

Research Article

OBSERVING THE EMERGENCE OF CONSTRUCTIONAL KNOWLEDGE

VERB PATTERNS IN GERMAN AND SPANISH LEARNERS OF ENGLISH AT DIFFERENT PROFICIENCY LEVELS

Ute Römer*

Georgia State University

Cynthia M. Berger

Georgia State University and Duolingo

Abstract

Based on writing produced by second language learners at different proficiency levels (CEFR A1 to C1), we adopted a usage-based approach (Ellis, Römer, & O'Donnell, 2016; Tyler & Ortega, 2018) to investigate how German and Spanish learner knowledge of 19 English verb-argument constructions (VACs; e.g., “V with n,” illustrated by *he always agrees with her*) develops. We extracted VACs from subsets of the Education First-Cambridge Open Language Database, altogether comprising more than 68,000 texts and 6 million words. For each VAC, L1 learner group, and proficiency level, we determined type and token frequencies, as well as the most dominant verb-VAC associations. To study effects of proficiency and L1 on VAC production, we carried out correlation analyses to compare verb-VAC associations of learners at different levels and different L1 backgrounds. We also correlated each learner dataset with comparable data from a large reference corpus of native English usage. Results indicate that with increasing proficiency, learners expand their VAC repertoire and productivity, and verb-VAC associations move closer to native usage.

The authors would like to thank Georgia State University's Scholarly Support Grant Program for sponsoring the project “Language use, acquisition, and processing: Cognitive and corpus investigations of Construction Grammar”; Nick Ellis for his encouragement and support during the initial stages of the project; Matt O'Donnell and Stephen Skalicky for their help with data preparation for and data processing in R; and Luke Plonsky as well as two anonymous reviewers for their constructive comments on an earlier draft of this article.

*Correspondence concerning this article should be addressed to Ute Römer, Department of Applied Linguistics and ESL, Georgia State University, 25 Park Place NE, Suite 1500, Atlanta, GA 30303, USA. E-mail: uroemer@gsu.edu

Copyright © Cambridge University Press 2019

INTRODUCTION

A central aim of second language acquisition (SLA) research is to better understand how language items develop in learners over time (e.g., Kramsch, 2000; Ortega, 2012). While this includes items at all levels of linguistic description (e.g., sounds, morphemes, words, grammatical structures), recent phraseology research has highlighted the importance of learning more about items at the interface of lexis and grammar. It has been shown that such items, variably referred to as formulaic sequences, phrases, multiword units, or constructions, are indicators of learner proficiency and facilitators of fluent use of language (Ellis et al., 2016; Goldberg, 2006; Hunston & Francis, 2000; Nattinger & DeCarrico, 1992; Pawley & Syder, 1983; Schmitt & Carter, 2004).

Recent research in usage-based linguistics has indicated that we learn language by learning constructions (Ellis, 2002, 2003; Ellis, O'Donnell, & Römer, 2013; Ellis et al., 2016; Goldberg, Casenhiser, & Sethuraman, 2004; Tomasello, 2003). Constructions, defined as form-meaning pairings that are entrenched as language knowledge in the speaker's mind and conventionalized in the speech community (Bybee, 2010; Goldberg, 1995, 2003, 2006; Trousdale & Hoffmann, 2013), can be considered the building blocks of language. Among the thousands of constructions a language is made up of, verb-argument constructions (VACs) form a particularly important set. They form the core of sentences and are important meaning-carrying units. Examples of VACs include the "V n n" or ditransitive construction (consisting of a verb, the central element, followed by two nouns or noun phrases, as in *he baked her a cake*), and the "V over n" construction (a verb, followed by the preposition *over*, followed by a noun or noun phrase, as in *she jumped over the fence*).

While the development of language learners' verb constructional knowledge has been a focus in L1 acquisition research (Ambridge & Lieven, 2015; Behrens, 2009; Goldberg, 1999, 2014; Goldberg et al., 2004; Lieven, Pine, & Baldwin, 1997; Ninio, 1999, 2006; Perfors, Tenenbaum, & Wonnacott, 2010; Tomasello, 1992, 2003), research on the development of L2 constructions is comparatively sparse and has relied almost exclusively on small sets of data, often collected from only one learner or a small number of learners (Eskildsen, 2012, 2014, 2017; Roehr-Brackin, 2014; Tode & Sakai, 2016). One major reason for this scarcity of research is that, until recently, cross-sectional and longitudinal L2 learner corpora were not easily available (Meunier, 2015). The study we report on here uses data from a large new corpus of learner writing at different proficiency levels (EFCAMDAT, described in more detail in the following text) to trace the emergence of VACs in L2 learners of two typologically different first language backgrounds. Our goal is to gain a better understanding of second language learners' developing "construction" (at least as it pertains to a set of frequent VACs) as their proficiency in the target language increases. We are also interested in the role that the learners' first language may play in this context.

THE DEVELOPMENT OF CONSTRUCTIONAL KNOWLEDGE IN L2 LEARNERS

We noted previously that, compared to first language acquisition research, constructions have received considerably less attention in SLA. While we know from previous work on VACs in L2 learner production (Gries & Wulff, 2005; Römer, O'Donnell, & Ellis, 2014; Römer, Skalicky, & Ellis, 2018; Römer, Roberson, O'Donnell, & Ellis, 2014) that

advanced learners of English, like native speakers, have constructional knowledge, that advanced learners' VAC knowledge differs in systematic ways from that of native speakers, and that advanced learners' verb-VAC associations differ across L1 groups, we know relatively little about how this constructional knowledge unfolds over time. Even though, as Ortega and Byrnes (2008) point out, “[m]any researchers believe that longitudinal findings can uniquely bolster our knowledge of language acquisition processes” (p. 3), there is a general dearth of studies that systematically track the development of linguistic features in the language of L2 learners as they advance to higher proficiency levels. As the review of quantitative longitudinal studies in Ortega and Ibarra-Shea (2005) indicates, existing work that traces English learners' development has focused predominantly on features of L2 morphology, the acquisition order of tense forms, and L2 phonology. This situation has now somewhat improved and researchers have started to become more interested in the development of phenomena at the interface of lexis and grammar and explored L2 learners' phraseological or constructional competence from a quantitative perspective.

Crossley and Salsbury (2011), for example, studied the development of spoken bigrams (e.g., *I think, you know*) in six low-level L2 English learners of various L1 backgrounds over the course of one year. They found that bigram accuracy improved as the learners' time studying English increased. Using a method called CollGram, which assigns association scores to bigrams, Bestgen and Granger (2014) traced how contiguous two-word sequences develop in a longitudinal corpus of L2 learner writing. They observed that learners moved toward using fewer bigrams consisting of high-frequency words at higher proficiency levels. Finally, in a study based on a cross-sectional corpus of English texts written by beginner to advanced L1 German learners, Garner (2016) found that phrase-frames (e.g., *on the * hand, in the * of*) became more variable and less predictable at higher levels of proficiency. All these studies try to better understand learner language development by extracting constructions or phraseological sequences from learner production data in large groups (e.g., all contiguous two-word sequences, or all four-word phrase-frames) and focusing on those groups of items.

There have also been a few studies that present more fine-grained analyses that zoom in on individual constructions and discuss how those emerge in L2 learners. Eskildsen (2009) studied the emergence of constructions of the modal verb *can* in a small longitudinal corpus of oral classroom language produced by one L1 Spanish learner of English. He observed that “*can*-patterns become increasingly varied” over time but did not find evidence in his data for a “movement towards fully abstract constructions” (p. 350). In a study based on data from the same L1 Spanish learner, Eskildsen and Cadierno (2007) found that English negation constructions follow a similar trend, developing from the initially fixed *I don't know* to increasingly abstract patterns (see also Eskildsen, 2012). Li, Eskildsen, and Cadierno (2014) focused on the same learner's developing realizations of motion constructions around high-frequency verbs such as *come* and *go*. The authors found that, similar to what was observed in the negation study, the learner's inventory of constructions that express directed motion grew from a small set of fixed constructions to a larger set of more productive ones. Frequent motion verbs are used in increasingly varied expressions. In a follow-up study, the same group of authors gather data from four L2 English learners (two L1 Chinese and two L1 Spanish) to trace how their inventory of motion constructions develops over the span of three years (Eskildsen,

Cadierno, & Li, 2015). Their findings confirm trends reported in Li et al. (2014) and highlight a few differences between the Spanish and Chinese learners that are likely due to crosslinguistic transfer effects.

