Construction Learning as a Function of Frequency, Frequency Distribution, and Function. — Source link

Nick C. Ellis, Fernando Ferreira-Junior

Institutions: Universidade Federal de Minas Gerais

Related papers:

- Frequency Effects in Language Processing: A Review with Implications for Theories of Implicit and Explicit Language Acquisition.

- Learning argument structure generalizations.

- Constructions at Work: The Nature of Generalization in Language

- Constructions: A Construction Grammar Approach to Argument Structure

- Language Acquisition as Rational Contingency Learning

Construction Learning as a Function of Frequency, Frequency Distribution, and Function

NICK C. ELLIS
University of Michigan
500 E. Washington Street
Ann Arbor, MI 48104
Email: ncellis@umich.edu

FERNANDO FERREIRA–JUNIOR
Federal University of Minas Gerais
Av. Antonio Carlos 6627, Pampulha
UFMG/FALE/PosLin, Sala 4039
Belo Horizonte, MG, CEP 31270-901
Brazil
Email: fernandoufmg@gmail.com

This article considers effects of construction frequency, form, function, and prototypicality on second language acquisition (SLA). It investigates these relationships by focusing on naturalistic SLA in the European Science Foundation corpus (Perdue, 1993) of the English verb–argument constructions (VACs): verb locative (VL), verb object locative (VOL), and ditransitive (VOO). Goldberg (2006) argued that Zipfian type/token frequency distributions (Zipf, 1935) in natural language constructions might optimize learning by providing one very high-frequency exemplar that is also prototypical in meaning. This article tests and confirms this proposal for naturalistic English as a second language. We show that VAC type/token distribution in the input is Zipfian and that learners first use the most frequent, prototypical, and generic exemplar (e.g., *put* in the VOL VAC, *give* in the VOO ditransitive, etc.). Learning is driven by the frequency and frequency distribution of exemplars within constructions and by the match of their meaning to the construction prototype.

HERE WE EXPLORE SECOND LANGUAGE (L2) acquisition of verb–argument constructions (VACs) from a cognitive linguistic, constructionist perspective. We investigate the degree to which three linguistic constructions—verb locative (VL), verb object locative (VOL), and ditransitive (VOO)—are acquired following general cognitive principles of category learning, with abstract schematic constructions being induced from concrete exemplars.

The constructionist framework (e.g., Bates & MacWhinney, 1987; Ellis, 1998, 2003, 2006a; Goldberg, 1995, 2003, 2006; Lakoff, 1987; Langacker, 1987; Ninio, 2006; Robinson & Ellis, 2008; Tomasello, 2003) holds that learning a language involves the learning of its constructions—the units of the linguistic system, accepted as convention in the speech community and entrenched

as grammatical knowledge in the speaker's mind. Constructions specify the morphological, syntactic, and lexical form of language and the associated semantic, pragmatic, and discourse functions. Constructions form a structured inventory of a speaker's knowledge. They are useful because of the symbolic functions that they serve. It is their communicative functions that motivate their learning. Goldberg (1995) claimed that verb-centered constructions are likely to be salient in the input because they relate to certain fundamental perceptual primitives. It has been argued that basic-level categories (e.g., hammer, dog) are acquired earlier and are more frequently used than superordinate (tools, canines) or subordinate (ball-pein hammer, Weimaraner) terms because, in addition to their frequency of use, this is the level at which the world is optimally split for function, the level at which objects within the class share the same broad visual shape and motoric function, and, thus, the level at which the categories of language most directly map onto

perceptual form and motoric function (Lakoff; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Rosch, Varela, & Thompson, 1991). Goldberg extended this notion to argument structure more generally:

> Constructions which correspond to basic sentence types encode as their central senses event types that are basic to human experience … that of someone causing something, something moving, something being in a state, someone possessing something, something causing a change of state or location, something undergoing a change of state or location, and something having an effect on someone. (1995, p. 39)

From these concrete seeds, abstract constructions eventually emerge. For example, the caused motion VAC (X causes Y to move Z $_{path/loc}$ [Subj V Obj Obl$_{path/loc}$]) exists independently of particular verbs; hence, "Tom sneezed the paper napkin across the table" is intelligible despite "sneeze" being usually intransitive (Goldberg, 1995).

How might these verb-centered constructions develop these abstract properties? One suggestion is that they inherit their schematic meaning from the conspiracy of the particular types of verb that appear in their *verb island* (Tomasello, 1992). The verb is a better predictor of sentence meaning than any other word in the sentence and plays a central role in determining the syntactic structure of a sentence. There is a close relationship between the types of verbs that typically appear within constructions (in this case, *put, move, push*, etc.); hence, their meaning as a whole is inducible from the lexical items experienced within them. Ninio (1999) likewise argued that in child language acquisition, individual *pathbreaking* semantically prototypic verbs form the seeds of verb-centered argument structure patterns, with generalizations of the verb-centered instances emerging gradually as the verb-centered categories themselves are analyzed into more abstract argument structure constructions.

Categories have graded structure, with some members being better exemplars than others. In the prototype theory of concepts (Rosch & Mervis, 1975; Rosch et al., 1976), the prototype as an idealized central description is the best example of the category, appropriately summarizing the most representative attributes of a category. As the typical instance of a category, it serves as the benchmark against which surrounding, less representative instances are classified. The greater the token frequency of an exemplar, the more it contributes to defining the category, and the greater the likelihood that it will be considered the prototype. The best way to teach a concept is to show an example of it. So the best way to teach a category is to show a prototypical example. Research on category learning suggests that acquisition is optimized by the introduction of an initial, low-variance sample centered on prototypical exemplars (Cohen & Lefebvre, 2005; Elio & Anderson, 1981, 1984; Murphy, 2003; Posner & Keele, 1968, 1970). This allows learners to get a "fix" on what will account for most of the category members.

Constructionist accounts thus hold that the acquisition of grammar involves the distributional analysis of the language stream and the parallel analysis of contingent perceptual activity. Goldberg, Casenhiser, and Sethuraman (2004) tested the applicability of these general cognitive principles of category learning to the particular case of children acquiring natural language constructions by investigating whether the frequency distribution of verb exemplars in different VACs might optimize learning by providing one very high-frequency exemplar that is also prototypical in meaning. They demonstrated that in samples of child language acquisition, for a variety of constructions there is a strong tendency for one single verb to occur with very high frequency in comparison to other verbs used:

1. The VOL [Subj V Obj Obl$_{path/loc}$] construction was exemplified in children's speech by *put* 31% of the time, *get* 16% of the time, *take* 10% of the time, and *do/pick* 6% of the time, a profile mirroring that of the mothers' speech to these children (with *put* appearing 38% of the time in this construction that was otherwise exemplified by 43 different verbs).
2. The VL [Subj V Obl$_{path/loc}$] construction was used in children's speech with *go* 51% of the time, matching the mothers' 39%.
3. VOO [Subj V Obj Obj2] was filled by *give* between 53% and 29% of the time in five different children, with mothers' speech filling the verb slot in this frame by *give* 20% of the time.

Thus, although phrasal form–meaning correspondences (such as X causes Y to move Z$_{path/loc}$ [Subj V Obj Obl$_{path/loc}$]) do exist independently of particular verbs, there is a close relationship between the types of verbs that appear therein (*put, get, take, push*, etc.). Furthermore, the frequency profile of the verbs in each family follows a Zipfian profile (Zipf, 1935), whereby the highest frequency words account for the most linguistic tokens. Zipf's law states that in human language, the frequency of words decreases as a power function of their rank. If p_f is the proportion of words whose frequency in a given language sample is f, then $p_f \sim f^{-b}$, with b, the exponent of the power

law ≈ 1. Zipf (1949) showed this relation holds across a wide variety of language samples. He proposed that it arose from the Principle of Least Effort, whereby natural languages are constrained to minimize speaker effort (optimized by having fewer words to be learned and accessed in speech production) and, simultaneously, the cost of ambiguity of interpretation in speech comprehension (optimized by having many words, one for each different meaning minimizes ambiguity). Only by balancing these efforts, he suggested, can effective communication be realized. Goldberg (2006) argued that Zipfian distributions of constructions of natural language, like natural categories, optimize them for learning by providing one very high-frequency exemplar that is also prototypical in meaning. The learner can readily apprehend these meanings that are frequently experienced and clear in their interpretation, and the prototypical form is sufficient for basic communication of VAC meaning. Because the same communicative and functional concerns motivate both first (L1) language and L2 (Robinson & Ellis, 2008), we expect a similar pattern for L2 acquisition. This article tests this proposal for naturalistic L2 learners of English. Our specific hypotheses are as follows:

1. The first-learned verbs in each VAC will be those that appear more frequently in that construction in the input.
2. The first-learned pathbreaking verb for each VAC will be much more frequent than the other members, and the distribution as a whole for the types constituting each construction will be Zipfian.
3. The first-learned verbs in each VAC will be prototypical of that construction's functional interpretation.

