Figure 4.3.

Visual representation of lemmas, their frequencies, and preference for the (non)progressive



First, when exploring the distinctive collexemes in the L1 Japanese L2 Korean data, it is clear that there is a variety of verb types which are attracted to both the progressive and non-progressive construction. For example, the most common verb overall was *saenggagha* (*think*), a mental verb which exhibits a strong preference for the non-progressive construction. In addition to *saenggaggha*, other mental verbs with a preference for the non-progressive include *moreu* (*not know*), *boi* (*be visible*), *bo* (*see*), and *neuggi* (*feel*).

(25) sample_6334.txt

Korean: 큰일이 생기면 나중에 후회할지도 모른다.

English: If a big issue crops up later, I **don't know** if I'll regret it.

In the case of the verb *bo* (*see*), learners usage of this verb to express their views and the way they observe the world were found in the data, for example, when discussing societal issues (the writer in the example below was discussing gender issues in Japan):

(26) sample_32001.txt

Korean: 아마도 여자의 의식이 남자보다 앞서 있다고 본다.

English: **The way I see it**, women probably have more consciousness/awareness than men do.

The non-progressive also features two communication verbs, namely *malha* (*speak*) and *iyagiha* (*talk*), two verbs which are often interchangeable and share similar ranking and collostructional strength scores (ranked fourth and fifth distinctive collexemes, with coll.strength scores of 22.02 and 19.33, respectively). Similar ranking suggests learners may use these verbs interchangeably and that they have similar levels of entrenchment in the learner language. Activity verbs (denoting actions and events associated with someone's choice or own volition) are represented among the distinctive collexemes for the non-progressive as well, including verbs such as *ga*, *meog* (*eat*), *ju* (*give*), *sa* (*buy*), and *sigsaha* (*have a meal*). Put simply, distinctive collexemes associated with the non-progressive construction in the L1 Japanese L2 Korean variety include mental, communication, and activity verbs only.

Diving into the progressive exhibits a richer array of verb types and larger variety of lemmas, suggesting that verbs which appear in both the (non)progressive may tend towards the progressive *-ko iss* construction in learner Korean. Additionally, a larger variety of verb types are found in the distinctive collexemes for the progressive, including activity verbs (e.g., *ilha* – *work*), mental verbs (e.g., *gominha* – *worry*), occurrence verbs (e.g., *baggui* – *be changed*; *baldalha* – *develop*). Therefore, it appears that in terms of overlapping verb types (based on Biber 1999 and 2021 semantic domains) overlap occurs for activity and mental verbs.

(27) sample_9277.txt

Korean: 지금의 직장엔 스트레스를 많이 받긴 하는데 즐겁게 일하고 있다.

English: At my current workplace I do get stressed but I am **working** joyfully.

(28) sample_33595.txt

Korean: 근데 어학당을 졸업한 후에 한국에서 일을 할지 일본에서 일을 할지
고민하고 있다.

English: But now, I am **worrying** about whether to work in Korea or Japan after
graduating from the Korean language school.

(29) sample_31394.txt

Korean: 최근 세계적으로 가족의 형태가 바뀌고 있다.

English: Recently, the form of the family unit is **changing** globally.

(30) sample_32874.txt

Korean: 한국에서는 인터넷을 이용한 배달이나 택배가 많이 발달하고 있다.

English: Online delivery and shipping services are **developing** in Korea.

Distinctive collexemes for communication verbs were only present in the non-
progressive, and occurrence verbs were only present in the progressive. Both constructions
yielded distinctive collexemes for stative verbs, with the stative verb with the highest
collostructional strength in the L1 Japanese L2 Korean data being *gaji* (*have*, coll.strength
138.10), followed by *gominha* (*worry*, coll.strength 32.34), *gidaeha* (*expect*, coll.strength 24.46),
and *mid* (*believe*, coll.strength 11.12).

(31) sample_15460.txt

Korean: 사람에 따라 다른 생각을 가지고 있다.

English: Depending on the person, the thoughts they **have/hold** (progressive marked in Korean) differ.

(32) sample_34551.txt

Korean: 그중에서도* 교토 사투리는 표준어에서는 찾아볼 수 없는 독특하고 부드러운 어감을 가지고 있다.

English: Even among them*, the Kyoto dialect **has** (progressive marked in Korean) a unique and soft sense of language that cannot be found in standard language.

*그중에서도 refers to the various dialects of the Japanese language.

(33) sample_6752.txt

Korean: 보통 특히 여성들은 명품에 관심이 있는 듯싶다. 나도 관심이 있고 갖고 싶은 욕심을 가지고 있다.

English: Usually, and especially, women have an interest in designer products. I also **have** (progressive marked in Korean) an interest in and greediness for designer products.

Apart from *gaji* (*have*), notable examples of the stative progressive in the L1 Japanese data were found for the distinctive collexeme *mid* (*believe*) in students' writing on superstitions:

(34) sample_13364.txt

Korean: 그것은 찻잎이 서면 집에 기운도 선다고 생각해서 근웅이
좋아지거나 좋은 일이 일어난다고 믿고 있다.

English: If the tea leaf stands, it is **believed** (progressive marked in Korean) that
good fortune will come.

In some instances, learners used the progressive with *mid* (*believe*) to express beliefs they currently hold:

(35) sample_35879.txt

Korean: 친구, 가족, 아는 사람들을 모든 사람들을 사랑하는 노력을 서로 할 수
있으면 다들 행복하게 살 수 있다고 믿고 있다.

English: I **believe** (marked with progressive in Korean) if people who know each
other put in an effort to love each other, then everyone can live happily.

Overall, results show promising acquisition of the progressive construction in Korean, its usage with various verbs of various semantic domains (not limited to physical actions), and the development of the usage of stative progressives in Korean.

4.2.4. Comparisons of L1 and L2 corpora and discussion of potential interlanguage transfer effects

As one goal of this dissertation is to illuminate how the progressive *-ko iss* construction is used in L1 and L2 Korean, in this section, I will briefly cover some of the differences observed in the progressive and its usage across the L1 and L2 varieties thus far. While the focus is not necessarily to discuss language learning mechanisms, the results from the collostructional

analysis show which lemmas prefer *-ko iss*, which can be a starting point for discussing differences in usage based on a language user's L1.

In terms of interlanguage transfer effects, one of the ways to identify whether typological differences could be at play is to look for stative verbs appearing in the progressive *-ko iss* construction across the L1 and L2 varieties. The three varieties, L1 Korean, L1 English L2 Korean, and L1 Japanese L2 Korean, lend themselves well to such analysis as both Korean and Japanese are known to allow perfective readings to be associated with the grammatical constructions associated with the progressive, whereas in English the progressive is never perfective and also describes an ongoing action or even a futurate reading in some cases (e.g., Lee, 2006; McLure, 1994; Yeon & Brown, 2011). Essentially, what this means is that the progressive constructions in Korean and Japanese can be used with stative verbs which would not normally be expected to take the progressive construction in a language such as English. Prime examples of such verbs include *to know*, *to have*, or *to believe*. For example, in the case of the verb *know* in Korean, it has been described as being able to take the progressive as, at one point, someone came to know the information, and they will remain in that state of 'knowing' that information until the moment they forget it, at which point, the state would be over (Lee, 2006). The Japanese progressive, *-te iru*, functions in a similar way particularly for stative verbs, thus it could be anticipated that L1 Japanese learners of Korean will incorporate more stative progressives in their writing than their L1 English counterparts, thus providing some potential evidence for interlanguage transfer effects. However, given that the present study is corpus-based and not experimental to test for specific interlanguage transfer effects or cross-linguistic influences, at most I will only discuss trends that appear in the data.

To compare the results of the distinctive collexeme analysis, I first normalized the coll.strength scores for each variety and visualized them using bar charts. Doing so makes identifying trends in lemma associations with the progressive across varieties easier and allows for an analyst to quickly identify stative progressives appearing in each variety.

Visual inspection of the normalized coll.strengths in Figures 4.4, 4.5, and 4.6, for all varieties shows a sharp drop in coll.strengths after the first two or three distinctive collexemes, showing that certain verbs may be more prototypically associated with the progressive *-ko iss* construction. Unsurprisingly, the L1 variety yielded the highest number of distinctive collexemes (256), followed by the L1 Japanese group (78) and the L1 English group (43).

While typological similarities could be one explanation as to why the L1 Japanese data including more distinctive collexemes than the L1 English data, it should also be noted that these differences could be due to the different number of learner essays available in each corpus, so this trend should be considered with caution.

A well attested stative progressive to appear in both Korean and Japanese is *know* (*al* in Korean; *shiru* in Japanese), and that verb appears in both the L1 and learner corpora. In the L1 data, *al* is the ranked 18th distinctive collexeme with a coll.strength score of 64.84, confirming the form-function mapping of *al* with the progressive *-ko iss* construction in Korean is well attested in the reference corpus. In the L1 English L2 Korean corpus, *al* also appears, ranked as the 29th distinctive collexeme with a coll.strength score of 3.3, a perhaps surprising finding as it was anticipated that L1 English speakers would likely not use the verb *know* in the progressive due to *know* rarely taking the progressive in English. However, more surprising is the finding that *al* was not a distinctive collexeme in the L1 Japanese learner data. From this, there are two points to highlight. First is that despite a relative lack of stative progressives in English (and

especially with the verb *know*) it does not appear to be impeding the uptake of this form-function mapping in L1 English learners of L2 Korean, which is a positive sign for second language learners. Second, while beyond the scope of this study, it could be the case that there are factors at play which determine the use of a stative progressive with *know* in Japanese that is not represented in the data. For example, while this study is limited to exploring written data across all varieties, it could be the case that certain stative progressives are more common in both Korean and Japanese spoken language, which could be a reason why *al* is not strongly associated with the progressive in the L1 Japanese learner data despite a similar form being well attested in the learners' L1. However, that is not to say that the Japanese data did not include *al (know)* at all. In fact, while *al (know)* was not a distinctive collexeme in the Japanese variety, normalizing the data revealed that Japanese speakers actually used *al* more frequently with the progressive - *ko iss* than both the English and L1 Korean varieties. This suggests that, in L1 Japanese learners' L2 Korean writing, they are choosing to use other verbs significantly more often than *an (know)*. Put simply, Japanese learners of Korean had more verbs co-occur with the progressive, and as such even some key verbs (including *al*) were not found to be distinctive collexemes of the construction despite an overall higher relative frequency.

While the L1 Japanese data lacked the verb *al (know)* as a distinctive collexeme, other common stative progressives were present in the data. For example, *gaji (to have/to hold)* appeared in the L1 data ranked 16th distinctive collexeme with a coll.strength of 69.5. The same verb was a distinctive collexeme for the L1 Japanese data, ranked 2nd distinctive collexeme with a coll.strength of 138.10. *Gaji* was not found to be a distinctive collexeme in the L1 English learner data. The verb *have* in Korean and Japanese (*gaji* and *motsu*) are both known to be used with the progressive construction, so it is unsurprising that it appeared in both the L1 data and

that it was a distinctive collexeme highly attracted to the progressive construction in the L1 Japanese learner data. A notable point about the way the verb *have* functions in both languages, and largely in Korean, is that it can be used to denote ongoing possession of not only physical objects but also intangible things including thoughts, feelings, backgrounds, impressions, and so on. In that way, Korean and Japanese are typologically similar, however, English differs in this regard as the verb *have* would rarely be used with the progressive to express similar sentiments.

Examples of intangible possession in L1 Korean data with progressive *gaji*:

(36) 4BH0005.txt

Korean: 그들은 '훌륭한 사회' (good society)에 대한 이미지를 가지고 있다.

English: They **hold/have** an image of a "good society."

(37) 4BH0004.txt

Korean: 그래서 이 세 가지는 항상 서로 통하는 의미를 가지고 있다.

English: So these three things always **have** a common meaning.

(38) 4BH00013.txt

Korean: 나는 시인들에게 깊은 존경심을 가지고 있다.

English: I **have** a deep respect for poets.

Example of tangible possession in L1 Korean data with progressive *gaji*:

(39) 5BA01B04.txt

Korean: 핵 시대에 접어들면 상황은 더 악화된다. '우리 핵무기 가지고 있다.' '너 뭐 줄래?' 하는 식이다.

English: If we enter a nuclear era/generation the situation will become much worse. It could be like 'we **have** nuclear weapons.' 'What can you give us?'

Examples of intangible possession in the L1 Japanese data with progressive *gaji*:

(40) sample_26742.txt

Korean: 나는 감시 카메라 설치 확대에 대한 찬성 의견을 가지고 있다.

English: About the expansion and installation of security cameras, I **have/hold** an opinion of agreement.

(41) sample_15454.txt

Korean: 나는 휴일마다 낮잠을 자는 습관을 가지고 있다.

English: I have/hold a habit of napping on my days off.

*have/hold marked with progressive *-ko iss*

(42) sample_18316.txt

Korean: 메테인 가스는 사실 이산화탄보다 약 10 배의 온실 효과를 가지고 있다.

English: In fact, methane gas **has/hold** approximately 10 times the greenhouse effects of carbon dioxide.

As can be seen in the examples above, L1 Japanese learners of L2 Korean are able to use the verb *gaji* (*have/hold*) with the progressive in ways that are in-line with how L1 speakers use the verb, namely with intangible objects including thoughts, feelings, effects, habits, and opinions. From an interlanguage standpoint, the fact that this verb in the L1 Japanese L2 Korean data was found to be highly associated with the progressive, but not so in the L1 English L2 Korean data according to the distinctive collexeme analysis, suggests that there are some interlanguage transfer effects at play. English lacking a system of using the equivalent English verb with the

progressive and with tangible/intangible objects may lead to learners having difficulty with uptake of this form-meaning association.

In terms of stative verbs, of particular interest are the stative verbs noted in Yeon and Brown (2011) as attested verbs associated with the progressive in Korean, namely *know (al)*, *not know (moreu)*, *love (sarangha)*, *believe (mid)*, *want (weonha)*, and *feel (neuggi)*. The distinctive collexeme analysis revealed the following trends in usage of these specific verbs with the progressive construction in Korean as shown in Table 4.1. Table 4.1 outlines the key verbs mentioned in Yeon and Brown, highlighting which construction each verb appeared in for each variety (the progressive or the non-progressive). If a verb appeared in and preferred either construction in each variety, it is marked with a plus sign '+'. For example, looking at the first row where *al (know)* is listed, it is clear that *al (know)* appeared in the L1 Korean data and that it exhibited a preference for the progressive *-ko iss* construction. Continuing on, the first major trend observed in the data is that all key verbs appeared in the L1 data. However, of note is that, among those verbs, only *al (know)*, *mid (believe)*, and *neuggi (feel)* were found to have a preference and association for the progressive construction in the distinctive collexeme analysis. Focusing on the learner data reveals some differences between L1 and L2 language. First of all, while English generally lacks the verb *know* being used with the progressive construction (e.g., it would be unnatural for a native speaker to say *I am knowing all about that* in English), the L1 English speakers exhibited usage of the verb with the progressive construction. Perhaps surprisingly, this verb was not found to be a distinctive collexeme of the progressive in the L1 Japanese learner data, and this is surprising as Japanese allows for the use of the progressive *-te iru* (Japanese progressive construction) with the verb *know*. For the L1 English learner data, after *al (know)*, the only other verb in the list that exhibited a preference for the non-progressive was

neuggi (*feel*), with a preference for the progressive construction. Despite having a low number of distinctive collexemes from the Yeon and Brown list, the distinctive collexemes that do appear in the L1 English data follow similar collostructional patterns with the L1 Korean data. For the L1 Japanese data, only two verbs from the list were associated with the progressive: *sarangha* (*love*) and *mid* (*believe*). Both of these verbs in Japanese are able to take the progressive construction in the Japanese language (e.g., *aishiteiru* “I am loving/I love you”, and *shinjiteiru* “I am believing”). As such phrases are possible in Japanese (though less common in English), this may be evidence for interlanguage transfer effects allowing for L1 Japanese speakers to acquire and use more stative verbs with the progressive *-ko iss* in their L2 Korean writing. However, notable is that in the L1 data, the verb *sarangha* is actually associated with the non-progressive. While this may be surprising at first, it is important to note that this dissertation only focuses on written language in all varieties. It is possible that such construction will appear more in spoken language (discussed more in limitations and future directions).

Table 4.1.

Summary of key verbs appearing in the progressive -ko iss construction as distinctive collexemes across varieties based on Yeon and Brown (2011)

Verb	L1 written corpus		L1 English L2 Korean		L1 Japanese L2 Korean	
	Prog	Non-prog	Prog	Non-prog	Prog	Non-prog
al – to know	+		+		N/A	N/A
moreu – to not know		+	N/A	N/A	N/A	N/A
sarangha – love		+	N/A	N/A	+	
mid – believe	+		N/A	N/A	+	
weonha – want		+	N/A	N/A	N/A	N/A
neuggi – feel	+		+			+

Finally, turning to specific verb choice in the L1 and L2 data reveals some trends which could be particularly helpful for teachers and textbook/materials developers. First, there were a few verbs that had similar meanings that appeared in the learner data, when another (perhaps more formal or academic) form appeared in the L1 data. One example of this is the difference between the verb *sal* (live) and *geojuha* (reside). While functionally similar, *sal* was found to be a distinctive collexeme for the progressive *-ko iss* in all varieties. On the other hand, *geojuha* was only found to be a distinctive collexeme in the L1 data. This difference exemplifies how learners may tend to rely on more common/less academic language, which is something pedagogues should be aware of. Another example of this was the pair of verbs *jeunggaha* (increase) and *manhaji* (increase). While both verbs have the same semantic meaning, *manhaji* was only found to be a distinctive collexeme in the L1 English learner data. In the L1 Korean data, only *jeunggaha* and *neuleona* (increase) were found. One possible explanation is that as *manhaji* is

used more commonly in speech and *jeunggaha* appears frequently in writing (such as articles or news reports) that learners have more frequent exposure to *manhaji*. As such, learners can benefit from being made aware of how semantically similar verbs in Korean are used differently depending on the modality. Distinctive collexeme analysis allows the analyst to quantitatively identify such variation between L1 and L2 language as discussed in this section.

In this section, I have covered the results from the distinctive collexeme analysis for three varieties: L1 Korean, L1 English L2 Korean, and L1 Japanese L2 Korean. I have discussed how the usage of the (non)progressive appears to vary across varieties, highlighting both key similarities and differences as it pertains to potential benefits for language teachers and materials developers. Additionally, despite this study being an exploratory corpus study in nature, I have offered some discussion of potential interlanguage transfer effects which may be impacting the usage of the progressive in the learner varieties, and highlighted how even learners who speak a language which is typologically dissimilar from Korean exhibit the ability to acquire and using distinctly Korean constructions (e.g., L1 English speakers usage of the progressive with the verb *know*). In what follows, I investigate the usage of the progressive in each variety using regression analysis to see if any identified predictors, or their interactions, appear to have an impact on the choice to use a (non)progressive.

Figure 4.4

Normalized coll.strengths in L1 Korean data.

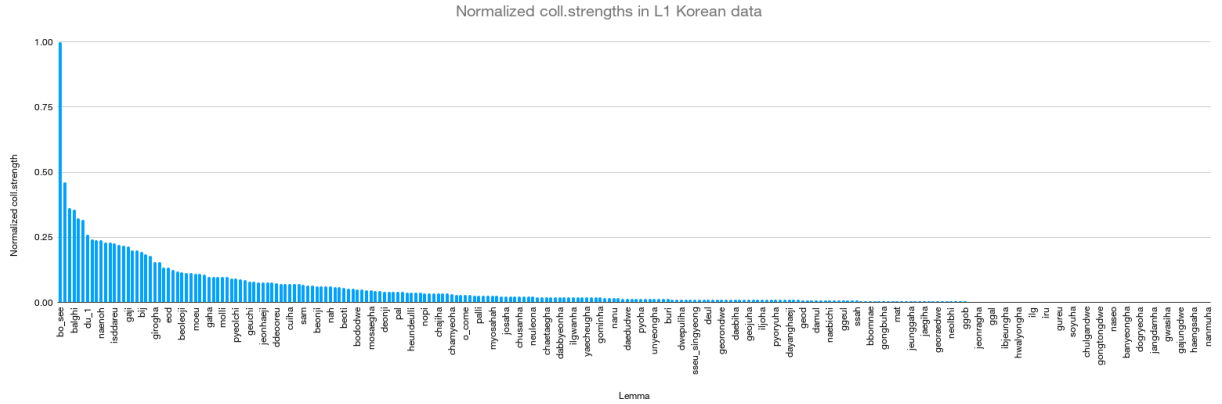


Figure 4.5

Normalized coll.strengths in L1 English L2 Korean data.

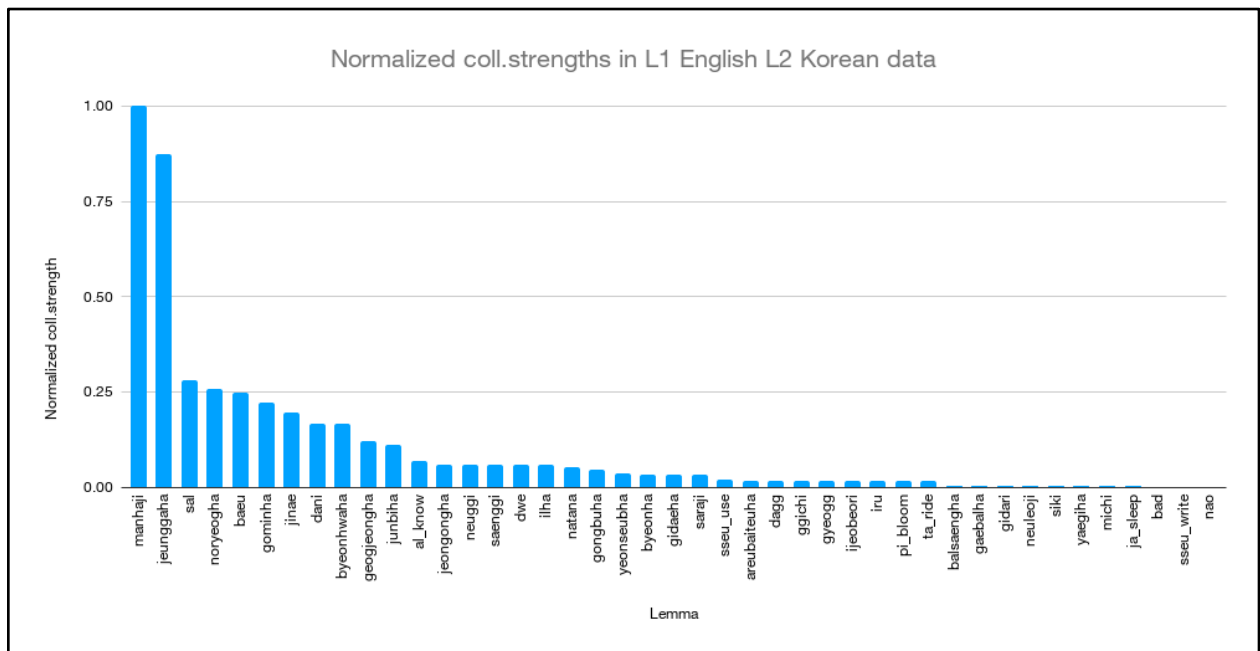
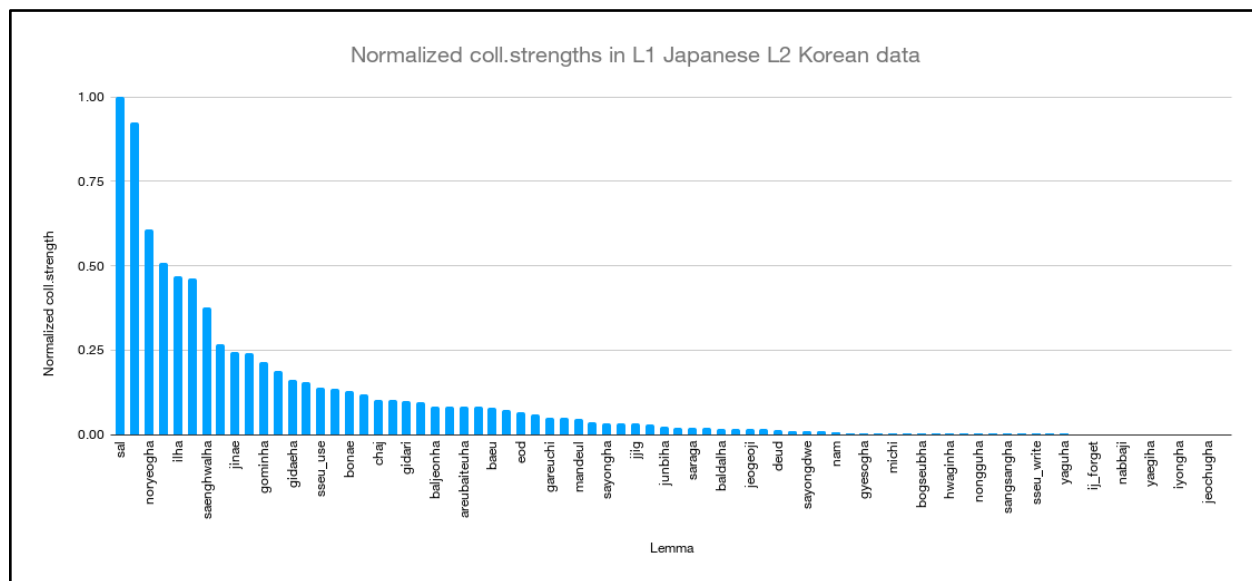


Figure 4.6

Normalized coll.strengths in L1 Japanese L2 Korean data.



4.3. Research question 2

4.3.1. Textbook analysis of *-ko iss*

The number of verbs appearing in the progressive *-ko iss* construction increased as the textbook level increased, as shown in Table 4.2, likely due to the increasing length and complexity of the texts featured in each series as the level progressed. In terms of the most frequent verbs featured in each series, there was some overlap across textbook series. For example, included in the top five most frequent verbs were *sal* (live), the most frequent verb used with the progressive *-ko iss* in both textbook series. Notably, the stative verb *al* (know) was also included in the top verbs of both textbook series (ranked #4 in the *New Sogang Korean* series and #2 in the *KLEAR Integrated Korean* series). In addition to *al* (know), the stative verb *gaji* (have/hold) was included in the *Integrated Korean* series as the fifth most frequently featured verb in the progressive overall.

Table 4.2.

Raw frequency of the -ko iss construction in L2 Korean textbooks.

	Level 1	Level 2	Level 3	Level 4	Total
Textbook 1	0	23	62	100	185
Textbook 2	18	25	99	147	289

Table 4.3.

Raw frequency and proportions of verbs appearing in the progressive -ko iss construction in New Sogang Korean and KLEAR Integrated Korean by level.

	New Sogang Korean		KLEAR Integrated Korean	
	Verb	Frequency (%)	Verb	Frequency (%)
Level 1	—	—	<i>ha-</i> ‘do’	4 (22.22)
	—	—	<i>ta-</i> ‘ride’	3 (16.67)
	—	—	<i>baeu-</i> ‘learn’	1 (5.56)
	—	—	<i>dani-</i> ‘attend’	1 (5.56)
	—	—	<i>deud-</i> ‘listen’	1 (5.56)
	—	—	Others	8 (44.44)
	—	—	Total	18 (100)
Level 2	<i>baeu-</i> ‘learn’	2 (8.70)	<i>sal-</i> ‘live’	3 (12)
	<i>junbiha-</i> ‘prepare	2 (8.70)	<i>chaj-</i> ‘find’	2 (8)
	<i>sal-</i> ‘live	2 (8.70)	<i>jinae-</i> ‘spend/pass’	2 (8)
	<i>yaegiha-</i> ‘talk’	2 (8.70)	<i>gaj-</i> ‘have/hold’	2 (8)

Table 4.3 (cont'd).

	<i>bo-</i> ‘see’	1 (4.35)	<i>al-</i> ‘know’	1 (4)
	Others	14 (60.87)	Others	15 (60)
	Total	23 (100)	Total	25 (100)
Level 3	<i>chaj-</i> ‘find’	7 (11.29)	<i>sal-</i> ‘live’	11 (11.11)
	<i>baeu-</i> ‘learn’	5 (8.06)	<i>al-</i> ‘know’	5 (5.05)
	<i>gidari-</i> ‘wait	5 (8.06)	<i>jinae-</i> ‘spend/pass’	4 (4.04)
	<i>sal-</i> ‘live’	5 (8.06)	<i>ga-</i> ‘go’	3 (3.03)
	<i>ul-</i> ‘cry’	3 (4.84)	<i>ggeul-</i> ‘pull/draw in’	3 (3.03)
	Others	37 (59.68)	Others	73 (73.74)
	Total	62 (100)	Total	99 (100)
Level 4	<i>sal-</i> ‘live’	7 (7.00)	<i>sal-</i> ‘live’	7 (4.76)
	<i>al-</i> ‘know’	4 (4.00)	<i>gaji-</i> ‘have/hold’	7 (4.76)
	<i>iyagiha-</i> ‘talk’	4 (4.00)	<i>ga-</i> ‘go’	6 (4.08)
	<i>bo-</i> ‘see’	3 (3.00)	<i>jui-</i> ‘take control’*	5 (3.40)
	<i>gaji-</i> ‘have/hold’	3 (3.00)	<i>jaesiha-</i> ‘suggest’	5 (3.40)
	Others	79 (79.00)	Others	117 (79.50)
	Total	100 (100)	Total	147 (100)

*쥐다 can also mean *have/hold/squeeze*, however, its use in Textbook 2-Level 4 related to taking control of economic power or finances (e.g., *한국의 가정에서 경제권을 쥐고 있는 사람이 누구일까요?* In Korean homes, who is the one who is has/takes control over finances?)

Table 4.4.

Five most frequent verbs used with the progressive -ko iss construction across all volumes of New Sogang Korean and KLEAR Integrated Korean

	New Sogang Korean		KLEAR Integrated Korean	
	Verb	Frequency # (%)	Verb	Frequency # (%)
1	sal- ‘live’	14 (7.67)	sal- ‘live’	22 (7.61)
2	chaj- ‘find’	9 (4.86)	al- ‘know’	11 (3.80)
3	baeu- ‘learn’	8 (4.32)	ga- ‘go’	9 (3.11)
4	gidari- ‘wait’	8 (4.32)	gaji- ‘have/hold’	9 (3.11)
5	al- ‘know’	6 (3.24)	ha- ‘to do’	8 (2.77)

4.3.2. Teaching of the progressive -ko iss in the textbooks

In addition to determining the frequencies of verbs appearing in the progressive across textbook levels, I also qualitatively explored how the progressive is taught and introduced in each series.

In both series, the progressive is introduced early on. In the *New Sogang* textbook series the progressive -ko iss is introduced in volume 2A, the third volume. In *Integrated Korean*, the construction is introduced in Beginner 2 (the second volume). In both series, the construction is only explicitly taught with verbs that denote physical actions, such as *bo* (watch), *deud* (listen), *cheongsoha* (clean), *drink* (masi), and *make* (mandeul), among others. Following the introduction of the progressive, both texts incorporate short dialogues that demonstrate the usage of the progressive. Below are short excerpts from both *New Sogang Korean* and *Integrated Korean*, respectively:

New Sogang Korean:

Excerpt from 2 과 말하기 대화 1 (Lesson 2 Speaking Dialogue 1): 전화를 받을 수 없는

이유 설명하기 – Explaining the reason you cannot take a phone call.

제니: 앤디 씨, 지금 통화할 수 있으세요?

Jenny: Andy, can you call now?

앤디: 미안해요. 제가 지금 친구하고 얘기하고 있어요.

Andy: Sorry. I'm talking with my friend now.

KLEAR Integrated Korean:

Excerpt from *Integrated Korean: Beginning Two*: Conversation 1 (차 한 잔 마실래요? –

Shall we have a cup of tea?)

유진: 어, 민지 씨 아니세요? 뭐 하세요?

Yujin: Hey, aren't you Minji? What are you doing?

민지: 차 마시고 있어요. 우진 씨도 차 한 잔 하실래요?

Minji: I'm drinking tea. Yujin, would you like a cup of tea, too?

유진: 네, 저도 마시고 싶었는데 잘 됐네요.

Yujin: Yes, I also wanted to drink one so this worked out well.

In both textbooks, the *-ko iss* construction is introduced in English as a construction used to denote a continuous action or an action in progress. In the *New Sogang Korean* series the construction is defined as follows:

Meaning: ‘-고 있다’ is used to express actions in progress or repeated actions. It has the same meaning as “to be doing (something)”.

Form: '고 있다' is always attached directly to the verb stem.'

Likewise, in *Integrated Korean*, the construction is introduced as the following:

~고 있다 expresses the continuation or progression of an action. Only verbs (not adjectives) can occur in this construction.

Of note in both textbooks is that the discussion of the usage of *-ko iss* outside of the 'action in progress' senses are not explicitly discussed. Usage of the progressive with stative and mental verbs comes in later volumes and is incorporated in readings and dialogues as it is used in daily conversation or written texts, though at low frequencies. For example, *New Sogang Korean* includes *al* (*know*) in the progressive when a character in a dialogue is talking to their friend whose dream to become a news anchor came true (from *New Sogang Korean*, volume 4B):

Korean: 나는 네가 유명한 앵커가 될 거라는 것을 10 년 전에 미리 알고 있었어.

English: I **knew** even ten years ago that you would become a famous news anchor.

As can be seen from the example from *New Sogang Korean*, *al* (*know*) and its use with the progressive *-ko iss* is represented. As the form is common in spoken Korean, the textbook incorporates it in the informal spoken form.

Examples of stative verbs in their written forms are also represented, for example, the verb *gaji* (*have/hold*) appears in the textbooks as a mental and stative verb to describe having or holding a meaning. Of note here, as well, is that we see the *-ko iss* with a stative progressive being used with an inanimate subject. Example below is taken from *Integrated Korean: High Intermediate II*:

Korean: ‘바쁜 사람들도, 굳센 사람들도, 바람과 같던 사람들도, 집에 돌아오면 아버지가 된다.’

‘연예인 이름보다 꽃 이름을 더 많이 아는 아이로 키우고 싶습니다.’

이러한 광고 카피에는 소비자들의 많은 호응과 칭찬이 이어졌다. 이렇듯 현대의 아파트는 우리가 사는 주거 공간 이상의 의미를 가지고 있다.

English: "The busy, the strong, like the wind, when they return home, they become fathers."

"I want to raise him to know more about flower names than celebrities."

This advertising copy was followed by a lot of response and praise from consumers. As such, modern apartments **have** more meaning than just the residential space we live in.

In the example above, *gaji* is used to express how apartments have or hold (sentimental) meaning in that they are important places for people and families to grow up in, with *gaji* (*have/hold*) being marked with the progressive *-ko iss*. This article in the textbook was describing apartments in Korea, and how Korean people generally prefer to live in apartments, and how apartments are advertised. Notably, this is a prime example of the textbooks using the *-ko iss* construction in a way exhibited in Korean, with an inanimate subject and a stative progressive.

Given the above, this analysis shows how *-ko iss* is used with a limited variety of verbs overall. It also shows that, despite each series only teaching the prototypical usage of the progressive explicitly, in later volumes of each series, *-ko iss* is incorporated with certain mental verbs such as *al* (*know*). That being said, the frequency analysis reveals that, overall, the progressive's usage in the textbooks is relatively low, with no more than a few hundred examples over the course of two textbook series and sixteen volumes. The progressive is used in a variety

of genres in the textbooks, though, including conversational and casual dialogue and articles, offering learners the opportunity to notice the form *-ko iss* with semantic meanings and usage cases beyond the prototypical ‘action in progress’ usage.

As far as grammar description goes, the textbooks appear to lack explicit instruction on the form-function mappings of the *-ko iss* construction as it is used beyond simply describing actions in progress and with stative and mental verbs. Including (i) more examples of the progressive in such usage cases and (ii) offering a description of how *-ko iss* can co-occur with stative and mental verbs, and be used with inanimate subjects, could help learners acquire and use the form in a way that is more in line with L1 speakers of Korean. As the prototypical *-ko iss* description appears early in the textbooks, textbook developers can incorporate descriptions of *-ko iss* beyond ‘action in progress’ starting from the intermediate textbook series and include explicit descriptions of how it is used in both spoken and written language.

4.3.3. Comparing L1 and L2 corpora with textbooks

To compare verbs appearing in the corpora (L1 and L2) and textbooks, I normalized the frequencies of key verbs in each to better compare general trends in the usage of each verb type based on variety. Figures 4.7, 4.8, and 4.9 visually represent the normalized frequencies of a few key verbs. First, turning attention to the most common verb associated with the progressive in the textbook data, *sal (live)*, it is clear that learners use the verb in the progressive in their writing much more than the verb appeared in the L1 corpus with the progressive. L1 English speakers used it about 5.5 times more frequently, and L1 Japanese speakers used it about 7.4 times as frequently. As Japanese allows for the verb of the same meaning (*sumu, to live*) to be used commonly with the progressive, it is not surprising that the Japanese data presents the

progressive construction being used commonly with *sal* (*to live*) in Korean. Also of note is that the verb is similarly frequent in both textbook series.

Moving on to the stative verbs that appeared in the top five most frequently used verbs in the textbooks, I explored *gaji* (*have/hold*) and *al* (*know*). *Gaji* can be used to express holding or having physical objects, but it can also be used to express having or holding opinions, thoughts, or feelings, and this more abstract usage might be difficult for learners to understand and acquire. However, the normalized frequencies suggest that, indeed, both varieties of learner language studied here exhibit a usage of this verb in the progress construction, with their normalized frequencies coming in at higher than the normalized frequencies of the verb in the L1 corpus. English learners of Korean used *gaji* with the progressive construction 4.5 times more, and the Japanese learners of Korean used *gaji* with the progressive construction 5.4 times more frequently than it was used in the L1 corpus. Again, the progressive usage with a stative verb is higher in the L1 Japanese data, perhaps as expected due to the typological similarities between Japanese and Korean.

Of important note is *al* (*know*), which is a mental and stative verb which commonly takes the progressive construction in Korean. *Al* was ranked fifth in *New Sogang Korean* series, and second in *KLEAR Integrated Korean* series in terms of its rate of co-occurrence with the progressive. *Al* is notable, as it is a prime example of how textbook frequencies differ from real-world usage frequencies. Figure 4.9 shows the normalized frequencies of *al*, it is obvious that its usage in the L1 corpus surpasses its usage in both textbook series, with *al* appearing 2.6 times less frequently in *New Sogang Korean*, and 2.1 times less frequently in *KLEAR Integrated Korean*, suggesting an underuse of the construction in the textbooks when compared with real-world corpus data. L1 English learners of L2 Korean exhibited a perhaps surprisingly high rate

of usage of *al* with the progressive, though their usage rate was still about 1.5 times less than the L1 data. However, given that English rarely, if ever, allows for the verb *know* to appear in the English progressive *be... ing* construction, this is a positive finding for English speaking learners of Korean. For L1 Japanese learners, the trends are different, as their usage of the progressive with *al* is about 1.4 times higher than the L1 data, 2.14 times higher than their L1 English counterparts, and 3.6 times higher than *New Sogang Korean* and 2.6 times higher than *KLEAR Integrated Korean*. As Japanese allows for stative progressives, and in particular, allows for the Japanese verb *know* (*shiru*) to co-occur frequently with the Japanese progressive construction (*-te iru*), there is evidence that even in the case of relatively sparse representation in the textbook data, a learner's L1 will play a role in their target-like usage of a construction. In this case in particular, that is borne out with the comparisons of the L1 English group underuse the verb with the progressive, and the L1 Japanese group using the construction more frequently. Overall, this shows a clear need for textbooks to be designed with their target audience in mind, for example, a textbook geared towards L1 English speakers may need more examples of *al* with *-ko iss* to help learners notice the form.

Figure 4.7.

Relative frequency of lemma sal (live) with -ko iss across corpora.

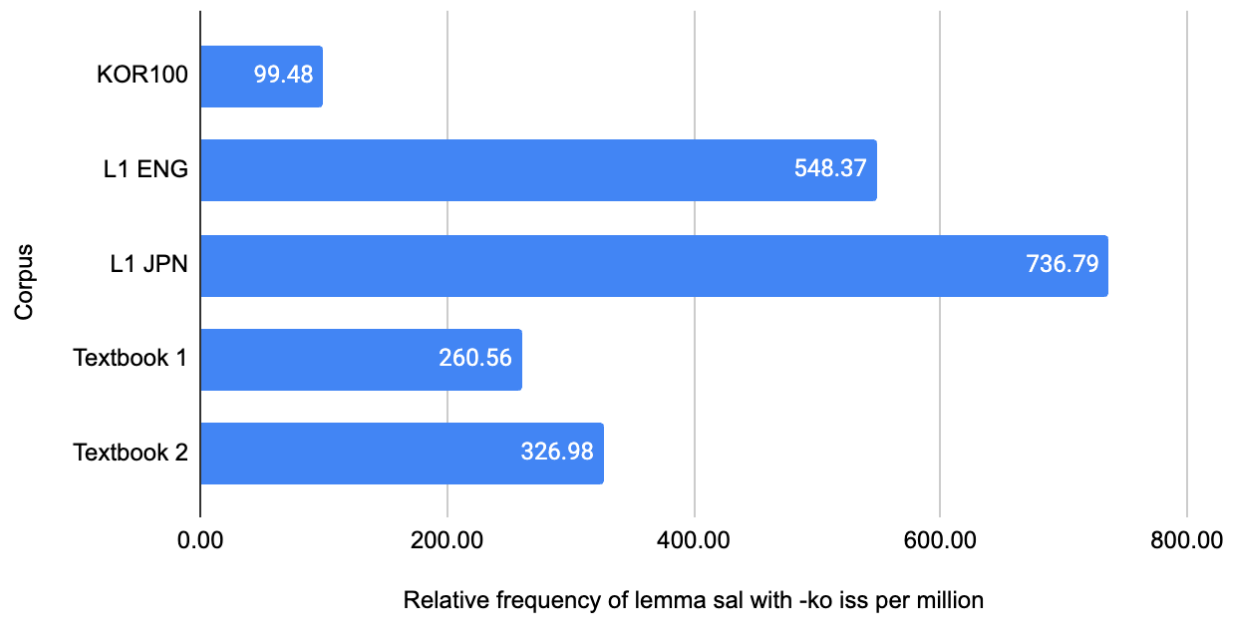


Figure 4.8.

Relative frequency of lemma gaji (live) with -ko iss across corpora.

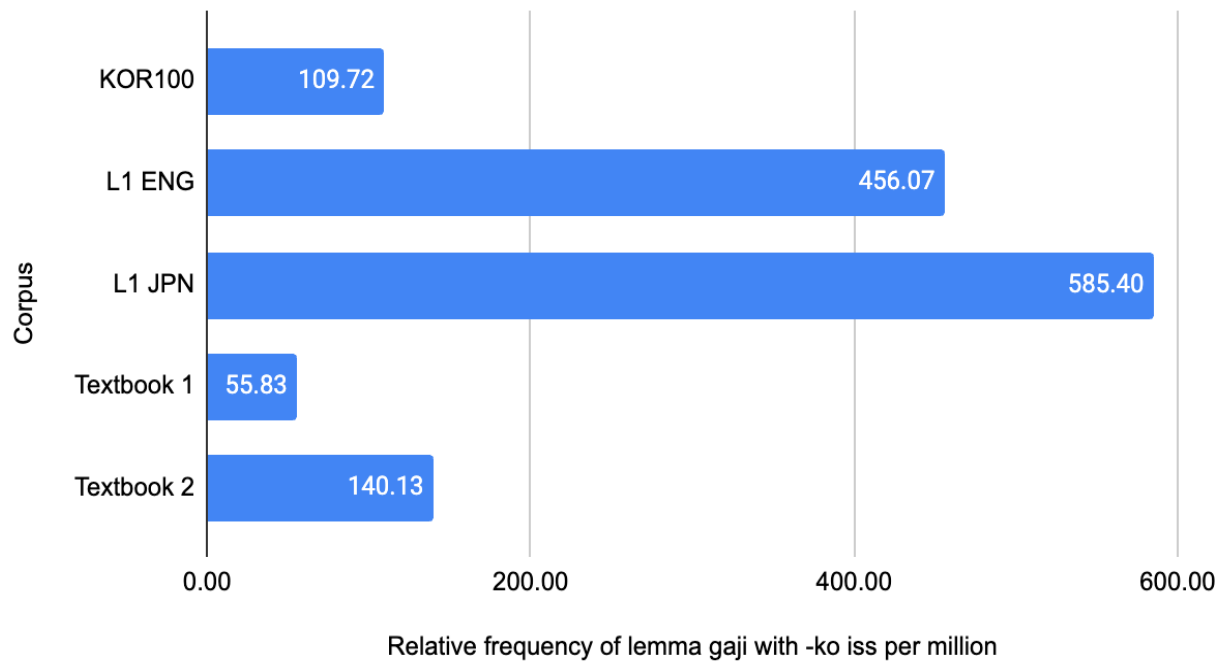
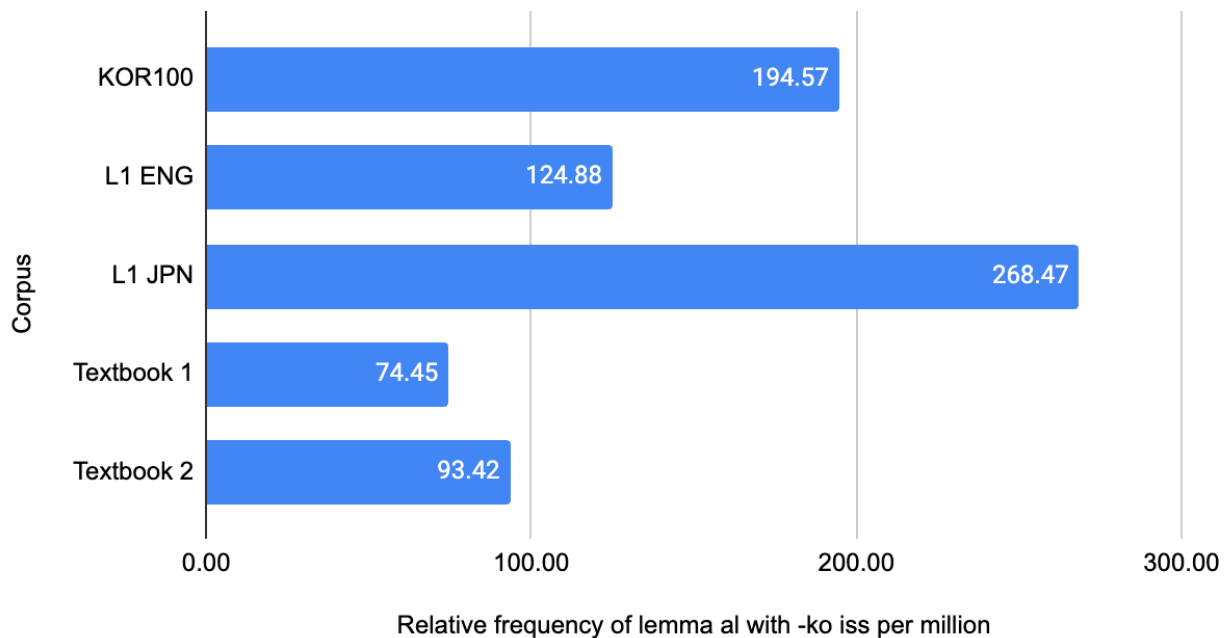


Figure 4.9.

Relative frequency of lemma al (know) with -ko iss across corpora.



4.4. Research question 3

To assess what explanatory variables may influence the choice of a progressive *-ko iss* or a non-progressive in the L1 and L2 data, I chose to run a binary logistic regression in JASP (2024) version 0.18.3. Logistic regression is used when the dependent variable consists of two possible outcomes. In this case, whether the construction choice is (a) the progressive *-ko iss* or (b) the non-progressive. In logistic regression, the predictors (explanatory variables) are categorical or scale (Brezina, 2018). Logistic regression can be of particular use in corpus linguistics as the method itself does not assume a linear relationship between the dependent and independent variables, nor do they need to be normally distributed or of equal variance within each group. Likewise, the residuals do not need to be normally distributed. However, the dependent variable must be dichotomous, and the categories for the dependent variable must be exhaustive in that

every case submitted to the binary logistic regression must be a member of only one group (in other words, either *progressive* or *non-progressive*). Multicollinearity must be checked to ensure that no predictor variables are highly correlated with each other, and this is assessed by checking the VIF, or Variance Inflation Factor³, multicollinearity diagnostics in JASP. As this is an exploratory analysis, entry method is used.

Approximately 500 random samples from each variety (L1 Korean, L1 English L2 Korean, L1 Japanese L2 Korean) were extracted from the overall dataset for manual annotation of predictors, totaling 1523 extractions. Random sampling was done using the randomize range feature in Google Sheets. In addition to variety and construction, the data were manually annotated for aktionsart (four levels: activity/process, accomplishment, achievement, stative), semantic domain (seven levels: activity, communication, mental, causative, occurrence, existence, aspectual), and animacy of the subject (three levels: animate, human, inanimate). In the present study, I did not include speaker/author as a fixed effect because in the L1 corpus that information is not known, though I acknowledge that the inclusion of such fixed effects can help in building the best model. In total, I completed approximately 4500 manual annotations.

According to Brezina (2018), block entry is “usually preferable” in corpus studies employing logistic regression as the predictor variables to include have been decided based on literature or theory (p. 123). As the predictor variables for the present study were chosen based on existing literature, I employed the block entry method when running the regression in JASP.

³ Variance Inflation Factor (VIF) is a statistic used to check for multicollinearity between predictor variables. Generally, VIFs larger than 10 are considered as a warning sign of multicollinearity issues. See <https://online.stat.psu.edu/stat462/node/180/> for a discussion on VIFs.

4.4.1. Inter-rater reliability

Approximately 10% of the annotations (about 450) were separately annotated by an L1 speaker of Korean for inter-rater reliability. The rater hired for this study, at the time of participation, held an advanced degree in linguistics and language education from a Korean university and was teaching Korean language courses in the North American University context. The rater was reimbursed at a rate of \$20 USD per hour.

Reliability statistics were calculated using JASP, with the reliability function installed. Cohen's kappa was calculated for each explanatory variable separately, and the output was interpreted considering guidelines for interpreting Cohen's kappa put forward by Landis and Koch (1977). For *aktionsart*, Cohen's kappa was .68, showing substantial agreement. For *animacy*, Cohen's kappa was .79, showing substantial agreement. For *semantic domain*, Cohen's kappa was .69, showing substantial agreement. *Variety* and *construction* did not require inter-rater agreement statistics. Given the adequate inter-rater reliability statistics between both raters, the data was further analyzed. Due to time constraints, the hired annotator and I were unable to discuss and re-annotate for areas of disagreement. For the present analysis, my annotations are used.

4.4.2. Results

Overview summary: As this dissertation is exploratory with the goal of identifying what factors may contribute to the choice to use a progressive in L1 and L2 Korean, I ran three models, with the goal of improving the explanatory power of the model each time. I will briefly summarize each model: For the first model, *model 1*, each predictor variable was added to the first model without interactions (block entry was used for all models). To investigate whether the influence of the L1 of the writer combined with other predictor variables (*aktionsart*, *semantic domain*,

animacy) influence the choice to use a progressive or non-progressive, I introduced interactions between variety and aktionsart, variety and semantic domain, and variety and animacy in *model 2*. While this model did show some statistically significant interactions, I found it had issues with Standard Errors larger than the estimates for some interactions between semantic domain and variety. Following Brezina's guidelines which warn that Standard Errors larger than the estimates suggest something is wrong with the model, I then explored the predictor semantic domain to identify the cause using contingency tables. I identified that extremely low rates of the aspectual, causative, and communication levels of semantic domain appeared to be causing this issue. I then ran a third model sans those levels of semantic domain, which remedied the issue with Standard Error. This model, *model 3*, will be used to discuss potential interactions that may influence the choice of the progressive in L1 and L2 Korean writing.

4.4.3. Logistic regression – first model

Multicollinearity was assessed for each of the explanatory variables by calculating the Variance Inflation Factor (VIF) for each explanatory variable in the logistic regression. Generally, a VIF of greater than 10 indicates multicollinearity. No explanatory variable had a VIF of greater than 10 (semantic domain: 3.36; animacy: 2.17; aktionsart: 2.08; variety: 1.58). I also used a Confusion Matrix as a performance diagnostic in JASP to assess the overall accuracy of the model's predictions. The Confusion Matrix yielded an Overall Correct Prediction Rate of 69.96%, indicating that the model is correctly predicting about seventy percent of the time.

Model 1 was statistically significant ($p = .001$) with Nagelkerke R^2 of 0.241 (Nagelkerke effect size computed between 0 and 1), so about 24.1% of the variance in the dependent variable can be explained by the model. Summaries of *model 1* are included in Table 4.1 and Table 4.2.

Several levels of semantic domain were significant predictors in the model, (aspectual: $OR = .18, p < .001$; communication: $OR = .54, p < .008$; existence: $OR = .31, p < .001$; mental: $OR = .18, p < .001$). Verbs in these categories appear more likely to be used in the non-progressive, or show a dispreference for the progressive *-ko iss*. The factor of animacy was also found to be statistically significant on the level of human ($OR = .22, p < .001$).

For aktionsart verb categorizations, statistically significant results were found for achievement verbs ($OR = 0.518; p = .007$), activity verbs ($OR = 1.92; p = .003$), and stative verbs ($OR = 1.829; p = .009$). These results show that when a verb falls into the activity or stative aktionsart category, it is likely to trigger use of *-ko iss* (progressive) in Korean. Likewise, when a verb falls into the achievement category, it is less likely to trigger a progressive.

Table 4.1

Model 1 summary.

Model Summary

Model	Deviance	AIC	BIC	df	X ²	p	Nagelkerke R ²
H ₀	2103.720	2105.720	2111.045	1517			
H ₁	1801.074	1829.074	1903.626	1504	302.646	< .001	0.241

Table 4.2.

Results for model 1.

Coefficients

	Estimate	Standard Error	Odds Ratio	z	Wald Test			95% Confidence interval (odds ratio scale)	
					Wald Statistic	df	p	Lower bound	Upper bound
(Intercept)	0.841	0.273	2.319	3.080	9.485	1	0.002	1.358	3.961
aktionsart (achievement)	-0.657	0.242	0.518	-2.716	7.378	1	0.007	0.323	0.833
aktionsart (activity)	0.656	0.218	1.928	3.013	9.076	1	0.003	1.258	2.954
aktionsart (stative)	0.604	0.232	1.829	2.605	6.787	1	0.009	1.161	2.881
animacy (human)	-1.516	0.232	0.220	-6.539	42.759	1	< .001	0.139	0.346
animacy (inanimate)	-0.104	0.236	0.901	-0.440	0.193	1	0.660	0.568	1.431
semantic_domain (aspectual)	-1.700	0.398	0.183	-4.270	18.237	1	< .001	0.084	0.399
semantic_domain (causative)	0.666	0.563	1.946	1.183	1.400	1	0.237	0.646	5.866
semantic_domain (communication)	-0.615	0.230	0.541	-2.669	7.125	1	0.008	0.344	0.849
semantic_domain (existence)	-1.176	0.330	0.309	-3.559	12.670	1	< .001	0.161	0.590
semantic_domain (mental)	-1.495	0.181	0.224	-8.266	68.325	1	< .001	0.157	0.320
semantic_domain (occurrence)	-0.022	0.207	0.978	-0.107	0.011	1	0.915	0.653	1.466
variety (L2_ENG)	0.517	0.169	1.677	3.063	9.382	1	0.002	1.205	2.334
variety (L2_JPN)	0.565	0.164	1.759	3.442	11.850	1	< .001	1.275	2.427

Note. progressive level 'yes' coded as class 1.

4.4.4. Logistic regression – second model

To explore potential interaction effects, a second logistic regression, *model 2*, was run.

Theoretically, it is assumed that variation could exist between L1 Korean and L2 Korean-English and Korean-Japanese varieties due to typological differences. As such, in the second logistic regression, I used JASP to explore interactions between the predictors and variety (i.e., L1 and Learner Language). The summary and results for *model 2* are listed in Table 4.3 and Table 4.4.

Table 4.3

Model 2 summary.

Model	Deviance	AIC	BIC	df	X ²	p	Nagelkerke R ²
H ₀	2103.720	2105.720	2111.045	1517			
H ₁	1692.669	1764.669	1956.375	1482	411.051	< .001	0.316

Table 4.4*Results for model 2.***Coefficients**

	Estimate	SE	Odds Ratio	z	Wald Test			95% Confidence interval	
					Wald Statistic	df	p	Lower bound	Upper bound
(Intercept)	1.150	0.353	3.158	3.255	10.593	1	0.001	0.457	1.842
aktionsart (achievement)	-0.488	0.323	0.614	-1.510	2.280	1	0.131	-1.122	0.146
aktionsart (activity)	0.128	0.312	1.137	0.411	0.169	1	0.681	-0.484	0.740
aktionsart (stative)	0.066	0.315	1.069	0.211	0.045	1	0.833	-0.550	0.683
animacy (human)	-1.075	0.283	0.341	-3.791	14.375	1	< .001	-1.630	-0.519
animacy (inanimate)	-0.087	0.277	0.917	-0.314	0.098	1	0.754	-0.630	0.456
semantic_domain (aspectual)	-1.877	0.449	0.153	-4.182	17.491	1	< .001	-2.757	-0.998
semantic_domain (causative)	0.855	0.681	2.351	1.255	1.575	1	0.210	-0.480	2.190
semantic_domain (communication)	-1.089	0.312	0.337	-3.493	12.199	1	< .001	-1.700	-0.478
semantic_domain (existence)	-1.119	0.440	0.327	-2.540	6.452	1	0.011	-1.982	-0.255
semantic_domain (mental)	-0.748	0.344	0.473	-2.175	4.730	1	0.030	-1.423	-0.074
semantic_domain (occurrence)	-0.866	0.314	0.421	-2.753	7.579	1	0.006	-1.482	-0.249
variety (L2_ENG)	0.779	1.067	2.179	0.730	0.533	1	0.465	-1.313	2.871
variety (L2_JPN)	-0.288	0.750	0.750	-0.384	0.147	1	0.701	-1.759	1.182
semantic_domain (aspectual) *	14.416	1455.398	1.824×10 ⁺⁶	0.010	9.812×10 ⁻⁵	1	0.992	-2838.111	2866.944
variety (L2_ENG)									
semantic_domain (causative) *	-2.958	1.408	0.052	-2.100	4.411	1	0.036	-5.717	-0.198
variety (L2_ENG)									
semantic_domain (communication) *	16.083	467.663	9.653×10 ⁺⁶	0.034	0.001	1	0.973	-900.520	932.686
variety (L2_ENG)									
semantic_domain (existence) *	-1.767	1.067	0.171	-1.656	2.741	1	0.098	-3.859	0.325
variety (L2_ENG)									

Table 4.4 (cont'd).

semantic_domain (mental) * variety (L2_ENG)	-1.061	0.456	0.346	-2.326	5.412	1	0.020	-1.955	-0.167
semantic_domain (occurrence) * variety (L2_ENG)	1.411	0.588	4.100	2.401	5.766	1	0.016	0.259	2.563
semantic_domain (aspectual) * variety (L2_JPN)	15.761	1026.010	6.998×10 ⁺⁶	0.015	2.360×10 ⁻⁴	1	0.988	-1995.182	2026.704
semantic_domain (causative) * variety (L2_JPN)	12.865	1026.476	386726.757	0.013	1.571×10 ⁻⁴	1	0.990	-1998.991	2024.721
semantic_domain (communication) * variety (L2_JPN)	0.450	0.541	1.568	0.833	0.693	1	0.405	-0.609	1.510
semantic_domain (existence) * variety (L2_JPN)	1.277	1.013	3.585	1.260	1.589	1	0.208	-0.709	3.262
semantic_domain (mental) * variety (L2_JPN)	-1.254	0.506	0.285	-2.479	6.143	1	0.013	-2.245	-0.262
semantic_domain (occurrence) * variety (L2_JPN)	2.074	0.582	7.956	3.566	12.719	1	< .001	0.934	3.214
aktionsart (achievement) * variety (L2_ENG)	-1.372	0.750	0.254	-1.830	3.350	1	0.067	-2.842	0.097
aktionsart (activity) * variety (L2_ENG)	0.545	0.646	1.725	0.845	0.714	1	0.398	-0.720	1.811
aktionsart (stative) * variety (L2_ENG)	1.032	0.671	2.805	1.538	2.365	1	0.124	-0.283	2.346
aktionsart (achievement) * variety (L2_JPN)	0.044	0.627	1.045	0.070	0.005	1	0.944	-1.184	1.272
aktionsart (activity) * variety (L2_JPN)	0.965	0.528	2.624	1.827	3.337	1	0.068	-0.070	2.000
aktionsart (stative) * variety (L2_JPN)	0.907	0.588	2.476	1.541	2.375	1	0.123	-0.246	2.060
animacy (human) * variety (L2_ENG)	-1.178	0.962	0.308	-1.224	1.499	1	0.221	-3.063	0.708
animacy (inanimate) * variety (L2_ENG)	-0.008	0.974	0.992	-0.008	6.289×10 ⁻⁵	1	0.994	-1.916	1.901
animacy (human) * variety (L2_JPN)	-0.203	0.639	0.816	-0.318	0.101	1	0.750	-1.455	1.049
animacy (inanimate) * variety (L2_JPN)	-0.130	0.717	0.878	-0.181	0.033	1	0.856	-1.534	1.275

Note. progressive level 'yes' coded as class 1.

However, a look at the model summary reveals some critical issues with *model 2*. While the model does show some statistically significant interactions, the model ultimately suffered from high Standard Errors for several interactions and several confidence intervals spanning 1, which serve as a warning sign that the *model 2* has some issues. Looking at the output, issues seem to occur when variety interacts with semantic domain on the level of aspectual, causative, and communication.

To address this, I took an exploratory approach using contingency tables in JASP to see the distribution of the annotations for semantic domain, taking care to explore the distribution of annotations in each variety (L1 and L2) separately. This analysis revealed that across L1 and L2 there were lower numbers of exemplars annotated as aspectual, causative, and communication. For example, in the L1 data, only 14 verbs co-occurring with *-ko iss* were annotated as causative, and likewise only ten verbs co-occurring with *-ko iss* were annotated as aspectual. L1 Japanese L2 Korean data followed a similar trend, with only two verbs co-occurring with *-ko iss* annotated with aspectual and causative semantic domains each; only nine verbs received an existence annotation for semantic domain. L1 English L2 Korean data, likewise, had one verb in *-ko iss* for aspectual, two for causative, and one for existence. To attempt to remedy this issue, I removed the levels of aspectual, causative, and communicative from the data and ran a final model to attempt to find potential interactions.

Table 4.5

Contingency table for levels of semantic domain in L1 Korean.

Contingency Tables

Semantic domain	progressive		Total
	no	yes	
activity	42	77	119
aspectual	27	10	37
causative	3	14	17
communication	50	32	82
existence	20	18	38
mental	51	42	93
occurrence	56	57	113
Total	249	250	499

Table 4.6

Contingency table for levels of semantic domain in L1 English L2 Korean.

Contingency Tables

Semantic domain	progressive		Total
	no	yes	
activity	117	129	246
aspectual	0	1	1
causative	2	2	4
communication	0	9	9
existence	8	2	10
mental	111	44	155
occurrence	11	78	89
Total	249	265	514

Table 4.7

Contingency table for levels of semantic domain in L1 Japanese L2 Korean.

Contingency Tables

Semantic domain	progressive		Total
	no	yes	
activity	71	121	192
aspectual	0	2	2
causative	0	2	2
communication	13	13	26
existence	2	9	11
mental	152	39	191
occurrence	11	74	85
Total	249	260	509

4.4.5. Logistic regression – third model

Because the focus of this section was to attempt to identify potential interactions between variety and semantic domain, variety and aktionsart, and variety and animacy, these predictors were put into *model 3* using entry method; interactions were added to the model manually in JASP.

Overall, *model 3* is an improvement over *model 2* and allows for cautious optimism in its interpretation. The *model 3* is statistically significant with a low p -value ($p < .001$) and decent effect size (Nagelkerke $R^2 = .32$). Similar to *model 1*, *model 3* shows that semantic domains on their own seem to predict the use of a non-progressive: existence (OR = .35, $p = .02$), mental (OR = .45, $p = .02$), occurrence (OR = .45, $p = .01$). Interaction effects were found between variety and semantic domain. According to this model, L1 English speakers are less likely to use a progressive when the verb's semantic domain is mental (OR = .37, $p = .03$). L1 Japanese speakers follow a similar pattern in terms of a verb's semantic domain being categorized as mental (OR = .28, $p = .01$). Both L1 English and L1 Japanese learners of L2 Korean appear to be more likely to use a progressive when the semantic domain of the verb is occurrence based on

the interactions between Variety (English)*Semantic domain (occurrence) and Variety (Japanese)*Semantic domain (occurrence) (L1 English: OR = 3.5, $p = .04$; L1 Japanese: OR = 8.00, $p < .001$). L1 English learners of L2 Korean are more less likely to use a progressive when the subject is animate (OR = .09, $p = .05$).

While these results may scratch the surface on what aspects of a verb (phrase) may influence the choice of a progressive or non-progress across L1 and L2 Korean, even *model 3* has some shortcomings which cannot go unstated. First, while *model 3* was able to address shortcomings that plagued *model 2*, such as the high Standard Errors, many Confidence Intervals in *model 3* were found to include 1, which suggests results may not be statistically significant. This highlights the difficulty of modeling corpus data, particularly when incorporating interactions. Going forward, the way to address this would be to (i) include more data from each variety and (ii) reconsider some categories for annotation. This is discussed in more detail in section 5.2.1 (Future directions).

Table 4.8

Model 3 summary.

Model	Deviance	AIC	BIC	df	X ²	p	Nagelkerke R ²
H ₀	1853.543	1855.543	1860.742	1337			
H ₁	1492.387	1546.387	1686.758	1311	361.156	< .001	0.316

Table 4.6*Results for model 3.***Coefficients**

	Estimate	SE	Odds Ratio	z	Wald Test			95% Confidence interval	
					Wald Statistic	df	p	Lower bound	Upper bound
(Intercept)	0.772	0.399	2.165	1.935	3.742	1	0.053	-0.010	1.555
variety (L2_ENG)	1.625	1.302	5.076	1.248	1.558	1	0.212	-0.926	4.176
variety (L2_JPN)	-0.265	0.817	0.767	-0.324	0.105	1	0.746	-1.866	1.336
aktionsart (achievement)	-0.414	0.378	0.661	-1.097	1.204	1	0.273	-1.155	0.326
aktionsart (activity)	0.406	0.381	1.501	1.067	1.140	1	0.286	-0.340	1.152
aktionsart (stative)	0.171	0.349	1.187	0.491	0.241	1	0.623	-0.513	0.855
semantic_domain (existence)	-1.057	0.452	0.348	-2.337	5.461	1	0.019	-1.943	-0.170
semantic_domain (mental)	-0.804	0.351	0.448	-2.286	5.228	1	0.022	-1.492	-0.115
semantic_domain (occurrence)	-0.809	0.324	0.445	-2.496	6.229	1	0.013	-1.444	-0.174
animacy (human)	-0.660	0.343	0.517	-1.921	3.689	1	0.055	-1.333	0.014
animacy (inanimate)	0.138	0.330	1.148	0.417	0.174	1	0.677	-0.510	0.785
variety (L2_ENG) * aktionsart (achievement)	-1.137	0.805	0.321	-1.412	1.995	1	0.158	-2.714	0.441
variety (L2_JPN) * aktionsart (achievement)	-0.010	0.667	0.990	-0.015	2.338×10 ⁻⁴	1	0.988	-1.318	1.297
variety (L2_ENG) * aktionsart (activity)	0.644	0.713	1.905	0.903	0.816	1	0.366	-0.753	2.042
variety (L2_JPN) * aktionsart (activity)	0.743	0.579	2.103	1.283	1.647	1	0.199	-0.392	1.879
variety (L2_ENG) * aktionsart (stative)	1.256	0.719	3.513	1.746	3.050	1	0.081	-0.154	2.666
variety (L2_JPN) * aktionsart (stative)	0.972	0.624	2.643	1.557	2.423	1	0.120	-0.252	2.195
variety (L2_ENG) * semantic_domain (existence)	-2.036	1.123	0.131	-1.814	3.289	1	0.070	-4.236	0.164
variety (L2_JPN) * semantic_domain (existence)	1.200	1.027	3.319	1.168	1.365	1	0.243	-0.813	3.212
variety (L2_ENG) * semantic_domain (mental)	-1.006	0.463	0.366	-2.174	4.727	1	0.030	-1.912	-0.099
variety (L2_JPN) * semantic_domain (mental)	-1.290	0.525	0.275	-2.459	6.047	1	0.014	-2.319	-0.262
variety (L2_ENG) * semantic_domain (occurrence)	1.253	0.597	3.502	2.101	4.413	1	0.036	0.084	2.422
variety (L2_JPN) * semantic_domain (occurrence)	2.080	0.597	8.003	3.485	12.145	1	<.001	0.910	3.250
variety (L2_ENG) * animacy (human)	-2.406	1.224	0.090	-1.965	3.860	1	0.049	-4.805	-0.006
variety (L2_JPN) * animacy (human)	-0.312	0.702	0.732	-0.444	0.197	1	0.657	-1.688	1.064
variety (L2_ENG) * animacy (inanimate)	-0.928	1.238	0.395	-0.750	0.563	1	0.453	-3.355	1.498
variety (L2_JPN) * animacy (inanimate)	-0.115	0.760	0.892	-0.151	0.023	1	0.880	-1.604	1.375

Note. progressive level 'yes' coded as class 1.

V. DISCUSSION

5.1. Addressing the research questions

This study, which is collostructional, frequency based, and multivariate, provides a view of the progressive construction *-ko iss* in Korean across varieties and textbooks. This study also shows the importance of considering constructions from multiple statistic perspectives (including collostructional analysis and regression modeling) and qualitative perspectives (exploring how a construction is introduced in textbooks in addition to quantifying the verbs with the construction) to develop a robust understanding of how a construction is used in terms of lemmas (dis)associated with it and to address linguistic factors that may influence the choice of one particular construction over another. These methodologies allow me to address three overarching research questions, which are (i) what are the distinctive collexemes for the progressive and non-progressive across L1 and L2 written Korean varieties, (ii) what linguistic factors influence the choice of a progressive or a non-progressive, and (iii) how are Korean language textbooks incorporating the progressive *-ko iss* construction overall and across levels?

Research question (i) addressed verbs and their preference for the progressive (or non-progressive) in terms of association strengths measured in collostructional strength, specifically using the distinctive collexeme analysis method. Overall, the results are promising in that both learner varieties exhibited wide variety in their choice of verbs in the progressive in the writing, much of which fell in-line with the L1 corpus data. Of particular interest was the fact that L1 English learners, despite the English language overall featuring fewer stative progressives than the Korean language does, key verbs which are highly associated with and well attested to co-occur with the progressive *-ko iss* construction. Verbs such as *sal* (*live*), *al* (*know*), and *neuggi* (*feel*), stative verbs which appearing in the progressive in Koreana, were found to be distinctive

collexemes in the L1 English learner data, showing a positive sign for uptake of typologically distinct forms from the learners' L1. In terms of sheer number of distinctive collexemes, L1 Japanese learners did overshadow the L1 English learners, including *sal (live)*, *gaji (have/hold)*, *gidae (hope)*, *mid (believe)*. The L1 Japanese learners using more progressives overall and more progressives that are particularly stative and mental verbs also found to be distinctive collexemes in the L1 data is not surprising as Japanese also allows for stative and mental verbs to take the progressive construction. This finding highlights the fact that potential entrenchment of the form (-*ko iss* progressive construction) with stative and mental verb readings can be a point of issue for learners of Korean whose L1s are typologically distant from Korean in their form-function mapping of progressive constructions and stative/mental meanings.

With research question (ii), I employed logistic regression analysis to look at the big picture of what linguistic factors may influence the choice of a progressive across varieties, to varying degrees of success. I created three models, one without interaction effects (*model 1*), and two including interaction effects (*model 2*; *model 3*), which led me to modify the levels of semantic domain included in the final *model 3*. While *model 3* has some weakness in terms of CIs including 1, as this study is exploratory, I will mention results from both *model 1* and *model 3*, with cautious optimism when discussing *model 3*.

The first model showed that on their own, activity and stative verbs were more likely to predict the usage of the progressive -*ko iss*, whereas achievements showed a trend towards preferring the non-progressive. As some multifactorial studies on the progressive across varieties of English have shown that achievement verbs can trigger a progressive construction in academic writing (see, for example, Rautionaho et al., 2018) I anticipated that L1 English learners of Korean might tend towards using achievement verbs in the progressive in Korean. However, this

was not the case. Considering *model 3* and the interactions found between variety and aktionsart and variety and semantic domain, results suggest learners may use the progressive *-ko iss* more commonly with occurrence verbs in their writing than L1 Korean speakers. Learners also appear to be using mental verbs less often in their writing, which may be in part due to typological differences. However, given that *model 3* suffers from some issues with Confidence Intervals, more data must be collected in order to confirm this apparent trend.

Considering the textbooks featured in this study leads to research question (iii), asking what verbs are most prevalent in the progressive in Korean across textbooks and stratified by level. Overall, the rate of usage of progressives was higher in *KLEAR Integrated Korean* than in *New Sogang Korean*. As textbook length increased, so too did the frequency of the use of the progressive *-ko iss*, though overall, the construction itself was still not used as frequently as expected. Diving into the verbs used in the progressive in the textbook corpora, the verbs used are largely action verbs (e.g., *see, watch, listen, do*). *KLEAR Integrated Korean* does incorporate a key stative verb, *al (know)*, starting from the second level. *New Sogang Korean* incorporates *al (know)* as one of the top five most frequent verbs from level 4. Considering the results of the collostructional analysis and regression in combination with the relatively low frequencies of stative and mental verbs, it seems clear that future iterations of the textbooks could better represent real-world Korean language with more frequent inclusion of stative and mental verbs in the progressive, as well as using more diverse verbs in the progressive, considering distinctive collexemes that were absent in the learner data and not found frequently in the textbooks. It is also suggested that learners with L1 backgrounds that are both typologically similar and dissimilar to Korean could benefit from inclusion of such verbs, particularly as both learner groups have a trend to use mental verbs less frequently (calling back to the logistic regression).

My results support findings from, for example, Jang (2005), who found that the progressive *-ko iss* was largely one note and focused on the ‘action in progress’ meaning of *-ko iss* in the textbooks I surveyed. Further, in light of the combined textbook analysis in tandem with analysis of L1 and L2 data, in conversation with Kim (2014) who looked at *-ko iss* diachronically and noted its increasing usage with stative progressives (and overall similar rates of usage with both ‘action in progress’ and ‘stative/resultative’ meanings), I think it is time for textbooks to teach present *-ko iss* to learners as a construction with several distinct meanings and usages. That is to say, textbooks would better serve learners were they to introduce the ‘action in progress’ *-ko iss* at the lower levels as they do now, and then at the intermediate or advanced level re-introduce *-ko iss* and its multitude of usages, include with stative and mental verbs, its usage in writing to denote changes over time particularly when the subject is inanimate, and the breadth of verbs that can co-occur with the *-ko iss* construction that L1 English speakers might not anticipate. While textbooks are only one source of input, they can have an impact. Recall that Northbrook and Conklin (2019) found that even low-level learners were able to respond faster to a phrasal judgement task when the lexical bundles they encountered matched those in their textbooks. As the authors put it, this “indicates that... students are sensitive to whether items appeared in their books” and thus, “input given to students matters” (p. 828). Bearing this in mind, the argument to incorporate more usages of the *-ko iss* construction with a variety of verbs can have a positive impact on student uptake of the construction. Relating this back to Gabrielatos (2006), clearly, as input from textbooks is effective for students, informing their development using corpora can only aid students in their language learning. By revamping the representation of *-ko iss*, textbooks would be better prepared to serve learners and provide them with accurate descriptions and examples of this complex Korean construction.

5.2. Implications for the field of (Korean) second language acquisition

This study also highlights the merit of using multiple statistical approaches to investigate large corpus data. Specifically, the results from the collocation analysis, the regression, and the frequency analysis in the textbook section revealed how looking at the data from one singular method may cause overgeneralization of certain findings. Take, for example, the verb *al* (*know*). In the collocation analysis, *al* was found to be a distinctive collexeme of the progressive in both the L1 Korean and L1 English L2 Korean varieties, but not in the L1 Japanese L2 Korean variety. This finding is initially surprising, as typological (dis)similarities across languages would suggest the Japanese learner language would be more likely to include mental, stative progressives such as *al* as distinctive collexemes. Further, the mental verb category was also found to cause a dispreference in the learner group. However, when considering the normalized frequencies of *al* across L1, L2, and textbook corpora, it became clear that in actuality *al* was used more (in terms of relative frequency) in the Japanese learner language than any other variety. Were this study to solely consider collocation strengths or the results of the regression analysis the conclusion may have been that Japanese learner language lacks/underuses stative progressives and mental verbs with *-ko iss*. In fact, this triangulation approach suggests a more nuanced result, that while Japanese learners, overall, may be using mental verbs and stative progressives less than we see in the L1 data, when a verb functions similarly in Japanese (*know* in Japanese functioning similarly) they are using the verb at a higher frequency in particular with the construction, potentially due to transfer effects leading to entrenchment of that particular verb. Considering the fact that *al* was not a distinctive collexeme and that mental verbs generally disfavored the progressive in Japanese learner data, this shows that analysts considering interlanguage effects will need to consider, based on the languages in question, if any

particularly linguistic elements will require a deep dive beyond what association strength measures and inferential statistics can provide separately.

5.2.1 Future directions

In this section, I hope to highlight issues that arose during this project and provide insights and suggestions that may benefit future corpus-based projects on Korean. Namely, I will discuss modeling Korean using logistic regression. In sections 4.4 through 4.4.5, I conducted a logistic regression with the intention of determining which predictors may influence the choice of a progressive in L1 and L2 Korean. My annotation scheme was based on previous literature, and particularly, recent corpus studies which focus on English. However, what my analysis has highlighted is that some annotation schemes, particularly semantic domain, need adjusting in a follow-up study. For example, while semantic domains such as *aspectual* and *causative* were included in the present study, they were rarely found during annotation. In retrospect, the *causative* category could have been omitted considering Koreans typology. Causative verbs include *allow* or *permit*, however, in Korean, rather than a singular verb, another construction can be used to express this meaning. Therefore, it is perhaps not surprising that verbs with a *causative* reading were so sparse in the Korean data across all varieties. This highlights the importance of considering Korean's distinct features when selecting annotations. A future study could eliminate *aspectual* and *causative* verbs from the annotation scheme altogether.

Going forward, there is also room to improve the annotation schemes to tease out the nuances in the usage of key stative and mental verbs. In this study, many stative and mental verbs were found to be distinctive collexemes in L1 and L2 varieties of Korean, and usage of such verbs, though limited, appeared in textbooks as well, and a verb being categorized as a stative verb (aktionsart) made it more likely to trigger the use of a progressive according to

model 1 as well. Thus, to make the analysis of stative and mental verbs more robust, Korean-specific aktionsart-esque categories can be introduced to a future multivariate analysis. For example, Lee (2006), in her paper on stative progressives in Korean takes the stance that ‘*know-type*’ verbs, such as *al* (know) are punctual, and emotion verbs such as *sarangha* (love) are durative. Thus, a future analysis could consider whether punctuality or durativity of a stative verb in Korean lend themselves more to the progressive, and under what context. To add to that, some scholars have even suggested that verbs such as *al* (know) may even be accomplishments (Hong, 1991) depending on the event description that led someone to come to know something. Thus, a careful analysis of key stative and even mental verbs in Korean, where more language specific annotations are employed may help tease apart what makes Korean stative and mental verbs so unique in their usage in the *-ko iss* construction. Findings would have clear implications for pedagogy and materials development.

5.3. Limitations

There are several limitations to this study, some of which may serve as a guide for future studies and demonstrate the need for further development of learner corpora of languages other than English. First, a critical limitation of this study is in fact the L1 and L2 corpora that were available for this study. While both L1/L2 corpora were compiled and made available by the National Institute of Korean Language (NIKL) and feature relatively large amounts of data, they are not directly comparable with each other. For example, the NIKL L1 corpus is akin to the British National Corpus (BNC) as the written corpus comprises novels, short stories, newspapers, articles, opinion pieces, etc. On the other hand, the learner essays submitted to NIKL when compiling the corpus ranged in topic, including argumentative essays, opinion essays, and personal narrative essays. However, as these are some of the largest and most widely

available Korean language corpora, they were selected for this study. A prime example of the type of corpus that the field of Korean corpus linguistics is in need of is the International Corpus Network of Asian Learners of English (ICNALE; Ishikawa, 2023). The ICNALE corpus consists of spoken and written data contributed by L1 and L2 speakers of English, so that the language data is controlled for genre, and well documented for a speaker's L1, proficiency, task, and other pertinent background information such as years studying English, all of which allows for robust interlanguage comparisons to be made. As of yet, such a corpus does not exist for the Korean language, and so the findings in the present study must be taken with the understanding that comparisons could shift should a more balanced corpus arise.

In terms of the data analysis itself, due to (i) the massive amount of data to be extracted and (ii) time and resource constraints, this study only focused on exploring the prototypical action in progress *-ko iss* progressive construction in Korean. However, as noted in the literature, there are several ways to express continuous aspect in Korean, including other constructions (such as *neun jung* or *a/eo iss*) which constitute full examinations in papers of their own to create a full understanding of the continuous aspect in L1, L2, and textbook Korean. In terms of the learner data itself, while proficiency was originally intended to be considered as a factor in the regression analysis, unlike ICNALE, the NIKL learner data does not provide verified information on a learner's language proficiency, such as a c-test or Test of Proficiency in Korean (TOPIK) score (ICNALE, in most cases, is able to provide c-test or TOEFL scores). Proficiency information in the NIKL learner corpus is based on level of Korean class (e.g., *level 1*, *level 2*), which as any classroom teacher can attest to, does not necessarily correspond to a learner's actual language proficiency. In existing literature, some corpus linguists have found interactions between explanatory variables such as genre or tense, however, due to time constraints for data

extraction, cleaning, and manual annotation by two raters, tense was not considered in this analysis. A follow up study should consider tense as one explanatory variable in the choice of a (non)progressive across varieties of Korean, keeping in mind the large dataset that this will lead to and the amount of time necessary for manual annotation. Genre can also be considered in future studies, provided a corpus of Korean language is created to be both comparable across speaker varieties and documented so that the genre is known to the analyst. Finally, future studies may be able to incorporate fixed effects into their models, such as speaker, which was unable to be tested in this study due to a lack of speaker information for the L1 data.

Finally, in terms of the textbook corpus, levels 1 through 4 of two of the commonly used textbook series for teaching Korean were included. Future studies could benefit from adding other textbook series to the corpus to compare the usage of the progressive across multiple volumes. Further, while levels 1 through 4 were included in the present study, in fact, both series have more advanced volumes (through 5 and 6). While these are not often used in teaching in the North American context due to each volume generally corresponding to an academic year of study, for Korean programs serving advanced learners who may use those advanced volumes, adding them to the analysis could provide useful information for textbook developers and language teachers.

5.4. Pedagogical Implications

5.4.1. For teachers

In terms of pedagogical implications, I will discuss them here in terms of implications for teachers of Korean and implications for language materials and textbook developers. First of all, it is clear that learner language differs from L1 language in terms of the verbs that are associated with the progressive, as well as the wide variety of semantic domains that those verbs can fall

into. Functionally, learners are limited in terms of their ability to use stative and mental verbs with the progressive in their writing. As such verbs also appear commonly in authentic written texts it is important for teachers to at the very least make learners aware of this form-function connection in the classroom. To facilitate this, genres which learners are interested in, such as manhwa (Korean comics) or clips from Korean shows can be used in lower levels as they will include examples of stative mental verbs in the progressive. Likewise, higher level learners can be exposed to news articles or short stories and novels, and teachers can modify the text complexity to accommodate their learners while maintaining examples including the progressive. Additionally, highlighting the progressive form with stative and mental verbs in class through discussion where learners are required to reproduce the form can help facilitate practice and uptake.

Empowering learners with authentic materials in the classroom has been found to motivate learners at all levels (Bahrani et al., 2014). As learners may be demotivated if the texts are too difficult (Sample, 2015), it is important for teachers to modify authentic materials for intermediate or emerging advanced learners of Korean. One actionable recommendation is for teachers to start by using authentic news articles about topics learners are familiar with. For example, news sites such as Huffington Post Korea often publish articles on topics learners are interested in and familiar with, including Korean pop culture but also extending to celebrities and headlines trending outside of Korea. Teachers can use such articles as a gateway to authentic materials while avoiding issues of topic unfamiliarity.

Additionally, the collocation analysis revealed that certain distinctive collexemes in the learner data may be being overused when compared with the L1 data. For example, *sal* (live) and *manhaji* (increase) were distinctive collexemes in learner data. However, in the L1 written

corpus data, equivalent but more academic terms, namely *geojuha* (*live/reside*) and *jeunggaha* (*increase*) were identified as distinctive collexemes. Learners may rely on simpler terms they have learned early on, and thus these terms are well entrenched in their mental lexicons. Thus, when incorporating authentic materials, Korean language teachers can help learners improve their Korean writing to appear more academic by making these equivalent verb forms salient and helping learners identify when to use each verb type.

5.4.2. For textbook developers and materials designers

For textbook developers, a major implication is that the use of the progressive needs to be more widespread in the textbooks, particularly in terms of mental verb representation as the analysis has found that both learner groups used mental verbs in their writing significantly less than the L1 Korean corpus, and the variety of verbs that learners associated with the progressive as a whole was significantly lower. As textbooks are one main source of input for learners, including such examples at all levels is critical. Lower-level textbooks can incorporate stative/mental verbs in the progressive to dialogues as they are used in spoken language, which learners will practice in the classroom, thus sowing the seeds for them to gain awareness of the form-function mapping and be more inclined to notice and acquire progressives when they appear in texts at the higher level. This is of particular note for textbook series which altogether lack examples of the progressive in the lower-level volumes as was found in this study.

In addition to incorporating more stative and mental verbs even at lower levels, textbooks could aid learners' uptake of the usage of the progressive by incorporating readings which include inanimate subjects of the progressive verb. The logistic regression analysis here revealed that, in L1 Korean writing, the progressive was more likely to be used when the subject was inanimate, and this was commonly seen in the L1 corpus. Further, while it was anticipated that

achievement verbs (which express punctual events) would be used in the progressive by L1 English speakers, this was not borne out in the results. In fact, L1 English speakers were far less likely to use an achievement verb with the progressive than was seen in the L1 corpus data. So, at the very least, incorporating readings in the texts which include achievement verbs that were associated with the progressive in the textbooks (e.g., *natana* - *appear*, *geuchi* - *stop/cease*; *ireugi* when *ireugi* is used with the semantic meaning of *trigger*).

Further, for textbooks specifically, incorporating grammar explanations of the variety of uses the progressive can have would benefit learners. The textbook series presented in this study, when they introduce the progressive, include grammar explanations as to how the progressive -*ko iss* describes an action in progress. For example, qualitative exploration of the textbooks revealed that when -*ko iss* was explicitly taught it was used with actions such as *watch*, *listen*, *wash*, *clean*, and so forth. Mental and stative verbs were not represented in the explicit teaching sections of the texts, and only appeared later on in the textbook series incidentally. In fact, it appears that textbooks, particularly at the lower levels, incorporate more instances of the progressive being used in tandem with the present to show learners how its usage is option (e.g., asking *what are you doing* with the main verb *do* in the simple present, and then responding *I am drinking tea* in the present progressive). While such distinctions are important, including grammar descriptions and examples with explanations of the progressive used with stative verbs and mental verbs in particular can help learners notice and acquire the forms.

To strengthen the linguistic description of the -*ko iss* construction in textbooks, I recommend introducing it at least twice at different levels. In the beginner levels, introducing -*ko iss* as ‘action in progress’ can help facilitate the acquisition of this prototypical form-function mapping that learners can easily practice in the classroom. At more advanced levels,

reintroducing the progressive as it is used with various semantic senses in both spoken and written language can also be beneficial and allow learners, particularly L1 English speakers, to notice the forms of the progressive which are less common in English. Namely, this amounts to teaching frequently taught chunks such as *al (know)*, *gaji (have/hold)*, *jeunggaha (increase)*, *bododwe (be reported)* among others which were found to be distinctive of the progressive in L1 writing to learners. At the very least, a re-examination of the *-ko iss* construction and its various semantic meanings beyond simply ‘action in progress’ is warranted and would be beneficial for learners in their Korean language learning.

REFERENCES

- Abbot, K., & Tomasello, M. (2006). Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *The Linguistic Review*, 23(3), 275-290. doi: 10.1515/TLR.2006.011
- Ahn, Y. (1995). The aspectual and temporal system of Korean: From the perspective of the two-component theory of aspect. Unpublished Doctoral Dissertation, University of Texas at Austin.
- Andersen, R. W. (1990). Models, processes, principles and strategies: Second language acquisition inside and outside the classroom. In B. VanPatten & J. F. Lee (Eds.), *Second language acquisition-Foreign language learning*, Multilingual Matters, 45-78.
- Andersen, R. W. (1991). Developmental sequences: The emergence of aspect marking in second language acquisition. In T. Huebner & C. A. Ferguson (Eds.), *Tense-aspect morphology in L2 acquisition*, 79-105. John Benjamins.
- Andersen, R. W., & Shirai, Y. (1994). Discourse motivations for some cognitive acquisition principles. *Studies in Second Language Acquisition*, 16, 133-156.
- Anderwald, L. (2012). "I'm loving it" – marketing ploy or language change in progress?. Presented at the Symposium: The pragmatics of aspect in varieties of English. <https://doi.org/10.1080/00393274.2016.1208536>
- Anthony, L. (2023). AntConc (Version 4.2.4) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>
- Anthony, L. (2022). TagAnt (Version 2.0.5) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>
- Bardovi-Harlig, K., & Comajoan-Colomé, L. (2020). The aspect hypothesis and the acquisition of L2 past morphology in the last 20 years: A state-of-the-scholarship review. *Studies in Second Language Acquisition*, 42, 1137-1167. doi:10.1017/S0272263120000194
- Bahrani, T., Tam, S. S., & Zuraidah, M. D. (2014). Authentic Language Input Through Audiovisual Technology and Second Language Acquisition. *Sage Open*, 4(3), 1-8. doi:10.1177/2158244014550611
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1-48.
- Belli, S. A. (2018). An analysis of stative verbs used with the progressive aspect in corpus-informed textbooks. *English Language Teaching*, 11(1), 120-135. doi: 10.5539/elt.v11n1p120
- Biber, D. (1999). Longman Grammar of Spoken and Written English. Longman.

- Biber, D., Johansson, S., Leech, J., Conrad, S., & Finegan, E. (2021). *Grammar of Spoken and Written English*. Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Biber, D., & Conrad, S. (2010). Corpus linguistics and grammar teaching. Available at: www.longmanhomeusa.com/content/pl_biber_conrad_monograph_lo_3.pdf
- Brown, L., & Yeon, J. (2010). Experimental research into the phases of acquisition of Korean tense-aspect: Focusing on the progressive marker “-ko issta.” *Journal of Korean Language Education*, 21(1), 151–173.
- Bybee, J. L. (2013). Usage based theory and exemplar representations of constructions. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 49-69). *The Oxford Handbook of Construction Grammar* (2013; online edn, Oxford Academic, 16 Dec. 2013), <https://doi.org/10.1093/oxfordhb/9780195396683.013.0004>
- Chae, H-R. (2018). The pseudo-resultative {V-ko (iss)} Construction in Korean. *Language Research*, 54(2), 157-200. <https://doi.org/10.30961/lr.2018.54.2.157>
- Davies, M. (2008-) *The Corpus of Contemporary American English (COCA)*. Available online at <https://www.english-corpora.org/coca/>.
- Deshors, S. C., (2011). A multifactorial study of the uses of *may* and *can* in French-English interlanguage. A University of Sussex DPhil theses, Available online via Sussex Research Online: <https://core.ac.uk/download/pdf/2710234.pdf>
- Deshors, S. C., & Gries, S. T. (2014). A case for the multifactorial assessment of learner language: The uses of *may* and *can* in French-English interlanguage. In D. Glynn and J. A. Robinson (Eds.). *Corpus Methods for Semantics: Quantitative studies in polysemy and synonymy*. John Benjamins Publishing Company.
- Deshors, S. C., & Gries, S. T. (2023). Using corpora in research on second language psycholinguistics. In A. Godfroid & H. Hopp (Eds.). *The Routledge Handbook of Second Language Acquisition*. Routledge.
- Flowerdew, L. (1998). Corpus linguistic techniques applied to textlinguistics. *System*, 26(4), 541-552. [https://doi.org/10.1016/S0346-251X\(98\)00039-6](https://doi.org/10.1016/S0346-251X(98)00039-6)
- Fokkema, M., Smits, N., Zeileis, A., Hothorn, T., & Kelderman, H. (2018). Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees. *Behavior Research Methods*, 50, 2016-2034. doi: <https://doi.org/10.3758/s13428-017-0971-x>
- Freund, N. (2016). Recent change in the use of stative verbs in the progressive form in British English: I'm loving it. *Language Studies Working Papers*, 7, 50-61.

Fuchs, R., & Werner, V. (2018). The use of stative progressives by school-age learners of English and the importance of the variable context. *International Journal of Learner Corpus Research*, 4(2), 195-224. <https://doi.org/10.1075/ijlcr.00004.int>

Gabrielatos, C. (2005). Corpora and language teaching: Just a fling or wedding bells? *The Electronic Journal for English as a Second Language*, 8(4).

Granath, S. & Wherrity, M. (2014). "I'm loving you - and knowing it too": Aspect and so-called stative verbs. *Rhesis: Linguistics and Philology*, 4(1), 2-22. <http://urn.kb.se/resolve?urn=urn:nbn:se:kau:diva-31699>

Granger, S. (2009). The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation. In K. Aijmer (Ed.), *Corpora and Language Teaching*. John Benjamins Publishing Company. Permalink: <http://digital.casalini.it/9789027289988>

Gries, S. T., & Stefanowitsch, A. (2004). Extending collocation analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics*, 9, 97-129. <https://doi.org/10.1075/ijcl.9.1.06gri>

Gries, S., Hampe, B. & Schönefeld, D. (2005). Converging evidence: Bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics*, 16(4), 635-676. <https://doi.org/10.1515/cogl.2005.16.4.635>

Gries, S. T. (2014). Coll.analysis 3.5: A script for R to compute perform collocation analyses.

Gries, S. T., & Deshors, S. C. (2014). Using regressions to explore deviations between corpus data and a standard/target: Two suggestions. *Corpora*, 9(1), 109-136. DOI: 10.3366/cor.2014.0053

Gries, S. T. (2015). The most underused statistical methods in corpus linguistics: Multi-level (and mixed effects) models. *Corpora*, 10(1), 95-125. DOI: 10.3366/cor.2015.0068

Hong, K-S. (1991). Argument selection and case marking in Korean. Unpublished Doctoral Dissertation, Stanford University.

Hothorn, T., & Zeileis, A. (2015). partykit: A modular toolkit for recursive partytioning in R. *Journal of Machine Learning Research*, 16, 3905-3909. Available from <https://jmlr.org/papers/v16/hothorn15a.html>

Hundt, M., & Vogel, K. (2011). Overuse of the progressive in ESL and learner Englishes - fact or fiction?. *Studies in Corpus Linguistics*, 44, 145-166. <https://doi.org/10.1075/scl.44.08vog>

Hundt, M., Rautioaho, P., & Strobl, C. (2020). Progressive or simple? A corpus-based study of aspect in World Englishes. *Corpora*, 15(1), 77-106–106. doi: 10.3366/cor.2020.0186

- Jang, M.-s. (2005). The improved plans for teaching Korean tense and aspect forms: Focusing on *hanta* type and *hako isssta* type. *Journal of Korean Language Education*, 16(3), 305–330.
- Jeong, S. J. (2011). A study on the description methods of the adverbial case postpositions for Korean education based on cognitive linguistics. *The Korean Language and Literature*, 112, 79–110.
- Jung, B. K. (2022). The nature of L2 input: Analysis of textbooks for learners of Korean as a second language. *Korean Linguistics*, 18(2), 182–208. <https://doi.org/10.1075/kl.20001.jun>
- Kim, H., Kang, B., & Hong, J. (2007). 21st Century Sejong Corpora (to be) completed. *The Korean Language in America*, 12, 31–42. JSTOR, <http://www.jstor.org/stable/42922169>.
- Kim, S. K. (2011). Education method for the adverb postpositions of ‘ey’, ‘eyse’, ‘lo’ in the Korean language. *Kwukhakyenkwulonchong*, 8, 199–236.
- Kim, Y., & Guo, J. (2016). A study on the acquisition of Korean adverbial case marker *ey* in spoken production by Chinese Korean L2 learners. *Korean Education Research*, 38, 1–26.
- Koprowski, M. (2005). Investigating the usefulness of lexical phrases in contemporary coursebooks. *ELT Journal*, 59 (4), 322–32.
- Kranich, S. (2010) Progressive in modern English: A corpus-based study of grammaticalization and related changes. Amsterdam: Rodopi. doi: 10.1163/9789042031449
- Lam, P. W. Y. (2009). Discourse particles in corpus data and textbooks: The case of *Well*. *Applied Linguistics*, 31(2), 260–281. <https://doi.org/10.1093/applin/amp026>
- Lee, E. (2006). Stative progressives in Korean and English. *Journal of Pragmatics*, 38, 695–717. doi:10.1016/j.pragma.2005.09.006
- Northbrook, J., & Conklin, K. (2019). Is what you put in what you get out?: Textbook-derived lexical bundle processing in beginner English learners. *Applied Linguistics*, 40(50), 816–833. doi:10.1093/applin/amy027
- Rautonaho, P. (2014). Variation in the progressive: A corpus-based study into World Englishes. Tampere: Tampere University Press.
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.
- Rautonaho, P., & Deshors, S. C. (2018). Progressive or not progressive?: Modeling the constructional choices of EFL and ESL writers. *International Journal of Learner Corpus Research*, 4(2), 225–252. <https://doi.org/10.1075/ijlcr.16019.rau>

- Rautavaaho, P. (2020). Revisiting the myth of stative progressives in world Englishes. *World Englishes*, 41, 183-206. DOI: 10.1111/weng.12520
- Römer, U. (2004). Comparing real and ideal learning input: The use of an EFL textbook corpus in corpus linguistics and language teaching. In G. Aston, S. Bernardini, & D. Stewart (Eds.). *Corpora and Language Learners*. John Benjamins Publishing Company. <https://doi.org/10.1075/scl.17.12rom>
- Römer, U. (2005). Progressives, Patterns, Pedagogy: A corpus-driven approach to English progressive forms, functions, contexts and didactics. John Benjamins Publishing Company. doi: <https://doi.org/10.1075/scl.18>
- Römer, U. (2006). Where the computer meets language, literature, and pedagogy: Corpus analysis in English studies. In A. Gerbig, A. Müller-Wood (Eds.). *How Globalization Affects the Teaching of English: Studying Culture Through Texts*. Lampeter: E. Mellen Press. 81-109.
- Römer, U. (2011). *Corpus research applications in second language teaching*. Annual Review of Applied Linguistics, 31, 205-225. doi: 10.1017/S0267190511000055
- Salaberry, R., & Shirai, Y. (2002). *L2 acquisition of tense-aspect morphology*. In Salaberry, R., & Shirai, Y. (Eds.). *L2 Acquisition of Tense-Aspect Morphology*. John Benjamins Publishing Company.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129-158.
- Sinclair, J. M. (Ed.) (1987). Looking up: An account of the COBUILD project in lexical computing. London: Collins ELT.
- Sinclair, J. (1997). *Corpus evidence in language description*. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and Language Corpora*, 27-39. London: Longman.
- Sinclair, J. M. (Ed.) (2004). *How To Use Corpora in Language Teaching*. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Stefanowitsch, A., & Gries, S. T. (2003). Collocations: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8, 209-243. <https://doi.org/10.1075/ijcl.8.2.03ste>
- Straka, M., Hajič, J., & Straková, J. (2016). UDPipe: Trainable pipeline for processing CoNLL-U files performing tokenization, morphological analysis, POS tagging and parsing. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Portorož, Slovenia, May 2016.
- Timmis, I. Corpora and materials: Towards a working relationship. In B. Tomlinson (Ed.). *Developing materials for language teaching*. Bloomsbury Publishing.

Vendler, Z. (1957). Verbs and Times. *The Philosophical Review*, 66, 143-160.
<https://doi.org/10.2307/2182371>

Virtanen, T. (1996). The progressive in NS and NNS student compositions: Evidence from the International Corpus of Learner English. In M. Ljung (Ed.), *Corpus-based Studies in English Papers from the Seventeenth International Conference on English Language Research and Computerized Corpora*, Rodopi B.V: Amsterdam.

Yeon, J., & Brown, L. (2011). *Korean, A Comprehensive Grammar*. New York: Routledge.

Zaenen, A., Carletta, J., Garretson, G., Bresnan, J., Koontz-Garboden, A., Nikitina, T., O'Connor, M. C., & Wasow, T. (2004). Animacy encoding in English: why and how. In *Proceedings of the ACL-04 Workshop on Discourse Annotation*. Available at:
<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.154.7>

APPENDIX A: DISTINCTIVE COLLEXEME ANALYSIS I

Table A-1.

Distinctive collexemes for the (non)progressive in L1 written Korean data

Progressive	Coll.strength	Non-progressive	Coll.strength
bo – ‘see’	319.52	moreu – ‘not know’	727.00
beoli – ‘start/begin’	148.47	boi – ‘be seen’	407.30
bad – ‘receive’	116.98	malha – ‘speak’	394.15
balghi – ‘light/brighthen’	114.51	saenggagha – ‘think’	195.83
gyeogg – ‘experience (esp. hardship)’	104.03	na – ‘happen’	160.38
banbalha – ‘oppose’	102.15	sijagha – ‘start’	107.96
du – ‘put/set/place’	84.31	deuleoga – ‘go in’	96.15
jarijab – ‘settle/situate’	78.29	yeolli – ‘open’	92.38
nopaji – ‘rise’	77.46	ireu – ‘reach/get to’	90.95
naenoh – ‘put/take out’	77.02	sihaengdwe – ‘go into effect’	78.30
sal – ‘live’	74.97	ju – ‘give’	76.51
jaegidwe – ‘be raised/made’	74.71	pyeolcyeoji – ‘spread’	63.27
isddareu – ‘occur in succession’	73.95	jumogdwe – ‘be watched’	61.56
hwagsandwe – ‘spread’	71.54	deulli – ‘be heard’	54.97

Table A-1 (cont'd).

nao – ‘come out’	70.77	ggobhi – ‘be in a range’	53.20
gaji – ‘have/hold’	69.49	geolli – ‘take time’	45.49
gajchu – ‘prepare/be equipped’	65.19	bara – ‘hope’	45.27
al – ‘know’	64.84	bulli – ‘be referred/called as’	41.44
bij – ‘come into/be in conflict/criticism’	63.14	manna – ‘meet’	41.20
allyeoji – ‘be known’	60.70	deuleoo – ‘come in’	37.52
alh – ‘suffer’	58.20	saenggi – ‘form’	37.41
girogha – ‘record/document’	51.03	bureu – ‘call’	36.82
sa – ‘buy’	51.03	neom – ‘over/excess’	33.57
dalli – ‘run’	44.30	ollaga – ‘go up’	31.42
eod – ‘gain’	44.16	gidaedwe – ‘expect’	30.52
geomtoha – ‘review/examine’	41.53	yeol – ‘open’	27.05
gaj – ‘have/hold’	39.21	jeonmangdwe – ‘view/predict’	26.50
beoleoji – ‘happen, take place’	37.88	jonjaeha – ‘exist’	25.78
naedabo – ‘predict’	37.42	jijeogha – ‘indicate’	25.40
sseu – ‘use’	36.86	dalha – ‘reach (e.g., level)’	25.12
moeu – ‘gather’	36.22	gweonha – ‘advise’	24.96

Table A-1 (cont'd).

sseu – ‘write’	35.98	pulidwe – ‘be explained’	22.59
gidari – ‘wait’	34.93	jujangha – ‘assert’	22.03
gaha – ‘apply, spur, cause’	32.68	dojeonha – ‘challenge’	20.65
jeonha – ‘tell, convey, pass on information’	32.68	bunseogdwe – ‘analyze’	18.75
nuri – ‘enjoy’	32.28	gyeoljeongha – ‘decide’	17.62
molli – ‘be driven to/into’	32.07	yeogyeoji – ‘be considered as’	15.71
olli – ‘raise’	32.07	ddeona – ‘depart’	15.56
nori – ‘seek, aim’	31.04	deungjangha – ‘appear’	15.17
pyeolchi – ‘spread’	30.55	haeseogdwe – ‘interpret’	15.06
naeri – ‘get off’	29.41	yogudwe – ‘request’	14.83
gojodwe – ‘tone up, enhance’	28.88	johaha – ‘like’	14.21
geuchi – ‘stop’	27.18	balgyeonha – ‘discover’	12.91
chujinha – ‘push ahead with sth, promote’	26.34	yeongyeoldwe – ‘connect’	12.85
boyuha – ‘posses’	26.17	seolmyeongha – ‘explain’	12.56
jeonhaeji – ‘be passed along/conveyed’	25.93	gangjoha – ‘emphasize’	12.44

Table A-1 (cont'd).

bul – ‘blow’	25.49	uryeodwe – ‘be concerned’	12.04
ta – ‘ride’	25.49	ggaedad – ‘realize’	12.03
ddeoreu – ‘rise, come up’	24.48	gubundwe – ‘sort’	11.91
uryeoha – ‘be concerned or fearful’	24.31	mandeul – ‘make’	11.18
hwaldongha – ‘do an activity’	24.26	deud – ‘listen’	10.74
cuiha – ‘be drunk or enraptured in something’	23.81	mud – ‘ask’	10.49
ginjangha – ‘worry’	23.81	ggeutna – ‘end’	10.32
junbiha – ‘prepare’	23.71	seonboi – ‘show/present’	9.70
sam – ‘be considered as’	23.05	chujeongdwe – ‘trace’	9.66
jiki – ‘protect’	22.13	jinaga – ‘pass’	9.33
eosgalli – ‘have a disagreement’	21.62	neomchi – ‘overflow’	9.14
beonji – ‘spread’	21.37	yeosboi – ‘get a sense of’	9.14
ssod – ‘spill, pour’	21.37	dolao – ‘return’	8.86
ssodaji – ‘pour, gush’	21.37	heureu – ‘flow’	8.14
nah – ‘produce, spawn, give birth’	20.76	salpyeobo – ‘examine/check’	8.10
ga – ‘go’	19.99	ddeu – ‘scoop’	7.93

Table A-1 (cont'd).

namgi – ‘save, set aside sth’	19.79	jeulgi – ‘enjoy’	7.79
beoti – ‘endure’	18.80	salpi – ‘look (as in see about something)’	7.79
umjigi – ‘move’	18.43	ggichi – ‘influence’	7.79
maej – ‘bear, sign, enter into contract’	18.20	chamgaha – ‘attend’	7.01
bododwe – ‘be reported’	17.15	balgyeondwe – ‘be discovered’	6.82
yeogi – ‘regard as’	17.15	ja – ‘sleep’	6.45
keoji – ‘get bigger’	16.72	nureu – ‘push’	6.45
mosaegha – ‘seek, find’	16.16	sarangha – ‘love’	6.45
injeongbad – ‘be recognized’	15.51	hwaginha – ‘confirm’	6.31
saenghwalha – ‘live, as in make a living or live your life’	15.14	gobaegha – ‘confess’	5.59
deonji – ‘throw’	14.75	sui – ‘rest’	5.58
geuri – ‘draw’	14.74	yeongebha – ‘do business’	5.58
deureonae – ‘expose’	14.57	jeogyongdwe – ‘get used to’	5.47
pal – ‘sell’	14.23	nolla – ‘be surprised’	5.44
yoguha – ‘request’	14.20	pyeonggadwe – ‘be rated’	5.25
dolli – ‘turn’	13.82	ddareu – ‘follow’	5.19
heundeulli – ‘shake’	13.82	gongyeonha – ‘perform’	4.73

Table A-1 (cont'd).

saenggyeona – ‘emerge, occur’	13.04	gieogdwe – ‘be remembered’	4.69
yaegoha – ‘notify previously/in advance’	13.04	heoyongdwe – ‘be permitted’	4.69
nopi – ‘increase’	12.87	punggi – ‘give off smell’	4.69
jibjungdwe – ‘be focused’	12.81	bumbi – ‘be overcrowded’	4.56
gyesogdwe – ‘be continued’	12.72	jindanha – ‘diagnose’	4.56
chajiha – ‘possess, or take possession’	12.70	pyeonggaha – ‘rate’	4.42
pyeonggabad – ‘receive a ranking’	12.33	jinae – ‘spend/pass’	4.33
geumjiha – ‘be prohibited’	12.27	tujaha – ‘invest’	4.22
chamyeoha – ‘attend’	11.27	meog – ‘eat’	3.95
bulanhaeha – ‘feel uneasy’	10.68	gusaha – ‘have command of’	3.91
pum – ‘brood’	10.68	kyeo – ‘turn on’	3.91
o – ‘come’	10.18	yeogseolha – ‘emphasize’	3.76
yujiha – ‘keep, maintain’	10.18	noneuiha – ‘discuss/debate’	3.57
badadeulyeoji – ‘accept something’	9.97	gamjidwe – ‘sense/detect’	3.26
palli – ‘be sold’	9.64	chusandwe – ‘be estimated’	3.13

Table A-1 (cont'd).

eongeubha – ‘mention’	9.21	jarangha – ‘brag’	3.11
insigha – ‘be aware’	9.21	dwe – ‘become’	3.10
myosahah – ‘describe’	9.21	jihyangha – ‘pursue’	2.95
ddi – ‘assume (as in take sth on)’	9.15	bunryudwe – ‘classify’	2.70
bultaeu – ‘burn’	9.11	ihaeha – ‘understand’	2.70
josaha – ‘investigate’	9.11	balpyoha – ‘present’	2.63
oichi – ‘shout’	9.11	gwancheugdwe – ‘be observed/predicted’	2.60
simhoadwe – ‘deepen’	9.11	naemil – ‘stick/hold out’	2.60
chusanha – ‘estimate’	8.95	olmgi – ‘move’	2.60
bandaeha – ‘oppose’	8.57	jaesiha – ‘suggest’	2.54
neuggi – ‘feel’	8.34	weonha – ‘want’	2.37
neuleona – ‘increase’	8.20	gongyuha – ‘share’	2.35
baeu – ‘learn’	7.80	neoh – ‘put in’	2.35
busangha – ‘float, emerge’	7.80	seo – ‘stand’	2.35
chaetaegha – ‘choose/adopt (as in a resolution etc.)’	7.80	seoneonha – ‘declare’	2.35
jibaeha – ‘rule, dominate’	7.80	teu – ‘open’	2.35

Table A-1 (cont'd).

chisos – ‘rise, soar, surge’	7.57	iyongha – ‘use’	2.27
dabbyeonha – ‘reply’	7.57	injeongha – ‘accept’	2.24
dwechaj – ‘take back’	7.57	dolaga – ‘go back’	2.21
ginjangsiki – ‘make nervous’	7.57	gieogha – ‘remember’	2.19
ilgwanha – ‘be consistent in doing something’	7.57	deuleoseo – ‘enter in’	2.19
jigmyeonha – ‘encounter’	7.57	jaeanha – ‘offer’	1.95
taeu – ‘burn, singe’	7.57	bunpoha – ‘distribute’	1.76
yaecheugha – ‘predict’	7.57	ilh – ‘lose/be deprived’	1.76
ilha – ‘work’	7.52	apseo – ‘get head’	1.76
bonae – ‘send’	7.46	ihaedwe – ‘be understood’	1.76
gominha – ‘worry’	7.28	naga – ‘go out’	1.69
jab – ‘grab’	6.52	sidoha – ‘try’	1.66
jibjungha – ‘focus’	6.43	banghwangha – ‘wander’	1.63
nanu – ‘distribute’	6.43	chugadwe – ‘be added’	1.63
seonjeonha – ‘propagate’	6.43	gajyeoga – ‘take’	1.63
ddeoleoji – ‘fall, decrease’	6.16	galli – ‘be changed/divided’	1.63

Table A-1 (cont'd).

daedudwe – ‘come to the fore, be on the rise’	6.06	ggojib – ‘pinch’	1.63
ddeolchi – ‘shake off, ride oneself of’	6.06	bbae – ‘subtract’	1.39
meomureu – ‘stay’	6.06	gueonyuha – ‘invite’	1.39
pyoha – ‘express’	6.06	heoyongha – ‘permit’	1.39
uihyeobha – ‘intimidate’	6.06	goreu – ‘choose’	1.32
jagyongha – ‘act, function’	6.01		
unyeongha – ‘manage (e.g., business)’	5.99		
gonggeubha – ‘supply, provide’	5.78		
gyesogha – ‘continue’	5.56		
buri – ‘manage, handle’	5.45		
geol – ‘count on hopes or expectations’	5.22		
alli – ‘tell’	5.11		
dwepuliha – ‘repeat’	5.11		
salaga – ‘live’	5.11		
sihaengha – ‘carry out, enforce’	5.11		

Table A-1 (cont'd).

sseu_singyeong – ‘care about something’	5.11
gareuchi – ‘teach’	5.09
sidalli – ‘suffer from something’	5.09
deul – ‘hold, pick up’	4.95
natanae – ‘show, present’	4.90
paagha – ‘identify’ figure out’	4.89
georondwe – ‘be mentioned, brought up’	4.64
balghyeoji – ‘be illuminated’	4.60
bichu – ‘shine’	4.60
daebiha – ‘prepare, be ready’	4.60
gamchu – ‘reduce’	4.60
ganjigha – ‘keep’	4.60
geojuha – ‘live’	4.60
ibjiha – ‘be positioned at’	4.60
ilheoga – ‘lose something, someone’	4.60
iljoha – ‘play a part, contribute’	4.60

Table A-1 (cont'd).

jeungpogsiki – ‘amplify’	4.60
nol – ‘hang out’	4.60
pyoryuha – ‘drift, float’	4.60
siinha – ‘admit, acknowledge’	4.60
beoseona – ‘get out, get free’	4.60
dayanghaeji – ‘become diverse’	4.60
simhaeji – ‘become severe’	4.60
saraji – ‘disappear’	4.49
geod – ‘walk’	4.28
ganghwaha – ‘reinforce’	4.07
barabo – ‘look, watch, stare’	3.94
damul – ‘keep quiet’	3.86
geomtodwe – ‘be examined’	3.86
jeomchi – ‘predict future’	3.86
naebichi – ‘hint at’	3.86
seongjangha – ‘grow up’	3.86
jeonmangha – predict	3.85

Table A-1 (cont'd).

ggeul – ‘pull’	3.68
georonha – ‘mention, bring up’	3.48
jusiha – ‘watch carefully’	3.48
ssah – ‘pile up’	3.48
gangjodwe – ‘be emphasized’	3.26
hoagboha – ‘secure’	3.26
bbomnae – ‘boast, show off’	3.21
gamsiha – ‘monitor’	3.21
gangguha – ‘take measures to do sth’	3.21
gongbuha – ‘study’	3.21
gunrimha – ‘dominate’	3.21
pyosiha – ‘express’	3.13
mat – ‘be in charge of something’	3.01
chireu – ‘pay out’	3.00
chujeongha – ‘estimate’	3.00
jeunggaha – ‘increase’	2.98
jis – ‘build, construct’	2.98
mid – ‘believe’	2.75

Table A-1 (cont'd).

jaegiha – ‘raise, bring up’	2.69
balghyeonae – ‘reveal or disclose’	2.68
binbalha – ‘occur frequently’	2.68
georaedwe – ‘be traded, dealt’	2.68
giul – ‘lean or tilt’	2.68
musiha – ‘ignore’	2.68
neolbhi – ‘make wide’	2.68
maryeonha – ‘prepare, arrange’	2.58
ileona – ‘get up’	2.52
ggob – ‘count (also count on fingers)’	2.47
geodu – ‘reap’	2.44
gugaha – ‘sing praises’	2.44
jeonragha – ‘fall (into ruin)’	2.44
byeonhwaha – ‘change’	2.32
ganjuha – ‘regard, consider as’	2.32
ggal – ‘spread, pave’	2.32
haemyeongha – ‘clarify’	2.32

Table A-1 (cont'd).

hoagdaecha –	2.32
‘expand, enlarge’	
ibjeungha – ‘prove’	2.32
insigdwe – ‘be	2.32
acknowledged’	
naebonae –	2.32
‘remove’	
hwalyongha – ‘use’	2.27
unyeongdwe – ‘be	2.27
run, managed’	
daebyeonha –	2.26
‘represent’	
ilg – ‘read’	2.26
seonhoha – ‘prefer’	2.26
ssodanae –	2.26
‘push/spill out’	
iru – ‘achieve’	2.20
bunseogha –	1.99
‘analyze’	
eongeubdwe – ‘be	1.93
mentioned’	
gureu – ‘stomp	1.93
feet’	
mangchi – ‘spoil,	1.93
ruin’	
neombo – ‘covet	1.93
(e.g., first place)’	
soyuha – ‘own’	1.93
baggu – ‘change’	1.91

Table A-1 (cont'd.)

jinhaengdwe – ‘proceed as’	1.79
chulgandwe – ‘be published’	1.63
chulsidwe – ‘be released, launched’	1.63
ggi – ‘cloud over’	1.63
gongtongdwe – ‘be common’	1.63
nonha – ‘discuss’	1.63
majiha – ‘receive, greet, welcome someone’	1.58
naseo – ‘take action’	1.55
gaebalha – ‘develop’	1.54
nae – ‘submit’	1.54
banyeongha – ‘reflect’	1.54
balb – ‘step (on)’	1.52
deohaega – ‘add’	1.52
dogryeoha – ‘encourage’	1.52
euisimha – ‘doubt or be suspicious’	1.52
silgamha – ‘feel, sometimes to the point of realizing’	1.52

Table A-1 (cont'd).

jangdamha –	1.51
‘guarantee’	
teoddeuri – ‘pop,	1.51
break, or burst’	
ganghwadwe – ‘be	1.49
strengthened’	
gwasiha – ‘show	1.49
off’	
naepoha – ‘involve’	1.49
bijeoji – ‘be made’	1.49
gajungdwe – ‘be	1.49
aggravated’	
gamdol – ‘hang’	1.49
gyeongjaengha –	1.49
‘compete, vie for’	
haengsaha –	1.49
‘invoke’	
mangraha –	1.49
‘include or cover	
everything’	
mojibha – ‘recruit’	1.49
nanmuha – ‘be rife’	1.49

APPENDIX B: DISTINCTIVE COLLEXEME ANALYSIS II

Table A-2.

Distinctive collexemes for the progressive (left) and non-progressive (right) in L1 English L2 Korean

Progressive	Coll.strength	Non-progressive	Coll.strength
manhaji – ‘increase’	48.37	saenggagha – ‘think’	163.08
jeunggaha – ‘increase’	42.29	bo – ‘see’	24.43
sal – ‘live’	13.57	meog – ‘eat’	8.53
noryeogha – ‘make effort	12.47	ju – ‘give’	3.50
baeu – ‘learn’	11.94	bandaecha – ‘oppose’	2.86
gominha – ‘worry/agonize’	10.79	bonae – ‘send’	2.37
jinae – ‘spend/pass time’	9.49	masi – ‘drink’	2.27
dani – ‘attend’	8.07	sayongha – ‘use’	2.27
byeonhwaha – ‘change’	8.03	deud – ‘listen’	1.72
geogjeongha – ‘worry’	5.86		
junbiha – ‘prepare’	5.4		
al – ‘know’	3.3		
jeonggongha – ‘major in’	2.92		
neuggi – ‘feel’	2.92		
saenggi – ‘be formed’	2.92		
dwe – ‘become’	2.84		
ilha – ‘work’	2.83		
natana – ‘appear’	2.50		
gongbuha – ‘study’	2.23		
yeonseubha – ‘practice’	1.72		
byeonha – ‘change’	1.66		
gidaecha – ‘expect/anticipate’	1.66		
saraji – ‘disappear’	1.66		

APPENDIX C: DISTINCTIVE COLLEXEME ANALYSIS III

Table A-3.

Distinctive collexemes for the progressive (left) and non-progressive (right) in L1 Japanese L2

Korean

Progressive	Coll.strength	Non-progressive	Coll.strength
sal – ‘live’	149.60	saenggagha – ‘think’	824.74
gaji – ‘have/hold’	138.10	ga – ‘go’	91.69
noryeogha – ‘make effort’	90.75	moreu – ‘to not know’	51.72
neuleona – ‘increase’	75.92	malha – ‘speak’	22.20
ilha – ‘work’	70.02	iyagiha – ‘talk’	19.33
gongbuha – ‘study’	69.28	boi – ‘be seen/visible’	16.13
saenghwalha – ‘live’	56.26	sogaeha – ‘introduce’	7.97
dani – ‘attend’	40.03	o – ‘come’	4.79
jinae – ‘spend time’	36.45	meog – ‘eat’	4.07
yeonseubha – ‘practice’	36.22	bo – ‘see’	3.43
gominha – ‘worry’	32.34	ju – ‘give’	3.23
dwe – ‘become’	28.46	sa – ‘buy’	1.67
gidaeha – ‘expect’	24.46	neuggi – ‘feel’	1.60
saenggi – ‘form’	23.47	sigsaha – ‘eat’	1.40

Table A-3 (cont'd).

sseu – ‘use’	20.80
balsaengha – ‘occur’	20.38
bonae – ‘send’	19.47
ggeul – ‘pull, attract’	18.07
chaj – ‘find’	15.33
eungweonha – ‘cheer’	15.33
gidari – ‘wait’	14.85
bad – ‘receive’	14.60
baljeonha – ‘develop’	12.64
moeu – ‘collect’	12.64
areubaiteuha – ‘work part-time job’	12.47
ileona – ‘get up’	12.30
baeu – ‘learn’	12.12
mid – ‘believe’	11.12
eod – ‘gain’	9.99
dallaji – ‘change/become different’	9.06
gareuchi – ‘teach’	7.41
saraji – ‘disappear’	7.41
mandeul – ‘make’	7.06

Table A-3 (cont'd).

baggui – ‘change’	5.73
sayongha – ‘use’	5.14
hwalyagha – ‘be active’	4.94
jjig – ‘take a picture’	4.94
natana – ‘appear’	4.44
junbiha – ‘prepare’	3.58
nol – ‘play’	3.22
saraga – ‘make a living’	3.22
gamsaha – ‘appreciate’	2.91
baldalha – ‘develop’	2.64
haengdongha – ‘act/behave’	2.64
jeogeoji – ‘diminish’	2.64
silgamha – ‘realize’	2.64
deud – ‘listen’	1.95
pal – ‘sell’	1.43
sayongdwe – ‘be used’	1.43