Also working with longitudinal learner data but focusing on different types of verb constructions, Ellis and Ferreira-Junior (2009b) examined how the verb locative (VL), verb object locative (VOL), and verb object object (VOO or ditransitive) constructions develop in the spoken language of seven ESL learners (L1 Italian and L1 Punjabi). For each VAC type, the authors observed strong correlations between the verb frequency profiles in learner production data and in the input the learners received, indicating a usage effect on speakers' verb-VAC associations. They also noted that learners initially used predominantly semantically general, high-frequency verbs (e.g., *go, put, give*) before expanding their repertoire to verbs that are more specific and less frequent in usage. While they have provided us with important insights into the development of L2 learners' constructional knowledge, these studies are limited in that they (a) are based on data produced by small numbers of learners (between one and seven), and (b) discuss a small number of construction types (with the exception of Eskildsen 2014 and 2017, which both attempted to describe the emergence of the entire construction inventory of a single learner). However, there are studies on learner constructional knowledge that are based on larger learner corpora and larger sets of VACs (Römer et al., 2014; Römer et al., 2018). These studies have demonstrated that, similar to native speakers, advanced L2 English learners have strong constructional knowledge that overlaps to some extent with that of L1 English speakers but also differs from it because of interferences with the constructional knowledge that learners have gained in their first languages. While important and informative, this work is limited in that it focuses exclusively on advanced L2 learners and does not include a developmental perspective.

THE CURRENT STUDY

The current study goes beyond previous work on L2 VAC knowledge and addresses its limitations by analyzing changes in learners' use of 19 verb-argument constructions in a large corpus of L1 German and L1 Spanish learner writing at five proficiency levels from beginner to advanced. It provides a fine-grained analysis of VAC development across levels and the distribution of verbs in VACs in learner production compared to native speaker usage data on the same VACs. The study aims to address the following research questions:

- RQ1: How do frequent VACs emerge in L2 learner writing?
- RQ2: How do learners' verb-VAC associations differ across proficiency levels?
- RQ3: Do learners' verb-VAC associations move closer to a native usage norm as language proficiency increases?
- RQ4: How do learners' verb-VAC associations differ across L1 groups if task and level are controlled for?

Addressing these questions will allow us to gain insights into the development of L2 learners' constructional knowledge and into how proficiency level and first language background affect the emergence of a selected group of VACs in written learner English.

METHOD

The analyses presented in this article are based on large sets of texts retrieved from the Education First-Cambridge Open Language Database (EFCAMDAT; Alexopoulou, Geertzen, Korhonen, & Meurers, 2015; Geertzen, Alexopoulou, & Korhonen, 2013). EFCAMDAT consists of student writing samples submitted to Education First's online language school, Englishtown (www.englishtown.com; now English Live), which has several million students worldwide. Incoming Englishtown students are required to take a test that places them in one of 16 proficiency levels. These 16 proficiency levels are aligned with the six levels of the CEFR (Common European Framework of Reference for Languages; Council of Europe, 2001; Hawkins & Buttery, 2010). For example, Englishtown levels 1–3 correspond to the lowest CEFR level A1 (“beginner” or “break-through”), Englishtown levels 4–6 correspond to CEFR level A2 (“elementary” or “waystage”), and so on up to Englishtown level 16, which corresponds to CEFR level C2 (“proficiency” or “mastery”).

The texts in EFCAMDAT are responses to writing tasks that are part of each of 16 teaching levels. Each level features eight task prompts for a total of 128 tasks. Learners are often asked to write a letter or an e-mail, summarize information provided in a text, or compose a short argumentative text in response to a topic prompt. Sample writing topics include introducing yourself by e-mail, writing a movie review, and covering a news story (Alexopoulou et al., 2015). In its first release, EFCAMDAT contained about half a million texts (more than 33 million words) written by learners from 172 nationalities. Because the database does not include information on each learner's language background, we used student nationality as a proxy for L1. Inspired by our earlier work on VACs in L2 production in which we included data produced by advanced L1 German and L1 Spanish learners of English and observed strong effects of the learners' L1 on their verb-in-VAC production (Ellis et al., 2016; Römer et al., 2014; Römer et al., 2014), we decided to again focus on these two groups and only selected learners from countries where German or Spanish is the official and dominant language. While having access to L1 background information for each learner would be desirable, recent EFCAMDAT-based studies have demonstrated strong national language effects in the database, as demonstrated by Alexopoulou et al. (2015), Murakami (2013), and Nisioi (2015). According to these studies, nationality information in EFCAMDAT is a reliable proxy for learner first language.

From EFCAMDAT, we selected all texts produced by learners in Germany (dominant L1 German) and learners in Mexico (dominant L1 Spanish). We included texts written at proficiency levels A1 through C1. For both learner groups, the numbers of texts produced at level C2 were extremely small. This level was hence excluded. Table 1 provides an overview of the EFCAMDAT subsets on which we based our analyses. Particularly large numbers of writing samples were available at CEFR levels A1 and A2. Overall we worked with more than 28,000 texts (2.8 million words) produced by German and more than 40,000 texts (3.2 million words) produced by Mexican learners of English. These texts were written by more than 12,000 learners. Even though EFCAMDAT (and the subsets we retrieved from it) contain(s) longitudinal data from the same learners at different proficiency levels, the majority of learners did not progress through all 16 levels (some may place into a higher Englishtown level when they start the course, others may

TABLE 1. Overview of EFCAMDAT subsets used in this study

Learner group	Number of texts	Number of learners	Number of words
Mexican A1	24,275	4,043	1,533,012
Mexican A2	10,572	1,596	1,012,049
Mexican B1	3,903	808	471,543
Mexican B2	1,158	273	178,907
Mexican C1	186	34	37,225
<i>Mexican all levels</i>	<i>40,094</i>	<i>6,754</i>	<i>3,232,736</i>
German A1	10,721	2,072	728,275
German A2	8,507	1,580	811,842
German B1	5,222	1,240	631,338
German B2	3,092	926	488,431
German C1	930	202	186,176
<i>German all levels</i>	<i>28,472</i>	<i>6,020</i>	<i>2,846,062</i>
Overall	68,566	12,774	6,078,798

drop out before reaching level 16). At the same time, because most learners complete more than one course level, EFCAMDAT is not a purely cross-sectional corpus either. For these reasons, the learner corpus used in the current study is best considered “pseudolongitudinal” (Jarvis & Pavlenko, 2008).

The 10 EFCAMDAT subsets listed in Table 1 were part-of-speech (POS) tagged with the help of the TagAnt software package (Anthony, 2014b). From these POS-annotated file sets, we exhaustively retrieved instances of the 19 target VACs listed in Table 2 using the concordance function in AntConc (Anthony, 2014a). These VACs constitute a portion of the constructions investigated by Ellis et al. (2016), which in turn were derived from the large catalog of verb patterns included in Francis, Hunston, and Manning (1996). They all follow the “V preposition n” pattern and have been shown to be frequent in usage (Ellis et al., 2016). We used combined POS and lexical item searches to retrieve sequences of any verb form followed by one of the prepositions listed in Table 2 (*about*, *across*, *after*, etc.). We carried out 190 searches altogether (19 VACs in 10 datasets). For each search, concordance results were saved as text files and filtered manually for true hits of each construction. For instance, for the “V *about* n” VAC, examples in which *about* was used as an adverb were excluded (e.g., “we invited about 30 people”). From the filtered concordances we then derived frequency-sorted lemmatized verb lists for each VAC and L1-level combination (e.g., “V *about* n” in German A1).