LONGITUDINAL CORPORA ANALYSES

The English as a second language (ESL) data from the European Science Foundation (ESF) project provided a wonderful opportunity for secondary analysis in pursuit of these phenomena (Dietrich, Klein, & Noyau, 1995; Feldweg, 1991; Perdue, 1993). The ESF study, carried out in the 1980s over a period of 5 years, collected the spontaneous L2 of adult immigrants in France, Germany, Great Britain, The Netherlands, and Sweden. There were in all five target L2s (English, German, Dutch, French, and Swedish) and six L1s (Punjabi, Italian, Turkish, Arabic, Spanish, and Finnish). Data were gathered longitudinally, with the learners being recorded in interviews ev-

ery 4–6 weeks for approximately 30 months. The corpus is available in CHILDES (MacWhinney, 2000a, 2000b) chat format from the Talkbank Web site (MacWhinney, 2007).

Participants

Our analysis is based on the data for 7 ESL learners living in Britain whose native languages are Italian (Vito, Lavinia, Andrea, & Santo) or Punjabi (Ravinder, Jarnail, & Madan). Details of these participants can be found in Dietrich et al. (1995). Data were gathered and transcribed for these ESL learners and their native-speaker (NS) conversation partners from a range of activities including free conversation, interviews, vocabulary elicitation, role-play, picture description, stage directions, film watching/commenting/retelling, accompanied outings, and route descriptions. The NS language data are taken to be illustrative of the sorts of naturalistic input to which the learners were typically exposed, although we acknowledge some limitations in these extrapolations. In all, 234 sessions involving these 7 participants and their conversation partners were analyzed.

Procedure

The transcription files were downloaded from the Max Planck Institute for Psycholinguistics Web site using the IMDI BCBrowser 3.0 (ISLE Metadata Initiative, 2009). Various Computerized Language Analysis (CLAN; MacWhinney, 2000a) tools were used to separate out the participant and interviewer tiers, to remove any transcription comments or translations, to do rough tagging to identify the words that were potentially verbs in these utterances, and to do frequency analyses on these. The resultant 405 forms served as our targets for semiautomated searches through the transcriptions to find tokens of their use as verbs and to identify the verb–argument constructions of interest. The tagging was conducted by the second author following the operationalizations and criteria described in Goldberg et al. (2004) to identify utterances containing examples of VL, VOL, or VOO constructions. For example:

1. Lavinia: you come out of my house. [come] [VL]
2. Madan: Charlie say # shopkeeper give me one cigar ## he give it ## he er # he smoking #. [give] [VOO]
3. Ravinder: no put it in front # thats it # yeah. [put] [VOL]

The coded constructions so identified were checked for accuracy by a native English speaker

FIGURE 1

Type–Token Frequency Distributions of the Verbs Populating the Interviewers' VL, VOL, and VOO Constructions
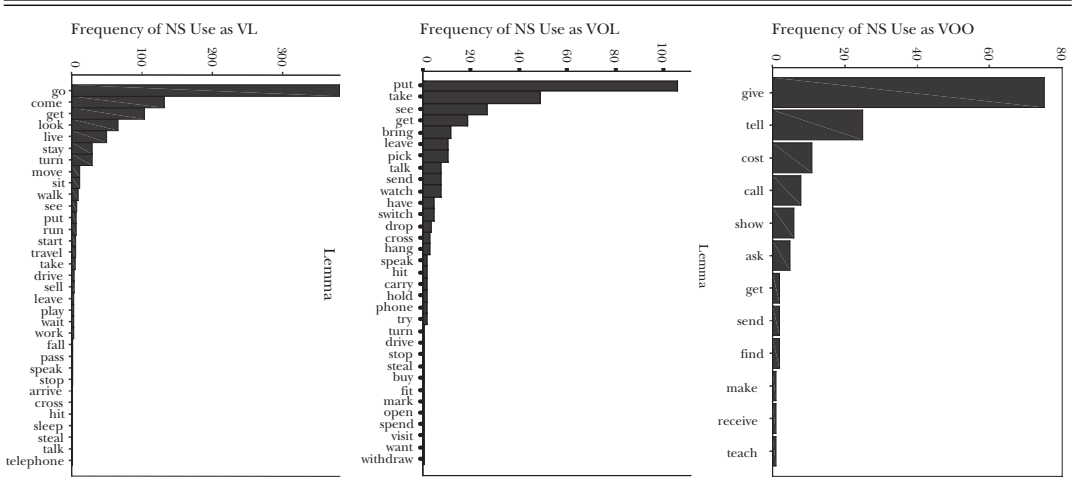
Note. NS = native speaker; VL = verb locative; VOL = verb object locative; VOO = ditransitive.

research assistant who served as an independent coder. Any disagreements were resolved through discussion. Each identified construction was also tagged for its speaker and for the number of months the speaker had been in the study at the time of utterance.

Results

For the NS conversation partners, we identified 14,574 verb tokens (232 types), of which 900 tokens were identified to occur in VL (33 types), 303 in VOL (33 types), and 139 in VOO constructions (12 types). For the non-native-speaker (NNS) ESL learners, we identified 10,448 verb tokens (234 types), of which 436 tokens were found in VL (39 types), 224 in VOL (24 types), and 36 in VOO constructions (9 types).

1. Are the Frequency Distributions Zipfian? NS interviewers. The frequency distributions of the verb types in the VL, VOL, and VOO constructions produced by the NS interviewers are shown in Figure 1. It can be seen that for each construction there is one exemplar that accounts for a substantial share of total productions of that construction. The 380 instances of *go* (/*going*/*went*) constituted 42% of the total tokens of VL; 106 instances of *put* (/*putting*) constituted 35% of the total tokens of VOL; and 75 instances of *give* (/*gave*) constituted 53% of the total tokens of VOO. After each leading exemplar, subsequent verb types decline in

frequency, confirming that Zipf's law holds: The frequency of any verb is inversely proportional to its rank in the frequency table for that construction. Figure 2 plots these frequency distributions as log verb frequency against log verb rank. The fact that these produce straight-line functions confirms that the relationship is a power function, as Zipf's law predicts.

NNS learners. The frequency distributions of the verb types in the VL, VOL, and VOO constructions produced by the NNS learners are shown in Figure 3. As with the interviewers, for each construction there is one exemplar that accounts for the majority of total productions of that construction. The 380 instances of *go* constituted 53% of the total tokens of VL, 153 instances of *put* constituted 68% of the total tokens of VOL, and 22 instances of *give* constituted 64% of the total tokens of VOO. Figure 4 shows that the type–token distributions of these learner constructions are also, like those of the interviewers, Zipfian.

