The nature of L2 input

Analysis of textbooks for learners of Korean as a second language

Boo Kyung Jung
University of Pittsburgh, USA

How language learners of Korean acquire knowledge on postpositions has been a long-standing research question in Korean language pedagogy due to their polysemous nature. The present study investigates the nature of input involving the locative function of the postposition *-ey*, one of the representative polysemous postpositions in Korean, through the frequency of its occurrence, types of verbs co-occurring with *-ey*, and keyness analysis. Sejong written and spoken corpora and two types of textbooks (eight volumes for each type: two volumes for four proficiency levels) for language learners of Korean are analyzed. Results show that *iss-* 'to be/exist' predominantly occurred with locative *-ey* in the Sejong corpora and a few verbs occupied a large proportion of the total usage. On the other hand, the most frequent verb was *ka-* 'to go' in all proficiency levels of the textbooks, with the exception of the fourth level of the second-type textbook. This suggests that, while the Sejong usage highlights its existential role, *-ey* for indicating destination is widely emphasized in the textbooks. Since the purpose of language learning is to learn the structure and usage of the target language, this study's findings can offer guidance in setting and building pedagogical goals and directions.

Keywords: Korean postposition, locative function, language textbook, input frequency

1. Introduction

Usage-based approaches to language acquisition argue that language is acquired through actual experiences of language use (Behrens 2009, Langacker 2008,

https://doi.org/10.1075/kl.20001.jun
Korean Linguistics 18:2 (2022), pp. 182–208. ISSN 0257-3784 | E-ISSN 2212-9731

Tomasello 2003).[1] While all the cognitive properties (e.g., processing cost, embodiment from experience, abstraction, categorization, generalization) affect the formation of the conceptual structure of language, studies have been manifesting the critical role of frequency-based language learning in first language (L1) (e.g., Abbot-Smith & Tomasello 2006) and second language (L2) (e.g., Ellis & Ferreira-Junior 2009a) acquisition due to human beings' innate sensitivity to frequency (Ellis 2002). As usage-based approaches broadly explore the process and enhancement of language acquisition through actual language use, corpus analysis serves as a useful tool in the examination of the properties of language from these perspectives. Following these approaches, the present study analyzes the use of *-ey* in the Korean corpora.

Korean is a Subject-Object-Verb (SOV) language, and this property renders the linguistic information carried by a verb (e.g., number of arguments, tense) inaccessible until the utterance reaches an end. In addition, as a situation-oriented language, sentential elements are easily omitted when they are retrievable from the context. For example, the subject is not realized in the question *mwe hani?* 'what do you do?' because it is situationally clear whom the person is asking. Thus, diverse linguistic strategies are utilized to reduce ambiguity in the interpretation of the utterance. One of them is the use of postpositions, which are function words that deliver the syntactic and semantic information to which they relate (Sohn 1999). For instance, in the dative construction *Tom-i Minho-eykey senmwul-ul cwu-ess-ta* 'Tom gave Minho a present', *-eykey* and *-ul* indicate the recipient and the theme respectively. Overall, information carried by postpositions is crucial in organizing and understanding a sentence (e.g., Kim 1999).

There are a number of polysemous adverbial postpositions in Korean: one form carries multiple meanings, or vice versa (Kim 2011). The polysemous nature of the adverbial postposition *-ey,* as the most commonly used adverbial postposition in Korean (Kang & Kim 2009), provides a good testing ground in Korean studies. It is most often used to denote a location, as in *cip-ey iss-ta* '(someone) is at home' and *cip-ey ka-ta* '(someone) goes home'. It is also used for time (*sey si-ey* 'at three o'clock'), unit (*hana-ey* 'per one'), cause (*palam-ey namu-ka ppophi-ess-ta* 'Due to wind, the tree was unrooted'), etc. There are other forms of postpositions used to mark locations (*-eyse* and *-(u)lo*) in Korean as well, and a number of studies have investigated the functions of *-ey* and its unique and shared properties. However, little research has paid close attention to corpus analysis which attests to the use of *-ey* by language users of Korean.

[1] The term 'language use' is generally used in usage-based approaches to include the experience of/exposure to language.

Due to their multiple form-function relations, the acquisition of postpositions has been one of the main research topics in Korean as a second/foreign language (KSL/KFL) studies. Studies on postpositions in KSL/KFL settings have mostly taken one of two forms: (1) analyzing errors produced by learners in relation to postposition usage; and (2) examining how textbooks introduce the functions of each postposition based on target proficiency levels. Most studies on error production have collected their own corpora, which limits their generalizability. Due to the increasing demand for utilizing corpora in developmental research on Korean, the efforts put forth by several domestic universities and national institutions to develop a learner corpus of language learners of Korean have been growing. In the case of analyses of textbook or input materials, however, few resources are publicly available. The absence of well-organized textbook corpora has limited the scope of most textbook studies to qualitative analyses. Overall, studies of *-ey* have not fully examined diverse aspects of its use in the L2 context, which would be much easier with the utilization of corpora. Thus, the present study aims to investigate how the locative *-ey* is presented in L2 language textbooks and how it is (dis)similar to the L1 corpora, together with stressing the necessity of building and utilizing language corpora.

As an initial step to address these issues, the present study analyzes 16 volumes of textbooks for adult learners of Korean (eight volumes for two types of textbooks) as corpora for the L2 learning of Korean. The Sejong written and spoken corpora provided by the National Institute of Korean Language (NIKL) will be used as reference L1 corpora. More specifically, this study focuses on examining the locative use of *-ey* in three ways. First, the frequency of *-ey* will be counted because frequency is a critical factor in facilitating language acquisition (Ellis 2002). Second, the verbs co-occurring with *-ey* will be examined because the decision on which locative postposition to use among multiple form-function relations heavily depends on the linguistic environment (e.g., types of verb) (Sohn 1999). Finally, keyness analysis will be conducted because it will show which items are used significantly more or less in one corpus compared to the reference corpus.

2. Background

2.1 Usage-based approaches to language development

Usage-based approaches (Behrens 2009, Bybee 2013, Langacker 2008) perceive language as being acquired through previous experiences of language use, with cognitive properties interworking simultaneously to handle input from the out-side world. Each linguistic token not only carries information about the form-function relationship but also transfers the token's contextual/social environment. During all steps of language processing, human cognitive mechanisms analyze, categorize, and generalize information about each token. Repeated use of language instances expedites the process, yielding automatization. Thus, knowledge of a language system emerges and develops from using the language.

Like other types of learning, each token of input affects the language user's memory. When a newly absorbed item is identical to the previous one in the sense of its form-function relations, phonetic properties, context, and so forth, the new item reinforces the representation of the previous one, whereas the new item clusters around a previous one with similar properties when there are any variations (Goldberg 2019). Then, the use of a sufficient number of tokens strengthens the representation/schema of the target feature (Abbot-Smith & Tomasello 2006). The accumulation of those structures holds a language, which is accepted and conventionalized through language use by the language users.

In this process of language acquisition, relative frequency and distributional property play key roles in shaping a language user's perception of language (Bybee 2013, Ellis 2002, Klein & Manning 2005, Tomasello 2000, 2003). To illustrate, Ellis and Ferreira-Junior (2009b) showed that the order of the frequency of English verbs in the constructions of L2 learners was broadly matched with that of L1 users in L2 acquisition studies. In a similar vein, Kyle and Crossley (2017) evidenced that novice learners relied on frequently occurring verb-argument constructions while language learners utilized both frequent and infrequent ones as their proficiency advanced. Several studies focused on the role of token/type frequency. The former indicates the number of occurrences of an item with the same lemma, while the latter refers to the frequency of a "distinct lexical item" (Ellis 2002, p. 166) in an utterance. In general, high token frequency strengthens memory representation and reduces the processing burden (Bybee 2008), and type frequency is considered to accelerate the productivity of a construction (Bybee & Hopper 2001).

With a general consensus having been reached on the pivotal role of input (and input frequency) in language learning, the demand for a close investigation of the properties and roles of input, such as what types of input are provided, how they are presented in pedagogical contexts, and how they have effects on language learners' perceptions of language, is increasing. Although it is almost impossible to precisely gauge all possible input source/types for L2 learners, textbooks for language learners are generally considered a major input source (Römer 2004, Tono 2004), and language use in textbooks need to be examined in relation to the L1 language use considering the purpose of language learning. Tyler (2012) addresses this issue clearly by citing Moder's (2010) study.

