English Teaching, Vol. 76, No. 1, Spring 2021, pp. 3-31
DOI: https://doi.org/10.15858/engtea.76.1.202103.3
http://journal.kate.or.kr

A Usage-Based Study of L2 Constructional Development: Combining Learner Corpus and Experimental Data

Rami Jo and Sun-Young Oh*

Jo, Rami, & Oh, Sun-Young. (2021). A usage-based study of L2 constructional development: Combining learner corpus and experimental data. *English Teaching*, *76*(1), 3-31.

By adopting a usage-based approach to language acquisition, this study investigated the emergence and development of L2 constructional knowledge. A total of 19 English verb-argument constructions (VACs) and their associated verbs were extracted from a learner corpus and three verbal fluency tasks, each conducted in L1 and L2 English and L1 Korean. We compared verb usage in the target VACs across proficiency levels between the L1 and L2 groups and between data types for VAC productivity and verb-VAC associations. The results identified three stages through which Korean learners' VAC knowledge develops in L2 English: emerging through the frequent use of a few general verbs, expanding the range of verbs associated with a VAC to include more specific and prototypical verb types, and then developing them into a creative constructional schema. Moreover, we determined similarities between L1 and L2 English VAC knowledge in higher L2 proficiency levels, as well as L1 Korean influences related to L1 typology and L1 collocational transfer.

Key words: usage-based language acquisition, constructional development, learner corpus, psycholinguistic experiment, L1 influence

---

This work was supported by the Seoul National University Research Grant in 2019 (700-20190044).

*First Author: Rami Jo, Graduate Student, Department of English Language Education, College of Education, Seoul National University
Corresponding Author: Sun-Young Oh, Professor, Department of English Language Education, College of Education, Seoul National University; 1, Gwanak-ro, Gwanak-gu, Seoul 08826, Korea; Email: sunoh@snu.ac.kr

Received 30 December 2020; Reviewed 30 January 2021; Accepted 16 March 2021

1. INTRODUCTION

Usage-based approaches to language acquisition have attracted a great deal of attention as the contrasting perspective to the generativist and nativist view that the language system is largely innate and uninfluenced by individual utterances (Ambridge & Lieven, 2011; Tomasello, 2003). Usage-based approaches posit that language acquisition is an emergent phenomenon (Hopper, 1987) that is constantly shaped by general cognitive processes such as categorization, inference, and entrenchment in language use (Ellis, Römer, & O'Donnell, 2016; Macwhinney & O'Grady, 2015; Tomasello, 2003). Rejecting a traditional dichotomy between lexis and grammar, these approaches argue that language is comprised of thousands of constructions, defined as form-meaning pairings, some aspects of which are not predictable from their component parts[1] (Ellis et al., 2016; Goldberg, 1995). Language learning is, as such, not about mastering rules and their proper application, but a gradual process of building a structured inventory of constructions (Langacker, 1987).

Constructions are existent at various levels of complexity and abstraction and differ in size from morphemes to full utterances (Croft, 2001; Langacker, 1987). Among them, verb-argument constructions (henceforth VACs), which consist of a verb and all the arguments it takes, form the core of sentences and represent prototypical event types that are basic to human experiences (Goldberg, 1995, 2006). VACs carry meaning themselves, independently of specific words in a sentence. A ditransitive construction (VOO), for example, carries the meaning of transfer from one entity to another, and the core meaning does not change even when different verbs are used in this VAC. VACs can be thus considered "good predictors of overall sentence meaning, useful for conveying 'who did what to whom'" (Goldberg, Casenhiser, & White, 2007, p. 73).

Given the crucial importance of VACs in language acquisition, there have been numerous studies on the ontological and epistemological status of VACs in the first language (L1) (Bencini & Goldberg, 2000; Ellis, O'Donnell, & Römer, 2014a; Goldberg, 1995, 2006; Goldberg, Casenhiser, & Sethuraman, 2004; Lieven & Tomasello, 2008; Ninio, 1999; Tomasello, 1992). Research on second language (L2) learners' constructional development, however, has been relatively sparse and has mostly repeated what has been investigated in L1 studies (Ellis et al., 2014b; Eskildsen, 2014, 2017; Gries & Wulff, 2005; Römer, O'Donnell, & Ellis, 2014; Shin, 2010). For example, following research methodologies employed in L1 acquisition studies (e.g., sentence-sorting tasks), Gries and

[1] Simple words like dog and idiomatic expressions such as make hay while the sun shines are both constructions in that the forms are conventionally associated with meanings (e.g., *dog*: a domesticated carnivorous mammal, *make hay while the sun shines*: take the opportunity to do something while the time and conditions are optimal), which are not predictable *per se*.

Wulff (2005) suggested that advanced L2 learners have a similar mental representation of VACs to that of native speakers. Ellis, O'Donnell, and Römer (2014b) extended the psycholinguistic experiments performed in L1 English (Ellis et al., 2014a) to L1 German, L1 Spanish, and L1 Czech learners of English and concluded that the advanced L2 learners had constructional knowledge comparable to that of native speakers.

Thus far, quite a few studies have established the existence of constructional knowledge in advanced L2 learners across various L1 groups. There is still, however, a paucity of information on L2 VAC development across different proficiency levels. Although Römer and Berger (2019) attempted to trace L2 learners' developing constructional knowledge over time, their corpus data yielded only small token numbers of verbs for the majority of selected VACs, and therefore, an additional examination was deemed necessary to back up their findings. Employing a sentence-sorting and a translation task, meanwhile, Kim, Rah, and Hwang (2020) reported stronger constructional sorting and higher translation accuracy in participants with higher L2 proficiency. Considering the few target VACs (i.e., four verbs crossed with four VACs) provided in the task, however, there is a need for further investigation covering more VACs and comprehensive verb usages that L2 learners commonly associate with them.

In response to the increasing need for research on L2 VAC development, the present study aims to investigate Korean learners' VAC knowledge in L2 English through different L2 developmental stages. We also compare their L2 VAC knowledge with the L1 VAC knowledge of native speakers of both English and Korean, not only to show crosslinguistic similarities and differences but also to identify the learners' L1 influence. Furthermore, given the value and necessity of combining data types in studying linguistic phenomena (Ellis & Simpson-Vlach, 2009; Gilquin & Gries, 2009; Gries & Wulff, 2005), this study takes a mixed-methods approach, combining learner corpus and psycholinguistic data. With a large learner corpus and complementary experimental data, we hope to determine the usage-based developmental path of L2 constructional knowledge, which to this point has been exclusively studied using small sets of data (Ellis & Ferreira-Junior, 2009a, 2009b; Eskildsen, 2009, 2012), and also demonstrate crosslinguistic influences at different stages of language acquisition, which have rarely been studied in usage-based second language acquisition (SLA) research.

2. LITERATURE REVIEW

2.1. Usage-based Approaches to Language Acquisition

Usage-based linguistics (UBL) is established on a range of theoretical backgrounds, including cognitive linguistics, construction grammar, corpus linguistics, psycholinguistics, child language acquisition, and more recently SLA (Dąbrowska & Divjak, 2015; Ellis et al., 2016; Langacker, 1987; MacWhinney & O'Grady, 2015; Robinson & Ellis, 2008; Tomasello, 2003). Among the set of ideas uniting these theories, the core principle is that language use is of key importance in language acquisition and development. People learn language by experiencing and using it within specific usage events, and from the accumulated experience of language use, they can statistically derive linguistic patterns as symbolic units, such as form-meaning pairings (Goldberg, 1995; Langacker, 1987). Since the emergence of a language system is perceived as a usage-driven process employing general cognitive skills, these approaches are in stark contrast with Universal Grammar (UG), which views linguistic knowledge as something inherently stored in the mental lexicon. UG proponents claim that language can be learned independently of language use, which can be affected by speech errors (Chomsky, 1988; Pinker, 1994). However, numerous UBL studies have provided empirical evidence for the fundamental role of language use in acquisition by taking advantage of various corpora and computational analyses, as well as psycholinguistic experiments (Bencini & Goldberg, 2000; Dąbrowska & Lieven, 2005; Ellis & Simpson-Vlach, 2009; Gries & Wulff, 2005; Lieven, Behrens, Speares, & Tomasello, 2003; Lieven & Tomasello, 2008).

UBL researchers in L1 acquisition have discovered that children start learning a language based on concrete examples and gradually expand that knowledge into a repertoire of abstract constructions, drawing on recurring experiences (Dąbrowska & Lieven, 2005; Lieven et al., 2003; Tomasello, 2003). Among those early-learned examples, UBL proponents argue that formulaic language plays a crucial role in acquisition (Behrens, 2009; Ellis, 2012). Children initially stick to formulaic language (e.g., *I wanna do it*, *I dunno*), but in their later utterances, those fixed phrases are manipulated and develop into schematic patterns. Dąbrowska and Lieven (2005) analyzed children's utterances using a 'traceback' method and found out that what seemed to be novel utterances originated from past utterances by substituting one content word in a sentence (e.g., *do you want some THING*?). These findings verify the developmental path from a formula via a low-scope pattern to a fully schematic construction in the L1 (Lieven & Tomasello, 2008).