A comparison of these VACs within and across NS and NNS learners also shows the generalized implications of Zipf's law for learning. The smaller the number of types in each category, the larger the degree to which the pathbreaking exemplar takes the lion's share. With the NSs there were just 12 types in VOO, with *give* constituting 53% of overall tokens; there were 33 types in VOL, with *put* constituting 35% of overall tokens; there were 33 types in VL, with *go* constituting 42% of overall

FIGURE 2

Zipfian Type–Token Frequency Distributions of the Verbs Populating the Interviewers' VL, VOL, and VOO Constructions
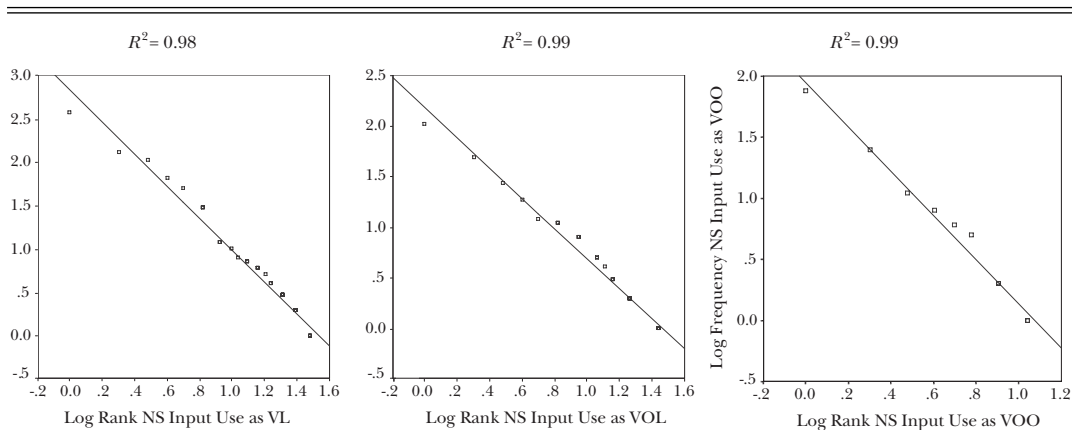
Note. NS = native speaker; VL = verb locative; VOL = verb object locative; VOO = ditransitive.

FIGURE 3

Type–Token Frequency Distributions of the Verbs Populating the Learners' VL, VOL, and VOO Constructions
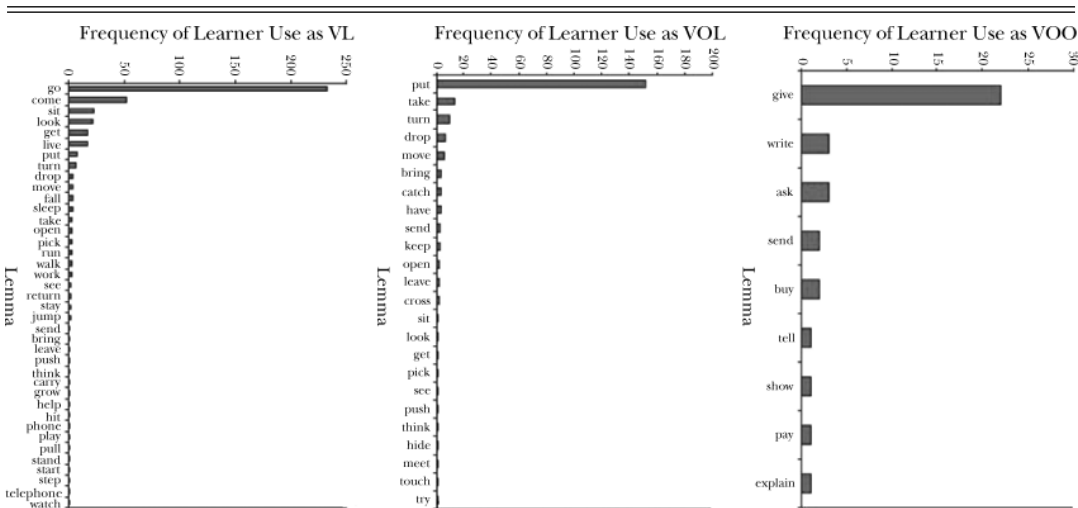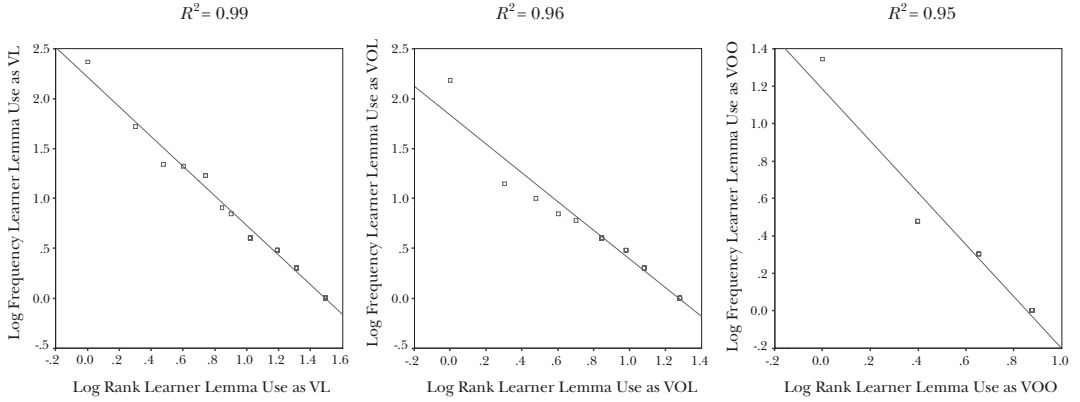

Note. VL = verb locative; VOL = verb object locative; VOO = ditransitive.

tokens. With the NNSs, who had fewer types per construction, the lead exemplars took larger majority shares (VL 52%, VOL 68%, VOO 64%). The limit is clearly at the very beginning of category learning, where one exemplar constitutes 100% of the category.

2. Does Learner Use Match the Relative Input Frequencies? Inspection of Figures 1 and 3 demonstrates that the rank order of verb types in the learner constructions is broadly similar to that in the interviewer NS data. Correlational analyses across all 80 verb types that are featured in any of the NS and/or NNS constructions confirm this to be so. For the VL construction, the frequency of lemma use by learner is correlated with the frequency of lemma use by the NS interviewer: $r(78) = 0.97$, $p < .001$. The same analysis for VOL results in $r(78) = 0.89$, $p < .001$. The same analysis for VOO results in $r(78) = 0.93$, $p < .001$.

FIGURE 4

Zipfian Type–Token Frequency Distributions of the Verbs Populating the Learners' VL, VOL, and VOO Constructions

Note. VL = verb locative; VOL = verb object locative; VOO = ditransitive.

3. Does Learner Use Follow the Input Frequencies? Does this matching of frequency reflect learner acquisition orders? Figures 5, 6, and 7 show cumulative learner use of the different verb types as a function of the number of months in the study at time of production respectively for the VL, VOL, and VOO constructions. It can be seen that the ESF study spanned the early stages of acquisition, from novice to intermediate proficiency, rather well. Relatively few verb types are used in these productions in the early months, particularly for the more difficult VOO construction. However, as time progresses, these constructions do show development in these learners. In each case, however, it is clear that there is one pathbreaking verb that seeds the construction and leads its development (*go* seeds VL, *put* seeds VOL, and *give* seeds VOO). Thereafter, with increasing experience of input, other verbs are recruited to the construction.

Analyses of variance were used to assess the degree to which particular verbs were associated with particular constructions across learners and across time (with each participant's involvement binned into four quarters).

Verb type use in VL was a significant factor by subjects: $F(39, 234) = 9.29$, $p < .001$, with *go* being used significantly more frequently than all other verbs on post hoc Tukey testing at $p < .01$ but all other contrasts insignificant, and by time $F(39, 119) = 25.58$, $p < .001$, with *go* being used significantly more frequently than all other verbs on post hoc Tukey testing at $p < .01$, *come* being used significantly more frequently than all other verbs

except *get* and *live*, and all other contrasts being insignificant.

Verb type use in VOL was a significant factor by subjects: $F(23, 115) = 3.70$, $p < .001$, with *put* being used significantly more frequently than all other verbs on post hoc Tukey testing at $p < .01$ and all other contrasts insignificant, and by time $F(23, 71) = 5.83$, $p < .001$, with *put* being used significantly more frequently than all other verbs on post hoc Tukey testing at $p < .01$ and all other contrasts insignificant.

Verb type use in VOO was a significant factor by subjects: $F(7, 35) = 4.88$, $p < .001$, with *give* being used significantly more frequently than all other verbs on post hoc Tukey testing at $p < .01$ and all other contrasts insignificant, and by time $F(7, 23) = 6.00$, $p < .001$, with *give* being used significantly more frequently than all other verbs on post hoc Tukey testing at $p < .01$ and all other contrasts insignificant.

We plotted these acquisition functions for each learner and each construction. There is insufficient space to show them here, although we are happy to send these graphs on request. It is clear from these individual learning curves that the developmental sequences summarized in Figures 5–7 are true of each of the learners. In every case the relevant pathbreaking verb seeds the construction, albeit with much earlier development in some learners than others.