These learner-group and VAC-specific verb lists allowed for various comparisons that enabled us to address our research questions. Verb frequency lists were compared across constructions, across proficiency levels (A1 to C1, keeping learner L1 constant), across L1 groups (L1 German vs. L1 Spanish, keeping proficiency level constant), and between native speaker usage and each learner group. The native speaker usage data came from previous analyses of the selected VACs in the 100-million-word British National Corpus (BNC; Burnard, 2007), described in Ellis et al. (2016). Analytic steps included type-token comparisons, construction growth analysis, and correlation analysis of verb-VAC associations. The construction growth analysis involved identifying the top-10 most

TABLE 2. Verb-argument constructions included in this study (in alphabetical order)

VAC	EFCAMDAT example (and subcorpus)
V <i>about</i> n	I always hear about a sore throat caused by nervous cough. (German B1)
V <i>across</i> n	The supermarket is across the train station. (German A1)
V <i>after</i> n	I go out to run after the square to buy clothes. (Mexican A1)
V <i>against</i> n	The Corleone family was fighting against the other criminal families in New York. (Mexican A2)
V <i>among</i> n	Today, his paintings rank among the most expensive in the world. (German A2)
V <i>around</i> n	I want to travel around the world. (Mexican B1)
V <i>as</i> n	I have been working as personal trainer since 1999. (Mexican B2)
V <i>between</i> n	Isn't it very complicated to differentiate between deliberate and accidental discrimination? (German B2)
V <i>for</i> n	I'm looking for a job as a Marketing Assistant. (German A2)
V <i>in</i> n	I am in four extra classes. (Mexican C1)
V <i>into</i> n	If there is a flood, do not go into a basement. (Mexican B1)
V <i>like</i> n	And a complete trim will make you feel like the king. (German C1)
V <i>of</i> n	I also dream of having enough time for painting. (German B1)
V <i>off</i> n	The girl grabbed his laptop off him and ran off the street. (Mexican B2)
V <i>over</i> n	I hope you get over your addiction. (Mexican A2)
V <i>through</i> n	Therefore we need to push through changes to improve. (German A1)
V <i>toward</i> n	This money can also count toward paying back the loan. (German B2)
V <i>under</i> n	The keys are under the plant in front the door. (Mexican A1)
V <i>with</i> n	He experimented with bold colors on his paintings. (Mexican C1)

frequent verbs in a VAC at each learner proficiency level and plotting their cumulative normalized frequencies (normalized per 100,000 words) on a line graph (see Ellis & Ogden, 2017; Ellis & Ferreira-Junior, 2009a). Doing so allowed us to not only see when a verb emerged as a frequent occupant of a VAC but also whether the verb continued to play a role as a strong lexical associate of a VAC across proficiency levels. The correlation analysis allowed us to systematically compare, for a specific VAC, the verbs that were produced in it by groups of learners of different L1 backgrounds and at different levels. It also allowed us to correlate each learner group's preferred verb-VAC associations with those derived from native English usage (comparing EFCAMDAT and BNC data). For all comparisons, we calculated Pearson correlation coefficients (*r*) in R (R Development Core Team, 2012) and generated scatterplots to visualize distributions of sets of verbs. All calculations were based on the log10 transformations of the verb token frequencies. To avoid missing responses as a result of logging zero, all values were incremented by 0.01.

RESULTS

DISTRIBUTION OF VACS ACROSS DATASETS

To address RQ1 (How do frequent VACs emerge in L2 learner writing?), we compared the frequencies of each included VAC in the EFCAMDAT subsets across learner

TABLE 3. Overview of VACs across L1 German learner datasets

	German A1			German A2			German B1			German B2			German C1		
	Types	Tokens	Tkns per 100K	Types	Tokens	Tkns per 100K	Types	Tokens	Tkns per 100K	Types	Tokens	Tkns per 100K	Types	Tokens	Tkns per 100K
<i>V about</i> n	24	259	35.6	26	242	29.8	50	501	79.4	49	865	177.1	25	131	70.4
<i>V across</i> n	2	2	0.3	2	3	0.4	3	4	0.6	4	7	1.4	-	-	-
<i>V after</i> n	3	4	0.5	16	32	3.9	17	50	7.9	7	14	2.9	2	4	2.1
<i>V against</i> n	2	17	2.3	5	69	8.5	10	15	2.4	24	72	14.7	4	7	3.8
<i>V among</i> n	-	-	-	2	2	0.2	-	-	-	4	12	2.5	-	-	-
<i>V around</i> n	3	5	0.7	10	32	3.9	14	93	14.7	6	11	2.3	4	5	2.7
<i>V as</i> n	11	76	10.4	19	239	29.4	23	121	19.2	27	190	38.9	16	28	15.0
<i>V between</i> n	5	90	12.4	6	10	1.2	9	52	8.2	11	15	3.1	6	7	3.8
<i>V for</i> n	55	330	45.3	82	705	86.8	92	908	143.8	89	1,387	284.0	61	416	223.4
<i>V in</i> n	93	3,381	464.2	115	2,830	348.6	138	1,051	166.5	125	619	126.7	88	336	180.5
<i>V into</i> n	9	23	3.2	18	76	9.4	19	99	15.7	29	68	13.9	18	58	31.2
<i>V like</i> n	5	7	1.0	15	309	38.1	17	140	22.2	14	89	18.2	11	23	12.4
<i>V of</i> n	10	24	3.3	20	46	5.7	25	70	11.1	25	102	20.9	11	68	36.5
<i>V off</i> n	-	-	-	10	13	1.6	6	9	1.4	12	73	14.9	7	10	5.4
<i>V over</i> n	3	4	0.5	12	14	1.7	22	28	4.4	23	36	7.4	9	11	5.9
<i>V through</i> n	3	3	0.4	11	62	7.6	13	32	5.1	16	39	8.0	13	29	15.6
<i>V toward</i> n	-	-	-	1	1	0.1	1	1	0.2	7	12	2.5	2	3	1.6
<i>V under</i> n	4	8	1.1	4	7	0.9	8	19	3.0	11	22	4.5	1	9	4.8
<i>V with</i> n	57	303	41.6	90	732	90.2	119	800	126.7	107	740	151.5	63	219	117.6

TABLE 4. Overview of VACs across L1 Spanish learner datasets

	Spanish A1			Spanish A2			Spanish B1			Spanish B2			Spanish C1		
	Types	Tokens	Tkns per 100K	Types	Tokens	Tkns per 100K	Types	Tokens	Tkns per 100K	Types	Tokens	Tkns per 100K	Types	Tokens	Tkns per 100K
<i>V about</i> n	20	134	8.7	35	299	29.5	45	505	107.1	34	323	180.5	11	26	69.8
<i>V across</i> n	2	2	0.1	3	3	0.3	-	-	-	2	2	1.1	-	-	-
<i>V after</i> n	7	8	0.5	3	9	0.9	2	4	0.8	1	1	0.6	-	-	-
<i>V against</i> n	1	3	0.2	6	64	6.3	2	4	0.8	6	11	6.1	1	2	5.4
<i>V among</i> n	-	-	-	1	1	0.1	1	2	0.4	2	3	1.7	-	-	-
<i>V around</i> n	5	12	0.8	8	38	3.8	9	80	17.0	4	4	2.2	1	1	2.7
<i>V as</i> n	5	24	1.6	13	45	4.4	14	65	13.8	13	47	26.3	5	5	13.4
<i>V between</i> n	6	261	17.0	3	9	0.9	7	24	5.1	3	3	1.7	1	1	2.7
<i>V for</i> n	45	352	23.0	76	627	62.0	54	605	128.3	44	457	255.4	15	60	161.2
<i>V in</i> n	81	7,847	511.9	144	3,342	330.2	128	918	194.7	79	295	164.9	28	75	201.5
<i>V into</i> n	4	4	0.3	21	46	4.5	22	66	14.0	14	24	13.4	5	7	18.8
<i>V like</i> n	6	19	1.2	13	425	42.0	23	96	20.4	9	30	16.8	2	4	10.7
<i>V of</i> n	17	56	3.7	47	125	12.4	27	64	13.6	14	25	14.0	4	7	18.8
<i>V off</i> n	2	3	0.2	6	11	1.1	8	12	2.5	7	27	15.1	1	2	5.4
<i>V over</i> n	-	-	-	12	13	1.3	7	9	1.9	4	10	5.6	-	-	-
<i>V through</i> n	2	2	0.1	7	44	4.3	9	13	2.8	4	6	3.4	5	9	24.2
<i>V toward</i> n	-	-	-	4	4	0.4	1	1	0.2	1	5	2.8	-	-	-
<i>V under</i> n	2	11	0.7	6	10	1.0	6	27	5.7	1	2	1.1	2	4	10.7
<i>V with</i> n	60	461	30.1	97	883	87.2	89	588	124.7	61	213	119.1	29	54	145.1