The usage-based language learning trajectory is also reported in SLA. Eskildsen and his colleagues carried out a string of investigations into the emergence of several constructions, for example, *can* constructions (Eskildsen, 2009), *auxiliary do* constructions (Eskildsen,

2012), and motion constructions (Li, Eskildsen, & Cardierno, 2014), as well as syntactic operations functioning in construction development (Eskildsen, 2014, 2017). In these studies, close inspection of an adult L1 Spanish speaker's classroom interaction in English identified the exemplar-based path for L2 learning, and 'utterance schemas,' which refer to semi-fixed linguistic patterns, were suggested to enhance the productivity of constructions. Adopting the 'traceback' method, they also confirmed that substitution among the various syntactic operations accounted for a great deal of the emergent creativity in L2 learning.

2.2. Previous Studies on Verb-argument Constructions

Some UBL researchers have shown particular interest in clause-level constructions, such as VACs, which are attested to be the basic units of expression of daily communicative functions. They have primarily addressed the ontological status of VACs and expanded the research into how they are processed, and how they emerge and develop in the minds of L1 and L2 speakers. As UBL is an umbrella term, embracing a variety of theoretical frameworks from psycholinguistics to SLA, researchers have followed numerous lines of investigation, from psycholinguistic experiments to natural language processing, which have resulted in a variety of empirical evidence.

In order to identify the psychological reality of VACs, Bencini and Goldberg (2000) employed a sentence-sorting task, in which participants were asked to sort sentence cards into construction- or verb-based groups by referring to the overall meaning of the sentence. This seminal work confirmed L1 English speakers' mental representations of constructions and has given rise to many replication studies that investigate L2 learners' reliance on constructional information in sentence-sorting tasks (Gries & Wulff, 2005; Kim & Rah, 2016; Kim et al., 2020; Shin, 2010; Valenzuela & Rojo, 2008). Although these studies varied in terms of learners' L1 backgrounds, L2 proficiency levels, age groups, and additional methodologies employed, the common findings determine that L2 learners also possess constructional knowledge, varying in its degree according to proficiency levels or verb semantics. For example, Kim and Rah (2016) and Kim et al. (2020) demonstrated L2 learners' high reliance on light verbs (i.e., semantically general verbs such as *go*, *do*, *make*, and *come*) in the early developmental stages, when processing L2 constructional knowledge.

While the sentence-sorting task demonstrated constructional processing during sentence comprehension, other researchers have examined production data collected from corpora and psycholinguistic experiments in order to identify various factors operating in VAC processing. Ellis et al. (2014a) employed a free association task and a verbal fluency task, requiring native English speakers to generate verbs that fit the verb slots in skeletal VAC frames. They compared the results of the experiments with the data from native speaker

corpus analyses and semantic network analyses; based on these comparisons, they concluded that VAC processing is influenced by verb frequency in the VAC, by verb-VAC contingency, and by verb prototypicality in general English usage.

A series of studies following that of Ellis et al. (2014a) have investigated L2 learners' VAC knowledge, focusing on different and specific research questions. Römer, O'Donnell, and Ellis (2014) revealed that learners whose L1 is typologically related to the L2 produce more target-like VACs than those whose L1 and L2 differ typologically. By combining data from learner corpora and psycholinguistic experiments, Römer, Roberson, O'Donnell, and Ellis (2014) and Römer, Skalicky, and Ellis (2018) provided converging evidence on advanced L2 learners' VAC knowledge in addition to insights into VAC-specific semantic verb groups. Given the strengths and limitations of the two data types, the researchers underscored the usefulness of mixed-methods research. More recently, Römer (2019) conducted a pseudo-longitudinal investigation on a large-scale corpus of learner writing. Based on various analytic methods from corpus linguistics and natural language processing, she concluded that L2 English learners show an increase in VAC repertoire, growth in VAC productivity, and a usage-based developmental sequence from lowest to highest proficiency levels.

Despite ample evidence of L2 VAC knowledge in the literature, little is known as yet about *how* L2 learners' VAC knowledge emerges and develops into a creative system due to the lack of conclusive longitudinal data. On top of that, a single data source may not show enough occurrences of specific linguistic phenomena. The present study, therefore, attempts to combine different data sources, specifically, a learner corpus and psycholinguistic experiments, incorporating various proficiency levels. In particular, we delve into Korean learners' VAC knowledge in L2 English in terms of VAC productivity and verb-VAC association. The VAC knowledge is further compared between L1 and L2 English as well as between L2 English and L1 Korean in order to demonstrate crosslinguistic similarities and differences and the learners' L1 influence. Specific research questions the present study intends to investigate are as follows:

1. How does Korean learners' VAC knowledge in L2 English emerge and develop across proficiency levels?
2. How does Korean learners' VAC knowledge in L2 English compare to native English speakers' VAC knowledge?
3. How is Korean learners' VAC knowledge in L2 English influenced by the L1, i.e., Korean?

3. METHODOLOGY

The present study collected data from two sources: (a) learner and native speaker (NS) corpora and (b) three verbal fluency tasks, each performed in L1 and L2 English and L1 Korean. The two data types were analysed quantitatively with type-token distributions and correlation analysis, and qualitatively by comparing the most frequent verb lists and performing construction growth analysis across proficiency levels, between L1 and L2 groups, and between data types. Table 1 provides an overview of the data types this study is based on.

TABLE 1

Overview of Corpus and Experimental Data

Corpus Data						
YELC (reclassified levels)	A1	A2	B1	B2	C	Total
Number of texts	226	684	1,878	459	39	3,286
Number of words	23,644	133,675	468,381	132,762	12,040	770,502
LOCNESS subsets	Am-argumentative		Am-mixed	Br-argumentative		Total
Number of texts	176		55	147		321
Number of words	149,574		18,826	79,228		247,628
Experimental Data						
Task type	Subject L1		Performed language	Number of subjects		
Verbal fluency task	English		English	50		
Verbal fluency task	Korean		English	150		
Verbal fluency task	Korean		Korean	75		

Note. Am = American; Br = British

3.1. Corpus

The learner corpus selected for this study is Yonsei English Learner Corpus (YELC, Rhee & Jung, 2014), which includes 3,286 argumentative essays collected from an English placement test conducted at a university in Korea. The written texts were rated according to the Common European Framework of Reference (CEFR, Council of Europe, 2001) and distributed into nine levels. These nine proficiency levels were reclassified in this study into the original six-point levels of CEFR, with the levels C1 and C2 grouped into level C due to the small number of texts. As a reference corpus, the Louvain Corpus of Native English Essays (LOCNESS, Granger, 1998) was selected. LOCNESS consists of argumentative essays and mixed essays written by American and British university students. American students' mixed essays were argumentative in nature, but British ones included answers in exams on literary works and expository-historical topics. The latter data was thus excluded from the current study.

A total of 19 VACs[2] were selected to study L2 learners' constructional knowledge, following Römer, Roberson, et al. (2014) and Römer and Berger (2019). To facilitate VAC extraction, the two corpora were part-of-speech (POS) tagged using CLAWS tagger, and the VAC lists were retrieved using the concord function in WordSmith Tools 7.0 (Scott, 2016). For concordancing, a preposition in a VAC (e.g., about <PRP>) was used as a search word, and its context was entered in the left span (e.g., *<V*>). The resulting concordances were manually filtered for true instances of each VAC by the two authors and a native speaker of English. The VACs in passive or relative clauses were excluded, considering that meanings carried in syntactic forms may affect the semantics of VACs. Lastly, the verbs used in the concordance lines were lemmatized (e.g., *goes*, *went*, *was going*, etc. → RUN) and frequency-sorted, and a verb list was generated for each VAC, proficiency level, and language group.

3.2. Verbal Fluency Task

The verbal fluency task adopted in this study was a psycholinguistic test for investigating semantic category knowledge and its fluency of access. Participants in the test were asked to produce as many words as possible from a category within a given time when presented with a certain stimulus. The results of such a test can provide insights into the prototypicality of a category since more typical items are produced by more participants and earlier within the time limit than less typical ones (Gruenewald & Lockhead, 1980). In this study, the verbal fluency task was designed to investigate participants' knowledge of English verbs commonly associated with a certain VAC, following Römer et al. (2018). To prevent a context from influencing the responses, a bare sentence frame including a VAC with the verb missing was given to participants (e.g., "She/He _____ about the …" or "It _____ with the …"). Before starting an actual task, participants were presented with an instruction to generate as many verbs as possible with which they might fill in the gap within one minute for each sentence.