4. Does the Match of Native Speaker and Learner Verb Use Reflect Conversational Priming? Language use exhibits recency effects in the "dance of

FIGURE 5
Learner Use of the Top 10 Most Frequent Verb Types in the VL Construction as a Function of Study Month

Cumulative Acquisition Curve VL
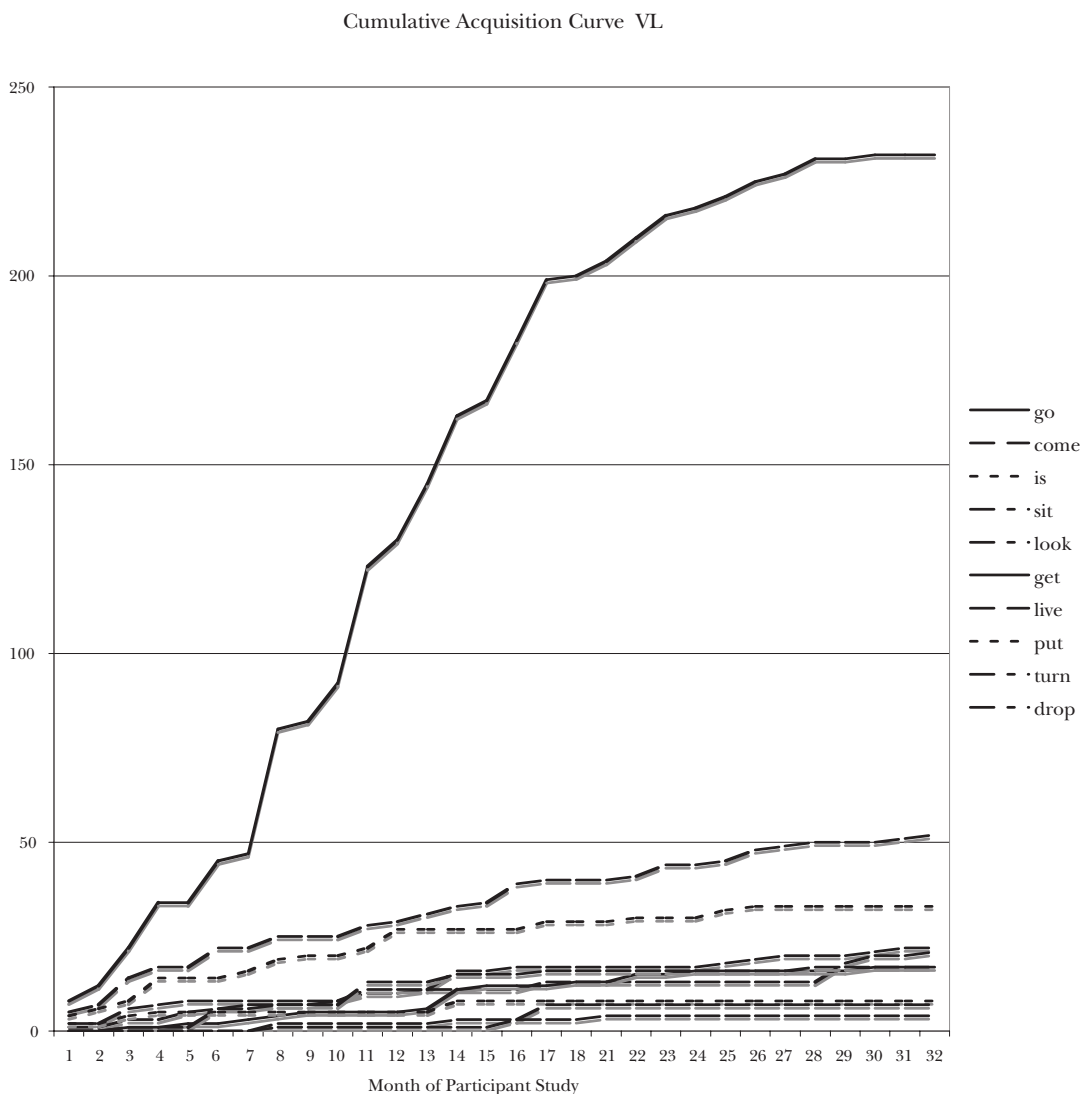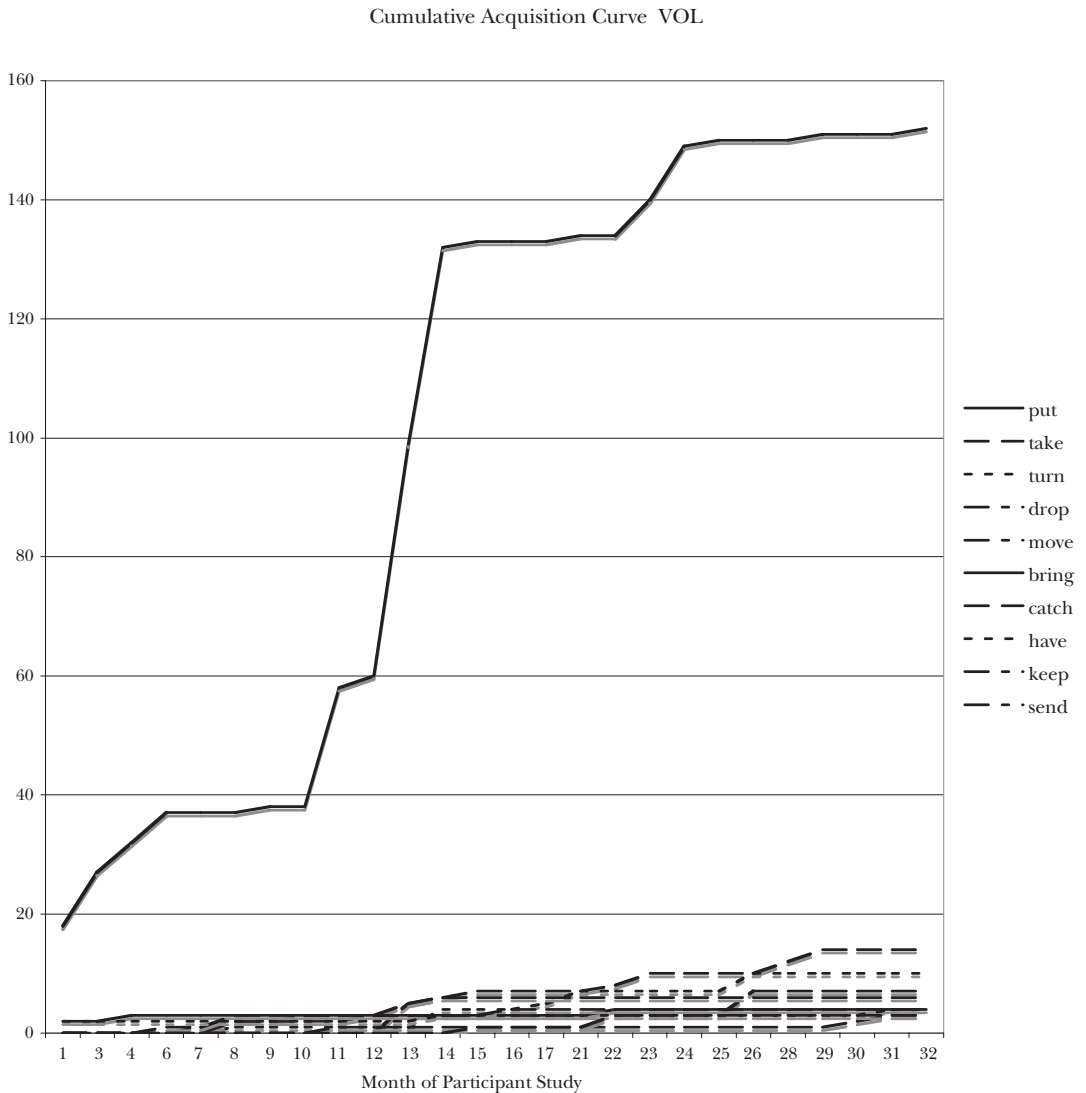
Note. VL = verb locative.

dialogue" whereby the constructions used by one speaker affect the use and availability of the same constructions in their conversation partner (e.g., Pickering, 2006; Pickering & Garrod, 2006). This phenomenon, known as *priming*, can be observed across phonology, conceptual representations, lexical choice, and syntax. We must ask, then, to what extent the matching profiles of constructional verb use in NSs and NNSs result from priming. To investigate this, we used the CHIP routine in CLAN (MacWhinney, 2000a) to look for this phenomenon in the highest frequency verbs in each VAC: To what extent do the NNS

uses of *go*, *put*, and *give* follow immediately from NS uses?