proficiency levels. Table 3 presents an overview of verb types and tokens across VACs and levels for the L1 German learner datasets. Table 4 gives the same overview for L1 Spanish learner data. The first thing we noticed were the rather low type and token frequencies of a majority of VACs in both L1 German and L1 Spanish learner writing. Given the large corpora sizes and given how frequent all these VACs are in native usage (Ellis et al., 2016), we had expected higher numbers overall.¹ Frequencies were particularly low for “V across n,” “V among n,” “V off n,” “V toward n,” and “V under n” but robust for “V about n,” “V for n,” “V in n,” and “V with n,” with normalized frequencies above 10 per 100,000 words across all EFCAMDAT subcorpora. At the lowest proficiency level (A1), a large group of VACs are either not attested at all or have token numbers in the single digits. This applies to 10 VACs in the German A1 and to nine VACs in the Spanish A1 dataset. At level A2, all 19 VACs are attested in both corpora. For some VACs, including “V for n,” “V into n,” and “V of n,” the type and token frequencies show a clear, though not always steady, increase from lower to higher proficiency levels, resulting in higher type-token ratios. This implies that these VACs become more productive as learners become more proficient. The lack of a steady increase can likely be attributed to smaller subcorpora sizes at higher proficiency levels. VACs such as “V about n” and “V with n” are very productive with high type numbers across levels, from beginner to advanced. In the following, we will focus on a subset of the 19 VACs for which token numbers were robust across datasets and sufficient for quantitative analyses. These focus VACs are “V about n,” “V for n,” “V in n,” and “V with n.”

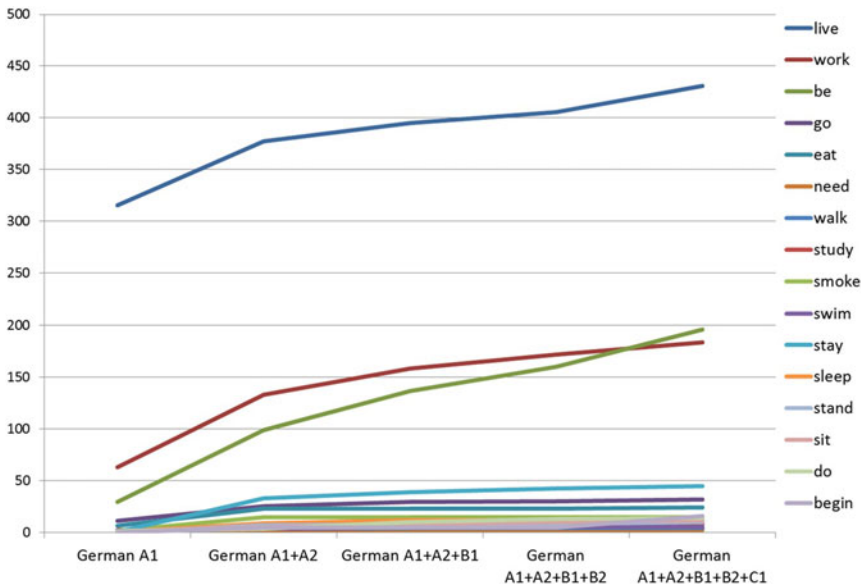


FIGURE 1. Emerging verbs in “V in n” in L1 German learners.

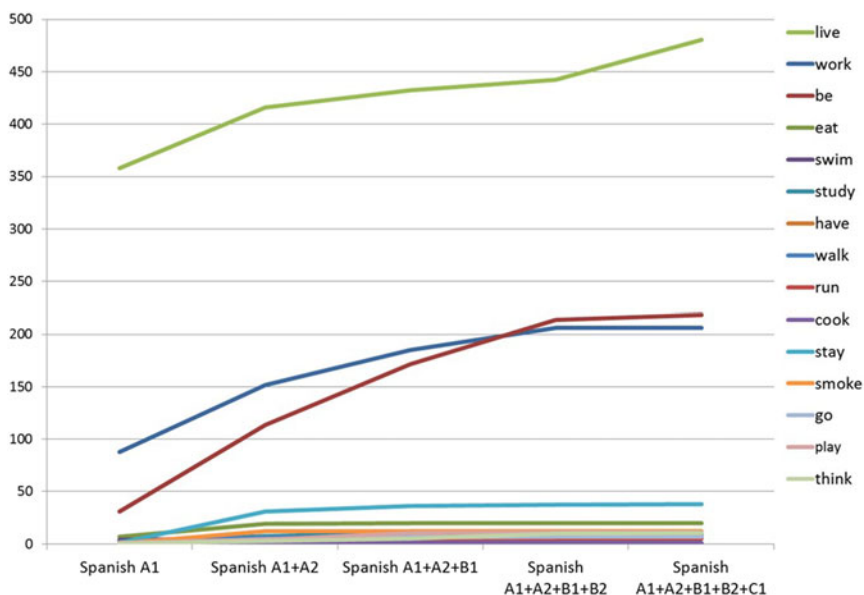


FIGURE 2. Emerging verbs in “V in n” in L1 Spanish learners.

EMERGING VERBS IN VACS

To study VAC lexical development in more detail and further address RQ1 as well as RQ2 (How do learners’ verb-VAC associations differ across proficiency levels?), we carried out a construction growth analysis that tracks emerging verbs within a VAC. We only applied this method to VACs with type numbers higher than 10 and sufficient token numbers across levels. Figures 1 and 2 show the construction growth results (cumulative normalized frequencies, following Ellis & Ogden [2017] and Ellis & Ferreira-Junior [2009a]) for “V in n” in L1 German and L1 Spanish learners. Figures 3 and 4 provide evidence of emerging verbs in the “V with n” VAC for the same two learner groups. For “V in n,” the most frequent verb at the lowest proficiency level (A1) in both L1 plots is *live*, followed by *work* and *be*. At this level, these three verbs are the primary occupants of the VAC. All other verbs occur with frequencies of 11.4 per 100,000 words or below in the L1 German and with frequencies of 6.9 or below in the L1 Spanish data. At proficiency levels A2 and B1 additional verbs emerge in this VAC (including *stay*, *sleep*, *stand*) but the most frequent ones remain *live*, *work*, and *be*. All three of them increase consistently across proficiency levels and continue to play an important role there, while the lines for the lower frequency verbs in this VAC quickly level out. Similar to what Ellis and Ferreira-Junior (2009a) observed for *go* in the verb-locative construction, we can say that in the “V in n” VAC for these learners the verb *live* functions as the “pathbreaking verb that seeds the construction and leads its development” (Ellis & Ferreira-Junior, 2009a, p. 375).

In terms of verb development, the growth of “V with n” appears to be more dynamic than that of “V in n,” and also exhibits more variation across the two L1 groups (see

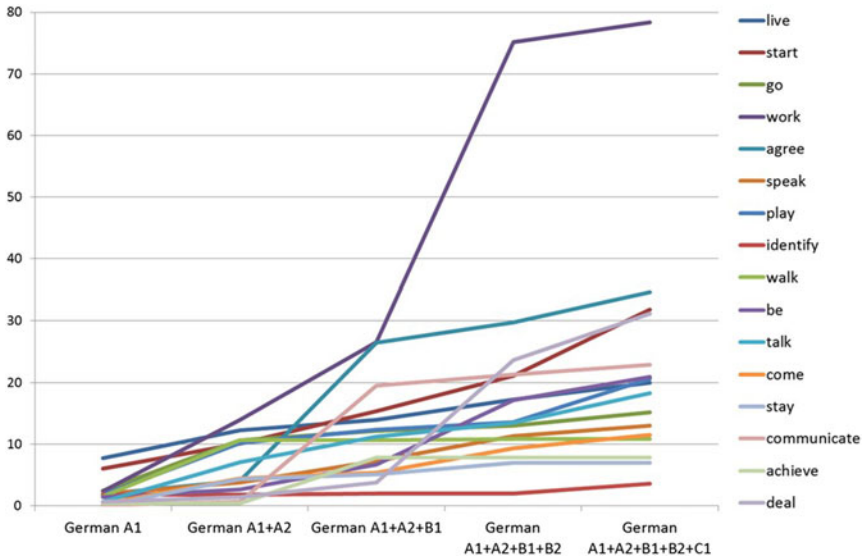


FIGURE 3. Emerging verbs in “V with n” in L1 German learners.