The task was distributed online via Qualtrics (www.qualtrics.com) to the participants, who were recruited from a crowdsourcing site (i.e., Amazon Mechanical Turk, henceforth AMT) and several Korean university websites. A total of 50 native speakers of English were selected from those qualified as US high school graduates by the AMT. The Korean cohort, meanwhile, completed an online participation form asking personal information,

[2] They were derived from the extensive lists of verb patterns in the COBUILD Grammar Patterns: Verbs volume (Francis, Hunston, & Manning, 1996): V *about* n, V *across* n, V *after* n, V *against* n, V *among* n, V *around* n, V *as* n, V *between* n, V *for* n, V *in* n, V *into* n, V *like* n, V *of* n, V *off* n, V *over* n, V *through* n, V *toward* n, V *under* n, V *with* n

including scores of standardized English proficiency tests such as Test of English for International Communication (TOEIC) and Test of English as a Foreign Language (TOEFL). Finally, 150 Korean learners of English were selected, 50 each from level A1-A2 (lower), B1-B2 (intermediate), and C1-C2 (advanced). The CEFR levels were estimated by converting the scores of the standardized English proficiency tests according to the criteria that the ETS research team suggests (Tannenbaum & Wylie, 2008). Three months later, the verbal fluency task, which was manipulated to investigate VAC knowledge in Korean, was distributed to 75 Korean learners of English who were randomly selected among those having participated in the same task in English. To compensate the participants for their time, a gift card worth five USD was offered. The verbs from the participants' responses were lemmatized and sorted by their frequency to make a verb list for each VAC, proficiency level, and language group.

3.3. Analysis

To address the study's research questions, we performed various analyses of the verb frequency lists extracted from the learner corpus and the experimental data. Firstly, verb types and tokens along with their ratios were compared across proficiency levels and between data types. Then, the most frequent verb lists were compared across proficiency levels, between L1 and L2 groups, and between data types to examine crosslinguistic similarities and differences as well as the influence of the learners' L1. A construction growth analysis was further performed by plotting cumulative normalized frequencies of high-ranked verbs on a line graph (see Römer & Berger, 2019). The analysis is intuitive in that it visualizes the VAC development, tracking emerging verbs within a VAC and emphasizing how a pathbreaking verb leads the constructional development. Lastly, a correlation analysis provided statistical evidence of Korean learners' VAC development in L2 English, and of a relationship between L1 and L2 VAC knowledge.

4. RESULTS

4.1. Type-Token Distribution of VACs

The verb lists extracted from a learner corpus (YELC) and a verbal fluency task were sorted by their frequency and analyzed in terms of types (i.e., the number of distinct verbs used in a VAC) and tokens (i.e., the total number of verbs that occurred within a particular VAC). The greater the type frequency of a VAC, the less likely it is that the VAC is associated with a particular verb, and the more likely it is that a general schema is formed

TABLE 2

Verb Type-Token Frequencies and Ratios Across Proficiency Levels in YELC

	A1			A2			B1			B2			C			Total Type /Token
	Type	Token	Ratio	Type	Token	Ratio	Type	Token	Ratio	Type	Token	Ratio	Type	Token	Ratio	
V *about* n	16	42	38%	41	212	19%	79	823	10%	38	156	24%	10	18	56%	184/1251
V *across* n	-	-	-	1	1	100%	1	3	33%	1	1	100%	-	-	-	3/5
V *after* n	-	-	-	3	5	60%	15	21	71%	9	9	100%	1	1	100%	28/36
V *against* n	3	5	60%	2	10	20%	13	51	25%	6	28	21%	3	4	75%	27/98
V *among* n	-	-	-	-	-	-	2	2	100%	-	-	-	-	-	-	2/2
V *around* n	1	1	100%	6	7	86%	14	39	36%	3	4	75%	-	-	-	24/51
V *as* n	-	-	-	5	7	71%	15	27	56%	7	17	41%	2	2	100%	29/53
V *between* n	-	-	-	1	1	100%	10	15	67%	2	2	100%	2	2	100%	15/20
V *for* n	11	15	73%	54	144	38%	111	501	22%	49	138	36%	10	10	100%	235/808
V *in* n	34	86	40%	96	490	20%	171	1,706	10%	76	485	16%	14	27	52%	391/2794
V *into* n	-	-	-	6	7	86%	20	35	57%	9	25	36%	3	3	100%	38/70
V *like* n	4	11	36%	22	59	37%	39	194	20%	14	37	38%	1	1	100%	80/302
V *of* n	9	10	90%	32	50	64%	55	150	37%	19	43	44%	3	3	100%	118/256
V *off* n	-	-	-	6	14	43%	8	23	35%	5	7	71%	-	-	-	19/44
V *over* n	2	2	100%	2	2	100%	17	24	71%	7	8	88%	1	1	100%	29/37
V *through* n	-	-	-	4	11	36%	14	34	41%	12	26	46%	1	1	100%	31/72
V *toward* n	-	-	-	1	1	100%	-	-	-	1	1	100%	1	1	100%	3/3
V *under* n	-	-	-	2	2	100%	6	17	35%	4	7	57%	1	1	100%	13/27
V *with* n	12	48	25%	60	323	19%	124	1,141	11%	60	310	19%	15	26	58%	271/1848

over the verbs. The type frequency, therefore, determines the productivity of the VAC (i.e., the degree to which the VAC extends to new verb types). On the other hand, the greater the token frequency of a verb in a VAC is, the more likely that verb is to be considered prototypical of the VAC, and to promote the entrenchment of the verb-VAC association (see Bybee & Hopper, 2001; Ellis et al., 2016).

Tables 2 and 3 illustrate type-token distributions of the verbs used in 19 VACs across proficiency levels from the learner corpus and the experimental data, respectively, with their ratios calculated to show how productive a VAC is (see Li et al., 2014; Römer et al., 2018). However, caution must be exercised in taking the ratios at face value when they have low token frequencies. In the learner corpus data (see Table 2), most of the VACs indicated very few or no instances of most verb types in levels A1 and C, probably due to the small sub-corpus size. Four VACs ("V *about* n," "V *for* n," "V *in* n," and "V *with* n"), however, showed robust frequencies across proficiency levels with normalized frequencies above 10 per 100,000 words, similar to the type-token distribution reported in Römer and Berger (2019). Analyzing verb type-token ratios of these four VACs in terms of VAC productivity across proficiency levels revealed that the learners produced more verb types at levels A1 and A2 than B1, and also more verb types at levels B2 and C than B1 (see Figure 1). In other words, as the learners' proficiency levels increased, the degree of productivity of the four VACs developed in U-shaped patterns.

FIGURE 1

VAC Productivity Across Proficiency Levels in YELC

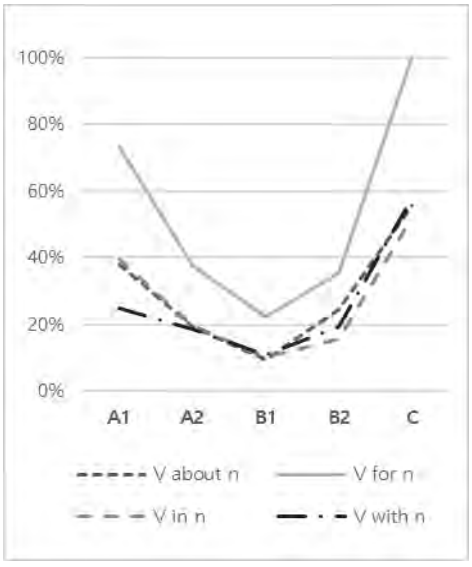
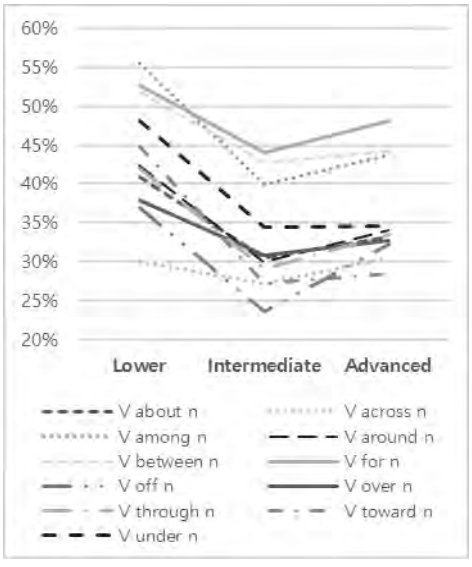


FIGURE 2

VAC Productivity Across Proficiency Levels in a Verbal Fluency Task



As can be seen in Table 3, the verbal fluency task, which was completed by 150 Korean learners in three levels (i.e., lower: A1-A2, intermediate: B1-B2, advanced: C1-C2), produced a considerable number of tokens across proficiency levels, on account of having the same number of participants (50) in each proficiency level. An analysis of verb type-token ratios of 19 VACs revealed V-shaped patterns in 11 of them (see Figure 2). The V-shaped patterns illustrate that the lower- and advanced-level learners produced a greater diversity of verbs than the intermediate learners. Consequently, the V-shaped developmental patterns in 11 VACs from the experimental data, together with the U-shaped patterns in four VACs from the learner corpus data, indicate the lowest VAC productivity of the intermediate learners.