Of 233 NNS uses of *go* in the VL construction, we found 17 that seemed to result from priming from the NS interview interaction. Examples include the following:

1. NS: did your [<1] wife go to the dentist ?
    NNS: uh yeah # she go [>1] to the dentist #
         for one ## previsit.

2. NNS: er # i dont know # to explain.
    NS: he is going.

FIGURE 6

Learner Use of the Top 10 Most Frequent Verb Types in the VOL Construction as a Function of Study Month

Cumulative Acquisition Curve VOL

Note. VOL = verb object locative.

NNS: he is going er.
NNS: down ?

3. NS: do you know where hes going ?

   NNS: er ## hes going er er at the [/?] on the window cooking [% means kitchen] [/?] cooking window.

   Of 152 NNS uses of *put* in VOL, 8 appeared to be directly primed in this way. Examples include the following:

4. NS: not put inside ?

   NS: < > [% rolls the magazine to put it inside the cup].
   NNS: not put inside # open.

5. NS: what does the shopkeeper do with the money ?

   NNS: ##.
   NS: does he put it in his pocket ?
   NNS: no put it in till or.

   Of the 22 NNS uses of give in VOO, 1 appeared to be primed in this way:

FIGURE 7

Learner Use of the Top 10 Most Frequent Verb Types in the VOO Construction as a Function of Study Month

Cumulative Acquisition Curve VOO
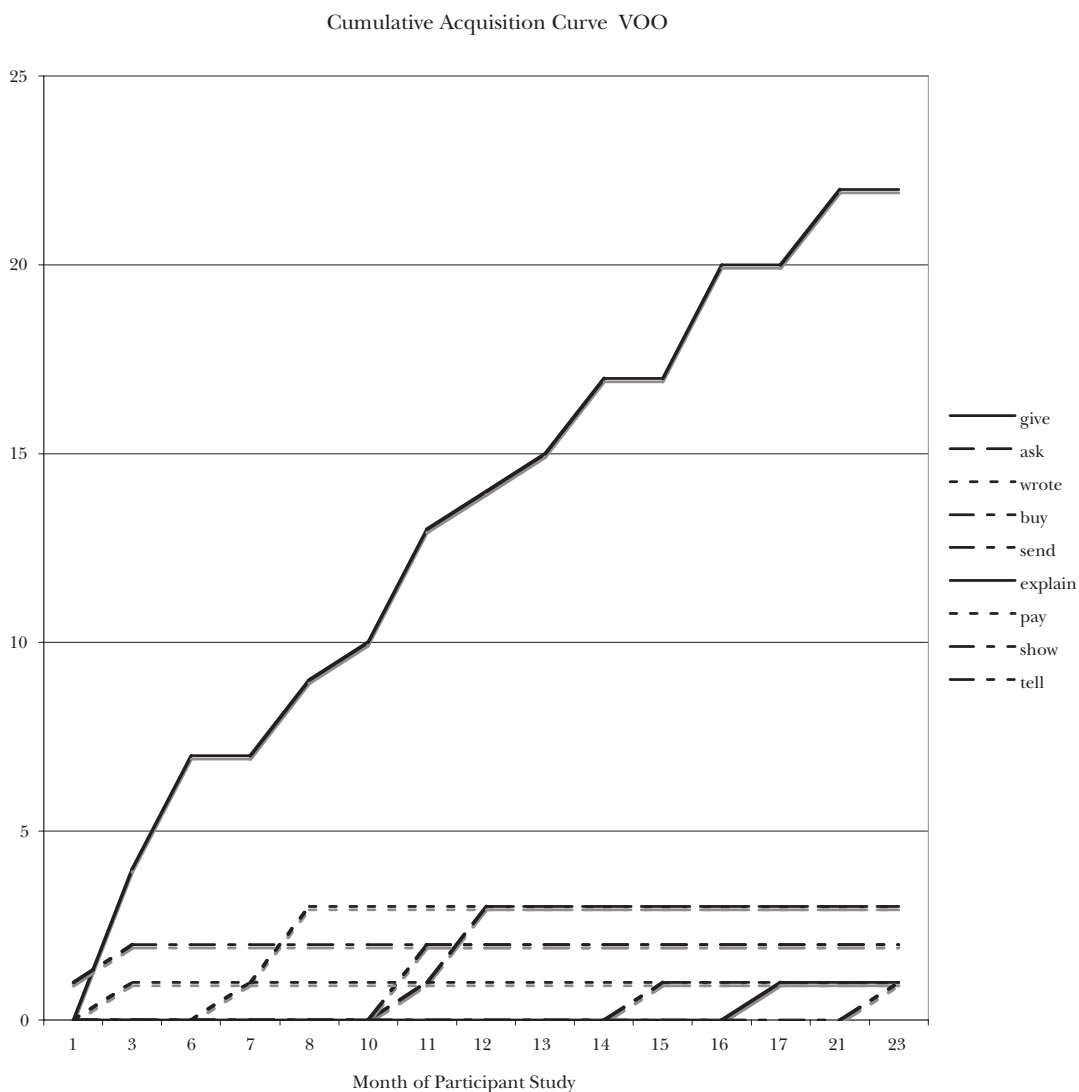
Note. VOO = ditransitive.

6.   NNS: she giving you know.
     NS: ## <she> [>1] gives him.
     NNS: she give me some money.
     NS: yeah.

   Interactional priming thus appears to account for some of the NNS use observed here, but by no means the majority.

VERB PROTOTYPICALITY RATINGS

   These analyses of the spoken ESF corpora demonstrate that the structure of VACs in NSs and NNSs alike is Zipfian. In each case there is one exemplar type that takes the lion's share of the tokens and seeds the category in development. However, what is the nature of this pathbreaking verb? Is it indeed close to the VAC prototype in meaning? To test this we had NSs of English rate the semantics of these verbs.

Participants

   Participants were 5 native English-speaking volunteers who responded to opportunistic requests

from the first author—his wife, his 13-year-old son, 1 graduate student, and 2 undergraduate students.

Procedure

The participants were presented with an Excel spreadsheet that contained as row headers in alphabetical order the 80 verb types that were used in the three constructions by the NSs and/or NNSs. Participants were instructed to rate the verbs for the degree to which they followed the core meaning of the VACs as expressed in the instructions in the column headers illustrated in Table 1. The instructions incorporated a visual image of the core meaning and asked participants to consider the verb's meaning while saying the verb in a sentence. Raters worked down and then across, rating every verb on one construction before moving to the next.

Albeit a rather diverse sample, these native English speakers were highly reliable in their rating of fit of the meaning of these verbs in the VACs. The interrater reliabilities as assessed using Cronbach's alpha were .86 for VL, .89 for VOL, and .91 for VOO.

Results

The mean (standard deviation) rating of the 80 verbs with regard to the core meaning of the VACs were respectively VL 4.03 (2.39), VOL 3.84 (2.48), and VOO 3.63 (2.49). To see how the degree to which the verb matched the prototype VAC meaning affected the degree to which each verb was used in the VAC by learners, we plotted verb prototypicality rating against log frequency of learner use. These scatterplots are shown in Figure 8.

We can see that in each case there is a significant positive association between prototypicality of verb meaning and log frequency of learner use: VL $\rho(78) = 0.44$, $p < .001$; VOL $\rho(78) = 0.29$, $p < .01$; VOO $\rho(78) = 0.34$, $p < .001$. These correlations are by no means perfect, however. Why is the most prototypical verb in a particular construction not the most frequently used? The answer lies in the scatterplots. Let us work from left to right.