Figures 3 and 4). At level A1, verb-in-VAC frequencies range between a low 0.1 and 9 per 100,000 words for the top-10 verbs in both L1s (with *live* being used most frequently). At level B1, several verbs, including *work*, *agree*, *communicate*, and *start*, become considerably more entrenched in this VAC. In the L1 German data (Figure 3), these verbs occur with cumulative normed frequencies of 15.3 to 26.6. All four verbs (particularly *work*) continue to be used frequently in this VAC at the higher proficiency levels with cumulative frequencies going up to 34.6 per 100,000 words for *agree* and 78.3 for *work* in the L1 German data. This growth is less pronounced in the L1 Spanish data (Figure 4) where a small number of verbs become more frequently used at proficiency levels B1 to C1 but only show maximum cumulative frequencies of 27.2 (*talk*), 35.6 (*agree*), 37 (*be*), and 44.1 (*work*) at the highest level. There is no distinct lead verb for this VAC as there is in the L1 German data even though both learner groups responded to the same task prompts at each level.

PROFICIENCY EFFECTS ON VERB-VAC ASSOCIATIONS

To further address RQ2 as well as RQ3 (Do learners’ verb-VAC associations move closer to a native usage norm as language proficiency increases?), we correlated verb frequencies in a VAC in two datasets at a time, calculated *r*-values, and plotted the log-transformed frequencies of verbs in VACs with the verbs used as data point labels. To better understand in what ways verb-VAC associations change with proficiency, we compared learner verb preferences across proficiency levels (e.g., German A1 vs. German B2); to see whether learners’ verb-VAC associations move closer to a usage norm as proficiency increases, we compared each proficiency level against native English usage (e.g., German A1 vs. BNC). *R*-values can range from 0 to 1, with 1

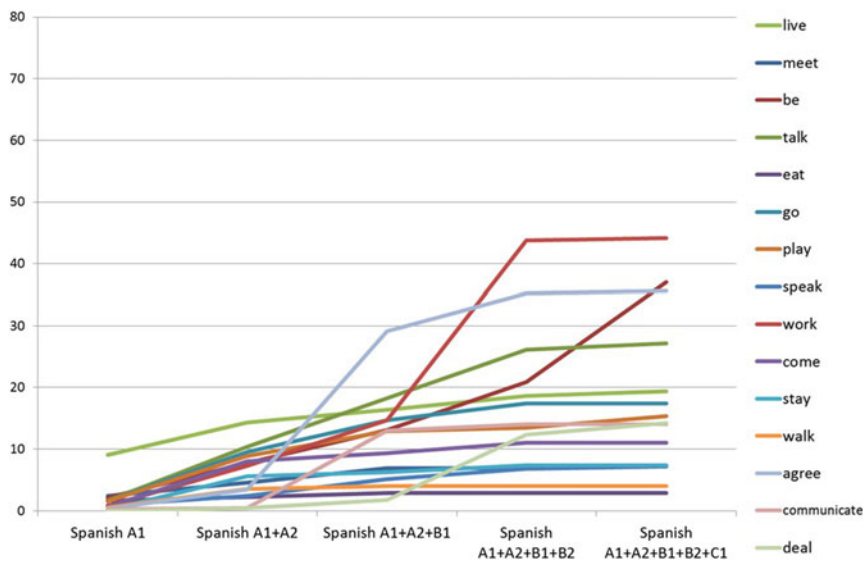


FIGURE 4. Emerging verbs in “V with n” in L1 Spanish learners.

indicating perfect and 0 indicating no correlation. The closer the value to 1, the stronger the correlation between two verb lists. For all selected focus VACs (“V about n,” “V for n,” “V in n,” “V with n”), correlation values for adjacent proficiency levels (e.g., A1 and A2, B1 and B2) were higher than those for nonadjacent ones (e.g., A1 and B1, B1 and C1), and correlation values for the BNC comparison increased from lowest to highest learner proficiency levels (A1 to C1). All *r*-values were significant at $p < .001$. We will discuss correlations in detail for the “V about n” construction.

Table 5 lists the correlation values for “V about n” for all 10 possible comparisons across EFCAMDAT proficiency levels for the L1 German and L1 Spanish datasets.

TABLE 5. Correlations (*r*-values) for verb usage in “V about n” across EFCAMDAT levels

Comparison	L1 German	L1 Spanish
A1 vs. A2	0.78	0.84
A1 vs. B1	0.62	0.76
A1 vs. B2	0.68	0.80
A1 vs. C1	0.66	0.68
A2 vs. B1	0.81	0.83
A2 vs. B2	0.78	0.80
A2 vs. C1	0.77	0.59
B1 vs. B2	0.78	0.76
B1 vs. C1	0.75	0.51
B2 vs. C1	0.72	0.65

These values are the result of comparing frequency verb lists for this VAC from two learner levels at a time, keeping learner L1 background controlled for. We see that at each level for which more than one comparison to a higher level is possible (A1, A2, B1), *r*-values are highest for adjacent levels (e.g., A2 vs. B1), and lower for nonadjacent levels (e.g., A2 vs. C1). This is true for both L1 groups. This means that learners at proficiency levels that are close together produce sets of verbs in this VAC that are more similar than learners at nonadjacent levels, indicating a development in learner verb-VAC associations from lower to higher proficiency levels.

Figures 5 to 9 show the correlation plots for verbs produced in “V about n” by learners (L1 German on the left, L1 Spanish on the right) at proficiency levels A1, A2, B1, B2, and C1 compared to verbs used in the same VAC in native English usage as captured in the BNC. In each scatterplot, the x-axis shows the logarithmic frequencies of verb types in L1 usage; the y-axis displays the logarithmic frequencies of verb types in a specific EFCAMDAT subset (e.g., German A1). Perfect overlap between two datasets (i.e., an *r*-value of 1) would place all verb labels neatly along the diagonal that runs through the middle of the plot (see the grey dotted lines in Figures 5 to 9). Verbs that appear above the diagonal are markedly more frequent in the EFCAMDAT than in the BNC data; verbs that are plotted below the diagonal are markedly less frequent in the EFCAMDAT than in the BNC data. The *r*-value is shown in the top left corner of each plot. All *r*-values were significant at $p < .001$. For the first comparison of verbs in “V about n” between the lowest (A1) level learner writing and native English usage data (Figure 5), correlations are fairly low for both L1s ($r = .46$ and $r = .49$). While considerably lower than those we observed between learner level groups (see Table 5), these correlations are still nontrivial and provide evidence for an effect of input frequencies on L2 VAC acquisition (in line with the findings reported by Ellis et al., 2016). Learners of both first language backgrounds, even at beginning proficiency levels, are influenced by verb-in-VAC token frequencies in L1 usage. Comparatively few verbs populate the plots in Figure 5 and all of them appear below the diagonal, indicating that they do appear in the learner data but not as frequently as in L1 usage. All other verbs that occur in this VAC in the BNC but

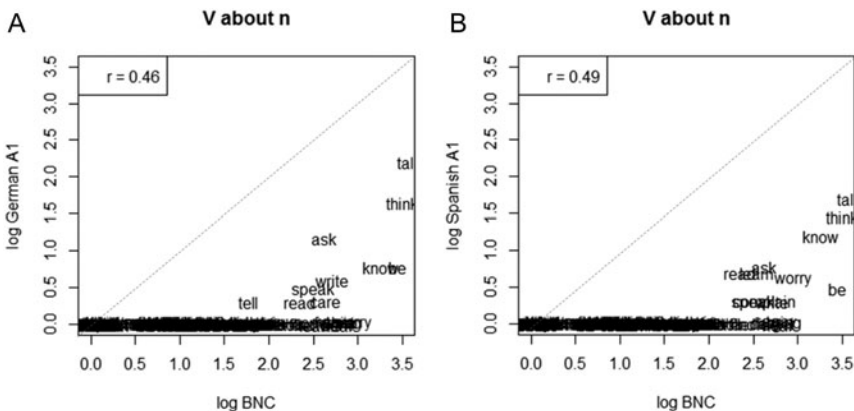


FIGURE 5. Correlations of verbs in learner writing at A1 “beginner” level (L1 German, left plot; L1 Spanish, right plot) and native usage (BNC) for “V about n”.

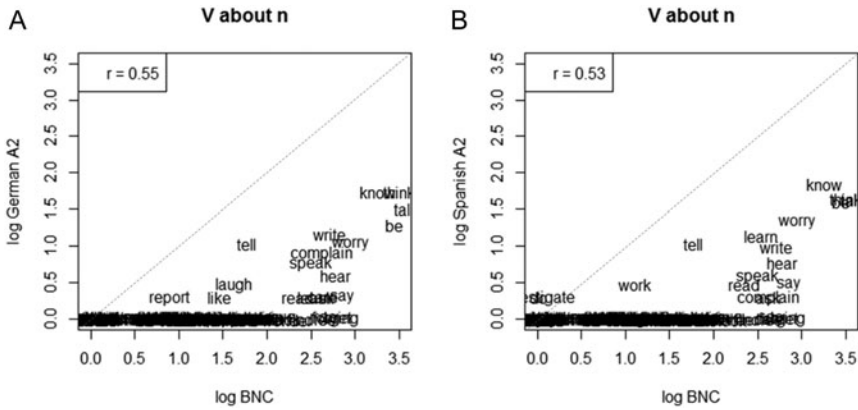


FIGURE 6. Correlations of verbs in learner writing at A2 “elementary” level (L1 German, left plot; L1 Spanish, right plot) and native usage (BNC) for “V about n”.

not in the EFCAMDAT data are plotted on top of each other at $y = 0$, resulting in a black bar at the bottom of each plot. The verbs that EFCAMDAT learners (both L1 German and L1 Spanish) at this lowest proficiency level repeatedly use in this VAC include *talk, think, know, ask, speak, and be*.

At the next proficiency level (A2; see Figure 6), both correlation plots become more populated, which indicates that additional verbs that were not part of the learners’ repertoire for this VAC at level A1 are now used. These include the verbs *hear, like, report, laugh, and complain*. Overall, verbs move closer to the diagonal, indicating that learner verb use in this VAC is becoming more similar to native usage—a trend that continues at the next highest proficiency level, B1 (see Figure 7). At level A2, we also notice an increase in r -values to $r = .55$ for L1 German and $r = .53$ for L1 Spanish. While r -values remain at this level as we move on to the intermediate proficiency groups (B1

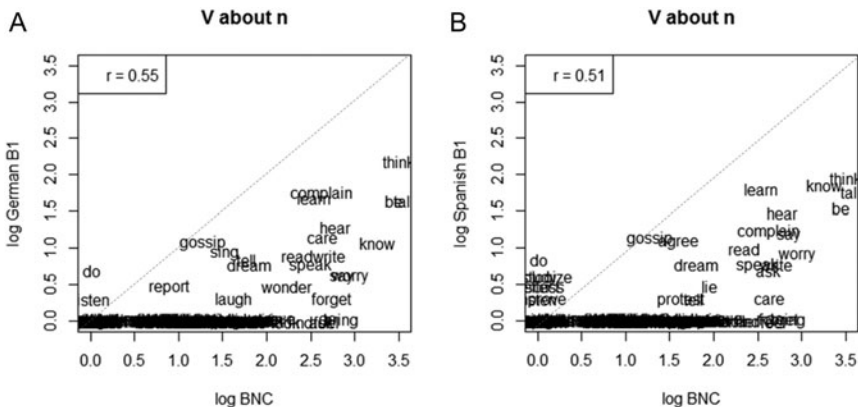


FIGURE 7. Correlations of verbs in learner writing at B1 “intermediate” level (L1 German, left plot; L1 Spanish, right plot) and native usage (BNC) for “V about n”.

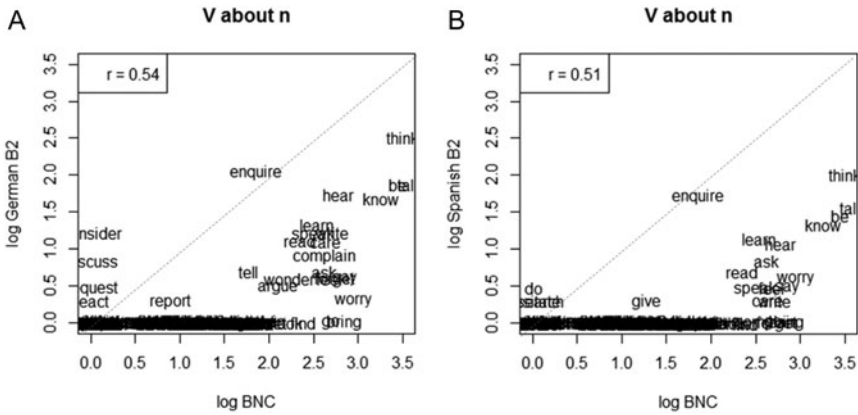


FIGURE 8. Correlations of verbs in learner writing at B2 “upper-intermediate” level (L1 German, left plot; L1 Spanish, right plot) and native usage (BNC) for “V about n”.

and B2), verb plots continue to change. As we can see in Figures 7 and 8, the plots become increasingly populated as this VAC gets more productive in learner writing. At the B1 and B2 levels, we also notice a few verbs that are plotted above the diagonal, which means that they are used relatively more frequently in learner production than in native usage. This applies to *do* and *listen* in the L1 German B1 data and to *do*, *apologize*, *improve*, and *discuss* in the L1 Spanish B1 data. In the L1 German comparison plot at level B2, *consider*, *discuss*, *request*, and *react* appear above the diagonal. Learners at these two intermediate levels use this VAC with verbs that are semantically related to one of the core “V about n” verbs (*talk*) but that are not common in this construction in native English usage. In some cases, this leads to realizations of the VAC deemed unidiomatic by the authors and not attested in the 560-million-word Corpus of Contemporary

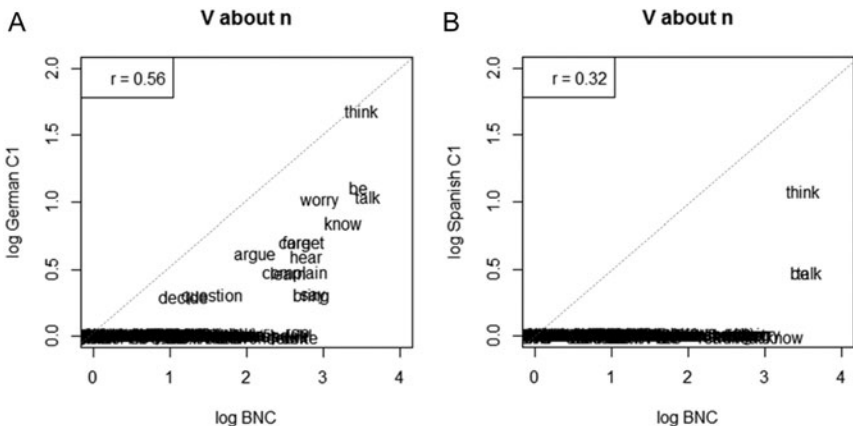


FIGURE 9. Correlations of verbs in learner writing at C1 “advanced” level (L1 German, left plot; L1 Spanish, right plot) and native usage (BNC) for “V about n”.

American English (COCA; Davies, 2008), such as “I urge you to consider about that job as a zookeeper,” “we request about a conversation to discuss all details” (both EFCAMDAT German B2), or “Would you like to join me and discuss about the contract?” (EFCAMDAT Spanish B1). This observation confirms findings based on German and Spanish learner data on the same VAC collected in verbal fluency tasks reported in Römer et al. (2018). In this study, the authors talk about “creative extensions of the core semantics of the VAC” (p. 18) in intermediate and advanced learner production data. At the highest proficiency level covered in our study, C1, these creative extensions no longer appear to play a role. The scatterplots in Figure 9 do not indicate any verbs that are used more frequently in the learner production data at this advanced level than in native usage. All verbs in the L1 German C1 plot appear underneath or, in the case of *think*, on the diagonal (for L1 Spanish, we have very little data from C1 level learners). There is no evidence of unidiomatic realizations of this VAC in the L1 German data at C1 level, which may indicate that learners at this level have figured out “what not to say” through statistical preemption (Boyd & Goldberg, 2011).

LEARNER L1 EFFECTS ON VERB-VAC ASSOCIATIONS

To address our final research question (RQ4; How do learners’ verb-VAC associations differ across L1 groups if task and level are controlled for?), we correlated verb distributions in selected VACs at each proficiency level between the two learner L1 groups (German vs. Spanish). For the selected VACs (“V about n,” “V for n,” “V in n,” “V with n”) we found high Pearson correlations throughout, ranging from $r = .72$ to $r = .92$. This indicates that for the selected constructions, learner L1 is not a strong variable that leads to the use of significantly different sets of verbs. As they respond to the same task prompts at each level, L1 German and L1 Spanish learners’ verb choices largely overlap. Correlations are particularly high at low proficiency levels (A1 and A2) and lower at levels B2 and C1, pointing to a more varied verb repertoire used by learners at higher levels of proficiency. The scatterplots in Figure 10 visualize two of these correlations, for

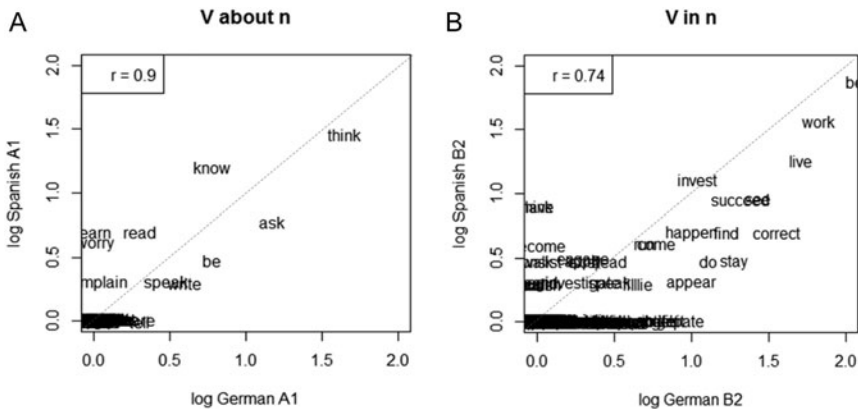


FIGURE 10. Correlations of verbs in learner writing across L1 groups for “V about n” at level A1 (left plot) and for “V in n” at level B2 (right plot).

“V about n” at level A1 and for “V in n” at level B2. In the left-hand plot (German A1 vs. Spanish A1, $r = .90$), the majority of verbs are plotted close to the diagonal. At this level, the learners in the two L1 groups show very similar verb associations for this VAC and use verbs such as *think*, *know*, and *ask* with similar frequencies. In the right-hand plot (German B2 vs. Spanish B2, $r = .74$), a number of verbs appear below the diagonal, indicating that they are used more frequently by L1 German than L1 Spanish learners at this level and contributing to a lower (but still high) overall correlation value.

DISCUSSION

The analyses presented in the previous sections have focused on comparing verb-argument construction use by L2 learners across proficiency levels and across L1 groups. They have enabled us to better understand how selected high-frequency VACs and the verbs in those VACs emerge in the written language production of L1 German and L1 Spanish learners of English. They have also enabled us to see how learners’ verb-VAC associations at different proficiency levels compare to verb-VAC associations in native speaker usage. Addressing RQ1, we observed that VACs increase in type number and overall also in productivity from beginning to more advanced proficiency levels. As learners become more proficient, their VAC repertoire grows and they learn to use a wider range of verbs within VACs. These findings are in line with previous studies on longitudinal L2 learner development of constructions (Eskildsen, 2009, 2012; Eskildsen & Cadierno, 2009; Li et al., 2014) and confirm these findings based on larger amounts of data produced by thousands of learners. The trends that these earlier studies observed in the language of a few individual learners, also hold true for large learner groups. In response to RQ2 and RQ3, we saw through comparisons with native speaker usage data and comparisons across levels that learners’ verb-VAC associations do not only become more varied, and hence more schematic, at higher proficiency levels but also move closer to a native usage norm. The noted development toward more variability and schematicity of constructions is in line with Garner’s (2016) observations on phrase frames in a cross-sectional learner corpus. The observed move in verb-VAC associations toward a native usage norm corroborates what Crossley and Salsbury’s (2011) found for bigrams in a small longitudinal learner corpus. A particularly interesting observation that resulted from our correlation analysis was that at proficiency level B2, learners are fairly creative in their verb use and produce verb-VAC combinations that appear to be semantic extensions of some of the core ones (e.g., *request/consider/discuss about* as extensions of *talk/think/speak about*). While these verbs are semantically related to the most prototypical verbs in the VAC (verbs of communication), they are not idiomatic and not (or not frequently) used by native English speakers. Finally, addressing RQ4, we found through comparisons of verb-VAC associations across L1 learner groups that for all target VACs with high enough token numbers for quantitative comparative analyses, there are strong similarities between L1 German and L1 Spanish learners’ uses of verbs in VACs. Verb-VAC associations are most similar at the lowest and less similar at the higher proficiency levels, indicating that learners move further apart in their verb choices as they become more proficient and their verb repertoire expands.

CONCLUSION

As Lowie and Verspoor (2015, p. 78) remind us, “All of the most relevant questions about SLA . . . , are implicitly or explicitly about change over time.” By comparing VACs in language data produced by L2 learners at different proficiency levels and correlating them with data on the same VACs from native English usage, this study has contributed new corpus-derived evidence in support of usage-based approaches to SLA with a focus on changing verb-VAC association over time. Confirming and expanding on results from previous usage-based SLA studies that focused on smaller sets of VACs and used data from smaller learner corpora (Bestgen & Granger, 2014; Ellis & Ferreira-Junior, 2009b; Eskildsen, 2009; Eskildsen & Cadierno, 2009; Römer, & Garner, in press), our findings indicate that the emergence of constructions in learners is characterized by the following: an expansion of the learners’ VAC repertoire in terms of VAC types; growth in the learners’ verb-in-VAC repertoire from small sets of high-frequency, general verbs to larger sets of more specific, lower-frequency verbs; an increase in the schematicity of selected high-frequency VACs; and a move in verb-VAC associations toward a native usage norm. We also observed that VAC acquisition appears to be strongly impacted by input frequencies, as indicated by the overlap of verbs that appear in usage and those used by learners.

We acknowledge that our work on the development of L2 constructions must not stop here as there are a number of tasks that need yet to be carried out. Future tasks should include an expansion of the set of focus VACs (ideally to a comprehensive analysis that looks at *all* VACs and their development systematically), the inclusion of data produced by learners of additional learner L1 backgrounds, and, if sufficient data can be extracted from EFCAMDAT, the detailed longitudinal analysis of data produced by individual learners who move through all (or most) levels. Also worth including in our agenda are an analysis of the 128 writing tasks that learners respond to trace potential task effects, and a qualitative exploration of cross-linguistic transfer effects from the learners’ L1s on their verb selection. We look forward to picking up some of these tasks in our future work and, through this work, hope to be able to contribute to an even better understanding of how constructions evolve in second language learners.

NOTE

¹To give a few examples, in the BNC “V across n” occurs 5.3 times per 100k words, “V into n” 50 times per 100k words, “V over n” 19.7 times per 100k words, and “V under n” 11 times per 100k words.

REFERENCES

- Alexopoulou, T., Geertzen, J., Korhonen, A., & Meurers, D. (2015). Exploring big educational learner corpora for SLA research: Perspectives on relative clauses. *International Journal of Learner Corpus Research*, 1, 96–129.
- Ambridge, B., & Lieven, E. (2015). A constructivist account of child language acquisition. In B. MacWhinney & W. O’Grady (Eds.), *The handbook of language emergence* (pp. 478–510). Malden, MA: Wiley-Blackwell.
- Anthony, L. (2014a). AntConc (Version 3.4.2) [Computer Software]. Tokyo, Japan: Waseda University. Retrieved from <http://www.laurenceanthony.net/software>.

- Anthony, L. (2014b). TagAnt (Version 1.1.0) [Computer Software]. Tokyo, Japan: Waseda University. Retrieved from <http://www.laurenceanthony.net/software>.
- Barlow, M., & Kemmer, S. (Eds.). (2000). *Usage-based models of language*. Stanford, CA: CSLI Publications.
- Behrens, H. (2009). Usage-based and emergentist approaches to language acquisition. *Linguistics*, 47, 383–411.
- Bestgen, Y., & Granger, S. (2014). Quantifying the development of phraseological competence in L2 English writing: An automated approach. *Journal of Second Language Writing*, 26, 28–41.
- 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.
- Burnard, L. (2007). Reference Guide for the British National Corpus (XML edition). Retrieved from <http://www.natcorp.ox.ac.uk/docs/URG/>.
- Bybee, J. (2010). *Language, usage and cognition*. Cambridge, UK: Cambridge University Press.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Crossley, S., & Salsbury, T. L. (2011). The development of lexical bundle accuracy and production in English second language speakers. *International Review of Applied Linguistics in Language Teaching*, 49, 1–26.
- Davies, M. (2008). The Corpus of Contemporary American English (COCA). 560 million words, 1990–present. Retrieved from <https://corpus.byu.edu/coca/>.
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24, 143–188.
- Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 63–103). Malden, MA: Blackwell.
- 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, 188–221.
- Ellis, N. C., & Ogden, D. (2017). Thinking about multiword constructions: Usage-based approaches to acquisition and processing. *Topics in Cognitive Science*, 9, 604–620.
- Ellis, N. C., O'Donnell, M. B., & Römer, U. (2013). Usage-based language: Investigating the latent structures that underpin acquisition. *Language Learning*, 63, 25–51.
- 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*. Language Learning Monograph Series. Malden, MA: Wiley.
- Eskildsen, S. W. (2009). Constructing another language: Usage-based linguistics in second language acquisition. *Applied Linguistics*, 30, 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.
- Eskildsen, S. W., & Cadierno, T. (2007). Are recurring multi-word expressions really syntactic freezes? Second language acquisition from the perspective of usage-based linguistics. In M. Nenonen & S. Niemi (Eds.), *Collocations and idioms 1: Papers from the first Nordic Conference on Syntactic Freezes* (Vol. 41) (pp. 86–99). Joensuu, Finland: Joensuu University Press.
- Eskildsen, S. W., Cadierno, T., & Li, P. (2015). On the development of motion constructions in four learners of L2 English. In T. Cadierno & S. W. Eskildsen (Eds.), *Usage-based perspectives on second language learning* (pp. 207–232). Berlin, Germany: Walter de Gruyter.
- Francis, G., Hunston, S., & Manning, E. (1996). *Grammar patterns 1: Verbs*. London: HarperCollins.
- Gamer, J. R. (2016). A phrase-frame approach to investigating phraseology in learner writing across proficiency levels. *International Journal of Learner Corpus Research*, 2, 31–67.
- Geertzen, J., Alexopoulou, T., & Korhonen, A. (2013). Automatic linguistic annotation of large scale L2 databases: The EF-Cambridge Open Language Database (EFCAMDAT). *Proceedings of the 31st Second Language Research Forum (SLRF)*. Pittsburgh, PA: Cascadia Press.

- Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago, IL: University of Chicago Press.
- Goldberg, A. E. (1999). The emergence of the semantics of argument structure constructions. In B. MacWhinney (Ed.), *The emergence of language* (pp. 197–212). Mahwah, NJ: Lawrence Erlbaum.
- Goldberg, A. E. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Science*, 7, 219–224.
- Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford, UK: Oxford University Press.
- Goldberg, A. E. (2014). Patterns of experience in patterns of language. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (Vol. 1, pp. 187–202). New York, NY: Psychology Press.
- Goldberg, A. E., Casenhiser, D. M., & Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive Linguistics*, 15, 289–316.
- 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, 182–200.
- Hawkins, J. A., & Buttery, P. (2010). Criterial features in learner corpora: Theory and illustrations. *English Profile Journal*, 1, 1–23.
- Hunston, S., & Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam, The Netherlands: Benjamins.
- Kramsch, C. (2000). Second language acquisition, applied linguistics, and the teaching of foreign languages. *The Modern Language Journal*, 84, 311–326.
- Jarvis, S., & Pavlenko, A. (2008). *Crosslinguistic influence in language and cognition*. New York, NY: Routledge.
- 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., Pine, J. M., & Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language*, 24, 187–219.
- Lowie, W., & Verspoor, M. (2015). Variability and variation in second language acquisition orders: A dynamic reevaluation. *Language Learning*, 65, 63–88.
- Meunier, F. (2015). Developmental patterns in learner corpora. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 379–400). Cambridge, UK: Cambridge University Press.
- Murakami, A. (2013). *L1 Influence and individual variation in the L2 accuracy development of grammatical morphemes: Insights from learner corpora* (Unpublished doctoral dissertation). University of Cambridge.
- Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford, UK: Oxford University Press.
- Ninio, A. (1999). Pathbreaking verbs in syntactic development and the question of prototypical transitivity. *Journal of Child Language*, 26, 619–653.
- Ninio, A. (2006). *Language and the learning curve: A new theory of syntactic development*. Oxford, UK: Oxford University Press.
- Nisioi, S. (2015). Feature analysis for native language identification. In A. Gelbukh (Ed.), *Computational linguistics and intelligent text processing: CICLing 2015*. Cham, Switzerland: Springer.
- Ortega, L. (2012). Interlanguage complexity: A construct in search of theoretical renewal. In B. Szmrecsanyi & B. Kortmann (Eds.), *Linguistic complexity in interlanguage varieties, L2 varieties, and contact languages* (pp. 127–155). Berlin, Germany: Walter de Gruyter.
- Ortega, L. & Byrnes, H. (2008). The longitudinal study of advanced L2 capacities: An introduction. In L. Ortega & H. Byrnes (Eds.), *The longitudinal study of advanced L2 capacities* (pp. 3–20). New York, NY: Routledge.
- Ortega, L., & Iberri-Shea, G. (2005). Longitudinal research in second language acquisition: Recent trends and future directions. *Annual Review of Applied Linguistics*, 25, 26–45.
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191–227). London, UK: Longman.
- Porfors, A., Tenenbaum, J. B., & Wonnacott, E. (2010). Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of Child Language*, 37, 607–642.

- R Development Core Team. (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Roehr-Brackin, K. (2014). Explicit knowledge and processes from a usage-based perspective: The developmental trajectory of an instructed L2 learner. *Language Learning*, 64, 771–808.
- Römer, U., & Garner, J. R. (in press). The development of verb constructions in spoken learner English: Tracing effects of usage and proficiency. *International Journal of Learner Corpus Research*.
- 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., 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>.
- 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, 59–79.
- Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 1–22). Amsterdam, The Netherlands: John Benjamins.
- 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.
- Tode, T., & Sakai, H. (2016). Exemplar-based instructed second language development and classroom experience. *International Journal of Applied Linguistics*, 167, 210–234.
- Tomasello, M. (1992). *First verbs: A case study of early grammatical development of cognition and action*. Cambridge, MA: MIT Press.
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Trousdale, G., & Hoffmann, T. (Eds.), (2013). *Oxford handbook of construction grammar*. Oxford, UK: Oxford University Press.
- Tyler, A. E., & Ortega, L. (2018). Usage-inspired L2 instruction. An emergent, researched pedagogy. In Tyler, A. E., Ortega, L., Uno, M., & Park, H. I. (Eds.). *Usage-inspired L2 instruction. Researched pedagogy* (pp. 3–26). Amsterdam: John Benjamins.