TABLE 3

Verb Type-Token Frequencies and Ratios Across Proficiency Levels in a Verbal Fluency Task

	Lower			Intermediate			Advanced			Total Type/ Token
	Type	Token	Ratio	Type	Token	Ratio	Type	Token	Ratio	
V *about* n	50	122	41%	51	167	31%	58	175	33%	159/464
V *across* n	31	103	30%	39	143	27%	48	157	31%	118/403
V *after* n	54	114	47%	56	146	38%	64	183	35%	174/443
V *against* n	54	97	56%	64	137	47%	65	161	40%	183/395
V *among* n	50	90	56%	48	120	40%	56	128	44%	154/338
V *around* n	47	111	42%	48	160	30%	61	179	34%	156/450
V *as* n	42	83	51%	69	122	57%	72	152	47%	183/357
V *between* n	55	106	52%	62	145	43%	68	154	44%	185/405
V *for* n	57	108	53%	79	179	44%	91	189	48%	227/476
V *in* n	49	102	48%	80	180	44%	71	184	39%	200/466
V *into* n	46	107	43%	48	170	28%	43	171	25%	137/448
V *like* n	39	96	41%	41	165	25%	47	209	22%	127/470
V *of* n	30	63	48%	37	78	47%	51	110	46%	118/251
V *off* n	34	92	37%	35	148	24%	51	158	32%	120/398
V *over* n	38	100	38%	51	165	31%	61	186	33%	150/451
V *through* n	42	100	42%	44	151	29%	58	173	34%	144/424
V *toward* n	44	98	45%	40	147	27%	49	172	28%	133/417
V *under* n	50	104	48%	52	151	34%	65	188	35%	167/443
V *with* n	55	107	51%	63	170	37%	73	197	37%	191/474

The high verb type-token ratio in this study represents a high degree of VAC productivity. The low verb type-token ratio, on the other hand, not only indicates a low degree of VAC productivity but may imply frequent uses of one or a few verb types within VACs, suggesting possible entrenchment of certain verbs in VACs (see Li et al., 2014). Thus, to account for the low verb type-token ratios of the VACs in the intermediate levels, token percentages of the first- and second-most frequent verbs in the corresponding VACs were calculated across proficiency levels.

FIGURE 3
Percentages of Two Most Frequent Verbs in VACs Across Proficiency Levels Extracted from YELC

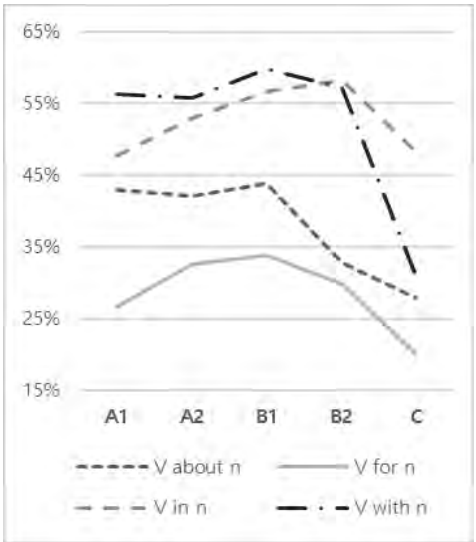


FIGURE 4
Percentages of Two Most Frequent Verbs in VACs Across Proficiency Levels Extracted from a Verbal Fluency Task

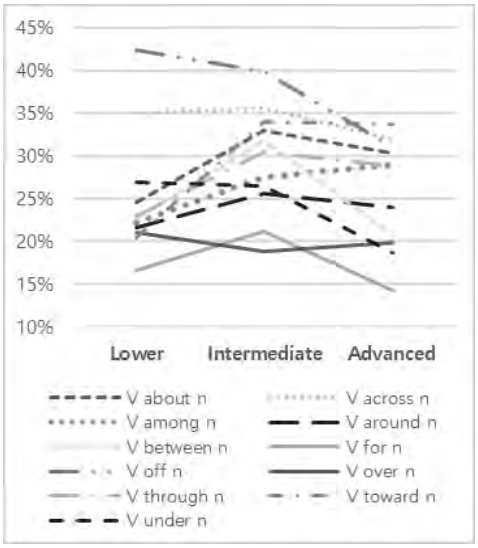


Figure 3 shows the percentages of the two most frequent verbs in four VACs at each proficiency level, which were extracted from YELC;[3] they roughly approximate inverted U-shapes. The high percentages of the two most frequent verbs at levels B1 and B2 show that the intermediate learners have established a strong association between those verbs and corresponding VACs. Figure 4 likewise displays the data from the verbal fluency task, in which ten VACs exhibit inverted V-shaped patterns while one VAC (i.e., "V *over* n") has a flattened V-shape. The token numbers of the two most frequent verbs in "V *over* n," however, increased from the lower to the intermediate level; we can thus conclude that the intermediate learners showed an entrenchment of the most frequent verb types in some VACs, confirming the result of the learner corpus data.

Analyses of type-token distribution and ratios in 19 VACs across proficiency levels identified U-shaped or V-shaped developmental patterns in VAC productivity. The

[3] The two most frequent verbs in four VACs at each proficiency level are as follows:

	A1	A2	B1	B2	C
"V *about* n"	THINK, KNOW	THINK, KNOW	THINK, TALK	THINK, BE	THINK, TALK
"V *for* n"	BE, DIE	BE, SMOKE	BE, WORK	BE, WORK	BE, APPLY
"V *in* n"	(A1~C) SMOKE, BE				
"V *with* n"	(A1~C) AGREE, DISAGREE				

percentages of the two most frequent verbs in VACs suggested an entrenchment of the most frequent verb types in some VACs at the intermediate level, providing an account for the L2 VAC developmental tendency. To be precise, the lower-level learners' high VAC productivity is attributed to low dependence on the most frequent verbs, possibly implying insufficient schematic knowledge. The intermediate learners, in contrast, exhibited a low degree of VAC productivity and great reliance on the most frequent verbs in VACs, which suggests an emerging constructional schema. Furthermore, high VAC productivity and decreasing use of the most frequent verbs in the advanced level may indicate developing schematic knowledge and increased mastery of VACs.

4.2. Comparing the Highest Frequency Verbs

In the previous section, L2 learners' low dependence on the most frequent verb types in VACs was interpreted differently at the lower and advanced levels, considering the developmental sequence of language. The present study further attempts to confirm this hypothesis by examining the most frequent verb lists among different proficiency levels. The highest frequency verb in a VAC plays the role of a pathbreaking verb that seeds the construction and carries prototypical and generic meaning (Ellis & Ferreira-Junior, 2009b; Goldberg, 2006; Ninio, 1999). Therefore, it is worth investigating which verb types Korean learners of English use the most frequently in VACs, and whether they are different across proficiency levels, between language groups, and between data types.

Table 4 provides the most frequent verb lists from the learner corpus (YELC) and the NS corpus (LOCNESS). The parenthesized verbs were excluded from the analysis, following Römer, Roberson, et al. (2014) since it would be misleading to talk about a 'lead' verb when the frequency is so low. The remaining verbs in the learner corpus data indicated an overlap of the lead verbs at different proficiency levels, except for "V *of* n." Meanwhile, "V *about* n," "V *in* n," and "V *with* n" identified the lead verbs as those related to argumentative writing (e.g., THINK, SMOKE, AGREE),[4] and seven VACs shared the lead verbs both in the learner and the NS corpora (e.g., THINK in "V *about* n"; GO in "V *through* n"; BE in "V *against* n," "V *for* n," "V *like* n," "V *over* n," and "V *under* n"). Given that the learner and the NS corpora are composed of argumentative essays, the text type seems to have affected the result, with similar verbs selected for the most frequent verb of each VAC regardless of proficiency levels and L1 or L2 status.

[4] Most instances of *think about* and *agree with* retrieved from YELC were in the form of THINK *about*/AGREE *with* (a topic) (e.g., *We do not think about animal's extinction*/*I agree with the physical punishment of children in schools*). More than half of the instances of SMOKE *in* retrieved from YELC included *smoke in public buildings,* a phrase contained in an essay prompt.

TABLE 4

The Most Frequent Verb Lists from YELC and LOCNESS

	YELC					LOCNESS
	A1	A2	B1	B2	C	
V *about* n	*THINK*	*THINK*	*THINK*	*THINK*	*THINK*	*THINK*
V *across* n	-	(GO)	(COME)	(JUMP)	-	(STUMBLE)
V *after* n	-	(DRIVE)	DRIVE	(COME)	(COME)	LOOK
V *against* n	(BE)	*BE*	*BE*	*BE*	(GO)	*BE*
V *among* n	-	-	(RANK)	-	-	(BE)
V *around* n	(LISTEN)	(WALK)	BE	(BE)	-	CENTER
V *as* n	-	(WORK)	WORK	WORK	(WORK)	SERVE
V *between* n	-	(FIGHT)	(CHOOSE)	(CHOOSE)	(CHOOSE)	BE
V *for* n	(BE)	*BE*	*BE*	*BE*	(BE)	*BE*
V *in* n	SMOKE	SMOKE	SMOKE	SMOKE	SMOKE	BE
V *into* n	-	(GROW)	GET	GET	(GET)	GO
V *like* n	*BE*	*BE*	*BE*	*BE*	(FEEL)	*BE*
V *of* n	(CONSIST)	CONSIST	THINK	GET RID	(GET RID)	BE
V *off* n	-	TURN	TURN	(PUT)	-	(PUT)
V *over* n	(BE)	(BE)	*BE*	(BE)	(WIN)	*BE*
V *through* n	-	*GO*	*GO*	*GO*	(EMPHASIZE)	*GO*
V *toward* n	-	(SHOUT)	-	(ACT)	(GROW)	(MOVE)
V *under* n	-	(BE)	*BE*	(BE)	(BE)	*BE*
V *with* n	AGREE	AGREE	AGREE	AGREE	(AGREE)	DEAL

Note. Italicized verbs indicate an overlap between learner and NS usages.
The verbs in parentheses have token numbers lower than 5.