Consider first the VL construction. The most used verb, *go*, was rated as 7.4 out of 9.0 in terms of the degree to which it described the movement of something or someone to a new place or in a new direction. It is very high. However, 10 other verbs surpassed it in this rating: *walk* (9.0), *move* (8.8), *run* (8.8), *travel* (8.8), *come* (8.4), *drive* (8.2), *arrive* (8.0), *jump* (8.0), *return* (8.0), and *fall* (7.8).

If prototypicality of meaning is the driving force, why did learners use these verbs to a lesser degree? We believe it is because they are less generic. *Walk*, *run*, and *jump* fit the change of location schema, but their specific requirements of manner of motion limit their general use. In contrast, *go* works for just about any change of place or direction.
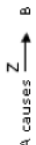
Next, we will look at the VOL construction. The most used verb, *put*, was rated 8.0 in terms of how well it described how someone causes the movement of something to a new place or in a new direction. However, it was surpassed in these rankings by *bring* (8.6), *move* (8.6), *send* (8.6), *take* (8.6), *carry* (8.4), *drive* (8.4), *drop* (8.4), *pass* (8.4), *push* (8.4), *hit* (8.2), and *pull* (8.2). Learners do tend to use these words, but with nothing like the frequency of *put*. Again, we believe the answer lies in generalizability of use. Verbs like *drive*, *drop*, and *send* are quite specific in their description of manner of transport, whereas *put* can be used to describe a much wider range of VOL circumstances.

Finally, let us consider the VOO construction. The most used verb, *give*, was rated 9.0 in terms of how well it fit the meaning of someone causing someone to receive something. It is indeed prototypical of this meaning. However, it tied in this rating with lesser used verbs like *send* (9.0) and was closely matched by other lesser used verbs like *buy* (8.0), *tell* (7.8), and *show* (7.0). Other verbs high in the VOO rating, like *pass* (8.6), *bring* (8.4), *sell* (8.4), and *take* (8.4), were not used by the learners at all. Again, g*ive* is prototypical in VOO meaning and, in contrast to other VOO verbs of specific manner like *tell*, *buy*, and *sell*, it is also much wider in its range of general application.

Discussion

These data confirm the association between frequency of learner verb use and prototypicality of VAC meaning, but they add qualification by showing that the relationship is moderated by the generality of verb semantics. The verbs that learners use first are prototypical *and* generic in function. The same has been shown for child language acquisition, where a small group of semantically general verbs, often referred to as *light verbs* (e.g., *go, do, make, come*) are learned early (Clark, 1978; Ninio, 1999; Pinker, 1989). Ninio argued that because most of their semantics consist of some schematic notion of transitivity with the addition of a minimum specific element, they are semantically suitable, salient, and frequent; hence, learners start transitive word combinations with these generic verbs. Thereafter, as Clark describes,

TABLE 1
Instructions in the VL, VOL, and VOO Conditions of the Prototypicality Rating Task

VL	VOL	VOO
In English some verbs are used to describe the movement of something or someone (X) to a new place (Y) or in a new direction.	In English some verbs are used to describe how someone (A) causes the movement of something (X) to a new place (Y) or in a new direction.	In English some verbs are used to describe how someone (A) causes someone (B) to receive something (Z).
Y X e.g., The fly buzzed into the room.	A causes X X Y e.g., Pat blew the feather into the drawer.	A causes $\xrightarrow{Z}$ B e.g., Pat faxed Tom the picture.
Rate the verbs in this column from 1 - 9 according to the degree to which they work like this, with 1 being not at all and 9 being absolutely!	Rate the verbs in this column from 1 - 9 according to the degree to which they work like this, with 1 being not at all and 9 being absolutely!	Rate the verbs in this column from 1 - 9 according to the degree to which they work like this, with 1 being not at all and 9 being absolutely!
It helps to say the verb as you do this. So if you were rating the word 'splog', you'd say "X splogged to/at/on Y" and judge whether it meant movement of an object to a new place like this.	It helps to say the verb as you do this. So if you were rating the word 'splog', you'd say "A splogged the X to/at/on Y" and judge whether it meant cause of movement of an object to a new place like this.	It helps to say the verb as you do this. So if you were rating the word 'splog', you'd say "A splogged B the Z" and judge whether it meant cause the transfer of something to a recipient like this.
Click in the appropriate cell, type a number from 1-9, then move down with the arrow key until you have rated all the verbs in this column.	Click in the appropriate cell, type a number from 1-9, then move down with the arrow key until you have rated all the verbs in this column.	Click in the appropriate cell, type a number from 1-9, then move down with the arrow key until you have rated all the verbs in this column.
Don't think too hard about it - make a rating and move right along.	Don't think too hard about it - make a rating and move right along.	Don't think too hard about it - make a rating and move right along.

Note. VL = verb locative; VOL = verb object locative; VOO = ditransitive.

FIGURE 8

Learner Use of Verb Types in the VL, VOL, and VOO Constructions as They Relate to the Fit of Those Verbs to the Prototypical Meaning of the Constructions as Judged by Native English Speakers
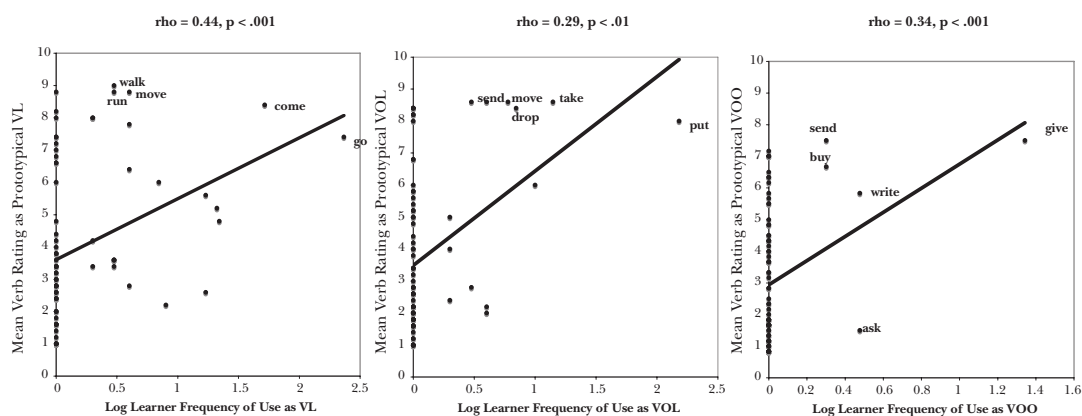
Note. VL = verb locative; VOL = verb object locative; VOO = ditransitive.

"many uses of these verbs are replaced, as children get older, by more specific terms. . . . General purpose verbs, of course, continue to be used but become proportionally less frequent as children acquire more words for specific categories of actions" (1978, p. 53).

Theakston, Lieven, Pine, and Rowland (2004) listed the range of light verbs defined according to the criteria as applied by Clark (1978), Pinker (1989), and Ninio (1999) (semantic generality, frequency, and tendency to grammaticalize cross-linguistically): *bring, come, do, get, give, go, make, put,* and *take.* Inspection of the NNS learner data in Figures 3 and 8 shows that *go, come,* and *get* are in the top five most frequent of the NNSs' VL verbs and *put, take,* and *bring* are in their top six most frequent VOL verbs. It seems clear, therefore, that for both naturalistic L1 acquisition and SLA, the first-used verbs in these VACs are frequent and both semantically prototypical and generic.

GENERAL DISCUSSION AND CONCLUSIONS

In sum, these data demonstrate the following phenomena:

1. The highest frequency exemplar within each VAC in the L2 input is much more frequent than the other members.
2. Zipf's law applies in the L2 input within individual VACs.
3. The highest frequency exemplar within each VAC is quite prototypical in meaning and semantically generic.

4. Learner use of the VAC is seeded by this highest frequency, rather prototypical, and generic exemplar across learners and VACs.

Construction type/token frequency distribution in natural language might thus optimize learning by providing one very high-frequency exemplar that is also prototypical in meaning and widely applicable. Although the current research is specific to VACs and there is a pressing need for investigation of the generalizability of these conclusions for a wide range of other construction types, there is some research that already points in this direction. We will give two examples for each of the broad claims.