> She (Moder) notes that, of course, frequency of input is crucial to L2 learning, but if the end goal is to achieve L2 speakers whose language approaches that of the NS (native speaker), the input has to reflect how NSs use the language. I would add that this is particularly true for instructed L2 learners whose exposure is often limited to unnatural, textbook language and faulty explanations of various components of the language. (p. 85)

Recently, a number of studies have examined textbooks' linguistic features. Previous studies on language textbooks have generally considered the lexicon and sentence structure. Alsaif and Milton (2012) and Davis and Face (2006) examined the sizes of the vocabularies in L2 textbooks in English and Spanish, respectively. They found that the proportion of frequently used vocabularies in L1 contexts decreased as the proficiency level of the textbooks increased. That is, lower level textbooks featured more commonly used words and upper level textbooks introduced less commonly used words. Studies that investigated sentence structure generally indicated that certain types of sentence patterns were over- or underused when compared to L1 use (see Gilsan and Drescher (1993) for L2-Spanish textbook analysis and Römer (2004) for L2-English textbook analysis).

Taken together, studies in favor of the usage-based approach indicated a positive relationship between input and output, and they generally agreed that language learners' perceptions and understanding of a language were greatly affected by how they were exposed to it. Textbook studies also addressed the need for an examination and utilization of an L1 corpus in making textbooks so that L2 language learners would be more engaged in interpreting their use of language, such as what types of lexicon and patterns are used in which contexts. Setting against this backdrop, the present study investigates how the locative postposition -*ey* is used in L1-Korean and L2-Korean textbooks.

2.2 Corpus studies on Korean postposition -*ey* in L1

In terms of studying the usage of postpositions, a limited number of studies have examined the frequency of postpositions in L1-Korean (e.g., Kang & Kim 2009, Seo 2006, 2014). Kang and Kim (2009), for instance, analyzed 15 million *eojeols* in the Sejong written corpus.[2] They showed that the most commonly used postposition was -*(l)ul* (accusative, 19.88%), followed by -*i/ka* (nominative, 14.39%). Topic/contrastive marker -*(n)un* (14.1%) and genitive marker -*uy* (12.97%) were also frequently used. Postposition -*ey* was the fifth most frequent postposition and

[2] An *eojeol* is a white-space-based unit. It serves as the minimal unit composing a sentence in Korean (Lee 2011).

the first among adverbial postpositions, representing about 9.9% of the total postposition usage.

Because *-ey* is a popular adverbial postposition, a number of studies have defined the fundamental concept and/or multiple functions of *-ey* (Chae & Lim 2013, Kang 2012, Ko 2011, Ko & Ku 2008, Lee 1981, Lee 2004, Lim 2017, Maeng 2016, Nam 1993, Park 2012, Sohn 1999). They showed that *-ey* is used to indicate a location when it is attached to a locative noun, as *pang-ey iss-ta* '(someone) is in one's room' and *hakkyo-ey ka-ta* '(someone) goes to school'. When it is combined with a stative verb, such as *iss-* 'to be/exist', the location is the place where the subject locates/exists, while the location is the destination a subject is heading toward with a motion verb. Other functions include time (*twu si-ey* 'at two o'clock' and *swuyoil-ey* 'on Wednesday'), cause (*pi-ey* 'because of rain'), scope (*kkochpath-ey mwul-ul cwu-ta* '(someone) gives water to the flowers/waters the garden'), means (*khal-ey tachi-ta* '(someone) hurts with the knife'), environment (*i nalssi-ey naka-ta* '(someone) goes out in this weather'), etc.

Among those studies, Nam's (1993) research was one of the earliest to utilize L1 corpora to picture *-ey*'s usages. It analyzed 3,875 examples of *-ey* in the Yonsei written corpus and delineated detailed functions of *-ey* and their characteristics. The major functions of *-ey* in the study were similar to those found in previous studies: location, time, cause, etc. In addition to the functions, Nam's study examined the characteristics of the use of *-ey* in various ways. First, the transitivity of a verb that co-occurred with *-ey* was studied. For instance, both intransitive and transitive verbs can be used for locative functions as in *Mina-ka hakkyo-ey ka-ta* 'Mina goes to school' and *Minho-ka yelsoy-lul chayksang-ey twu-ta* 'Minho puts keys on the table'. Second, sentence patterns in relation to the verb were studied. A transitive pattern with locative information such as [NP1-*i/ka* + NP2-*(l)ul* + NP3-*ey* + Verb (NP1: subject, NP2: object, NP3: adverbial phrase for location)] can occur with various types of verbs, including 'attach'-type (e.g., *tay-*, *pwuthi-*, *is-*) and 'add'-type (e.g., *sekk-*, *tha-*, *pipi-*) verbs. Third, it studied whether *-ey* can be replaced with other postpositions or not; another type of postposition, *-eytaka*, can be used in *pang-ey/eytaka twu-da* 'to put (something) in the room'. Nam's research is remarkable for its utilization of an L1 corpus to thoroughly investigate functions and their syntactic characteristics along with verb types that occur with *-ey*. However, it did not touch upon the quantitative aspect of the usages. For instance, the proportional usage of each function will show which function has a dominant status over the others. In addition, considering the numerous verb types used in the L1 environment, an examination of the percentage of each verb type could suggest which verb type held a prototypical position with the use of *-ey*.

Türker (2005) examined the proportions of each function of *-ey* in the L1 environment. The study categorized four *domains* (Locational, Temporal, Abstract, and Logical) of *-ey*, each of which was composed of a number of *senses* (e.g., proto-goal sense, temporal-location sense, additive sense, etc.). To examine the proportion of each sense, it randomly extracted one part (about 4,000 sentences) of the *CetCon* corpus, which is a tagged L1 corpus created by Korea University, and collected tokens of *-ey* in that part. The results showed that two major senses were dominant: the proto-goal sense (23%) and the proto-location sense (18%). Türker's (2005) study is meaningful because it tried to show the proportional use of each function (sense) of *-ey* by L1-Korean users. The weakness of the study is the lack of information it provides about the corpus type and size. It is not clear how the *CetCon* corpus was composed, i.e., whether the institution collected spoken conversations or academic writings. How many tokens of *-ey* were found in the 4,000 sentences is not stated clearly either, which leaves a question regarding the size of the corpus.

Using corpora in addressing linguistic inquiries is a growing trend in that corpus analysis reveals the actual language use relating to the inquiries. Corpus studies of the postposition *-ey* have thus far shown diverse aspects of its use. In general, however, a close examination of the linguistic environment, such as what types of verb are frequently used in a certain function, is either missing or limited in those studies. To address this issue, the present study plans to report the frequency and types of verbs that co-occur with the locative *-ey* in the Sejong written and spoken corpora. The results will be used as a reference to examine the use of *-ey* in textbooks for L2-Korean learners, which is the main purpose of this study.

The current study focuses on the locative function as a prototypical one. Some recent studies have mostly examined the relations among various functions of *-ey* by setting a prototypical meaning and considering how other functions deviate from it (Kang 2012, Lee 2004, Maeng 2016, Türker 2005). As was shown in Kang's (2012) three *clusters* (Location, Goal, and Inclusion) and Türker's (2005) four *domains*, the locative function (including Goal cluster in Kang) works as a primitive function of *-ey*. In addition, the use of the locative function is strongly associated with linguistic environments (e.g., verb type), unlike other functions; time expression with *-ey* such as *seysi-ey* 'at 3:00' can occur regardless of verb type. However, quantitative investigations of verb use with *-ey* are rare. Thus, this investigation of various aspects of the locative function as a prototypical function from which other functions originated and its environment will offer a clearer understanding of the use of *-ey*.

2.3 Corpus studies on the Korean postposition *-ey* in L2

In the field of KSL/KFL, postposition studies have mostly concentrated on either error analysis or textbook analysis. Research on error analysis has generally shown that postposition error is one of the most common errors (Ko et al. 2004, Lee 2003) made by language learners of Korean, and it has been also noted that even advanced learners still produce a considerable number of errors (Han 2014, Lee, 2003, Min 2002). Among various postpositions, *-ey* was one of the common postpositions around which errors were made (Kim 2002). For instance, Han (2014) analyzed 738 postposition errors by advanced learners in spoken discourse, finding that errors related to *-ey* formed the second most common postposition error.

Amongst error-focused studies, Kim and Guo (2016) designed two types of speech tasks to analyze learners' use of multiple functions of *-ey*. They examined 1,374 tokens of *-ey,* which were collected from a total of 47,807 *eojeols*, in 80 intermediate- and advanced-level learners' utterances. In the same tasks, a total of 15,040 *eojeols* were produced by 20 L1 users, among which 260 tokens of *-ey* were found. To determine the appropriateness of the usages, the accuracy rate was calculated by dividing the number of correct usages by the total number of tokens per function. The average rate of correct usage of 13 functions for L2 learners was 72.72%, whereas it was 96.92% for L1 users. In particular, the average rate of accurate use of the 'location/position/existence' function was second to lowest (49 accurate use out of 97 total use: 50.52%) among all functions. The rate of correct usage for 'direction' was 82.05% (128 accurate use out of 156 total use). The accuracy rates of the two functions for L1 users were both 100% (32 tokens for 'location/position/existence' and 23 tokens for 'direction'). That is, L2 learners used the 'directional' function more often and more accurately than the 'location/position/existence' function. In all, location postposition errors continued to be made by advanced learners, and acquiring knowledge on the correct use of the locative function (both locative-existential and locative-goal/direction) was a challenging task for language learners of Korean.

While many of the studies on errors collected their own learner production data, the need for and attempt to establish a systematic learner corpus of language learners of Korean has been raised in the literature (Ahn & Han 2011, Kang 2011, Kim 2002, Kim 2016, Kim 2017). Studies in general have pointed out that several factors, such as the standardization of annotation methods and learner information, need to be considered to build a reliable and sizable learner corpus for Korean. According to Kang (2011), two Korean universities (Yonsei and Korea) have established their own sizable learner corpora, both of which were composed of about 500,000 *eojeols* at the time of the study. As those corpora were mostly from assignments and writings produced by learners in each university, Kang emphasized the importance of diversifying the sources as well as increasing the size. NIKL also started a nationwide project to establish a learner corpus of Korean in 2015. Kim (2017) summarized how much of the project had been completed and how the corpus could be used in developmental research on Korean.

The significance of error analysis studies is that they show how language learners understand the target language. However, they hardly investigate the nature and tendency of the input that learners receive, which can directly affect learners' understanding of the language. Thus, interest is growing in the effects of input on language learning. Several recent studies in KSL/KFL have examined input materials, most of which were textbooks (Jeong 2011, Kang 2015, Kim 2011, Lee & Ko 2013). Textbook analysis studies have generally agreed on three points. First, that textbooks introduce only a limited number of functions across all levels of learning. Second, that textbooks do not systematically suggest the relationships between functions. And third, that they provide no clear distinction to show the differences among postpositions that share the same function (e.g., *-eyse* and *-(u)lo*).

For instance, Jeong (2011) analyzed three postpositions (*-ey, -eyse* and *-(u)lo*) in six types of textbooks: four types used in universities in Korea and two types of textbooks published by NIKL for use by immigrants to Korea. The study showed that each type of textbook introduced *-ey* in the first-level volumes. In addition to textbooks, Kim (2011) included three types of dictionaries as a reference corpus used to analyze how textbooks reflect the usages of L1. The study showed that there was a total of 14 functions of *-ey* in the dictionaries, of which only a limited number was suggested in the textbooks.

Some studies have tried to connect learners' task results to input materials (Lee & Ko 2013, Kang 2015). Kang (2015) first examined seven types of dictionaries and grammar books for language learners to summarize the functions of three postpositions (*-ey, -eyse*, and *-(u)lo*). Through a comparison of the functions found in those seven references to the functions found in three types of textbooks (two published and used in China and one published and used in Korea), the study indicated that diverse functions of *-ey* were not fully introduced in textbooks. As a second step, the study conducted a cloze test to examine L2 learners' understanding of the functions. Through an analysis of which postpositions were used in place of the correct one and how errors were made, the study concluded that several factors, such as the learner's native language and polysemy of spatial postpositions, affected error production. While the study examined both textbooks and learners' production data, it did not closely link textbooks to the results of the experiment.

The abovementioned textbook studies are noteworthy in that they recognized the importance of language input and suggested ways to examine input materials. However, they mostly focused on qualitative research, presumably due to the lack of an instructional input corpus. As a result, the scope of research was limited to investigating what functions were (not) introduced in the textbooks, not revealing various aspects of input. As was mentioned earlier, input frequency and the environment in which each token appears are crucial factors that define language learners' understanding of the target language system (Ellis 2002). Thus, the present study conducts a quantitative textbook analysis by investigating the distribution of the locative *-ey* in language learner textbooks and considering how it changes as the proficiency level increases. The frequency difference of some verb types between L1 corpora and L2-Korean textbooks will be examined as well.

3. Methods

The development of techniques for the establishment and analysis of large corpora has enabled in-depth investigations of linguistic environments. Myriad inquiries can be made by using corpora, including asking questions about the linguistic environment such as the context in which a certain item occurs more/less often or with which element an item occurs. Several scholars have proposed that recognizing contextual diversity is a powerful way to reveal the property of an item (cf. Divjak & Caldwell-Harris 2015) and that considering the environment provides a better understanding of the token (Gries et al. 2005, Jurafsky 1996).

In this vein, this study investigates language use from a corpus-based approach. To examine the characteristics of language input that are provided to language learners of Korean, which is the main purpose of this study, two types of textbooks are analyzed. Each type of textbook has two volumes for each of the first four proficiency levels. From the fifth level, the topics covered by Type-1 textbooks (T1) diverge to focus on specific language skills such as reading, speaking, grammar, and vocabulary. For a comprehensive investigation of how textbooks introduce the use of postpositions, rather than specific language use skills, a total of 16 volumes (eight per each type) used in the first four proficiency levels for T1 and Type-2 textbooks (T2) were considered. The size of the textbooks by level is shown in Table 1. T1 displays a number of listening-oriented practices (e.g., true/false questions after listening, summarizing, or writing about what learners heard) in each lesson. Scripts of these listening sources were given as appendices in each volume. The current study included those scripts in the analysis as the tasks/practices in each main lesson were based on the listening content.

Table 1. Data breakdown for T1 and T2 (number of *eojeols*)

	Level 1	Level 2	Level 3	Level 4	Total
T1	10,403	22,539	22,761	34,356	90,059
T2	13,708	20,360	19,599	38,977	92,644

All the sentences in the textbooks were manually converted to a machine-readable format. Tokenization (separating each *eojeol* in a sentence) and Part-of-Speech (POS) tagging, which enable researchers to select a specific sentential element such as nouns and verbs easily, are completed using *UDpipe* (Straka et al. 2016).[3] Before extracting the postposition *-ey*, complex sentences were divided into phrases with a single subject and a verb based on the existence of a verb in the middle of the sentences.[4] For instance, sentence (1) was divided into two phrases, *Minsu-ka hakkyo-ey ka-ss-ko* 'Minsu went to the school' and *hakkyo-eyse chinkwu-lul manna-ss-ta* '(he) met (his) friends in the school', based on the verb *ka-* 'to go' being in the middle of the original sentence.

(1) Example of complex sentence
Minswu-ka hakkyo-ey ka-ss-ko, hakkyo-eyse chinkwu-lul manna-ss-ta.
Minswu-NOM school-ey go-PST-and, school-LOC friend-ACC meet-PST-SE
Minswu went to the school, and (he) met (his) friends in the school'

Next, phrases that contained the postposition *-ey*, like the first phrase in (1), were selected using Python 3.7. Adverbial postpositions such as *-ey, -eyse*, *-eykey*, *-(u)lo*, and *-hanthey* are marked as JKB according to the Sejong tagging scheme (Kim et al. 2007). Thus, phrases that contained a JKB-tagged word were extracted first. Among them, instances that had a syllable *-ey* (e.g., *-ey, -eyse,* and *-eykey*) were selected, followed by removing other postpositions (such as *-eyse* and *-eykey*) one by one except for *-ey*.

The locational function of all the phrases that contained the postposition *-ey* was manually checked. The function was examined in accordance with Nam's (1993) criteria. Broadly, the book grouped locations into five categories: existential location (e.g., *-ey iss-* 'to be/exist (somewhere)'), activity location (e.g., *-ey palsayngha-* 'to occur (in a place)'), destination (e.g, *-ey ka-* 'to go (somewhere)'), contact location (*-ey pwuth-* 'to attach'), and source (e.g., *-ey nao-* 'to appear (somewhere)/to be based on (something)'). Nam also classified verb types into several sections and listed examples. For instance, put-type verbs include *twu-*,

[3] http://ufal.mff.cuni.cz/udpipe

[4] This idea is based on the definition of a clause as a language unit that contains a verb (Kroeger 2005).

namki-, *sewue- cinyelha-* and *paychiha-*, which all have a basic meaning of 'put.' Those lists were used as a reference to designate a verb as a locative function with *-ey* in the present study. As for the phrases that were finally selected as presenting the locative *-ey*, verb-tagged *eojeols* were collected via Python.

As a final step, verbs that were collected were lemmatized and counted manually. In the present study, the lemma form of a verb (token frequency) was counted without consideration of other markers such as tense or modality. For instance, both *ka-* (present form of 'to go') and *ka-ss-* (past form of 'to go') were counted as instances of the verb *ka-* 'to go'. The total number of locative *-ey* extracted from each proficiency level of T1 and T2 is shown in Table 2.

Table 2. Raw frequency of the locative *-ey* in L2-Korean textbooks

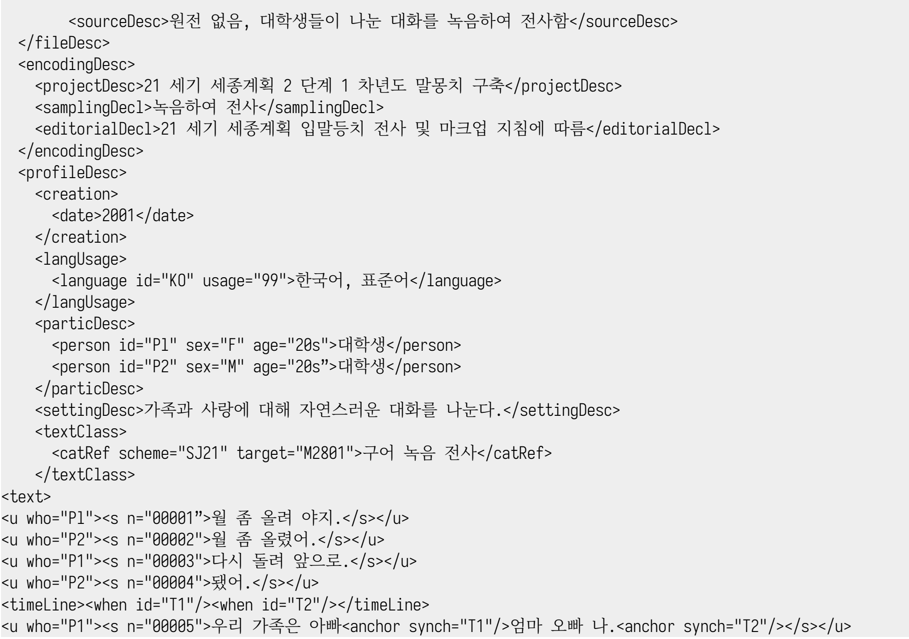
	Level 1	Level 2	Level 3	Level 4	Total
T1	121	152	169	203	645
T2	83	148	98	296	625

For L1 reference corpora in relation to L2 textbook corpora, the Sejong written and spoken corpora were used.[5] The Sejong written corpus comprises various genres of printed books as well as newspaper/magazine articles from different sectors. In the current study, part of the *UCorpus* (Lim et al. 2015) provided by the University of Ulsan was used as a written corpus because it provided a set of POS tags and words from the Sejong written corpus, which eased the process of extracting the target feature from the corpus. A total of 1,124,000 *eojeols* from the *UCorpus* were analyzed in the current study. Extraction of the locative *-ey* in the written corpus was nearly the same as in the textbook corpus except for the POS tagging process.

The Sejong spoken corpus is composed of conversations between friends, lectures/ presentations, sports broadcasts, sermons, etc. About 800,000 *eojeols* from the spoken corpus were analyzed. Extraction of the locative *-ey* in the Sejong spoken corpus was mostly done manually for time management purposes. More specifically, the extraction of the locative *-ey* from the corpus involved several (semi-) automatic and manual steps, as was shown in the textbook and L1 written corpora cases. To complete the tokenization and POS tagging, the Sejong spoken corpus texts needed to be cleaned first. That is, in the original form of the corpus, all the information about the text (e.g., where it occurred and who was participating) was included in each file along with various symbols, as in Figure 1. They had

[5] Sejong written and spoken corpora are managed and provided by NIKL in a text format. They can be found at https://www.korean.go.kr.

to be removed before the automatic extraction of target features from the corpus. Even after the cleaning, other works, including checking the functions of each instance of *-ey* and the lemmatization of verbs, required a manual check. What made the process more complicated was the frequent use of scrambled order in a spoken context, which might prevent an automatic program from detecting the right verb for the postposition. For example, sentence (2) is composed of two phrases: '*i-ke neh-eyo, kapang-ey* 'put this one, inside the bag', and *kuliko ka-yo* 'and let's go'. However, the automatic process will divide it into 'put this one' and 'inside the bag, and let's go' due to its scrambled order. Then, the verb *neh-*'to put', which occurred with *kapang-ey* 'inside the bag' will not be detected.



```
        <sourceDesc>원전 없음, 대학생들이 나눈 대화를 녹음하여 전사함</sourceDesc>
  </fileDesc>
  <encodingDesc>
    <projectDesc>21 세기 세종계획 2 단계 1 차년도 말뭉치 구축</projectDesc>
    <samplingDecl>녹음하여 전사</samplingDecl>
    <editorialDecl>21 세기 세종계획 입말등치 전사 및 마크업 지침에 따름</editorialDecl>
  </encodingDesc>
  <profileDesc>
    <creation>
      <date>2001</date>
    </creation>
    <langUsage>
      <language id="KO" usage="99">한국어, 표준어</language>
    </langUsage>
    <particDesc>
      <person id="Pl" sex="F" age="20s">대학생</person>
      <person id="P2" sex="M" age="20s”>대학생</person>
    </particDesc>
    <settingDesc>가족과 사랑에 대해 자연스러운 대화를 나눈다.</settingDesc>
    <textClass>
      <catRef scheme="SJ21" target="M2801">구어 녹음 전사</catRef>
    </textClass>
<text>
<u who="Pl"><s n="00001”>월 좀 올려 야지.</s></u>
<u who="P2"><s n="00002">월 좀 올렸어.</s></u>
<u who="P1"><s n="00003">다시 돌려 앞으로.</s></u>
<u who="P2"><s n="00004">됐어.</s></u>
<timeLine><when id="T1"/><when id="T2"/></timeLine>
<u who="P1"><s n="00005">우리 가족은 아빠<anchor synch="T1"/>엄마 오빠 나.<anchor synch="T2"/></s></u>
```

Figure 1. Original text in the Sejong spoken corpus

(2) Example of complex sentence (scrambled)
i-ke neh-eyo, kapang-ey, kuliko ka-yo.
this-one put-SE, bag-ey, and go-SE
Put this one in(side) the bag, and let's go.

Due to all these considerations involved in examining the spoken corpus, the present study directly read the L1 spoken corpus using a text-readable program (MS Word). While searching for the postposition *-ey*, each verb's locative function and lemma form were checked at the same time. When the instance matched with the location function, the lemma form of the verb was listed in a separate file. Then, the number of tokens was finally calculated. The total raw frequency of the locative *-ey* extracted from L1 corpora was 10,024 (written) and 2,573 (spoken). The normalized frequency (per million *eojeols*) of the locative *-ey* from four corpora is shown in Table 3.

Table 3. Normalized frequency of locative *-ey* (per million *eojeol*)

Sejong written	Sejong spoken	TI	T2
8,919	3,217	7,162	6,747

In addition to the examination of verb types and frequencies, keyness analysis was conducted to understand which items (verbs) were significantly frequent in a certain corpus in comparison to a reference corpus. Keyness analysis is commonly used in frequency comparison (Fraysse-Kim 2010, Kilgarriff 2001, 2005, Leone 2010), and either log-likelihood (*LL*) or Chi-square statistics (Biber et al. 2007) is represented. The present study used *LL*, which is calculated by using the observed (raw) frequency and expected frequency of a target item. For instance, in a contingency table (Table 4) of a target item (*word*) in Corpus 1 (C1) and Corpus 2 (C2), 'a' and 'b' are raw frequencies of *word*, where 'c' and 'd' are the total number of tokens in each corpus. In this case, the expected frequency of *word* in C1 (E_{11}) and C2 (E_{12}) equals c*(a+b)/(c+d) and d*(a+b)/(c+d) each (see 3).

Table 4. Contingency for the log-likelihood of '*word*' (adapted from Kilgarriff, 2001)

	Corpus 1	Corpus 2	Totals
word	*a*	*b*	*a+b*
not *word*	*c–a*	*d–b*	*(c+d)–(a+b)*
Totals	*c*	*d*	*c+d*

Note. a, b, c, and *d* represent frequencies.

(3) $O_{11}=\mathrm{a}$, $O_{12}=\mathrm{b}$, $E_{11}=\mathrm{c}*(\mathrm{a}+\mathrm{b})/(\mathrm{c}+\mathrm{d})$, $E_{12}=\mathrm{d}*(\mathrm{a}+\mathrm{b})/(\mathrm{c}+\mathrm{d})$
When *i*: row #, *j*: column #, O*ij* = Observed frequency of *i*th row and *j*th column, and *Eij* = Expected frequency of *i*th row and *j*th column

With raw and expected frequencies, *LL* is calculated using the formula in (4). For calculation purposes, zero occurrence was changed to one quadrillionth (Gabrielatos 2018). The higher the *LL*, the more likely a significant frequency difference between two corpora will be. The significance of *LL*, which is shown through its *p*-value, is the probability that the keyness is incidental (Biber et al. 2007).

(4) $LL = 2\sum(o_{ij} \ln \frac{Oij}{Eij})$

Note. ln = natural log

As even a weak relationship reveals statistical significance when the corpus size is large (Howell 2010, Kilgarriff 2005), effect size (the size of a frequency difference; Rosenfeld & Penrod 2011) was calculated as well. The present study adopted *%DIFF*, which considers the normalized frequency of an item in both compared and reference corpora. Using the normalized frequency in Corpus 1 (NFC1) and the normalized frequency in Corpus 2 (NFC2), *%DIFF*=(NFC1–NFC2)*100/NFC2, when C1 is a compared/study corpus and C2 is a reference corpus (adapted from Gabrielatos 2018). The value of *%DIFF* shows how often a certain item is used compared to the reference corpus. For instance, a *%DIFF* value of 300 means C1 (study corpus) used the item four times as often as C2. To illustrate, when NFC1 is 400 and NFC2 is 100 (showing four times more use in C1), *%DIFF* is 300 via the formula (=(400 – 100)*100/100). A negative *%DIFF* value shows less use of an item in C1 (e.g., *%DIFF*=–60 means 60% less use in C1). The negative limitation of *%DIFF* is –100 (no occurrence in C1), and there is no limitation on a positive *%DIFF* (no occurrence in C2).

The threshold for statistical significance in keyness analysis varies across studies (Gabrielatos 2018). Considering the sizes of corpora and the number of *Candidate Key Items* (CKIs) that the analysis will yield, the keyness analysis in the current study was conducted with a confidence level of 99.9% ($p < 0.001$, critical value of *LL*: 10.83). L1 Sejong corpora were used as reference corpora and L2-Korean textbook corpora were used as study corpora.

4. Results and discussion

4.1 L1: Sejong corpora

In the Sejong written corpus, a total of 894 verb types occurred in 10,024 locative *-ey* tokens. Among them, a few dominant verbs occupied a huge proportion (33 verb types occupied about 50%) and more than 800 verbs were used fewer than 20 times, following Zipf's law (1935).[6] The most frequent verb was *iss-* 'to be/exist' (9.45%) followed by *ka-* 'to go' (4.37%). In the Sejong spoken corpus, 256 verb

[6] Zipf's (1935) law explains that one item (or a limited number of items) occupies a larger proportion in natural language use and the proportions of the following items decrease exponentially. The tendency of Zipfian distribution of verbs with locative *-ey* in L1-Korean and L2-Korean textbooks can be found at Jung (2020).

types were used in 2,573 locative *-ey* tokens. Among these verb types, 30 verbs occurred more than 10 times, accounting for 82.55% of the total number of tokens, and the remaining 226 verb types occurred fewer than 10 times. The frequency distribution of the verbs in the spoken corpus also followed the Zipfian distribution. The most frequent verb was *iss-* 'to be/exist' (31.68%) followed by *ka-* 'to go' (14.46%). Tables 5 presents the 10 most frequently used verbs in the L1 written and spoken corpora along with their proportions. The 10 most frequently used verbs accounted for about 30% and 70% of the total use in the written and spoken corpora, respectively.

Table 5. Raw frequency and proportion of verbs with locative *-ey* in the L1 corpora

	L1 written		**L1 spoken**	
	Verb	**Token # (%)**	**Verb**	**Token # (%)**
1	*iss-* 'to be/exist'	947 (9.45)	*iss-* 'to be/exist'	815 (31.68%)
2	*ka-* 'to go'	438 (4.37)	*ka-* 'to go'	372 (14.46%)
3	*tuleka-* 'to enter/go in'	291 (2.90)	*tuleka-* 'to enter/go in'	138 (5.36%)
4	*anc-* 'to sit'	269 (2.68)	*o-* 'to come'	132 (5.13%)
5	*ilu-* 'to arrive'	251 (2.50)	*nao-* 'to come out'	99 (3.85%)
6	*se-* 'to stand'	218 (2.17)	*sal-* 'to live'	52 (2.02%)
7	*o-* 'to come'	186 (1.86)	*naka-* 'to go out'	51 (1.98%)
8	*sal-* 'to live'	158 (1.58)	*anc-* 'to sit'	50 (1.94%)
9	*ppaci-* 'to fall into'	146 (1.46)	*neh-* 'to put (inside)'	49 (1.90%)
10	*neh-* 'to put (inside)'	144 (1.44)	*nam-* 'to remain'	36 (1.40%)

4.2 L2: Textbooks for learners

The number of locative *-ey* in the T1 proportionally increased as the proficiency level increased, as Table 6 shows (level1: 121, level 2: 152, level 3: 169, level 4: 203). The number of verb types utilized also increased (level1: 8, level 2: 22, level 3: 33, level 4: 61) as the level changed. The most frequently used verb that occurred with *-ey* was *ka-* 'to go' across all the levels, followed by *iss-* 'to be/exist' (level 1, level 3, level 4) or *o*-'to come' (level 2). T2 generally showed a similar tendency with T1 in relation to verb use with the locative *-ey*. Token numbers of the locative *-ey* in the T2 also increased as the textbook proficiency level increased (level1: 83, level 2: 148, level 3: 98, level 4: 296) except in the level-3 volumes, as Table 6 indicates. The number of verb types also increased from level 1 to level 4 (level1: 7, level 2: 29, level 3: 36, level 4: 114). The two most frequently used verbs were *ka-* 'to go' and *iss-* 'to be/exist' for the first three levels of textbooks, and *iss-* 'to be/exist' and *ssu-* 'to write' in the fourth level.

Table 6. Raw frequency and proportion of verbs occurring with locative *-ey* in T1 and T2 by level

	T1		**T2**	
	Verb type	**Token # (%)**	**Verb type**	**Token # (%)**
Level1	*ka-* 'to go'	85 (70.25)	*ka-* 'to go'	38 (45.78)
	iss- 'to be/exist'	17 (14.05)	*iss-* 'to be/exist'	21 (25.30)
	o- 'to come'	8 (6.61)	*o-* 'to come'	13 (15.66)
	Others	11 (9.09)	Others	11 (13.25)
	Total	**121 (100)**	**Total**	**83 (100)**
Level2	*ka-* 'to go'	70 (46.05)	*ka-* 'to go'	74 (50.00)
	o- 'to come'	16 (10.53)	*iss-* 'to be/exist'	11 (7.43)
	iss- 'to be/exist'	13 (8.55)	*o-* 'to come'	9 (6.08)
	Others	53 (34.87)	Others	54 (36.49)
	Total	**152 (100)**	**Total**	**148 (100)**
Level3	*ka-* 'to go'	59 (34.91)	*ka-* 'to go'	24 (24.49)
	iss- 'to be/exist'	23 (13.61)	*iss-* 'to be/exist'	10 (10.20)
	o- 'to come'	12 (7.10)	*nao-* 'to come out'	5 (5.10)
			o- 'to come'	5 (5.10)
	Others	75 (44.38)	Others	54 (55.10)
	Total	**169 (100)**	**Total**	**98 (100)**
Level4	*ka-* 'to go'	31 (15.27)	*iss-* 'to be/exist'	30 (10.14)
	iss- 'to be/exist'	29 (14.29)	*ssu-* 'to write/use'	15 (5.07)
	o- 'to come'	15 (7.39)	*ka-* 'to go'	13 (4.39)
	Others	128 (63.05)	Others	238 (80.41)
	Total	**203 (100)**	**Total**	**296 (100)**

Note. Verb types included in 'Others' in T1 by level were as follows: 5 in Level 1; 19 in Level 2; 30 in Level 3; 58 in Level 4. Verb types included in 'Others' in T2 by level were as follows: 4 in Level 1; 26 in Level 2; 32 in Level 3; 111 in Level 4.

Table 7 lists the five verbs most frequently used with the locative *-ey* when all the volumes were considered together in T1 and T2, respectively. As the table shows, *ka-* 'to go' was the most frequently used verb and *iss-* 'to be/exist' the second in both types. In addition, the frequency of verbs decreases almost exponentially, following the Zipfian distribution. In all, both L1-Korean corpora and L2-Korean textbook corpora showed similar trends in that there was one dominantly employed verb with the locative *-ey* and the verb distribution followed Zipf's law. However, there were differences between L1 and L2 corpora regarding verb frequency and which verb type was more distinctively used, which will be discussed further in Section 4.3.

Table 7. Five most frequently used verbs with locative *-ey* across all volumes of T1 and T2

	T1		T2	
	Verb type	**Token # (%)**	**Verb type**	**Token # (%)**
1	*ka-* 'to go'	245 (37.98)	*ka-* 'to go'	149 (23.84)
2	*iss-* 'to be/exist'	82 (12.71)	*iss-* 'to be/exist'	72 (11.52)
3	*o-* 'to come'	51 (9.91)	*o-* 'to come'	36 (5.76)
4	*nao-* 'to come out'	23 (3.57)	*sal-* 'to live'	25 (4.00)
5	*tuleka-* 'to enter/go in'	23 (3.57)	*ssu-* 'to write/use'	19 (3.04)

4.3 Comparison of L1 and L2 corpora

To examine if token numbers show significant differences between L1 corpora and L2 textbooks, keyness analysis was conducted for the 10 most frequent verbs with a locative *-ey* in the L1 Sejong corpora (written and spoken; Table 5). Using the L1 written and spoken corpora as references, Table 8 shows which verbs (among 10 verbs) were over- or underused in both textbooks. Among the CKIs, some verbs show similar tendencies in both textbook types.

First of all, in comparison with the Sejong written corpus, *ka-* 'to go' and *o-* 'to come' were overused in both T1 and T2. As for *ka-* 'to go', it was used about five times (T2) and eight times (T1) more. The verb *o-* 'to come' was used three to four times more often in both textbook types. In contrast, *ilu-* 'to arrive' did not occur in either type, while the proportion of the verb occurring in the L1 written corpus was comparatively high. Second, compared to the Sejong spoken corpora, *ka-* 'to go' was overused (1.5~2.5 times) and *iss-* 'to be/exist' was underused (around 60% less use than in L1 spoken) in both textbook types. Finally, the over- or underuse of some verbs (e.g., *anc-* 'to sit', *sal-* 'to live', *naka-* 'to go out', and *tuleka-* 'to go in') seemed to have been caused by the specific topics covered in each type of textbook, which potentially shows the unique characteristics of each of them.

Changes in the proportion of the three most frequently used verbs in both T1 and T2 (*ka-* 'to go', *iss-* 'to be/exist', and *o-* 'to come'; Table 7), as they were also determined to be CKIs, were examined in more detail. Figure 2 shows how the proportions of these three verbs changed according to the textbooks' proficiency levels, together with percentages of those verbs in the L1 corpora. In general, the proportions of verbs decreased as the textbook proficiency changed, which seems natural when we consider that textbooks introduce more verb types as the proficiency level increases. In the case of *ka-* 'to go', the proportion of instances found in books of level-4 proficiency approached that of the L1 spoken (T1) and L1 written (T2) corpora. The proportion of *iss-* 'to be/exist' in level 4 was similar to L1 written and far less than L1 spoken proportions in both types of textbooks. The

Table 8. CKI verbs used with locative *-ey* in the Sejong corpora and the textbooks

Textbook type	Reference corpus	Verb type	*LL*	%DIFF
T1	Sejong written	*ka-* 'to go'	620.02	768.44
		o- 'to come'	64.80	325.70
		ilu- 'to arrive'	31.72	−100
		anc- 'to sit'	14.56	−76.91
	Sejong spoken	*ka-* 'to go'	178.13	162.73
		iss- 'to be/exist'	84.63	−59.87
T2	Sejong written	*ka-* 'to go'	257.83	445.05
		ilu- 'to arrive'	30.77	−100
		o- 'to come'	30.74	210.11
		sal- 'to live'	15.26	153.52
	Sejong spoken	*iss-* 'to be/exist'	93.67	−63.63
		ka- 'to go'	36.80	64.89
		naka- 'to go out'	22.19	−100
		tuleka- 'to enter/go in'	10.96	−55.25

Note. The *p-value* of *LL* was < .001

use of *o-* 'to come' in level-4 T1 showed a higher proportion than L1 spoken and written corpora. The proportion of *o-* 'to come' in T2 was higher than that of L1 written and lower than that of L1 spoken corpora at level-4 proficiency.

This section showed that there was a dominant verb used with the locative function of *-ey* in L1-Korean corpora and L2-Korean textbook (encompassing all levels) corpora, and token numbers decreased exponentially following the Zipfian distribution. However, the verbs most frequently found in the Sejong corpora and textbooks were different. Dominant use of *iss-* 'to be/exist' in the L1 corpora showed the locative-existential function (e.g., *cip-ey iss-ta* '(someone) is at home') as the prototypical function of the locative *-ey* in L1 use. On the other hand, the locative-destination function of *-ey* represented through *ka-* 'to go' and *o-* 'to come' (e.g., *cip-ey ka-ta* '(someone) goes home') was emphasized in the textbooks as the main function of *-ey* across all textbook levels. Usage-based language acquisition studies have shown the critical role of input properties. In this regard, Kim and Guo's (2016) study is worth revisiting. The study calculated the rate of correct usage of *-ey* in a learners' production corpus, showing that the accuracy rate was lower for the 'location/position/existence' function and higher for the 'direction' function than average. Learners also employed more 'directional' function tokens than 'location/position/existence' function ones. Verb use in the textbooks (dominant use of *ka-* 'to go' over *iss-* 'to be/exist') can explain the results, which can be further investigated in a future study.

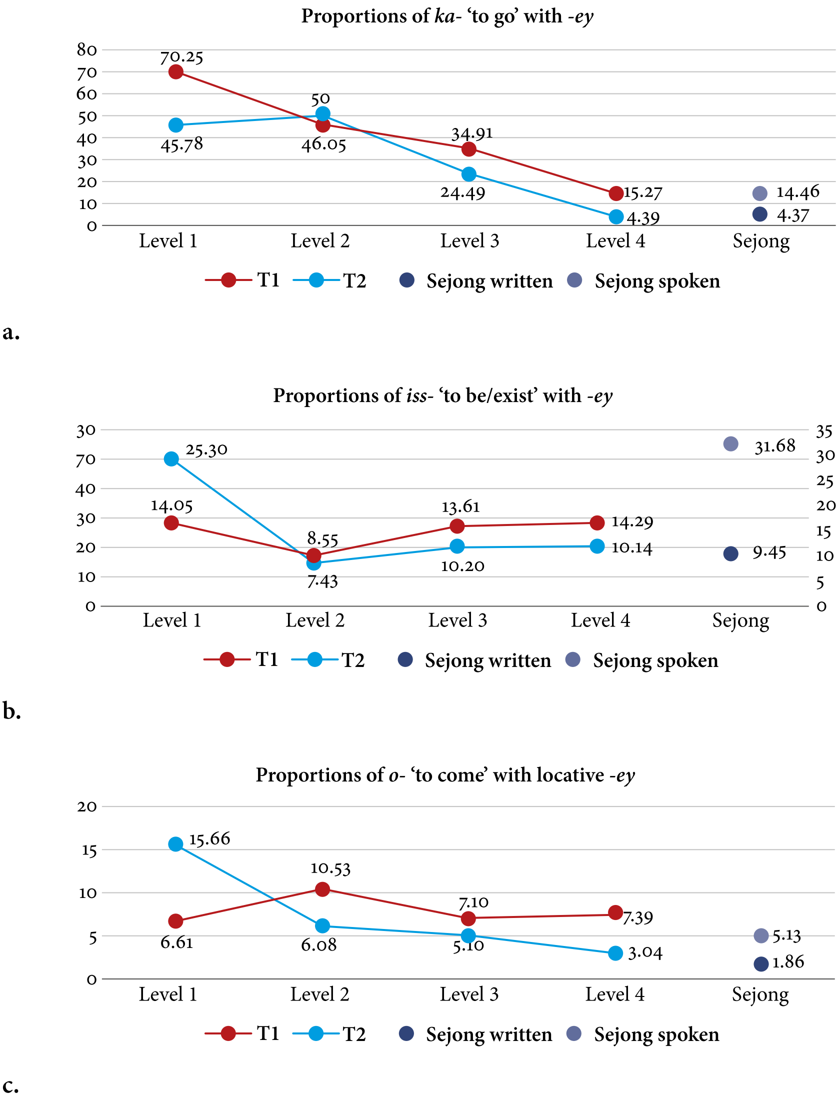


Figure 2. Proportions of three verbs in the L2 textbooks and in the L1 Sejong corpora

5. Implications of the use of Korean corpora for developmental research on Korean language acquisition

The present study investigated the use of the locative *-ey* in Korean L1 corpora (as a proxy for target language use) and Korean language textbooks, which are assumed to be the main input source for language learners in terms of shaping their understanding of the language. By demonstrating how *-ey* is suggested in L1 and different textbook proficiency levels by way of frequency (and proportion) and a keyness analysis of the postposition and co-occurring verbs, this study intended to take a step toward an analysis of how much language textbooks quantitatively reflect native use of the Korean language.

A corpus is a useful source for developmental research on Korean for L2 learners. It has been noted that the dynamism of a verb is one of the main criteria that differentiates *-ey* from other postpositions with similar functions (e.g., *-eyse* and *-(u)lo*) (Ko & Ku 2008, Sohn 1999). However, not all of the instances can be explained by dynamism. For instance, L1 corpora analysis in relation to verb use with *-ey* in the present study indicated that a number of verbs are used for both *-ey* and other locative postpositions (e.g., *na-* 'to appear/sprout' in *yeki-ey/eyse na-ta* '(something) appears here/from here'). As previous studies on textbooks have indicated, most textbooks introduce *-ey* for a static place/destination and *-eyse* for dynamic location/starting point (Jeong 2011, Kim 2011). This dichotomy of postpositions that share similar functions in the textbooks may confuse language learners when they are exposed to the natural language use environment, which is not the consistent with what is presented in the textbooks. Thus, corpus-based analysis is invaluable in considering how to approach linguistic features and what pedagogical strategies can be used for the language development of L2 learners.

In addition, (the analysis of) a corpus can be used practically in language learning-teaching contexts. First, the results in the present study have shown that a small number of verbs accounts for a large proportion of the total usage in the L1 corpora, but that the verbs in L1 were not fully reflected in the textbooks. With this understanding, the verbs that were frequent in L1 but not often used in the textbooks can be added to class. The results from frequency-driven research can further be utilized in designing language textbooks as they can introduce and prioritize frequently occurring words or language patterns. Second, the results of the present study pointed to the skewedness of the locative function of *-ey* toward the destination function, which possibly affects learners' perceptions of the locative *-ey* (e.g., Kim & Guo 2016). Based on this understanding of the properties of textbooks, instructors can utilize activities in class to enhance the existential function of *-ey* and provide more opportunities for learners to be exposed to those verbs. Third, well-developed corpora that show the representative characteristics of diverse linguistic environments can also be used as (supplementary) class materials to demonstrate how language use changes in different contexts. For instance, instructors may compare verbs with *-ey* in the written and spoken corpora to explain how they change in informal conversation and formal writing. Taken together, the establishment of linguistic resources based on corpora will enable a systematic approach to language presentation in textbooks and instructional material.

One reviewer of the present study commented that textbooks cannot be considered natural language because they serve specific educational purposes and expressed strong doubts regarding the idea that textbooks could reflect the natural language use of L1, further arguing that reflecting all of the linguistic information (e.g., frequency) about all the linguistic features of language in textbooks is impossible and ineffective. We generally agree that it is not feasible to reflect all the properties of all the linguistic features of Korean in textbooks (which is not an argument that the present study is making either). The present study does not argue that we should not consider educational purposes as well. However, when the goal of learning a second/foreign language is to acquire knowledge of the conventional use of the language by L1 users (Tyler 2012: Section 2.1), the reason for considering the 'educational purpose' is to facilitate language learners' acquisition of the target language as L1. Lee (2017) also reviewed L2-Korean studies conducted under the general premise that language education should be based on the natural use of L1. In this regard, L1 use serves as a reference when writing L2 textbooks. Thus, 'impossibility' (the term from the reviewer's comments) does not indicate that we should stop identifying the colors of language inputs and making efforts to improve language textbooks. In addition, the importance of properties of input (including textbooks) in language acquisition has been emphasized in several previous studies (see Section 2.1). Undertaking an investigation of the properties of input is a critical step in understanding how language learners perceive and shape their linguistic knowledge. By studying language learners' linguistic behavior, all researchers and practitioners involved in Korean pedagogy will be able to find ways to better serve their 'educational purposes.'

Although this study tried to examine language use by relying on some Korean corpora, there are a few limitations. First, this study does not specify the characteristics of the genres of each register. For instance, the L1 Sejong spoken corpus is composed of conversations between friends, church sermons, formal presentations, lectures, and the like, all of which may have different properties. The present study treated them as one corpus because not all of the genres of the corpus were sufficiently large. Examining the (dis)similarities among different genres in a follow-up study would paint a more complete picture of the use of *-ey* in each specific context in Korean. Second, this study only considered language textbooks as an input source. Although an increasing number of researchers have indicated the importance of examining input in studies on language development, not many input resources are available. Moreover, the sources had to be manually converted into a machine-readable format. Thus, the type and amount of input corpora were limited to textbooks in the present study. Finally, this study investigated the properties of input from a frequency perspective. However, various factors (e.g., instructional order and L1 background) affect L2 learners' perceptions of language. Further research should seek to uncover the relations between the individual roles of each factor and L2 language acquisition.

Corpus-driven approaches have been widely adopted in studies of language development research across diverse languages, but few Korean resources are publicly available. Accordingly, demand is growing to establish reliable and sizable Korean language corpora that encompass all aspects of language use from different documents (e.g., an academic written corpus and a casual conversation corpus across generations) and different users (e.g., a language learner corpus with different L1 backgrounds). With these sources, textbook authors and publishers will be able to utilize them in accordance with the focus and purpose of their textbooks. In addition, building up and utilizing more textbooks and input-related corpora, ranging from general purpose textbooks to textbooks for specific skills (e.g., speaking and writing) and input sources used in specific learning environments (e.g., KSL and KFL), will be useful resources in carrying out studies on the acquisition of Korean. Overall, developing language corpora for Korean will eventually expand the scope of research on how Korean is used and the role of linguistic input in Korean language development.

Abbreviation

ACC	accusative marker
LOC	locative marker
NOM	nominative marker
PST	past tense marker
SE	sentence ender

References

Abbot-Smith, Kirsten & Michael Tomasello. 2006. Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *The Linguistic Review* 23.3: 275–290. https://doi.org/10.1515/TLR.2006.011

Ahn, Eui-jeong & Songhwa Han. 2011. A study on the construction and application of YS-KLI corpus 1. *Enesasilkwa kwancem* 28: 153–189.

Alsaif, Abdullah & James Milton. 2012. Vocabulary input from school textbooks as a potential contributor to the small uptake gained by English as a foreign language learners in Saudi Arabia. *The Language Learning Journal* 40.1: 21–33. https://doi.org/10.1080/09571736.2012.658221

Behrens, Heike. 2009. Usage-based and emergentist approaches to language acquisition. *Linguistics* 47.2: 383–411. https://doi.org/10.1515/LING.2009.014

Biber, Douglas, Ulla Connor, Thoams A. Upton, Molly Anthony & Kostyantyn Gladkov. 2007. Rhetorical appeals in fundraising. In *Discourse on the move: Using corpus analysis to describe discourse structure* ed by Douglas Biber, Ulla Connor and Thomas A. Upton, 121–151. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.28

Bybee, Joan. 2008. Usage-based grammar and second language acquisition. In *Handbook of cognitive linguistics and second language acquisition* ed by Peter Robinson and Nick Ellis, 216–236. New York: Routledge.

Bybee, Joan. 2013. Usage-based theory and exemplar representations of constructions. In *Oxford handbook of construction grammar* ed by Thomas Hoffmann and Graeme Trousdale, 49–69. Oxford: Oxford University Press.

Bybee, Joan & Paul Hopper. 2001. *Frequency and the emergence of linguistic structure*. Amsterdam: Benjamins. https://doi.org/10.1075/tsl.45

Chae, Hee-Rahk. & Eunsuk Lim. 2013. An analysis of locative expressions [NP-ey] and [NP-eyse] in Korean. *Korean Journal of Linguistics* 38.4: 997–1026. https://doi.org/10.18855/lisoko.2013.38.4.010

Davis, Mark & Timothy L. Face. 2006. Vocabulary coverage in Spanish textbooks: How representatives is it? In *Selected Proceedings of the 9th Hispanic Linguistics Symposium* ed by Nuria Sagarra and Almeida Jacqueline Toribio, 132–143. Somerville, MA: Cascadilla Proceedings Project.

Divjak, Dagmar & Catherine Caldwell-Harris. 2015. Frequency and entrenchment. In *Handbook of cognitive linguistics* (Vol. 39) ed by Ewa Dąbrowska and Dagmar Divjak, 53–75. De Gruyter Mouton. https://doi.org/10.1515/9783110292022-004

Ellis, Nick & Fernando Ferreira-Junior. 2009a. Construction learning as a function of frequency, frequency distribution, and function. *The Modern Language Journal*, 93.3: 370–385. https://doi.org/10.1111/j.1540-4781.2009.00896.x

Ellis, Nick & Fernando Ferreira-Junior. 2009b. Constructions and their acquisition: Islands and the distinctiveness of their occupancy. *Annual Review of Cognitive Linguistics* 7.1: 188–221. https://doi.org/10.1075/arcl.7.08ell

Ellis, Nick. 2002. Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition* 24.2: 143–188. https://doi.org/10.1017/S0272263102002024

Fraysse-Kim, Soon Hee. 2010. Keywords in Korean national consciousness: A corpus-based analysis of school textbooks. In *Keyness in text: Corpus linguistic investigations* ed by Marina Bondi and Mike Scott, 219–233. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.41.16fra

Gabrielatos, Costas. 2018. Keyness analysis: nature, metrics and techniques. In *Corpus approaches to discourse: A critical review* ed by Charlotte Taylor and Anna Marchi, 225–258. Oxford: Routledge. https://doi.org/10.4324/9781315179346-11

Glisan, Eileen. W. & Victor Drescher. 1993. Textbook grammar: Does it reflect native speaker speech? *The Modern Language Journal* 77.1: 23–33. https://doi.org/10.1111/j.1540-4781.1993.tb01941.x

Goldberg, Adele E. 2019. *Explain me this*. Princeton University Press.

Gries, Stephan Th, Beate Hampe & Doris Schönefeld. 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

Han, Sang-Mee. 2014. An analysis of errors on the usage of postpositions in the discussions of advanced Korean language learners. *Bilingual research* 57: 223–255.

Howell, David. C. 2010. *Statistical methods for psychology* (7th ed.). Belmont, CA: Cengage Wadsworth.

Jeong, Su-jin. 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, Boo Kyung. 2020. Verb use for the locative functions of three adverbial postpositions (*-ey*, *-eyse*, and *-(u)lo*) in Korean: Analysis of L1-Korean corpora and L2-Korean textbooks. Unpublished doctoral dissertation. University of Hawai'i.

Jurafsky, Daniel. 1996. A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science* 20: 137–194. https://doi.org/10.1207/s15516709cog2002_1

Kang, Bum Mo & Hong Gyu Kim. 2009. *Hankwuke sayong pindo* [Usage Frequency of Korean language]. Seoul, Korea: Hankwuk Mwunhwsa.

Kang, Hyoun-Hwa. 2011. A study on designing of Korean learner corpus construction. *Journal of Korealex* 17: 7–42.

Kang, Yue. 2015. A study on education of Korean postposition 'e', 'eseo', 'ro' for Chinese Korean learners. Unpublished master dissertation. Seoul National University.

Kang, Yunkyoung. 2012. Cognitive linguistics approach to semantics of spatial relations in Korean. Unpublished doctoral dissertation. Georgetown University.

Kilgarriff, Adam. 2001. Comparing corpora. *International Journal of Corpus Linguistics* 6.1: 97–133. https://doi.org/10.1075/ijcl.6.1.05kil

Kilgarriff, Adam. 2005. Language is never ever ever random. *Corpus Linguistics and Linguistic Theory* 1.2: 263–276. http://search.proquest.com/docview/85643706/

Kim, Han-saem. 2017. Factors and practice of Korean learner corpus annotation. *Paytalmal* 61: 149–173.

Kim, Hung-gyu, Beom-mo Kang, & Jungha Hong. 2007. 21seyki seycongkyeyhoyk hyentaykwuke kichomalmwungchi sengkwawa cenmang [21st century Sejong modern Korean corpora: Results and expectations]. In *Proceedings of Annual Conference on Human and Language Technology* 31: 311–316.

Kim, Il Hwan. 2016. Hankwuke haksupca malmwungchiuy cwusek kwacengkwa hwalyong pangpep [Annotation process and its use of Korean learner corpus], In *Kwukceyhankwukekyoyukhakhoy Chwunkyeyhakswulpalpyononmwuncip* [Proceedings of Spring conference of the International Association for Korean Language Education], 233–239.

Kim, Seok Ki. 2011. Education method for the adverb postpositions of 'ey', 'eyse', 'lo' in the Korean language. *Kwukhakyenkwulonchong* 8: 199–236.

Kim, Young-joo & Jin Guo. 2016. A study on the acquisition of Korean adverbial case marker *ey* in spoken production by Chinese Korean L2 learners. *Kwukekyoyukyenkwu* [Korean Education Research] 38: 1–26.

Kim, Youngjin. 1999. The effects of case marking information on Korean sentence processing. *Language and Cognitive Processes* 14.5–6: 687–714. https://doi.org/10.1080/016909699386239

Kim, Yu-Mi. 2002. A study of error analysis of Korean learners by using 'Learner Corpus'. *Teaching Korean as a foreign language* 27: 141–168.

Klein, Dan & Christopher D. Manning. 2005. Natural language grammar induction with a generative constituent-context model. *Pattern Recognition* 38.9: 1407–1419. https://doi.org/10.1016/j.patcog.2004.03.023

Ko, Seok-Ju, Mi Ok Kim, Je Yeol Kim, Sang Gyu Seo, Hee Jeong Jung & Songhwa Han. 2004. *'Hankwuke haksupca' malmwungchiwa olyu pwunsek* ['Korean learners' Corpus and error analysis]. Seoul, Korea: Hankwukmwunhwasa.

Ko, Seok-Ju. 2011. A reconsideration of Korean particle 'e's meaning. *Journal of Korean Linguistics* 61: 93–115.

Ko, Young Gun & Pon Gwan Ku. 2008. *Wuli-mal mwunpep-lon* [A grammar theory of Korean]. Seoul: Cipmwuntang.

Kroeger, Paul. 2005. *Analysing grammar: An introduction.* Cambridge. UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511801679

Kyle, Kristopher & Scott Crossley. 2017. Assessing syntactic sophistication in L2 writing: A usage-based approach. *Language Testing* 34.4: 513–535. https://doi.org/10.1177/0265532217712554

Langacker, Ronald. W. 2008. *Cognitive grammar: A basic introduction.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195331967.001.0001

Lee, Chan-kyu & Ye-jin Ko. 2013. The degree of advanced Korean learners' recognition of postposition. *Emwunnoncip* 56: 485–511.

Lee, Ikseop. 2011. *Kwukehakkaysel* [Introduction to Korean linguistics]. Seoul: Hakyensa.

Lee, Jeong-Hwa. 2004. A cognitive account of the Korean locative postpositions -ey and -eyse. *Discourse and Cognition* 11.1: 237–251.

Lee, Jung Hee. 2003. *Hankwuke haksupcauy olyu yenkwu* [Error analysis of Korean learners]. Seoul, Korea: Pakiceng.

Lee, Ki Dong. 1981. The meaning of the postpositions ey and eyse. *Hangul* 173–174: 9–34.

Lee, Su-mi. 2017. Hankwuke haksupcauy malhakiwa ssukiey nathanan ehwi sayonguy congtancek yenkwu [A longitudinal study of vocabulary usage presented in speaking and writing of Korean learners]. *wulimalkul* 74: 183–214.

Leone, Paola. 2010. General spoken language and school language. In *Keyness in text: Corpus linguistic investigations* ed by Marina Bondi and Mike Scott, 235–248. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.41.17leo

Lim, Dong-Hoon. 2017. How to express local concepts in Korean. *Journal of Korean Linguistics* 82: 101–125.

Lim, Soojong, Minjung Kwon, Junsu Kim & Hyunki Kim. 2015. Korean proposition bank guidelines for ExoBrain. In *Proceeding of the 27th Annual Conference on Human & Cognitive Language Technology*, 250–254. Human and Language Technology.

Maeng, Kyung Hum. 2016. Hyentay hankwuke cosa 'ey'uy inciuymilon [Cognitive understanding of modern Korean postposition 'ey']. *The Journal of Korean Studies* 41: 325–366.

Min, Jin Young. 2002. Hankwuke kokup haksupcauy cosa olyu pwunsek [Postposition error analysis of advanced Korean learners]. Unpublished master dissertation. Yonsei University, Korea.

Moder, Carol Lynn. 2010. *To learn effectively: Grammatical constructions for second language grammar instruction.* Paper presented at AAAL: Atlanta, GA.

Nam, Ki Sim. 1993. *Kwuke cosauy yenkwu: 'ey'wa 'ro'lul cwungsimulo* [Study on Korean postpositions: focused on 'ey' and 'lo']. Seoul, Pakiceng.

Park, Jun Seok. 2012. The meaning and functions of particle '-e' in Korean. *Dong-ak Society of Language and Literature* 59: 427–455.

Römer, Ute. 2004. Comparing real and ideal language learner input: The use of an EFL textbook corpus in corpus linguistics and language teaching. In *Corpora and language learners* ed by Guy Aston, Silvia Bernardini, and Dominic Steward, 151–168. Amsterdam; Philadelphia: John Benjamins Pub. https://doi.org/10.1075/scl.17.12rom

Rosenfeld, Barry & Steven Penrod. 2011. *Research methods in forensic psychology*. John Wiley & Sons Inc.

Seo, Sang Gyu. 2006. *Oykwukinul wihan hankwuke haksup sacen* [Learner's dictionary of Korean]. Seoul, Korea: Sinwon Prime.

Seo, Sang Gyu. 2014. *Hankwuke kiponehwi uymi pinto sacen* [Frequency dictionary of Korean basic lexicon and meaning]. Seoul, Korea: Hankwuk Mwunhwasa.

Sohn, Ho Min. 1999. *The Korean language*. Cambridge University Press.

Straka, Milan, Jan Hajič & Jana Straková. 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.

Tomasello, Michael. 2000. First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics* 11.1: 61–82. https://doi.org/10.1515/cogl.2001.012

Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Tono, Yukio. 2004. Multiple comparisons of IL, L1 and TL corpora: The case of L2 acquisition of verb subcategorization patterns by Japanese learners of English. In *Corpora and language learners* ed by Guy Aston, Silvia Bernardini, and Dominic Steward, 45–66. Amsterdam/Philadelphia: John Benjamins Publishing Company. https://doi.org/10.1075/scl.17.05ton

Türker, Ebru. 2005. Locative expressions in Korean and Turkish: A cognitive grammar approach. Unpublished doctoral dissertation. University of Hawaii.

Tyler, Andrea. 2012. *Cognitive linguistics and second language learning: Theoretical basics and experimental evidence*. UK: Routledge. https://doi.org/10.4324/9780203876039

Zipf, George Kingsley. 1935. *The psyhco-biology of language*. Boston: Houghton Mifflin.

Address for correspondence

Boo Kyung Jung
2714 Cathedral of Learning
4200 Fifth Avenue
Pittsburgh, PA 15260
USA

boj11@pitt.edu

Publication history

Date received: 26 February 2020
Date accepted: 2 January 2021