To compensate for the limited corpus data, three verbal fluency tasks, each performed in L1 and L2 English and L1 Korean, were analyzed, the most frequent verb lists from which are presented in Table 5. In the L2 English data, 10 out of 19 VACs shared the lead verbs across proficiency levels,[5] while the remaining VACs varied in lead verbs at different levels. Particularly, as the proficiency level increased, the lead verbs changed from general to specific verbs among these VACs (e.g., lower – BE, intermediate – FIGHT, advanced – FIGHT in "V *against* n"; lower – BE, intermediate – LOOK, advanced – ACT in "V *as* n"). Ellis and Ferreira-Junior (2009b) claimed that L2 learners initially use semantically generic and high-frequency verbs and expand their verb repertoire into more specific and less frequent ones with increasing proficiency. Based on the data presented here, we can conclude that Korean learners of English follow a similar developmental sequence of L2 VAC usage regarding the lead verbs for VACs.

[5] Among the 10 VACs that shared the lead verbs across proficiency levels, seven took general verbs (e.g., BE or GO), which seem to be influenced by the Korean L1 typology, while the remaining three VACs took BE as the second-most frequent verb by a small difference of token frequency at the lowest level, with specific verbs noted in other proficiency levels.

TABLE 5

The Most Frequent Verb Lists from Three Verbal Fluency Tasks

	L2 English			L1 English	L1 Korean
	Lower	Intermediate	Advanced		
V *about* n	THINK	*TALK*	*TALK*	*TALK*	malhata (*TALK*)
V *across* n	WALK	WALK	WALK	RUN	kata (GO)
V *after* n	LOOK	LOOK	LOOK	RUN	seta (STAND)
V *against* n	BE	*FIGHT*	*FIGHT*	*FIGHT*	ssawuta (*FIGHT*)
V *among* n	*BE*	*BE*	*BE*	*BE*	seta (STAND)
V *around* n	*BE*	*WALK*	*WALK*	*WALK*	issta (*BE*) / kata (GO)
V *as* n	BE	LOOK	*ACT*	SPEAK	hayngtonghata (*ACT*)
V *between* n	*BE*	*BE*	*BE*	WALK	issta (*BE*) / seta (STAND)
V *for* n	BE	LOOK	WORK	RUN	nolyekhata (TRY)
V *in* n	*BE*	*BE*	*BE*	SIT	issta (*BE*)
V *into* n	GO	GO	GO	RUN	tulekata (GO *entering*)
V *like* n	*LOOK*	*LOOK*	*LOOK*	*LOOK*	boita (*LOOK*)
V *of* n	BE	*THINK*	*THINK*	*THINK*	sayngkakhata (*THINK*)
V *off* n	TURN	TAKE	TURN	JUMP	tomangkata (GO *escaping*)
V *over* n	LOOK	LOOK	*GO*	JUMP	kata (*GO*)
V *through* n	GO	GO	GO	WALK	cinakata (GO *passing*)
V *toward* n	*GO*	*GO*	*GO*	RUN	kata (*GO*)
V *under* n	*BE*	*BE*	*BE*	HIDE	issta (*BE*) / seta (STAND)
V *with* n	BE	BE	GO	RUN	nolta (PLAY)

Note. Italicized verbs indicate an overlap between different language groups (e.g., L1-L2 English, L2 English-L1 Korean).

To identify crosslinguistic similarities and differences between L1 and L2 VAC knowledge, the most frequent verb list of L2 English data was compared with that of L1 English and L1 Korean data. In the comparison between L1 and L2 English, two VACs (i.e., BE in "V *among* n"; LOOK in "V *like* n") shared their lead verbs; meanwhile, between L2 English and L1 Korean, five VACs (i.e., BE in "V *between* n," "V *in* n," and "V *under* n"; GO in "V *toward* n"; LOOK in "V *like* n") shared their lead verbs. Although this finding at first seems to suggest a stronger influence of the learner's L1 (i.e., Korean) on the L2 (i.e., English) VAC knowledge, in a detailed comparison of different L2 proficiency levels with L1 English, the lead verbs produced by higher-level learners turned out to approximate the lead verbs produced by native English speakers more closely than they approximated the productions of lower-level learners (e.g., TALK in "V *about* n"; FIGHT in "V *against* n"; WALK in "V *around* n"; THINK in "V *of* n"). In other words, the more L2 learners' proficiency improves, the more likely their L2 VAC knowledge is to resemble that of native speakers.

In the L1 Korean data, furthermore, some VACs revealed a language typology effect. Drawing on Talmy's (1985) typological framework, Korean is a verb-framed language, in which the path and motion of verbs are encoded in the verb root, while the manner of motion is less commonly expressed or, if at all, expressed as an adjunct such as a

participial form. Reflecting this characteristic, the lead verbs of "V *into* n," "V *off* n," and "V *through* n" in the L1 Korean data were "tulekata" (GO *entering*), "tomangkata" (GO *escaping*), and "cinakata" (GO *passing*), which encompass the manner of motion in an adjunctive inflectional form of "kata" (GO). Since manner of motion is less salient in the minds of Korean speakers, the lead verbs of the corresponding VACs in the L2 English data seem to leave out this concept, rather exhibiting general motion verbs such as GO in "V *into* n" and "V *through* n" or idiomatic phrases such as TURN or TAKE in "V *off* n." Additionally, the VACs expressing motion events (i.e., "V *between* n," "V *in* n," "V *under* n," and "V *toward* n"), which previously reported the same lead verbs both in L1 Korean and L2 English data, seem to demonstrate the L1 typology effect in that the shared lead verbs are semantically general verbs, BE or GO.

4.3. Tracking Dominant Verb-VAC Associations

A construction growth analysis was performed to look beyond the lead verbs and show how a specific VAC develops in terms of its dominant verb-VAC associations as learners' proficiency improves. The analysis was conducted on the four VACs that have sufficient type numbers in each proficiency level ("V *about* n," "V *for* n," "V *in* n," and "V *with* n"), but due to space considerations, "V *in* n" will form the focus here. The frequencies of 10 high-ranked verbs in each proficiency level (normalized per 100,000 words in the learner corpus data) were cumulated and plotted on a line graph to visualize the VAC development. This analysis enables an in-depth discussion of emerging verbs and their contribution to the overall construction development.

Figures 5 and 6 are line graphs for the construction growth analyses of "V *in* n" from the learner corpus and the verbal fluency task data. The uppermost lines in both graphs are located at a considerably higher position than the lines below, so the verbs creating the lines seem to lead the growth of "V *in* n." In other words, "V *in* n" could have been entrenched in the minds of Korean learners of English as a result of frequent experiences using these verbs with the VAC. However, the verb of the uppermost line in the learner corpus data is SMOKE, which apparently comes from an essay prompt as discussed in the previous section. It is thus inappropriate to take SMOKE as leading the growth of "V *in* n." Rather, it seems reasonable to consider the verb of the second uppermost line (i.e., BE) as a pathbreaking verb of the construction, especially given that BE is the verb of the uppermost line in the verbal fluency task data.

FIGURE 5

Cumulative Normalized Frequencies of Verb Types in "V *in* n" from YELC

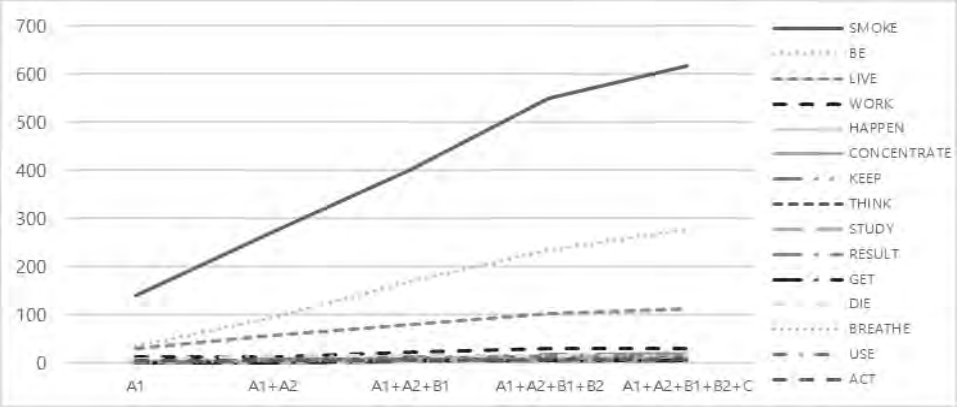
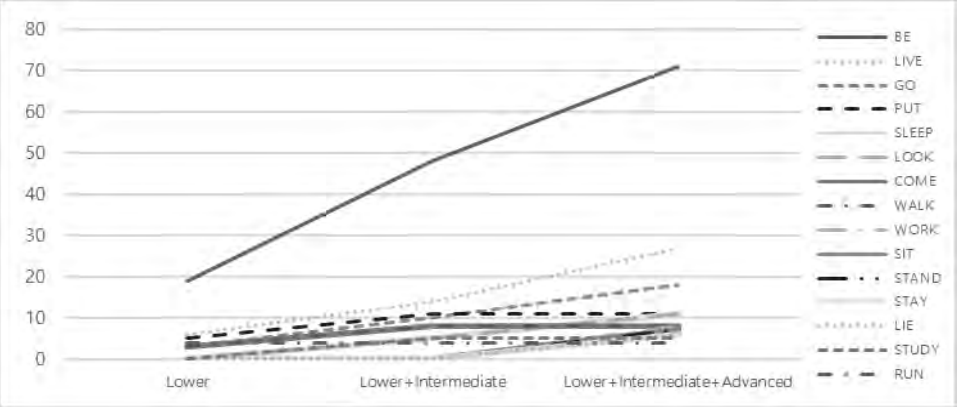


FIGURE 6

Cumulative Frequencies of Verb Types in "V *in* n" from a Verbal Fluency Task



Ninio (1999) described verbs that ease the spread of a VAC combination to other verbs as pathbreaking verbs, which tend to be "generic verbs that express the relevant combinatorial property in a relatively undiluted fashion" (p. 646). The verb BE of "V *in* n" indicates a state of someone or something (e.g., *When I was in high school*; *Korea is in dangerous situation*). Accordingly, since the verb BE is used as a pathbreaking verb that seeds "V *in* n" in the minds of Korean learners of English, there seem to be more stative verbs (e.g., LIVE) in the usage of "V *in* n" by Korean learners than native speakers of English, who frequently expressed directed motion through the top-ranked verbs of the "V *in* n" VAC in the verbal fluency task (e.g., SIT, WALK, and RUN).

Careful observation on the dominant verb-VAC association of "V *in* n" in the learner corpus data reveals few verbs at the lowest level (A1), with most verbs except SMOKE, BE, and LIVE showing low frequencies under 12.7 per 100,000 words. At the next proficiency level (A2), additional verbs (e.g., BREATHE, THINK, and USE) emerge, but their frequencies remain low across proficiency levels. Some verbs (e.g., HAPPEN, KEEP and RESULT), on the other hand, emerge at A2 or B1 and show gradually increasing cumulative frequencies, with the highest frequencies seen at B2. These verbs showed low frequencies of 5.2 per 100,000 words at the beginning, but each added up to 31, 24.9, and 24.1 per 100,000 words in total. In summary, a couple of frequent verbs (i.e., BE and LIVE) lead the growth of "V *in* n," while more specific verbs (i.e., HAPPEN, KEEP, and RESULT) appear in the VAC as the learners' proficiency develops.

A similar developmental path of "V *in* n" is observed in the experimental data (see Figure 6), yet the semantically specific verbs of "V *in* n," such as HAPPEN, KEEP, and RESULT, do not feature. Instead, motion verbs such as GO, PUT, COME, WALK and, RUN are used, albeit in low frequencies. It seems that due to the bare decontextualized frame in the verbal fluency task, "V *in* n" is subject to the verbs incorporating the motion sense that is conveyed by the path *in*.

4.4. Comparing L2 Learner and Native Speaker Verb Usages

Unlike the frequency analyses adopted in the previous sections, a correlation analysis provides statistical evidence of constructional development between different L2 proficiency levels, and of a relationship between L1 and L2 constructional knowledge (see Römer & Berger, 2019). We thus performed correlation analyses on "V *about* n" to compare verb usages between two learner levels and between subsets of the learner corpus (YELC) and the NS corpus (LOCNESS). Before performing the correlation analysis, verb frequencies in the learner and the NS corpora were normalized per 100,000 words and log transformed to make the Zipfian distribution of the verb frequencies comparable. The resulting correlations were statistically significant ($p < .01$).

TABLE 6

Correlation Values for Verb Usage in "V *about* n" Between Two Proficiency Levels in YELC

Comparison	Pearson's *r*	Comparison	Pearson's *r*	Comparison	Pearson's *r*
A1 vs. A2	0.73**	A2 vs. B1	0.87**	B1 vs. B2	0.87**
A1 vs. B1	0.72**	A2 vs. B2	0.68**	B1 vs. C	0.58*
A1 vs. B2	0.39**	A2 vs. C	0.36**	B2 vs. C	0.53**
A1 vs. C	0.21				

Table 6 lists Pearson correlation coefficients for all possible comparisons between five CEFR levels in YELC. A hypothesis underlying the comparisons is that if there is a constructional development from lower to higher proficiency levels, adjacent levels need to have higher correlations than non-adjacent levels since there are more overlaps between the verbs used in the adjacent than the non-adjacent levels. As expected, the *r* - values are highest for adjacent levels (e.g., A1 vs A2; A2 vs. B1) and lower for non-adjacent levels (e.g., A1 vs B1; A2 vs. B2), suggesting a sequential development towards more diversity of verbs in higher proficiency levels.

Figure 7 displays the results of the correlation analyses between subsets of the learner corpus and the NS corpus in a scatterplot, with the verbs used as data point labels. The *r* - values provided at the top-left corner of each graph indicate medium-to-high correlation with the native speaker benchmark. As the learners' proficiency increases, the correlation coefficient also increases, suggesting that the learners' proficiency level correlates with closer approximation of their verb usages in "V *about* n" to those of native speakers. Moreover, in the scatterplot, verbs resting on a diagonal line or surrounding it show the same or similar frequencies in the learner and the NS data. Thus, the positioning of THINK and TALK close to the diagonal line at the top-right corners of each graph indicates that the Korean learners of English in each proficiency level and the native speakers alike used those verbs most frequently, and that they therefore constitute the prototypical verbs of "V *about* n."

Verbs plotted at $x = 0$ or $y = 0$ occur only in the Korean learner or the NS data, respectively. As shown in the correlation plots in Figure 7, while many expressions are only attested in the learner data at levels A1 to B1, there are far fewer expressions that are preferred by the learners at levels B2 and C. The expressions unique to the lower-level learners include AGREE and DISAGREE (e.g., *Most of men doesn't agree about my opinion* from A1; *I disagree about physical punishment* from A2), whereas the higher-level learners preferred CONCERN, CONSIDER, and DISCUSS (e.g., *We also have to consider about the people who deny completing military service* from B1; *The board committee discuss about punishing children thoroughly* from B2). Although AGREE *about* and DISAGREE *about* are not ungrammatical *per se*, the instances retrieved from the learner corpus show that the lower-level learners tend to use the expressions inaccurately. On the other hand, although CONCERN *about,* CONSIDER *about*, and DISCUSS *about* are ungrammatical expressions, the fact that they are semantically related to THINK and TALK, the prototypical verbs of "V *about* n," seems to suggest that the higher-level learners understand the core meaning of the VAC, producing "creative extensions of the core semantics of the VAC" (Römer et al., 2018, p. 18).

FIGURE 7

Correlation of Verb-VAC Associations in "V *about* n" Between Each Level of YELC and LOCNESS

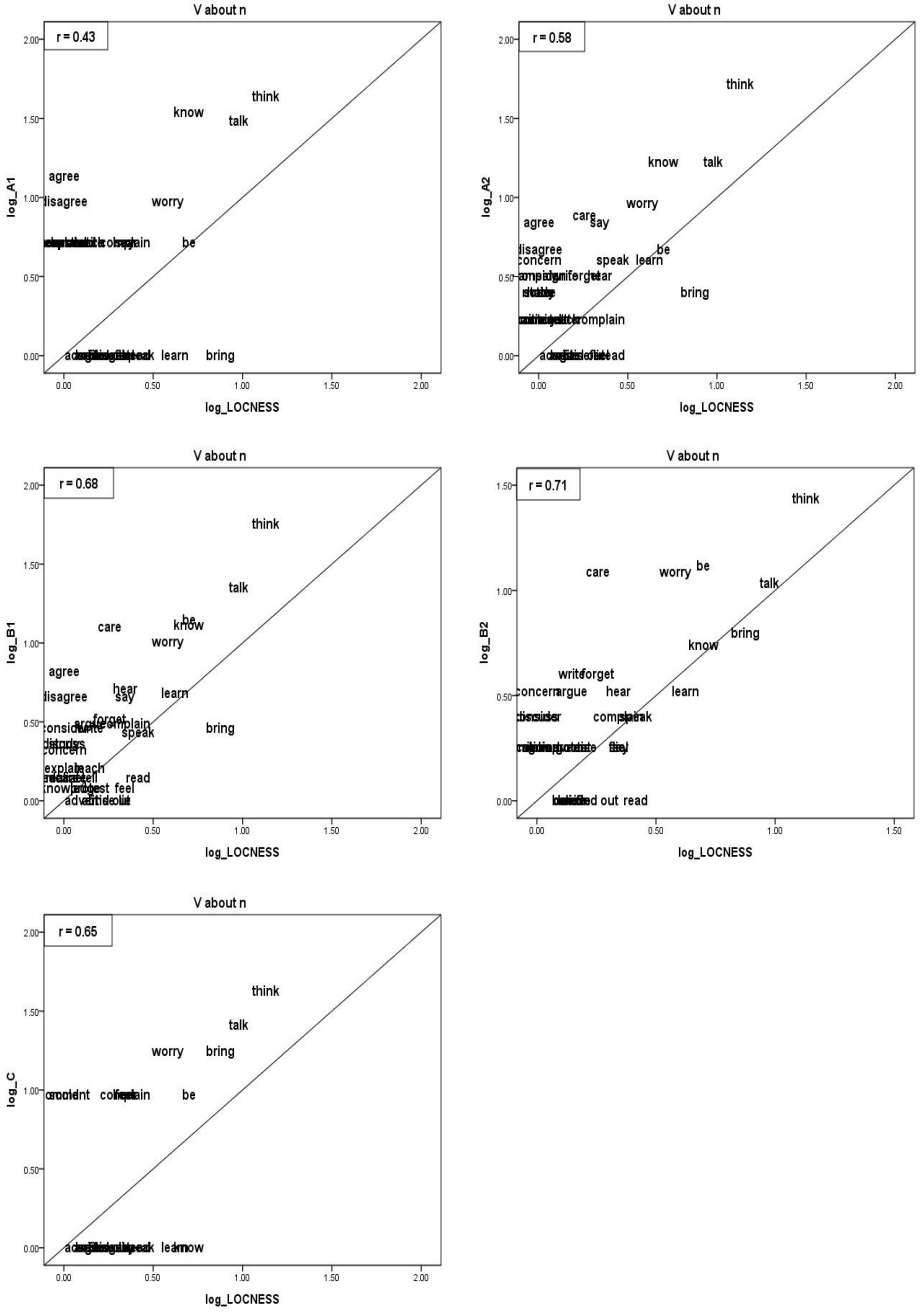


TABLE 7

Top 10 Verb List of "V *about* n" from a Verbal Fluency Task Performed in Korean

Verb	Frequency	Verb	Frequency
malhata (TALK)	32	kongpuhata (STUDY)	12
sayngkakhata (THINK)	28	cosahata (INVESTIGATE)	10
alta (KNOW)	27	ssuta (WRITE)	9
mutta (ASK)	13	selmyenghata (EXPLAIN)	9
kominhata (CONCERN)	12	nonuyhata (DISCUSS)	8

Lastly, in order to examine crosslinguistic transfer from L1 Korean to L2 English VAC knowledge, the result of the verbal fluency task performed in Korean was investigated. Among the top 10 Korean verbs of "V *about* n" produced by native Korean speakers (Table 7), the equivalents of some English verbs (i.e., CONCERN, STUDY, and DISCUSS) that are unique to the learner corpus were included (e.g., *We may concern about attack from other countries*; *I'll go abroad to study about the subject*). Given that these verbs are hardly used at all with "V *about* n" by the native speakers who contributed to LOCNESS, they are likely the result of collocational transfer from the participants' Korean VAC knowledge. In other words, since Korean speakers commonly collocate those verbs with the Korean equivalent of *about* (i.e., "ey tayhay"), they are prone to being literally translated into English, often resulting in ungrammatical expressions.

5. DISCUSSION AND CONCLUSION

The present study investigated the emergence and development of L2 learners' VAC knowledge by comparing verb usages of the target VACs across proficiency levels and between language groups. In order to do this, different data sources, specifically a learner corpus and psycholinguistic experiments, were examined. The mixed-data approach and various analytical methods (i.e., analyses on type-token distribution and the highest frequency verb lists, construction growth analysis, and correlation analysis) have contributed to answering the three research questions, as follows.

First, lower-level learners (i.e., A1 and A2) relied on the most frequent verb types for English VACs, which were semantically general verbs, less than the learners at other proficiency levels. They used only a few of the verb-VAC associations dominant in higher proficiency levels, while producing lots of unique expressions that were not common in native English usages. We concluded, therefore, that the lower-level learners have underspecified schematic knowledge in L2 English, which is activated on a verb-by-verb basis (see Eskildsen, 2009, 2012; Kim et al., 2020; Li et al., 2014). The lower-level learners, meanwhile, displayed high VAC productivity, which may in part be attributed to

the small sub-corpus size and the overproduction of verbs for VACs in the verbal fluency task; when the lower-level learners were asked to generate as many verbs as possible with which they might fill in a gap in a sentence within one minute, they may have submitted guesses with relatively low degrees of certainty, which resulted in the inflated VAC productivity. This seems a likely explanation for the plethora of ungrammatical expressions observed in the lower proficiency level.

With the various analyses taken together, the L2 constructional schema seemed to emerge at around the intermediate level (i.e., B1 and B2). The intermediate learners showed heavy reliance on the most frequent verbs for English VAC usages and confirmed the entrenchment of particular verb-VAC associations that many UBL researchers have attested in the early stages of construction development (Ellis & Ferreira-Junior, 2009b; Goldberg et al., 2004; Ninio, 1999; Tomasello, 1992). The highest-frequency verbs produced by the intermediate learners were prototypical of the VAC semantics, such as TALK *about*, FIGHT *against*, and WALK *around,* identical to those produced by native speakers. The intermediate learners, in addition, included more specific verbs (e.g., HAPPEN, KEEP and RESULT in "V *in* n") for VAC usages. Above all, the appearance of creative verb usages in VACs such as CONCERN *about,* CONSIDER *about*, and DISCUSS *about*, which are related to the core semantics of the VAC (i.e., cognition and communication), corroborated the establishment of the abstract constructional representations suggested by Römer et al. (2018).

Although learner corpus data revealed a limited number of verb types at the advanced level (i.e., level C), rich verbal fluency task data demonstrated high VAC productivity and decreasing use of the most frequent verbs in VACs, which is in line with the development of motion constructions discussed by Li et al. (2014). Since the advanced learners have established mental representations of VACs, they seemed to be able to expand the range of verbs used with VACs without recourse to the most frequent exemplars. The learners at this highest level hardly used ungrammatical expressions, and if any, they were creative extensions of the core verbs of the VAC (e.g., COMMENT/SCOLD *about* as extensions of TALK *about*). The Korean learners at this highest proficiency level thus seem to be creatively generating candidate verb types for VACs by extrapolating from their existing knowledge, although their ability to screen them for grammaticality – knowing "what not to say" (Boyd & Goldberg, 2011, p. 80) – is not yet fully developed.

To address the second research question, the present study compared the lists of the most frequent verbs between the L1 and L2 groups and revealed an overlap of the lead verbs in VACs between higher-level learners and native English speakers. A correlation analysis further confirmed the developmental tendency that Korean learners' verb usages in VACs increasingly approximate those of native English speakers with higher proficiency [e.g., from r (A1:LOCNESS) = 0.43 to r (C:LOCNESS) = 0.65], in line with Römer and Berger

(2019). The comparison of VAC knowledge between L2 learners and native English speakers, on the other hand, exhibited differences in some verb usages. The higher-level learners used some verbs that were not common in native usage (e.g., CONSIDER *about*), demonstrating learner creativity. Although these creative verb usages are not grammatical, they are important clues to a developing abstract construction.

The current study finally compared VAC knowledge across L2 English and L1 Korean to examine cross-linguistic transfer. By comparing the L1 Korean results of the verbal fluency task with the L2 English results, we identified influences of L1 typology and L1 collocational transfer on L2 VAC usage. Due to the verb-framed nature of their L1, Korean learners of English tended to favor general verbs such as BE and GO in motion constructions across L2 proficiency levels, whereas native English speakers commonly used verbs expressing a specific manner of motion (e.g., RUN, JUMP, and WALK). Our findings thus concur with those of Cadierno (2010) and Römer et al. (2014), whose investigations found preferences for general verbs or semantically bleached verbs in L2 VAC use by learners whose L1 is verb-framed. We also found that some verbs (e.g., CONCERN, STUDY, and DISCUSS) unique to L2 learners' VAC usage ("V *about* n") were attributed to collocational transfer from the corresponding L1 VAC. Moreover, considering that collocational transfer was observed at higher proficiency levels as well as lower, we posit a strong L1 influence on L2 VAC knowledge.

Based on the findings of the present study, some pedagogical implications can be made. As suggested by usage-based approaches and demonstrated in this study, L2 VAC knowledge develops as a result of cumulative information regarding verb usages from language experiences. In the early developmental stage, general and high frequency verbs play a role in VAC seeding, but eventually constructional schema, rather than individual verb usages, lead to VAC productivity. Unfortunately, however, English instruction in the Korean EFL context has often focused on discrete learning of grammatical rules and an extensive vocabulary list, which does not help to develop VAC knowledge in L2 English. Even when grammatical rules are presented in context, highlighting their communicative functions, the amount of exposure to repeated constructions with their pathbreaking verbs is still insufficient. Since VACs are often associated with the meanings of these particular verbs (Ellis & Ferreira-Junior, 2009a; Goldberg et al., 2004), it is important to offer plenty of opportunities to encounter these verb usages. Once the most entrenched verb-VAC associations are established, the existing schema constantly recruits new items, finally expanding the verb repertoire. Regarding crosslinguistic influences, on the other hand, Korean learners of English may benefit from explicit instruction. In the current study, the Korean learners failed to overcome L1 influences (e.g., L1 typology) even at the higher proficiency level, so it may be necessary to raise their awareness of alternate ways of thinking for speaking (Cadierno, 2008; Slobin, 1996).