Zipfian distributions of types have been confirmed for constructions as diverse as tense–aspect (TA) and compliments. For the case of TA, Wulff, Ellis, Römer, Bardovi-Harlig, and LeBlanc (2009) show that there are frequency biases in the available input whereby certain verbs are much more frequent in different TA categories and that the distribution of the verb types in the different TA categories is Zipfian. The same is also true for compliments. Manes and Wolfson (1989) examined a corpus of 700 examples of compliments uttered in day-to-day interactions. Just three constructions accounted for 85% of these: [NP <*is/looks*> (*really*) ADJ] (53%), [*I* (*really*) <*like/love*> NP] (16%), and [PRO *is* (*really*) (*a*) ADJ NP] (15%). Eighty percent of these depended on an adjective to carry the positive semantic load. Although the number of positive adjectives that could be used is virtually unlimited,

in fact two thirds of all adjectival compliments in the corpus used only five adjectives: *nice* (23%), *good* (20%), *pretty* (9%), *beautiful* (9%), and *great* (6%). Nonadjectival compliments were focused on a handful of semantically positive verbs, with *like* and *love* accounting for 86%. All of these adjectives and verbs are "light" and semantically generic.

For the case of acquisition sequences for TA, Wulff et al. (2009) show that the verbs first learned by adults in the progressive are also frequent in the progressive in the input, distinctively associated with the progressive in the input, and highly *atelic* (i.e., significantly less telic than verbs that are frequent and associated with past tense in the input). Likewise, the verbs first learned in past tense are frequent in past tense in the input, highly distinctive for past tense in the input, and highly telic.

The pattern whereby learners overuse the prototype exemplar and initially fail to show the range and specificity characteristic of fluent NSs has also been shown for collocational restrictions. Collocations are patterns of preferred co-occurrence of particular words, they pervade language use (Sinclair, 1991), and their learning poses great difficulties for SLA (Pawley & Syder, 1983). Kennedy (2003) analyzed amplifier patterns—the particular ways in which adverbs of degree modify adjectives and verbs—in the British National Corpus. His research clearly demonstrated that adjectives are very restrictive in their selection of particular boosters and maximizers, as shown in the following examples:

1. absolutely diabolical   *entirely diabolical
2. fully fledged   *absolutely fledged
3. entirely blameless   *fully blameless
4. badly mauled   *particularly mauled
5. deeply engrained   *badly engrained
6. particularly apposite   *deeply apposite

When Granger (2001) analyzed corpora of nonnative writing for their use of amplification patterns, she identified a very restricted range of "general purpose" types led by "a highly significant overuse of *very* as the all-round amplifier *par excellence*" (p. 151, emphasis in the original). As with the VACs, the pathbreaking type in the amplifier construction here is highly generic.

The high correlation of frequency and semantic generality has prompted disagreement in child language research over whether it is their semantic generality that drives acquisition (Clark, 1978; Ninio, 1999), perhaps tapping a small set of innately specified semantic elements (Pinker, 1989), or frequency (Theakston et al., 2004), or both

(Goldberg et al., 2004). The functions of language in human communication have resulted in the evolution through usage of a communicative system wherein factors such as frequency, generality, prototypicality, and distinctiveness interact and are positively associated. Therefore, distinguishing any one factor as the root cause of category acquisition is problematic and probably naïve. The data-driven and quantitative perspective adopted here suggests instead that, as in the acquisition of other categories, it is the conspiracy of these several different factors working together that drives acquisition of linguistic constructions.

There is good evidence, too, that these factors first play out in learning to comprehend the L2. The analyses of NNSs here are done irrespective of total accuracy of form in production. Although learner productions of the simpler VL construction are usually correct, the structurally more complex VOL and VOO constructions are often produced in a simplified form (i.e., the Basic Variety so clearly identified and analyzed in the original ESF project; Klein & Purdue, 1992; Perdue, 1993). This typically involves a pragmatic topic–comment word ordering, where old information goes first and new information follows. Examples for the VOL include the following:

1. yeah this television put it up the # book #
2. this bag <he put him> [/?] put in the st [/?] er floor # <bag> [>1]
3. a horse # put in there < > [$ laughs]
4. you know which block put down
5. yeah keep it money ## put the table [/?] # put in the table

Comprehending which verbs go with which arguments in which VACs is the start of the process. Learning to produce these arguments in their correct order is a slower process, one which, in these data, seems to start with highly generic formulaic phrases such as "put it there."

Readers of prior drafts of this article asked whether NNS verb use was a simple result of priming from prior NS use, hence the inclusion of our analyses in the Results section, which showed that interactional priming accounted for some of the NNS use observed here, but by no means the majority. To compartmentalize learner language as a result of the cognitive phenomenon of priming, or the discourse phenomenon of shared topic, seems naïve. SLA researchers who take an interactionist view of learning (Ellis, 2008; Gass, 1997, 2003; Gass & Mackey, 2007; Gass & Varonis, 1994; Long, 1980; Mackey & Gass, 2006) look to these interactions, where scaffolding, negotiation, and priming of form occur while conversation

partners focus on shared meanings, as the very process of acquisition.

One other suggested explanation of these findings is that they are statistical artifacts because high-frequency items will be sampled earlier than low-frequency ones; thus, the fact that high-frequency items are found in these samples for NSs and NNSs alike is uninformative with regard to NNSs' knowledge of these forms (Tomasello & Stahl, 2004). Again, there is no denying these sampling phenomena as a potential force behind the data observed by the ESF researchers and analyzed here. Yet at the same time, we are reluctant to relegate these observations as artifactual. This is because these same sampling phenomena are equally potent in the data observed and analyzed by NNSs whose experience also is more likely to sample high-frequency items that are prototypical and generic in meaning.

The present analyses of L2 VACs are just part of the story. There is more relating (a) to the distinctiveness of the verb as a cue for a particular VAC—some verbs are associated with several different VACs and so are less reliable markers of any one VAC than are verbs which inhabit just one—and (b) to the types that occupy the other islands in a construction (and their type/token frequency distributions). Whereas the present analysis concentrates solely on verb islands, there are other important cues to be had from the occupants of the other islands that come together in each construction archipelago. Our companion work (Ellis & Ferreira-Junior, in press) uses various measures of contingency from collostructional analysis (Gries & Stefanowitsch, 2004; Stefanowitsch & Gries, 2003) and from the human associative learning literature (Allan, 1980; Ellis, 2006b; Shanks, 1995) to investigate the degree to which the reliability of verb-construction mapping affects acquisition. It also analyzes the differential contributions of the types that inhabit the other thematic roles of the subcategorization frames of the VACs studied here.

Finally, it is clear that the acquisition of linguistic constructions is affected by a wide range of factors. For each of the slots in the construction frame, there is the following:

1. the frequency, the frequency distribution, and the salience of the form types;

2. the frequency, the frequency distribution, the prototypicality and generality of the semantic types, their importance in interpreting the overall construction;

3. the reliabilities of the mapping between 1 and 2; and

4. the degree to which the different elements in the VAC sequence (such as Subj V Obj Obl) are mutually informative and form predictable chunks.

Many of these factors are positively correlated. It is very difficult, therefore, to investigate their independent contributions or their conspiracy without formal modeling. Computer simulations allow investigation of the contributions of these factors to language learning, processing, and use and of the ways that language as a complex adaptive system has evolved to be learnable (Christiansen & Chater, 2008; Ellis & Larsen-Freeman, in press-b). Ellis and Larsen-Freeman (in press-a) present various connectionist simulations of the acquisition of these VACs.

Nevertheless, there are many limitations to the present research. The sample of learners is small; the sample of VACs is small; the sample of language is small; the degree to which the language produced by the NS interviewers as a proxy of the typical input received by learners is questionable. Analyses of these factors in other learners, other VACs, and other languages are clearly an important next step. The ESF project itself provides a variety of L2s that might be exploited here. However, the samples are relatively small, and we are left, as ever, echoing the need for dense longitudinal corpora of SLA (Ortega & Iberri-Shea, 2005) to match those of the L1 (Tomasello, 2003; Tomasello & Stahl, 2004). Denser corpora would allow proper longitudinal analyses, dividing verb use by NSs and NNSs over several stages of acquisition to look for stage by verb use by NS/NNS status interactions—the profile of verb uses by NSs should remain stable over time, that by NNSs should show any pathbreaker appearing earlier than the rest. In the present analyses, the small number of verbs and their relevance to different types of interview and/or elicitation task made such analyses impossible. There is also reason to believe that not all constructions follow a Zipfian type/token distribution—the transitive verb–object and subject–verb–object constructions that are acquired early by all learners have no semantic prototype (Goldberg, 1999; Ninio, 2006), and it will be important in future research to investigate a wide range of constructions in native language corpora and, separately, in acquisition to assay the generality of Zipfian distributions within constructions. So, too, this work is all correlational. However ecologically valid, it needs supplementing with experimental investigations that compare the effects of skewed versus even frequencies of type/token distributions, as well as

prototypicality versus nonprototypicality of lead exemplar, upon early acquisition in training studies of the sorts reported by Goldberg et al. (2004) and Year and Gordon (this issue).