Despite all the findings and implications, there are some limitations to our study. The psycholinguistic experiment that we conducted required some level of cognitive and linguistic ability to complete a bare decontextualized frame with appropriate verbs, which accounts for the scarcity of the results among lower proficiency levels. As such, future research needs to make an effort to ascertain lower-level learners' VAC knowledge more accurately. In addition, the present study investigated 19 VACs which are mostly verb-locative (VL) constructions, so our understanding of L2 constructional development remains limited. We therefore encourage future studies to include a comprehensive range of VACs from other constructions, such as verb object locative (VOL) and verb object object (VOO). Lastly, a longitudinal study tracking individual learners' L2 usages over time in relation to their overall L2 experiences would contribute to an even better understanding of L2 constructional development. It is our hope that this study has provided second language teachers and researchers with useful insights into usage-based approaches to language acquisition and development, and that it inspires them to take a step forward in their fields.

Applicable level: Tertiary

REFERENCES

Ambridge, B., & Lieven, E. (2011). *Child language acquisition: Contrasting theoretical approaches*. Cambridge, UK: Cambridge University Press.

Behrens, H. (2009). Usage-based and emergentist approaches to language acquisition. *Linguistics*, *47*, 383-411.

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

Boyd, J. K., & Goldberg, A. (2011). Learning what not to say: The role of statistical preemption and categorization in a-adjective production. *Language*, *87*, 55-83.

Bybee, J., & Hopper, P. (Eds.). (2001). *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.

Cadierno, T. (2008). Learning to talk about motion in a foreign language. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language*

acquisition (pp. 239-275). New York: Routledge.

Cadierno, T. (2010). Motion in Danish as a second language: Does the learner's L1 make a difference? In Z.-H. Han & T. Cadierno (Eds.), *Linguistic relativity in second language acquisition: Thinking for speaking* (pp. 1-33). Clevedon, UK: Multilingual Matters.

Chomsky, N. (1988). *Language and problems of knowledge: The Managua lectures.* Cambridge, MA: MIT Press.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge, U.K: Press Syndicate of the University of Cambridge.

Croft, W. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. Oxford, UK: Oxford University Press.

Dąbrowska, E., & Divjak, D. (Eds.). (2015). *Handbook of cognitive linguistics*. Berlin, Germany: Mouton De Gruyter.

Dąbrowska, E., & Lieven, E. (2005). Towards a lexically specific grammar of children's question constructions. *Cognitive Linguistics*, *16*, 437-474.

Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics*, *32*, 17-44.

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

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

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

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

Ellis, N. C., & Simpson-Vlach, R. (2009). Formulaic language in native speakers: Triangulating psycholinguistics, corpus linguistics, and education. *Corpus Linguistics and Linguistic Theory*, *5*, 61-78.

Ellis, N. C., Römer, U., & O'Donnell, M. B. (2016). *Usage-based approaches to language acquisition and processing: Cognitive and corpus investigations of construction grammar*. Malden, MA: Wiley.

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

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

372.

Eskildsen, S. W. (2014). What's new? A usage-based classroom study of linguistic routines and creativity in L2 learning. *International Review of Applied Linguistics*, *52*, 1-30.

Eskildsen, S. W. (2017). The emergence of creativity in L2 English: A usage-based case study. In N. Bell (Ed.), *Multiple perspectives on language play* (pp. 281-316). Berlin, Germany: Mouton de Gruyter.

Francis, G., Hunston, S., & Manning, E. (Eds.). (1996). *Grammar patterns 1: Verbs*. London: HarperCollins.

Gilquin, G., & Gries, S. T. (2009). Corpora and experimental methods: A state-of-the-art review. *Corpus Linguistics and Linguistic Theory*, *5*(1), 1-26.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago, IL: University of Chicago Press.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language.* Oxford, UK: Oxford University Press.

Goldberg, A. E., Casenhiser, D. M., & Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive Linguistics*, *15*, 289-316.

Goldberg, A. E., Casenhiser, D., & White, T. R. (2007). Constructions as categories of language. *New Ideas in Psychology*, *25*, 70-86.

Granger, S. (1998). The computer learner corpus: A versatile new source of data for SLA research. In S. Granger (Ed.), *Learner English on Computer* (pp. 3-18). London: Addison Wesley Longman.

Gries, S. T., & Wulff, S. (2005). Do foreign language learners also have constructions? Evidence from priming, sorting, and corpora. *Annual Review of Cognitive Linguistics*, *3*(1), 182-200.

Gruenewald, P. J., & Lockhead, G. R. (1980). The free recall of category examples. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 225-241.

Hopper, P. J. (1987). Emergent grammar. *Berkeley Linguistics Society*, *13*, 139-157.

Kim, H., & Rah, Y. (2016). Effects of verb semantics and proficiency in second language use of constructional knowledge. *The Modern Language Journal*, *100*(3), 716-731.

Kim, H., Rah, Y, & Hwang, H. (2020). Testing usage-based approaches to assessing EFL learners' development of English argument structure constructions. *English Teaching*, *75*(s1), 55-78.

Langacker, R. W. (1987). *Foundations of cognitive grammar: Theoretical prerequisites* (Vol. 1). Stanford, CA: Stanford University Press.

Li, P., Eskildsen, S. W., & Cadierno, T. (2014). Tracing an L2 learner's motion constructions over time: A usage-based classroom investigation. *The Modern Language Journal*, *98*, 612-628.

Lieven, E., Behrens, H., Speares, J., & Tomasello, M. (2003). Early syntactic creativity: A

usage-based approach. *Journal of Child Language*, *30*(2), 333-370.

Lieven, E., & Tomasello, M. (2008). Children's first language acquisition from a usage-based perspective. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 168-196). New York: Routledge.

MacWhinney, B., & O'Grady, W. (Eds.). (2015). *The handbook of language emergence*. Oxford, UK: Wiley-Blackwell.

Ninio, A. (1999). Pathbreaking verbs in syntactic development and the question of prototypical transitivity. *Journal of Child Language*, *26*, 619-653.

Pinker, S. (1994). *The language instinct*. New York: William Morrow.

Rhee, S. C., & Jung, C. K. (2014). Compilation of the Yonsei English Learner Corpus (YELC) 2011 and its use for understanding current usage of English by Korean pre-university students. *Journal of the Korea Contents Association*, *14*(11), 1019-1029.

Robinson, P., & Ellis, N. C. (Eds.). (2008). *Handbook of cognitive linguistics and second language acquisition*. London: Routledge.

Römer, U. (2019). A corpus perspective on the development of verb constructions in second language learners. *International Journal of Corpus Linguistics*, *24*(3), 268-290.

Römer, U., & Berger, C. M. (2019). Observing the emergence of constructional knowledge: Verb patterns in German and Spanish learners of English at different proficiency levels. *Studies in Second Language Acquisition*, *41*(5), 1089-1110.

Römer, U., O'Donnell, M. B., & Ellis, N. C. (2014). Second language learner knowledge of verb-argument constructions: Effects of language transfer and typology. *The Modern Language Journal*, *98*, 952-975.

Römer, U., Roberson, A., O'Donnell, M. B., & Ellis. N. C. (2014). Linking learner corpus and experimental data in studying second language learners' knowledge of verb-argument constructions. *ICAME Journal*, *38*, 115-135.

Römer, U., Skalicky, S., & Ellis, N. C. (2018). Verb-argument constructions in advanced L2 English learner production: Insights from corpora and verbal fluency tasks. *Corpus Linguistics and Linguistic Theory*. Advance online publication. https://doi.org/10.1515/cllt-2016-0055.

Scott, M. (2016). *WordSmith Tools* (Version 7). Liverpool: Lexical Analysis Software.

Shin, G.-H. (2010). On the contribution of argument structure constructions to sentence meaning for Korean learners of English. *English Teaching*, *65*(4), 209-227.

Slobin, D. (1996). From "thought and language" to "thinking for speaking." In J. Gumperz & S. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70-96). Cambridge: Cambridge University Press.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical form. In T. Shopen

(Ed.), *Language typology and syntactic description: Grammatical categories and the lexicon* (pp. 57-149). Cambridge, UK: Cambridge University Press.

Tannenbaum, R. J., & Wylie, E. C. (2008). *Linking English language test scores onto the Common European Framework of Reference: An application of standard-setting methodology* (TOEFL iBT Research Report RR-08-34). Princeton, NJ: Educational Testing Service.

Tomasello, M. (1992). *First verbs: A case study of early grammatical development*. Cambridge: Cambridge University Press.

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

Valenzuela, J., & Rojo, A. (2008). What can language learners tell us about constructions? In S. De Knop & R. De Rycker (Eds.), *Cognitive approaches to pedagogical grammar* (pp. 198-230). Berlin: De Gruyter Mouton.