Meanwhile, the current research has demonstrated that VAC type/token distribution in the input is Zipfian and that learners acquire the most frequent, prototypical, and generic exemplars. Learning is driven by the frequency and frequency distribution of exemplars within constructions and by the match of their meaning to the construction prototype.

REFERENCES

Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society*, *15*, 147–149.

Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 157–193). Hillsdale, NJ: Erlbaum.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral & Brain Sciences*, *31*, 489–509.

Clark, E. V. (1978). Discovering what words can do. In D. Farkas, W. M. Jacobsen, & K. W. Todrys (Eds.), *Papers from the parasession on the lexicon, Chicago Linguistics Society April 14–15, 1978* (pp. 34–57). Chicago: Chicago Linguistics Society.

Cohen, H., & Lefebvre, C. (Eds.). (2005). *Handbook of categorization in cognitive science*. Mahwah, NJ: Elsevier.

Dietrich, R., Klein, W., & Noyau, C. (Eds.). (1995). *The acquisition of temporality in a second language*. Amsterdam: Benjamins.

Elio, R., & Anderson, J. R. (1981). The effects of category generalizations and instance similarity on schema abstraction. *Journal of Experimental Psychology: Human Learning & Memory*, *7*(6), 397–417.

Elio, R., & Anderson, J. R. (1984). The effects of information order and learning mode on schema abstraction. *Memory & Cognition*, *12*, 20–30.

Ellis, N. C. (1998). Emergentism, connectionism and language learning. *Language Learning*, *48*, 631–664.

Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. Doughty & M. H. Long (Eds.),

Handbook of second language acquisition. Oxford, UK: Blackwell.

Ellis, N. C. (2006a). Cognitive perspectives on SLA: The associative cognitive CREED. *AILA Review*, *19*, 100–121.

Ellis, N. C. (2006b). Language acquisition as rational contingency learning. *Applied Linguistics*, *27*, 1–24.

Ellis, N. C. (2008). The psycholinguistics of the interaction hypothesis. In A. Mackey & C. Polio (Eds.), *Multiple perspectives on interaction in SLA: Second language research in honor of Susan M. Gass* (pp. 11–40). New York: Routledge.

Ellis, N. C., & Ferreira-Junior, F. (in press). Slots, frames, and the distinctiveness of their occupants: Island populations in the acquisition of verb argument constructions. *Annual Review of Cognitive Linguistics*.

Ellis, N. C., & Larsen-Freeman, D. (in press-a). Emergent models of the acquisition of verb argument constructions. *Language Learning*, *59*(Suppl. 1).

Ellis, N. C., & Larsen-Freeman, D. (in press-b). *Language as a Complex Adaptive System* [Special Issue]. *Language Learning*, *59*(Suppl. 1).

Feldweg, H. (1991). *The European Science Foundation Second Language Database*. Nijmegen, The Netherlands: Max Planck Institute for Psycholinguistics.

Gass, S. (1997). *Input, interaction, and the development of second languages*. Mahwah, NJ: Erlbaum.

Gass, S. (2003). Input and interaction. In C. Doughty & M. Long (Eds.), *Handbook of second language acquisition* (pp. 224–255). Oxford, UK: Blackwell.

Gass, S., & Mackey, A. (2007). Input, interaction and output: An overview. *AILA Review*, *19*, 3–17.

Gass, S., & Varonis, E. (1994). Input, interaction and second language production. *Studies in Second Language Acquisition*, *16*, 283–302.

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

Goldberg, A. E. (1999). The emergence of argument structure semantics. In B. MacWhinney (Ed.), *The emergence of language* (pp. 197–212). Hillsdale, NJ: 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: Oxford University Press.

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

Granger, S. (2001). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications*. Oxford: Oxford University Press.

Gries, S. T., & Stefanowitsch, A. (2004). Extending collostructional analysis: A corpus-based perspective on "alternations." *International Journal of Corpus Linguistics*, *9*, 97–129.

ISLE Metadata Initiative. (2009). IMDI Browser 3.0.Retrieved from Max Planck Institute Web site: http://corpus1.mpi.nl/ds/imdi_browser/

Kennedy, G. (2003). Amplifier collocations in the British National Corpus: Implications for English language teaching. *TESOL Quarterly*, *37*, 477–486.

Klein, W., & Purdue, C. (1992). *Utterance structure: Developing grammars again*. Amsterdam: Benjamins.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

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

Long, M. H. (1980). *Input, interaction, and second language acquiistion*. Unpublished doctoral dissertation, University of California, Los Angeles.

Mackey, A., & Gass, S. (2006). Pushing the methodological boundaries in interaction research: An introduction to the special issue. *Studies in Second Language Acquisition*, *28*, 169–178.

MacWhinney, B. (2000a). *The CHILDES project: Tools for analyzing talk: Vol. 1. Transcription format and programs* (3rd ed.). Mahwah, NJ: Erlbaum.

MacWhinney, B. (2000b). *The CHILDES Project: Tools for analyzing talk: Vol. 2. The database* (3rd ed.). Mahwah, NJ: Erlbaum.

MacWhinney, B. (2007). The TalkBank Project. In J. C. Beal, K. P. Corrigan, & H. L. Moisl (Eds.), *Creating and digitizing language corpora: Synchronic databases* (Vol. 1, pp. 163–180). Houndmills, UK: Palgrave-Macmillan.

Manes, J., & Wolfson, N. (1989). The compliment formula. In F. Coulmas (Ed.), *Conversational routine*. The Hague, The Netherlands: Mouton.

Murphy, G. L. (2003). *The big book of concepts*. Boston: MIT 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: Oxford University Press.

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. 191–225). London: Longman.

Perdue, C. (Ed.). (1993). *Adult language acquisition: Crosslinguistic perspectives*. Cambridge: Cambridge University Press.

Pickering, M. J. (2006). The dance of dialogue. *The Psychologist*, *19*, 734–737.

Pickering, M. J., & Garrod, S. C. (2006). Alignment as the basis for successful communication. *Research on Language and Computation*, *4*, 203–228.

Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: Bradford Books.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*, 353–363.

Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology*, *83*, 304–308.

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

Rosch, E., & Mervis, C. B. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*, 192–233.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.

Rosch, E., Varela, F., & Thompson, E. (1991). *The embodied mind*. Boston: MIT Press.

Shanks, D. R. (1995). *The psychology of associative learning*. New York: Cambridge University Press.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Stefanowitsch, A., & Gries, S. T. (2003). Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics*, *8*, 209–243.

Talkbank. (2007). European Science Foundation corpus. Retrieved from Talkbank Web site: http://talkbank.org/data/BilingBank/

Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2004). Semantic generality, input frequency and the acquisition of syntax. *Journal of Child Language*, *31*, 61–99.

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

Tomasello, M. (2003). *Constructing a language*. Boston: Harvard University Press.

Tomasello, M., & Stahl, D. (2004). Sampling children's spontaneous speech: How much is enough? *Journal of Child Language*, *31*, 101–121.

Wulff, S., Ellis, N. C., Römer, U., Bardovi-Harlig, K., & LeBlanc, C. (2009). The acquisition of tense–aspect: Converging evidence from corpora and telicity ratings. *Modern Language Journal*, *93,* 354–369.

Year, J.-E., & Gordon, P. (2009). Korean speakers' acquisition of the English ditransitive construction: The role of verb prototype, input distribution, and frequency. *Modern Language Journal*, *93,* 399–417.

Zipf, G. K. (1935). *The psycho-biology of language: An introduction to dynamic philology*. Cambridge, MA: MIT Press.

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley.