

John Benjamins Publishing Company



This is a contribution from *Annual Review of Cognitive Linguistics 7*
© 2009. John Benjamins Publishing Company

This electronic file may not be altered in any way.

The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.

Permission is granted by the publishers to post this file on a closed server which is accessible to members (students and staff) only of the author's/s' institute, it is not permitted to post this PDF on the open internet.

For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: www.copyright.com).

Please contact rights@benjamins.nl or consult our website: www.benjamins.com

Tables of Contents, abstracts and guidelines are available at www.benjamins.com

Constructions and their acquisition

Islands and the distinctiveness of their occupancy

Nick C. Ellis and Fernando Ferreira-Junior

University of Michigan / Federal University of Minas Gerais, Brazil

This paper presents a psycholinguistic analysis of constructions and their acquisition. It investigates effects upon naturalistic second language acquisition of type/token distributions in the islands comprising the linguistic form of English verb-argument constructions (VACs: VL verb locative, VOL verb object locative, VOO ditransitive) in the ESF corpus (Perdue, 1993).

Goldberg (2006) argued that Zipfian type/token frequency distribution of verbs in natural language might optimize construction learning by providing one very high frequency exemplar that is also prototypical in meaning. Ellis & Ferreira-Junior (2009) confirmed that in the naturalistic L2A of English, VAC verb type/token distribution in the input is Zipfian and learners first acquire the most frequent, prototypical and generic exemplar (e.g. *put* in VOL, *give* in VOO, etc.). This paper further illustrates how acquisition is affected by the frequency and frequency distribution of exemplars within each island of the construction (e.g. [Subj V Obj Obl_{path/loc}]), by their prototypicality, and, using a variety of psychological and corpus linguistic association metrics, by their contingency of form-function mapping.

Keywords: construction learning, verb argument constructions, second language acquisition (SLA), frequency, prototypicality, contingency, type/token frequency, Zipfian distributions, skewed input, categorization, usage-based learning

This paper presents a psycholinguistic analysis of constructions and their acquisition, focusing upon how acquisition is affected by the frequency and frequency distribution in natural language usage of exemplars within each island of the construction, by their prototypicality, and by their contingency of form-function mapping. Our theoretical framework is informed by Cognitive Linguistics, particularly constructionist perspectives (e.g., Bates & MacWhinney, 1987; Goldberg, 1995, 2003, 2006; Lakoff, 1987; Langacker, 1987; Ninio, 2006; Robinson & Ellis, 2008; Tomasello, 2003), corpus linguistics (Biber, Conrad, & Reppen, 1998;

Sinclair, 1991, 2004), and psychological theories of cognitive and associative learning as they relate to the induction of psycholinguistic categories from experience (Ellis, 1998, 2002a, 2003, 2006a, 2006b, 2006c). The basic tenets are as follows: Language is intrinsically symbolic. It is constituted by a structured inventory of constructions as conventionalized form-meaning pairings used for communicative purposes. Usage leads to these becoming entrenched as grammatical knowledge in the speaker's mind, the degree of entrenchment being proportional to the frequency of usage (Bybee, 2005; Ellis, 2002a; Langacker, 2000). Constructions are of different levels of complexity and abstraction; they can comprise concrete and particular items (as in words and idioms), more abstract classes of items (as in word classes and abstract grammatical constructions), or complex combinations of concrete and abstract pieces of language (as mixed constructions). The acquisition of constructions is input-driven and depends upon the learner's experience of these form-function relations. It develops following the same cognitive principles as the learning of other categories, schema and prototypes (Cohen & Lefebvre, 2005; Murphy, 2003). Creative linguistic competence emerges from the collaboration of the memories of all of the utterances in a learner's entire history of language use and the frequency-biased abstraction of regularities within them (Ellis, 2002a). Cognitive linguistics, corpus linguistics, and psycholinguistics are alike in their realizations that we cannot separate grammar from lexis, form from function, meaning from context, nor structure from usage.

Constructions specify the morphological, syntactic and lexical form of language and the associated semantic, pragmatic, and discourse functions (Figure 1). Any utterance is comprised of a number of constructions that are nested. Thus the expression *Today he walks to town* is constituted of lexical constructions such as *today, he, walks*, etc., morphological constructions such as the verb inflection *s* signaling third person singular present tense, abstract grammatical constructions such as Subj, VP, and Prep, the intransitive motion Verb-Locative (VL: [Subj V Ob]_{path/loc}) verb-argument construction (VAC), etc. The function of each of these forms contributes in communicating the speaker's intention.

Psychological analyses of the learning of constructions as form-meaning pairs is informed by the literature on the associative learning of cue-outcome contingencies where the usual determinants include: factors relating to the form such as frequency and salience; factors relating to the interpretation such as significance in the comprehension of the overall utterance, prototypicality, generality, redundancy, and surprise value; factors relating to the contingency of form and function; and factors relating to learner attention, such as automaticity, transfer, overshadowing, and blocking (Ellis, 2002a, 2003, 2006b, 2008b). For example, as illustrated in Figure 1, some forms are more salient: '*today*' is a stronger psychophysical form in the input than is '*s*', thus while both provide cues to present time, *today* is much

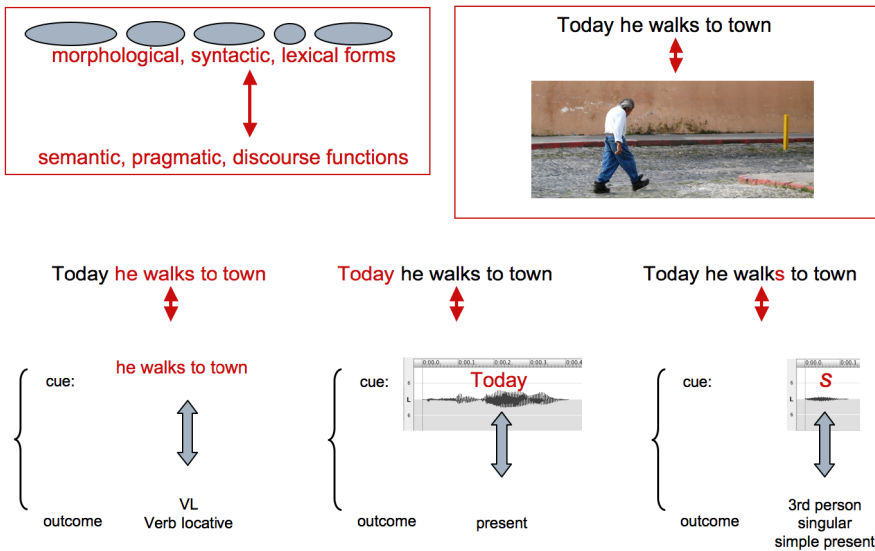


Figure 1. Constructions as form-function mappings. Any utterance comprises multiple nested constructions. Some aspects of form are more salient than others — the amount of energy in *today* far exceeds that in *s*.

more likely to be perceived, and *s* can thus become overshadowed and blocked, making it difficult for second language learners of English to acquire (Ellis, 2006c, 2008a). These various psycholinguistic factors conspire in the acquisition and use of any linguistic construction.

While some constructions, like *walk*, are quite concrete, imageable, and specific in their interpretation, others are more abstract and schematic. For example, the caused motion construction, (e.g. 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 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. 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 (Benigni & Goldberg, 2000; Tomasello, 1992). There is a close relationship between the types of verb 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) argues 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.

These are examples of *semantic bootstrapping* (Pinker, 1989) explanations of the acquisition of VACs whereby semantic categories are used to guide form-meaning correspondences — objects are nouns, actions are verbs, etc, and finer-grained action semantics guide particular VACs:

“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.” (Goldberg, 1995, p. 39).

Learning grammatical constructions thus involves the distributional analysis of the language stream and the contingent analysis of perceptual activity following general psychological principles of category learning. Categories have graded structures, with some members being better exemplars than others. The prototype is the best example, the benchmark against which surrounding “poorer,” more borderline instances are categorized. The greater the token frequency of an exemplar, the more it contributes to defining the category and the greater the likelihood it will be considered the prototype.

Frequency promotes learning, and psycholinguistics demonstrates that language learners are exquisitely sensitive to input frequencies of patterns at all levels (Ellis, 2002a). In the learning of categories from exemplars, acquisition is optimized by the introduction of an initial, low-variance sample centered upon prototypical exemplars (Elio & Anderson, 1981, 1984; Posner & Keele, 1968, 1970). This low variance sample allows learners to get a ‘fix’ on what will account for most of the category members. Then the bounds of the category can later be defined by experience of the full breadth of exemplars. Goldberg, Casenhiser & Sethuraman (2004) demonstrated that in samples of child language acquisition, for each VAC there is a strong tendency for one single verb to occur with very high frequency in comparison to other verbs used, a profile which closely mirrors that of the mothers’ speech to these children. In natural language, Zipf’s law (Zipf, 1935) describes how the highest frequency words account for the most linguistic tokens. Goldberg et al. show that Zipf’s law applied within VACs too, and they argue that this promotes acquisition: tokens of one particular verb account for the lion’s share of instances of each particular argument frame, and this pathbreaking verb is also the one with the prototypical meaning from which that construction is derived:

- The Verb Object Locative (VOL) [Subj V Obj Obl_{path/loc}] construction was exemplified in children’s speech by *put* 31% of the time, *get* 16%, *take* 10%, and

do/pick 6%), 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).

- The Verb Locative (VL) [Subj V Obj_{path/loc}] construction was used in children's speech with *go* 51% of the time, matching the mothers' 39%.
- The ditransitive (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.

Ellis & Ferreira-Junior (2009) replicated these patterns for adult language acquisition in naturalistic learners of English as a second language. They showed that VAC type/token distribution in the input is Zipfian, and that adult learners first acquired the most frequent, prototypical and generic exemplar (e.g. *put* in the VOL VAC, *give* in the VOO ditransitive, etc.). Learning was driven by the frequency and frequency distribution of exemplars within construction and by the degree of match of their meaning to the construction prototype.

Consider language as it passes, utterance by utterance, as illustrated in Figure 2. Learners with a history of exposure to this profile of natural language might thus successfully categorize the different utterances as examples of different VAC categories on the basis of the occupants of the verb islands.

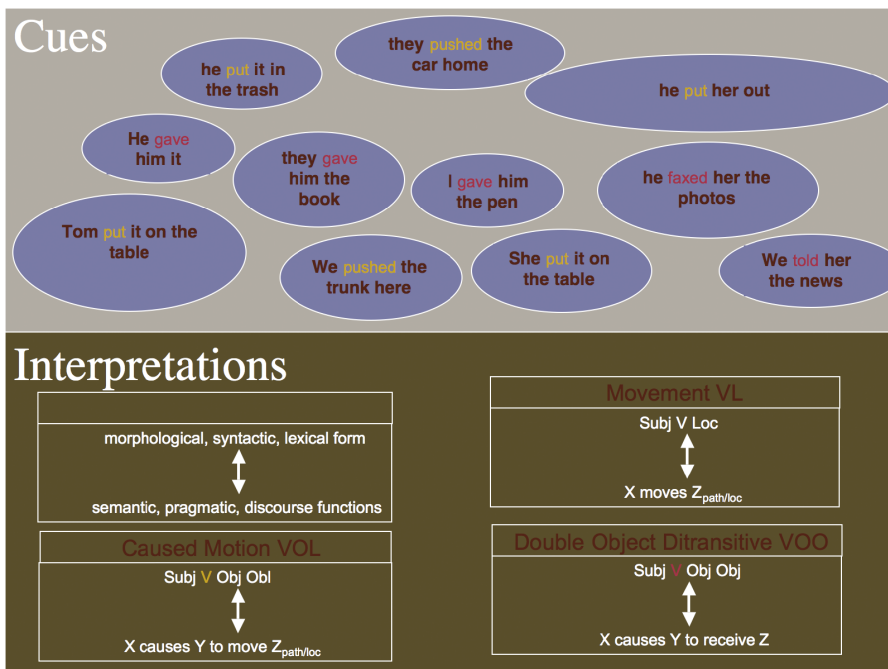


Figure 2. Verb island occupancy as cues to VAC membership

But if the verbs were the only cues that were available, then VACs could have no abstract meaning above that of the verb itself. For, ‘Tom sneezed the napkin across the table’ to make sense despite the intransitivity of *sneeze*, the hearer has to make use of additional information from the syntactic frame. In considering how children learn lexical semantics, Gleitman (1990) argued that they made use of clues from syntactic distributional information — nounlike things follow determiners, prepositions most often prepose a noun phrase in English, etc. The two alternatives of *semantic* and *syntactic bootstrapping* are by no means mutually exclusive, indeed, these two sources of information both reinforce and complement each other.

In the identification of the caused motion construction, (X causes Y to move $Z_{\text{path/loc}}$ [Subj V Obj Obj $_{\text{path/loc}}$]) the whole frame as an archipelago of islands is important. The Subj island helps to identify the beginning bounds of the parse. More frequent, more generic, and more prototypical occupants will be more easily identified. Pronouns, particularly those that refer to animate entities, will more readily activate the schema. As illustrated in Figure 3, the Obj island too will be more readily identified when occupied by more frequent, more generic, and more prototypical lexical items (pronouns like *it* rather than nouns such as *serviette*). So too the locative will be activated more readily if opened by a prepositional island populated by a high frequency, prototypical exemplar such as *on* or *in*. Activation of the VAC schema arises from the conspiracy of all of these features, and arguments about Zipfian type/token distributions and prototypicality of membership extend to all of the islands of the construction.

The role of pronoun islands in child language acquisition has been demonstrated by Childers and Tomasello (2001) and by Wilson (2003), that of prepositional islands by Tomasello (2003, p. 153). Before Powerpoint, in the days when overhead transparencies provided the heights of embellishment for conference papers, Tomasello used to illustrate a putative schematic for the acquisition sequence of VACs by overlaying sequences of exemplars and considering how their cumulative experience results in entrenchment and generalization. As approximated in Figure 4, a high frequency prototype VOL seeds the VAC as a formulaic phrase. Subsequent experience of other VOLs with high frequency prototypical occupants of the different constituent islands leads to generalization of the schema, with the different slots becoming progressively more defined as attractors. The verb island must indeed play a key role in the schema, given its importance in defining the semantics of the sentence as a whole, but the other islands make important contributions too.

So frequency of usage defines construction categories. However, there is one additional qualification to be borne in mind. Some lexical types are very specific in the VACs which they occupy, the vast majority of their tokens occur in just one

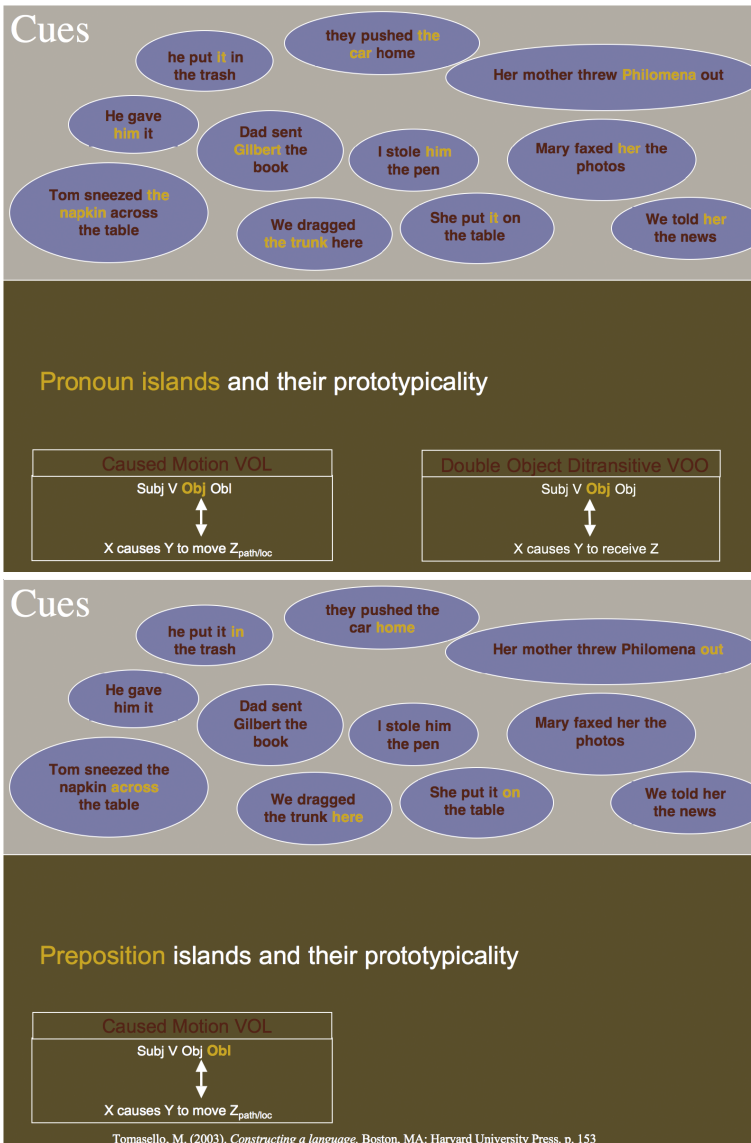


Figure 3. Other syntactic islands and their occupants as cues to VAC identity

VAC, and so they are very reliable and distinctive cues to it. Other lexical types are more widely spread over a range of constructions, and this promiscuity means that they are not faithful cues. *Put* occurs almost exclusively in VOL, it is defining in the acquisition of this VAC and a distinctive and reliable cue in its subsequent recognition. *Turn* however, occurs both in VL and VOL and is less distinctive in distinguishing between these two. Similarly, *send* is attracted to both the VOO

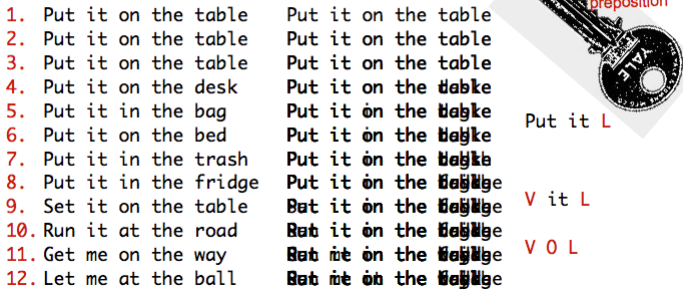


Figure 4. A schematic for the acquisition sequence of the VOL construction. Cumulative experience of VOL exemplars leads to entrenchment. A high frequency prototype VOL seeds the VAC as a formulaic phrase. Experience of other VOLs with high frequency prototypical occupants of the different islands leads to generalization of the schema, with the different slots becoming progressively defined as attractors.

and VOL constructions and so is a less discriminating cue for these categories. Think on the other islands too. It is clear that however useful they are at defining the beginning region of interest in the VAC parse, subject pronouns freely occupy any VAC with hardly any discrimination except that concerning animacy of agent. Prepositions are substantially selective for locatives, but as a class do not distinguish between the transitive and intransitive VACs. And so on.

The associative learning literature has long recognized that while frequency of form is important, so too is contingency of mapping. Consider how, in the learning of the category of birds, while eyes and wings are equally frequently experienced features in the exemplars, it is wings which are distinctive in differentiating birds from other animals. Wings are important features to learning the category of birds because they are reliably associated with class membership, eyes are neither. Raw frequency of occurrence is less important than the contingency between cue and interpretation. Distinctiveness or reliability of form-function mapping is a driving force of all associative learning, to the degree that the field of its study has been known as ‘contingency learning’ since Rescorla (1968) showed that for classical conditioning, if one removed the contingency between the conditioned stimulus (CS) and the unconditioned (US), preserving the temporal pairing between CS and US but adding additional trials where the US appeared on its own, then animals did not develop a conditioned response to the CS. This result was a milestone in the development of learning theory because it implied that it was contingency, not temporal pairing, that generated conditioned responding. Contingency, and its associated aspects of predictive value, information gain, and statistical association, have been at the core of learning theory ever since. It is central in psycholinguistic

theories of language acquisition too (Ellis, 2006b, 2006c, 2008b; Gries & Wulff, 2005; MacWhinney, 1987; Wulff, Ellis, Römer, Bardovi-Harlig, & LeBlanc, 2009).

Taken together, these considerations of language acquisition as the associative learning of schematic constructions from experience of exemplars in usage generate a number of hypotheses concerning VAC acquisition:

Verb islands

- H1. The frequency distribution for the types occupying the verb island of each VAC will be Zipfian.
- H2. The first-learned verbs in each VAC will be those which appear more frequently in that construction in the input.
- H3. The pathbreaking verb for each VAC will be much more frequent than the other members.
- H4. The first-learned verbs in each VAC will be prototypical of that construction's functional interpretation.
- H5. The first-learned verbs in each construction will be those which are more distinctively associated with that construction in the input.

The other islands in the VAC archipelago

We assume similar contributions from the other islands in each VAC, though perhaps to a lesser degree. We wish to determine the degree to which, for each constituent island:

- H6. The frequency distribution for the types occupying that island of each VAC will be Zipfian.
- H7. The first-learned types in each VAC island will be those which appear more frequently in that construction island in the input.
- H8. The pathbreaking type for each VAC island will be much more frequent than the other members.
- H9. The first-learned types in each VAC island will be prototypical of that island's contribution to the construction's functional interpretation.
- H10. The first-learned types in each VAC island will be those which are more distinctively associated with that construction island in the input.

We test these proposals for naturalistic second language learners of English VACs in the European Science Foundation (ESF) corpus (Perdue, 1993). In our earlier piece we tested hypotheses 1–4. This paper takes these beginnings further by addressing hypotheses 5–10.

Method

The 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 ½ years, collected the spontaneous second language of adult immigrants in France, Germany, Great Britain, The Netherlands and Sweden. Data was gathered longitudinally with the learners being recorded in interviews every 4 to 6 weeks for approximately 30 months. The corpus is available from the Max Planck Institute for Psycholinguistics (<http://www.mpi.nl/world/tg/lapp/esf/esf.html>) and alternatively in CHILDES (MacWhinney, 2000a, 2000b) chat format from the Talkbank website (<http://talkbank.org/data/SLA/>).

Participants

Our analysis is based on the data for seven ESL learners living in Britain whose native languages were Italian (Vito, Lavinia, Andrea, and Santo) or Punjabi (Ravinder, Jarnail, and Madan). Details of these participants can be found in Dietrich, Klein and Noyau (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, route descriptions, and role plays. The NS language data is 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 seven participants and their conversation partners were analyzed.

Procedure

The transcription files were downloaded from the MPI website using the IMDI BCBrowser 3.0. Various 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 semi-automated 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, Casenhiser & Set-

huraman (2004) to identify utterances containing examples of VL, VOL or VOO constructions, e.g.

- a. SLA: you come out of my house. [come] [VL]
- b. SMA: charlie say # shopkeeper give me one cigar ## he give it ## he er # he smoking # [give] [VOO]
- c. SRA: no put it in front # thats it # yeah [put] [VOL]

The coded constructions so identified were checked for accuracy by an English native research assistant who served as an independent coder. Any disagreements were resolved through discussion. Each identified construction was also tagged for their speaker and for the number of months they had been in the study at time of utterance.

Analysis of contingency / distinctiveness

A wide variety of measures are available to determine the degree of association between a cue and an outcome, or, in the case of language, between a linguistic form and its function. If the variables are categorical then all begin with a contingency table like that in Table 1, which shows the frequency of the number of observations that fall in to each of the cells.

Table 1. A contingency table showing the four possible combinations of events showing the presence or absence of a target Cue and an Outcome

	Outcome	No Outcome
Cue	<i>a</i>	<i>b</i>
No cue	<i>c</i>	<i>d</i>

a, b, c, d represent frequencies, so, for example, *a* is the frequency of conjunctions of the cue and the outcome, and *c* is the number of times the outcome occurred without the cue.

A good cue is one where, whenever it is present the outcome pertains, and whenever absent the outcome does not, i.e. where observations load on the diagonal in cells *a* and *d* rather than being randomly distributed about the table. Perhaps the most common statistic that is adopted to assess the association between a pair of events such as these is the chi squared test (χ^2). Within corpus linguistics, a suite of association measures like this have been developed for the particular case of determining the co-occurrences of words and other linguistic elements such as constructions. Within collostructional analysis (Gries & Stefanowitsch, 2004; Stefanowitsch & Gries, 2003), lexemes that are significantly associated with a construction are referred to as collexemes of that construction, where the association is quantified by means of the log to the base of 10 of the p-value of the Fisher

Yates exact test performed on such contingency data. This measure is preferred to χ^2 because it does not violate distributional assumptions. Distinctive collexeme analysis has mostly been applied to look into the association between words and constructional variants, such as the dative alternation or particle placement; for the purposes of the present study, we use it to investigate the association between verbs and the VACs they occur in. All computations were done with Stefan Gries' R script *coll.analysis* 3.2 (Gries, 2007). The script uses an exact binomial test to quantify the association strength between the verbs and the VACs they occur in. It provides a p-value for each verb with each VAC and log transforms it such that highly positive and highly negative values indicate a large degree of attraction and repulsion respectively, while 0 indicates random co-occurrence. An (absolute) p_{\log} value that is equal to or higher than 1.3 corresponds to a probability of error of 5% or less.

The Fisher-Yates exact test, like χ^2 , is a measure of the two-way dependency between a pair of events. But associations are not necessarily reciprocal in strength. Recall how 'bird' cues 'eyes', but eyes are not distinctive cues for the category 'bird'. These directional relations therefore need to be separately assessed. Ellis (2006b), reviewing relevant associative learning literature, proposes that for the case of construction learning, the directional association between a form and a function is best measured using the one-way dependency statistic ΔP (Allan, 1980):

$$\begin{aligned}\Delta P &= P(O|C) - P(O|-C) \\ &= a/(a+b) - c/(c+d) \\ &= (ad - bc)/[(a+b)(c+d)]\end{aligned}$$

ΔP is the probability of the outcome given the cue ($P(O|C)$) minus the probability of the outcome in the absence of the cue ($P(O|-C)$). When these are the same, when the outcome is just as likely when the cue is present as when it is not, there is no covariation between the two events and $\Delta P = 0$. ΔP approaches 1.0 as the presence of the cue increases the likelihood of the outcome and approaches -1.0 as the cue decreases the chance of the outcome — a negative association. Shanks' (1995) review of the human associative learning literature shows that ΔP is a good predictor of cue learnability. It can thus be used as a measure of the degree to which a particular type, for example the verb occupying a verb island, is distinctive in signaling a particular VAC, or, in turn, the degree to which a VAC selects a particular type in that slot. We will use ΔP in these ways to investigate the degree to which lexical types in islands are predictive of particular VACs and, separately, the degree to which particular VACs are predictive of particular lexical types in their various islands. Again, these relationships are not necessarily reciprocal.

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 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).

Ellis & Ferreira-Junior (2009) present the detailed methods and findings with relation to hypotheses 1–4. We summarize them in very brief synopsis here to set the stage for the subsequent hypotheses.

H1. The frequency distribution for the types occupying the verb island of each VAC will be Zipfian.

The frequency distributions of the verb types in the VL, VOL and VOO

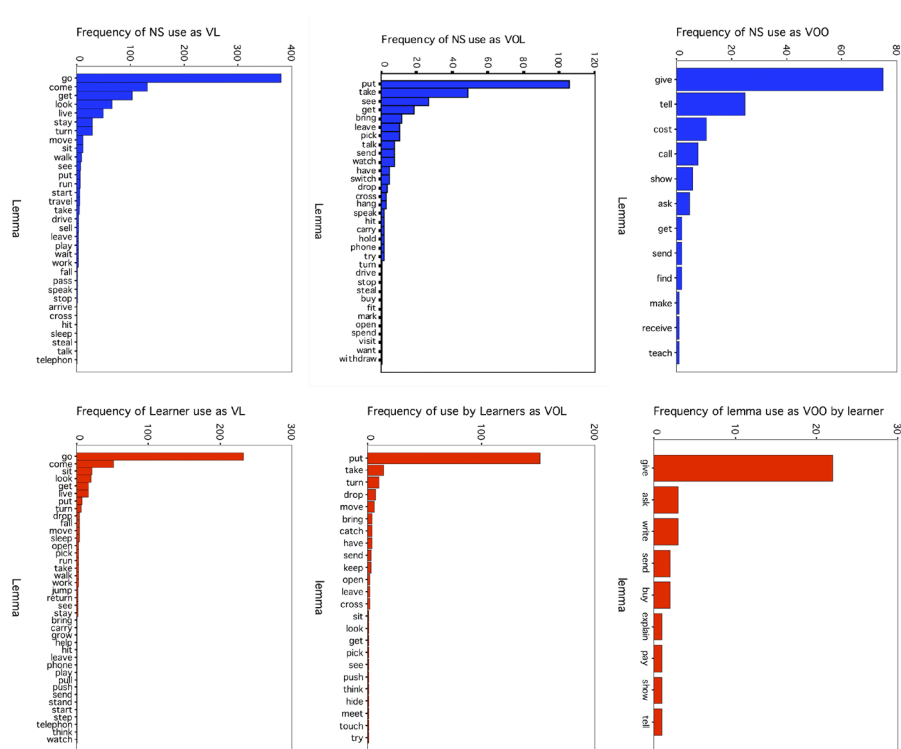


Figure 5. Zipfian type-token frequency distributions of the verbs populating the Interviewers' and Learners' VL, VOL, and VOO constructions. Note the similar rankings of verbs across Interviewers and Learners in each VAC.

constructions produced by the NS interviewers and the NNS learners are shown in Figure 5.

For the NS interviewers *go* constituted 42% of the total tokens of VL, *put* constituted 35% of VOL use, and *give* constituted 53% of VOO. After this leading exemplar, subsequent verb types decline rapidly in frequency.

For the NNS learners, again, for each construction there is one exemplar that accounted for the lion's share of total productions of that construction: *go* constituted 53% of VL, *put* 68% of VOL, and *give* 64% of VOO.

Plots of these frequency distributions as log verb frequency against log verb rank produce straight line functions explaining in excess of 95% the variance thus confirming that Zipf's law is a good description of the frequency distributions with the frequency of any verb being inversely proportional to its rank in the frequency table for that construction, the relationship following a power function.

H2. The first-learned verbs in each VAC will be those which appear more frequently in that construction in the input.

The rank order of emergence of verb types in the learner constructions followed the frequencies in the interviewer NS data. Correlational analyses across all 80 verb types which featured in any of the NS and/or NNS constructions confirmed this to be so. For the VL construction, frequency of lemma use by learner correlated with the frequency of lemma use by NS interviewer $r(78)=0.97, p<.001$. The same analysis for VOL resulted in $r(78)=0.89, p<.001$, and for VOO resulted in $r(78)=0.93, p<.001$. The acquisition functions are illustrated in Figure 6.

H3. The pathbreaking verb for each VAC will be much more frequent than the other members.

Go was the first-learned verb for VL, *put* for VOL, and *give* for VOO. The Zipfian frequency profiles (Figure 5) for the types/tokens confirm H3.

H4. The first-learned verbs in each VAC will be prototypical of that construction's functional interpretation.

In order to assess the degree to which different verbs matched the prototypical semantics of the three VACs, Ellis & Ferreira-Junior (2009) had native English speakers rate the verbs on a 9 point scale for the degree to which they matched a VL schema (the movement of someone or something to a new place or in a new direction), a VOL schema (someone causes the movement of something to a new place or in a new direction), or a VOO schema (someone causes someone to receive something). Ellis & Ferreira-Junior then assessed the association between verb-acquisition order and prototypicality so measured.

For the VL construction the most used verb, *go*, was rated as 7.4 out of 9 in terms of the degree to which it matched the prototypical schematic meaning. The

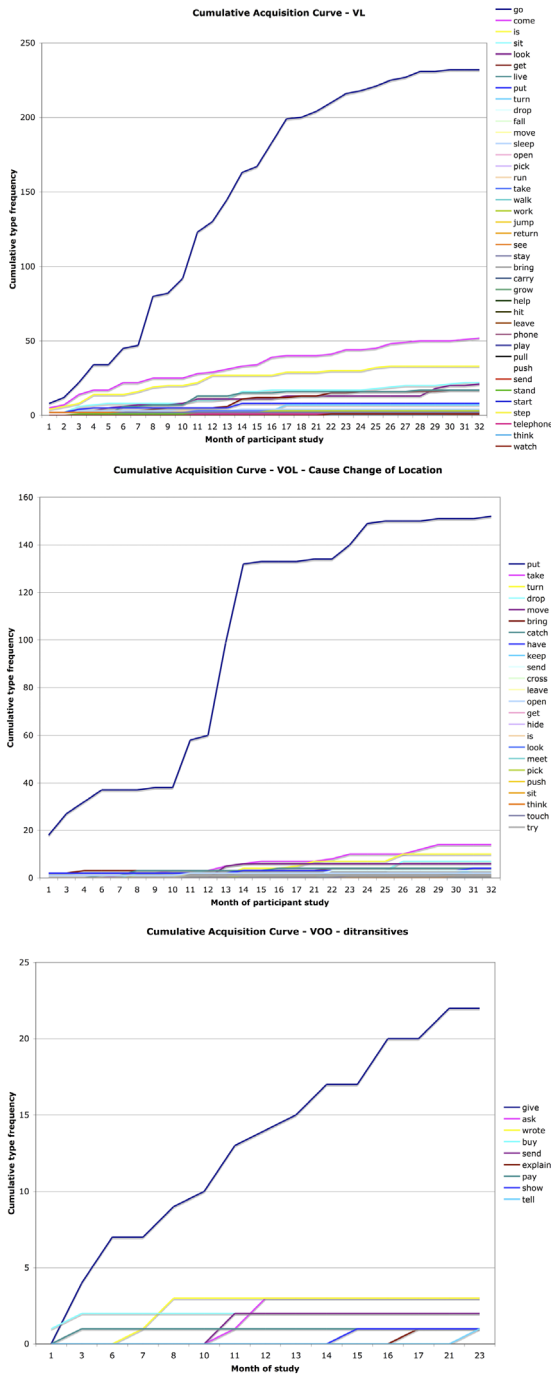


Figure 6. Learner use of verb types in the VL, VOL and VOO constructions as a function of study month.

correlation between prototypicality of verb meaning and log frequency of learner use was VL $\rho(78) = 0.44, p < .001$. We had expected a higher correlation than this but realized that ten other verbs surpassed *go* 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). These match the schema very well, but their additional specific action semantics limit the generality of their use. What is special about *go* is that it is prototypical and generic — thus widely applicable.

The same pattern held for the other constructions. For VOL, the most used verb *put* was rated 8.0 in terms of how well it described the construction schema. The correlation between prototypicality of verb meaning and log frequency of learner use was VOL $\rho(78) = 0.29, p < .01$. *Put* 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) which are more specific in their action semantics. *Put*, as the pathbreaking exemplar is both prototypical and generic. For the VOO construction, the most used verb *give* was rated 9.0 in terms of how well it described the VOO schema. The correlation between prototypicality of verb meaning and log frequency of learner use was VOO $\rho(78) = 0.34, p < .001$.

With regard hypotheses 1–4, in sum, learner VAC acquisition is seeded by the highest frequency, prototypical, and generic exemplar across learners and VACs.

H5. The first-learned verbs in each construction will be those which are more distinctively associated with that construction in the input.

Table 2 shows the top ten lexical types that occupied the verb islands for the three different VACs VL, VOL, VOO ordered by (1) by frequency in the NNS learners' speech, (2) frequency in the NS speech, (3) collexeme strength p_{\log} , (4) contingency (ΔP Construction- \rightarrow Word), (5) contingency (ΔP Word- \rightarrow Construction). When computing indices of contingency we use the frequency of these verbs in the whole corpus in calculating expected frequencies, and thus we are measuring the degree to which each verb is associated with these constructions in the language as a whole.

As we have already seen under H2, learner uptake frequency is strongly associated with frequency in the NS speech (over the 80 verbs, VL: $r = 0.97$; VOL $r = 0.89$; VOO $r = 0.93$). It can also be seen that learner uptake is predicted extremely well by collexeme strength (Fisher-Yates) in the NS speech (over the 80 verbs, VL: $r = 0.96$; VOL $r = 0.97$; VOO $r = 0.97$), by contingency (ΔP Construction- \rightarrow Word) in the NS speech (over all 80 verbs, VL: $r = 0.95$; VOL $r = 0.89$; VOO $r = 0.93$) and, to a lesser degree, by contingency (ΔP Word- \rightarrow Construction) in the NS speech (over the 80 verbs, VL: $r = 0.26$; VOL $r = 0.18$; VOO $r = 0.75$).

These different measures of association are themselves highly correlated, and with such multicollinearity it is difficult to separate the predictor variables.

Table 2. Top ten Verb island types for the three different VACs VL, VOL, VOO ordered by (1) by frequency in the NNS learners' speech, (2) frequency in the NS speech, (3) collexeme strength p_{log} , (4) contingency (ΔP Construction->Word), (5) contingency (ΔP Word->Construction).

	Frequency in that VAC in the NNS Learners		Frequency in that VAC in the NS input		Collocation Strength Fisher Yates Exact p_{log}		Delta P Construction > Word		Delta P Word->Construction
VL									
go	233	go	380	go	192.52	go	0.36	fell	0.61
come	52	come	132	come	75.37	come	0.13	turn	0.59
sit	22	get	104	look	42.58	get	0.09	stay	0.57
look	21	look	66	get	36.63	look	0.07	sit	0.48
get	17	live	50	live	32.00	live	0.05	pass	0.44
live	17	stay	30	turn	24.90	turn	0.03	look	0.42
put	8	turn	30	stay	24.08	stay	0.03	live	0.42
turn	7	move	12	sit	8.98	sit	0.01	move	0.38
drop	4	sit	12	move	7.68	move	0.01	come	0.36
fall	4	walk	10	walk	5.28	walk	0.01	run	0.33
VOL									
put	152	put	106	put	154.25	put	0.35	mark	0.98
take	14	take	49	take	40.06	take	0.15	hang	0.98
turn	10	see	27	pick	14.54	see	0.05	drop	0.98
drop	7	get	19	bring	11.90	bring	0.04	switch	0.81
move	6	bring	12	switch	7.65	get	0.04	put	0.74
bring	4	leave	11	leave	7.60	pick	0.04	pick	0.63
catch	4	pick	11	drop	6.74	leave	0.03	fit	0.48
have	4	send	8	send	5.69	send	0.02	cross	0.48
send	3	watch	8	hang	5.05	watch	0.02	bring	0.34
keep	3	talk	8	cross	3.77	talk	0.02	hit	0.31
VOO									
give	22	give	75	give	139.57	give	0.54	give	0.75
ask	3	tell	25	cost	16.31	tell	0.16	cost	0.47
write	3	cost	11	tell	14.81	cost	0.08	receive	0.32
send	2	call	8	show	7.63	call	0.05	show	0.29
buy	2	show	6	call	7.14	show	0.04	call	0.13
explain	1	ask	5	ask	1.62	ask	0.02	teach	0.08
pay	1	get	2	receive	1.55	send	0.01	tell	0.07
show	1	send	2	send	1.22	find	0.01	send	0.04
tell	1	find	2	teach	1.00	receive	0.01	ask	0.02
		receive	1	find	0.70	teach	0.01	find	0.01

However, it is clearly the case that NS collexeme strength (Fisher-Yates) is a very strong predictor of NNS acquisition, as is ΔP (Construction->Word). What is less predictive is ΔP (Word->Construction). When a construction cues a particular word, that word occurs very often in that construction and, as we can see in Table 2, it tends to be very generic. When a word cues a particular construction, it may be a lower frequency word, quite specific in its action semantics and thus very selective of that construction (e.g. *fell*, *turn*, and *stay* for VL, *mark*, *hang*, and *drop* for VOL).

The very strong correlations between learner uptake and contingency (Fisher-Yates and ΔP Construction->Word) confirm H5.

H6. The frequency distribution for the types occupying each of the islands of each VAC will be Zipfian.

We determined the frequency distributions of the types occupying each (non-verb) island in the VL (Subj, Prep, Locative), VOL (Subj, Obj, Prep, Locative), and VOO (Subj, Obj₁, Obj₂) constructions produced by the NS interviewers and the NNS learners. The frequency distribution for each island appeared Zipfian. There are too many graphs to be able to include them here, so we restrict ourselves to one example of each island type for illustration: those for the NS interviewers and

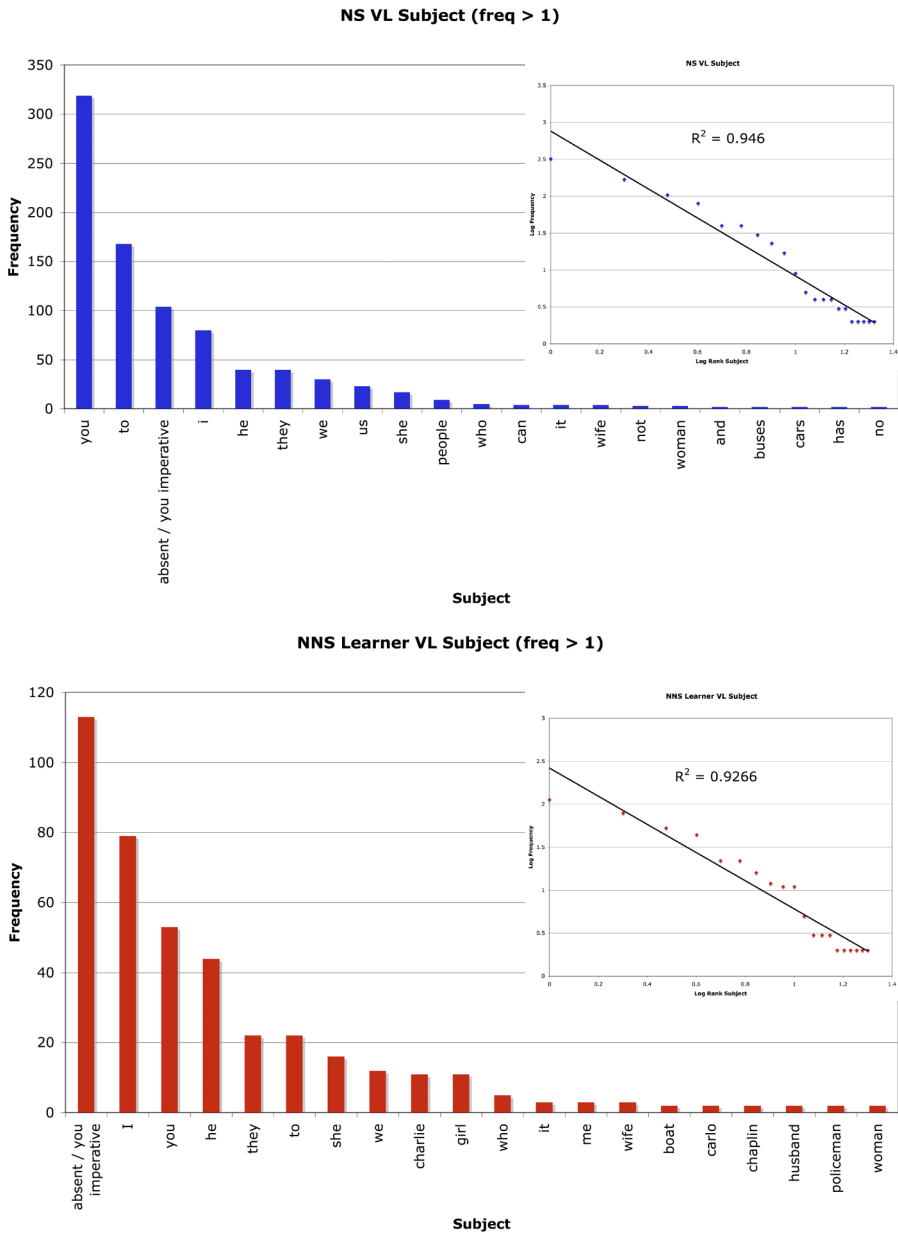


Figure 7. VL Subj island occupancy in NS and NNS learners. Inset shows log frequency vs. log rank plots to be linear, and thus a Zipfian power law relationship.

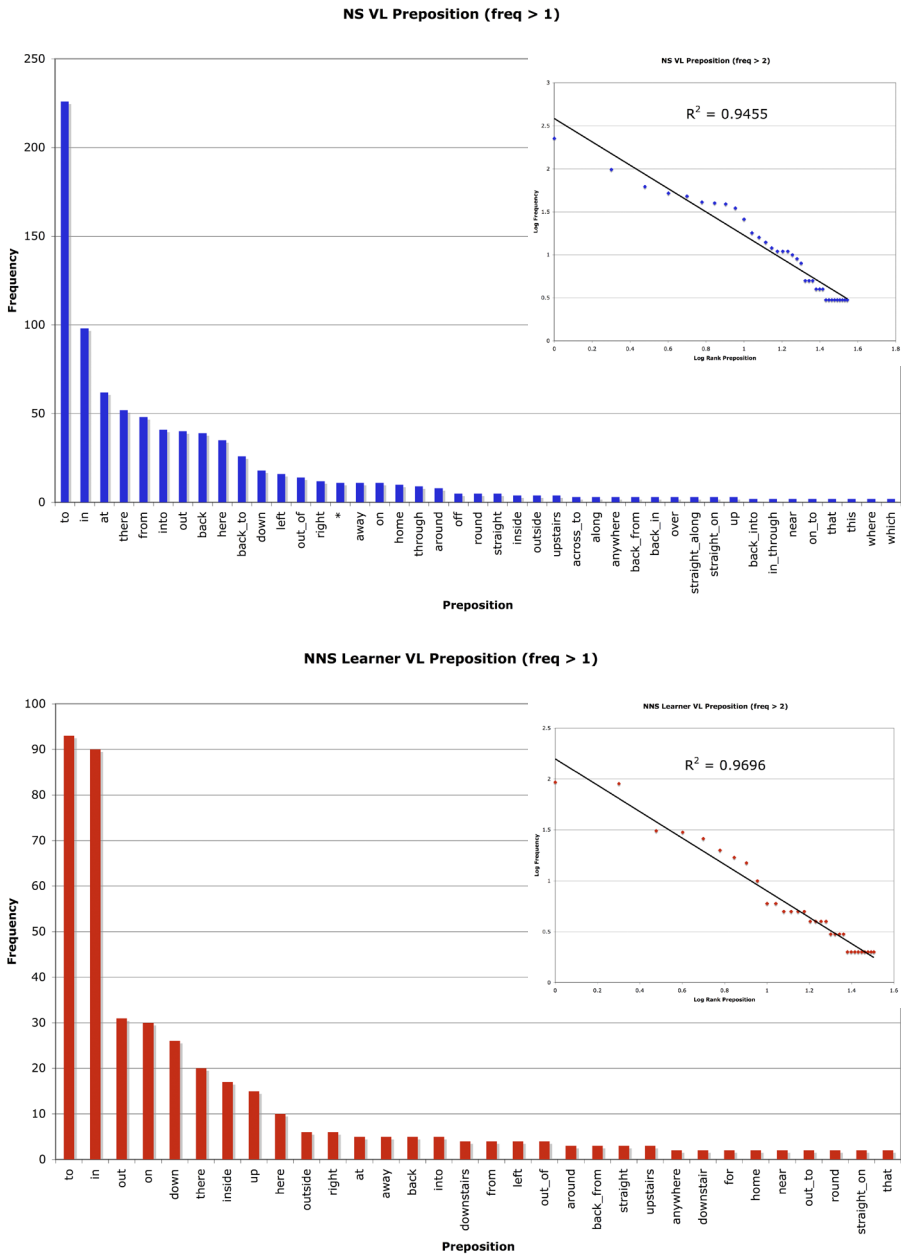


Figure 8. VL Prep island occupancy in NS and NNS learners. Inset shows log frequency vs. log rank plots to be linear, and thus a Zipfian power law relationship.

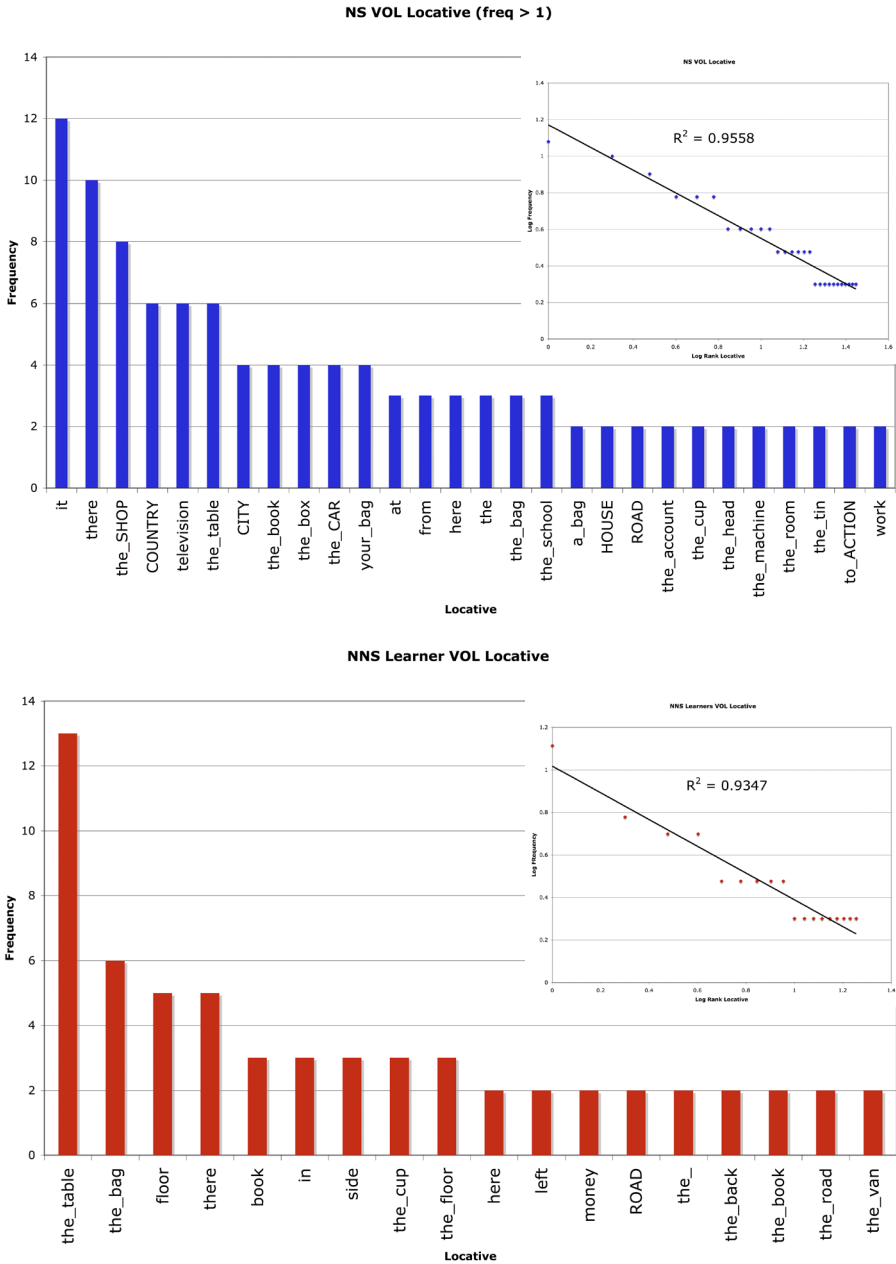


Figure 9. VOL Locative island occupancy in NS and NNS learners. Inset shows log frequency vs. log rank plots to be linear, and thus a Zipfian power law relationship.

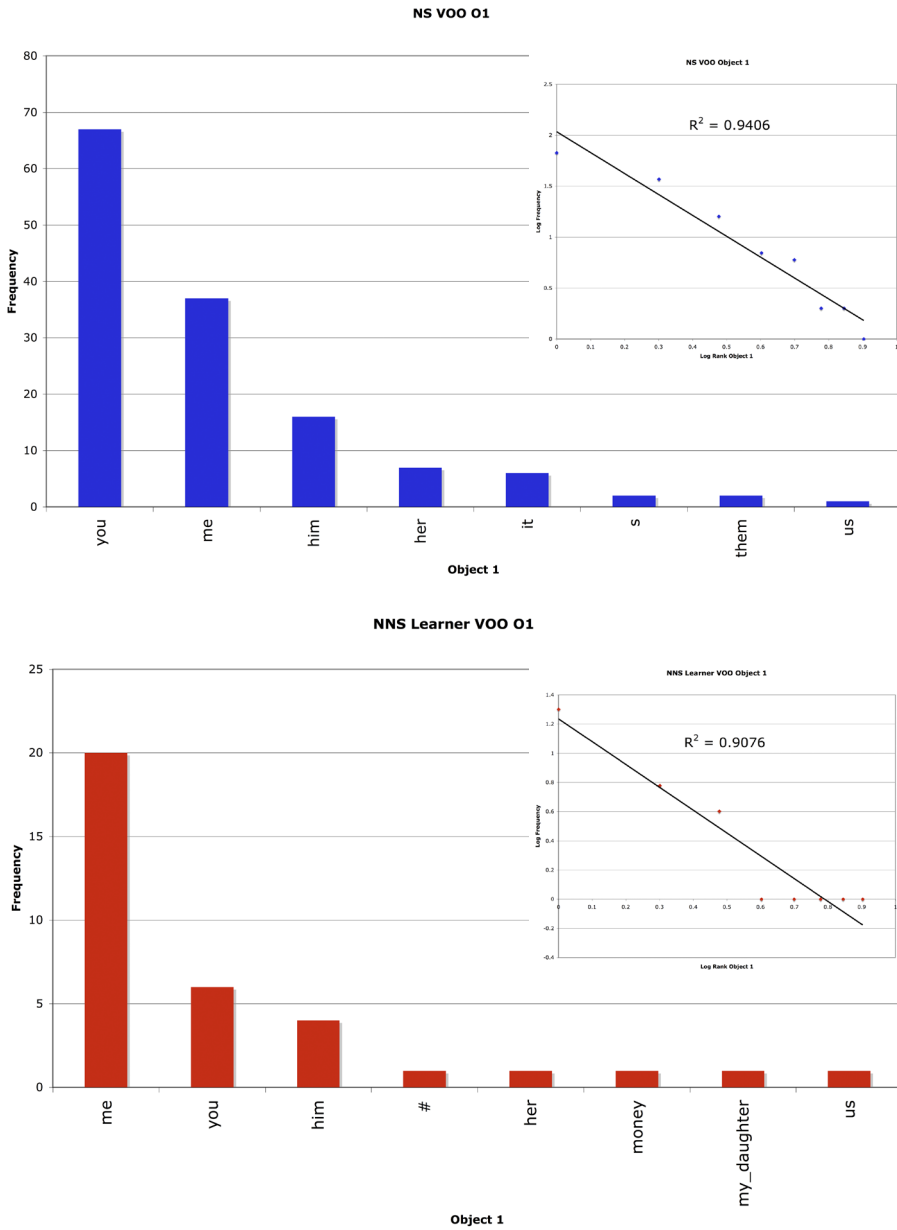


Figure 10. VOO Obj₁ island occupancy in NS and NNS learners. Inset shows log frequency vs. log rank plots to be linear, and thus a Zipfian power law relationship.

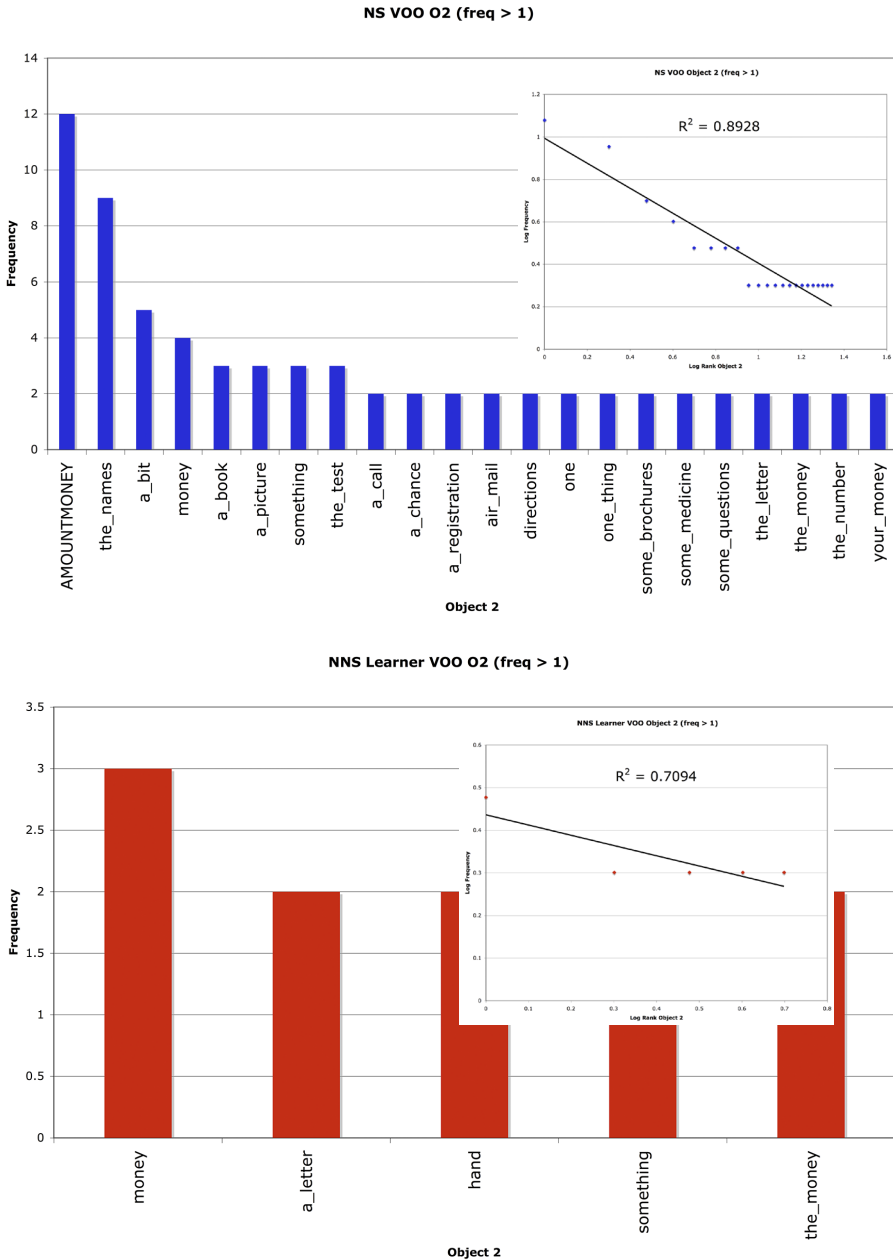


Figure 11. VOO Obj₂ island occupancy in NS and NNS learners. Inset shows log frequency vs. log rank plots to be linear, and thus a Zipfian power law relationship.

NNS learners for the Subj island of VL (Figure 7), the Prep island of VL (Figure 8), the Locative island of VOL (Figure 9), the Obj₁ island of VOO (Figure 10), and the Obj₂ island of VOO (Figure 11).

In each case, for NS and NNS both, there is one lead exemplar that takes the lion's share of instances in that island, and, as shown in the inset graphs, the distribution is a power function as indexed by the regression of log frequency vs. log type rank being linear and explaining a substantial part of the variance. This also held true for the other islands that space prevents us from illustrating here: the R² for the NS and NNS log-log regressions are, respectively, VL locative island (0.98, 0.93), VOL Subj (0.88, 0.90), VOL Obj (0.96, 0.90), VOL Prep (0.96, 0.98), and VOO Subj (0.81, 0.92).

H7. The first-learned types in each VAC island will be those which appear more frequently in that construction island in the input.

Inspection of the graphs in Figures 7–11 shows a clear correspondence between the types used in each island by the NNSs and the types that occupy them in the speech of the NS Interviewers.

The NS interviewers filled the Subj island of VL with the following top 8 types, in decreasing order: *you*, *to* [verb in infinitive phrase], implied *you* [imperative], *I*, *he*, *they*, *we*, *us*. The corresponding list for the NNS learners was: implied *you* [imperative], *I*, *you*, *he*, *they*, *to* [verb in infinitive phrase], *she*, *we*. A similar profile was found for the Subj island for VOL: NS (*you*, implied *you* [imperative], *to*, *I*, *they*, *he*, *we*, *she*), NNS (implied *you* [imperative], *I*, *you*, *to* [verb in infinitive phrase], *he*, *the*, *bag*, *they*), and for VOO: top 4 NS (*I*, *you*, implied *you* [imperative], *to* [verb in infinitive phrase]), NNS (*they*, *I*, *she*, implied *you* [imperative]). Although a potentially infinite range of nouns could occupy the Subj islands in these different constructions, in NS and learner alike, it is populated by far by a few high frequency generic forms, the pronouns.

The top 8 occupants of the Prep island in VL were for the NS speakers (*to*, *in*, *at*, *there*, *from*, *into*, *out*, *back*), and for the NNS Learners (*to*, *in*, *out*, *on*, *down*, *there*, *inside*, *up*). Similar profiles occurred for the Prep island of VOL: NS (*in*, *on*, *there*, *off*, *out*, *up*, *from*, *to*), NNS (*in*, *on*, *there*, *the_table*, *up*, *from*, *the_bag*, *down*). Although a wide range of directions or places could occupy the post-verbal island in these two constructions, in NS and learner alike, it is occupied by far by a few high frequency generic prepositions.

The rest of the locative in VL and VOL was filled with a wider range of occupant types than for the Prep slot — there was a much longer tail of low frequency items. Nevertheless, a few common stereotypical locations prevailed in the top 8 : NS VL (null, *COUNTRY* [*country* or specific example, e.g., Italy, England, India, etc.], *CITY* [London, Birmingham, etc.], *the_SHOP* [shop, or specific, e.g.

post office, newsagent, supermarket, etc.], *town*, *the_station* [station, police station, etc.], *the_BUILDING* [e.g., Court, nursery], *work*), NNS VL (null, *the_SHOP*, *CITY*, *HOUSE*, *school*, *COUNTRY*, *the_floor*, *the_ROOM*), NS VOL (*it*, *there*, *the_SHOP*, *COUNTRY*, *television*, *the_table*, *CITY*, *the_book*, *the_box*), and NNS VOL (*the_table*, *the_bag*, *floor*, *there*, *book*, *in*, *side*, *the_cup*).

Finally, the Obj islands of VOO. For Obj₁, the NS interviewers' top 5 occupants were (*you*, *me*, *him*, *her*, *it*), the NNS learners' the top 3 were (*me*, *you*, *him*). For Obj₂, the NS top 8 were (*AMOUNTMONEY* [like twenty pounds, three pounds, etc.], *the_names*, *a_bit*, *money*, *a_book*, *a_picture*, *something*, *the_test*), the NNS top 8 were (*money*, *a_letter*, *hand*, *something*, *the_money*, *a_bill*, *a_cheque*, *a_lot*).

The general pattern then, for each island of each VAC, is that there is high correspondence between the top types used in each island by the NNS learners and the types that occupy them in NS input typical of their experience.

H8. The first-learned pathbreaking type for each VAC island will be much more frequent than the other members.

Inspection of Figures 7–11 along with the qualitative patterns summarized under H7 demonstrates that, unlike for the verbs which centre the semantics of each VAC, there is no single pathbreaker that initially takes over each of the other islands of the VAC exclusively. Nevertheless, for each construction, the frequency distributions for each island is Zipfian, and there is a high overlap between NS and NNS use of the top 5–10 occupant types which together make up the predominance of its inhabitation.

H9. The first-learned types in each VAC island will be prototypical of that island's contribution to the construction's functional interpretation.

The 5–10 major occupant types for each island do indeed seem to be prototypical in role. We do not have native speaker ratings for the prototypicality of meaning of the other island inhabitants as we had for the verbs, however, the qualitative data described so far is highly consistent with this hypothesis.

Although a potentially infinite range of nouns could occupy the Subj islands in the VL, VOL and VOO constructions, in NS and NNS learner alike, these were occupied by the most frequent, prototypical and generic forms for this slot — pronouns such as *I*, *you*, *it*, *we*, etc.

For the Prep islands in VL and VOL, while other fillers such as the NS *up_on*, *straight_along*, and *round_to*, and the NNS learner *the_back_side*, *other_side*, and *downstairs*, were indeed to be found at lower frequencies, in the main this island was clearly identified with high frequency prototypical generic prepositions such as *in*, *on*, *there*, *to*, and *off*.

For the remainder of the locatives in VL and VOL, there is wider scope in this island, but nevertheless the normal conventions of everyday inquiry that relate to

Table 3. The two most frequent inhabitants of the islands constituting the VL, VOL, and VOO constructions in NS Interviewer and NNS Learner utterances (Percentages reflect total island occupancy)

VAC	Speaker		%		%		%		%		
VL	NS	Subj		Verb		Prep		Locative			
		you	35	go	42	to	25	_	31		
		to [infinitive]	19	come	14	in	11	Country	8		
	NNS	implied	26	go	53	to	21	_	34		
		I	18	come	12	in	21	the_shop	3		
	VOL	NS	Subj		Verb		Obj		Prep		Locative
you			36	put	35	it	21	in	17	it	4
		implied you	22	take	16	them	5	on	15	there	3
NNS		implied you	63	put	68	in	27	in	16	the_table	6
		I	5	take	6	it	19	on	13	the_bag	3
VOO		NS	Subj		Verb		Obj1		Obj2		
	I		21	give	54	you	48	AmountMoney	9		
		you	19	tell	18	me	27	the_names	6		
	NNS	they	22	give	62	me	56	money	8		
		I	17	write	8	you	17	a_letter	6		

our comings and goings typically stem or end up at countries or cities of origin or destination, or, depending on degree of zoom and scale, blocks, buildings or rooms of commerce, officialdom, or dwelling.

Likewise for the VOO VACs, these are very stereotypic in their functional interpretations, and there is broad overlap between NS and NNS use: people (as pronouns) routinely give people (as pronouns) money, letters, bills, or books.

Indeed if we put all of these data together, and simply choose the two lead exemplars, the most popular / populating types in each island in each VAC for the NS and NNSs in turn, we compose the following utterances shown in Table 3.

Selecting either of the top two alternatives and moving left to right as in a finite state grammar, we generate from these alternatives such prototypical VL sequences as “come in”, “I went to the shop”, and “to go to [Country]”, such prototypical VOL sequences as “you put it in it”, “take them in there”, “put it in the bag”, and “I put it on the table”, and such prototypical VOO sequences as “I gave you AmountMoney”, “you tell me the names”, “they wrote me a letter”, and “I’ll give you money”.

H10. The first-learned types in each VAC island will be those which are more distinctively associated with that construction island in the input.

Table 4 shows the top ten lexical types that occupied the Subj islands for VL, VOL, VOO ordered by (1) by frequency in the NNS learners’ speech, (2) frequency in the NS speech, (3) collexeme strength p_{log} , (4) contingency (ΔP Construction->Word),

Table 4. Top ten Subj island types for the three different VACs VL, VOL, VOO ordered by (1) by frequency in the NNS learners' speech, (2) frequency in the NS speech, (3) collexeme strength p_{\log} , (4) contingency (ΔP Construction->Word), (5) contingency (ΔP Word -> Construction)

	Frequency in that VAC in the NNS Learners		Frequency in that VAC in the NS input		Collocation Strength Fisher Yates Exact p_{\log}		Delta P Construction > Word		Delta P Word- >Construction
VL									
absent / you	113	you	319	us	1.93	you	0.05	who	0.33
i	79	to	168	to	1.71	to	0.05	girls	0.33
you	53	absent_you	104	you	1.38	he	0.02	buses	0.33
he	44	i	80	he	1.30	us	0.02	cars	0.33
they	22	he	40	people	0.97	she	0.01	has	0.33
to	22	they	40	she	0.90	people	0.01	no	0.33
she	16	we	30	who	0.87	who	0.01	not	0.33
we	12	us	23	not	0.52	we	0.00	woman	0.33
charlie	11	she	17	woman	0.52	not	0.00	people	0.23
girl	11	people	9	we	0.38	woman	0.00	us	0.22
VOL									
absent_you	142	you	109	absent_you	4.93	absent_you	0.10	absent_you	0.15
ashtray	2	absent_you	68	you	0.69	you	0.03	and	0.11
bag	4	to	46	i	0.48	i	0.01	you	0.02
block	2	i	35	and	0.27	and	0.00	i	0.02
book	1	they	14	they	0.26	they	0.00	they	0.00
chair	1	he	9	us	1.35	wife	0.00	wife	-0.03
girl	1	we	6	we	0.94	girls	0.00	to	-0.03
he	5	she	3	it	0.85	buses	0.00	he	-0.05
i	11	us	2	to	0.76	cars	0.00	she	-0.08
it	1	people	1	he	0.60	has	0.00	we	-0.09
VOO									
they	8	i	29	it	5.86	i	0.11	that	0.90
i	6	you	26	i	3.84	it	0.06	it	0.55
she	5	absent_you	22	that	2.96	that	0.02	i	0.11
*	3	to	16	we	0.83	we	0.02	can	0.10
v	3	it	9	they	0.51	absent_you	0.02	we	0.06
you	3	they	8	absent_you	0.45	they	0.01	they	0.03
to	2	we	7	can	0.38	can	0.00	absent_you	0.01
chaplin	1	that	3	you	4.59	girls	0.00	to	-0.04
electricity-bi	1	he	2	to	1.44	and	0.00	she	-0.06
shopkeeper	1	she	1	he	1.06	buses	0.00	us	-0.07

(5) contingency (ΔP Word->Construction). When calculating indices of association under H10, in computing expected frequencies, unlike for the verbs under H5, here we used the frequency of these lexemes across the three VACs under study rather than in the interviewer speech corpus as a whole, and thus we are measuring the degree to which each verb is more distinctively associated with one of these constructions over the others.

Collexeme strengths greater than 1.3 are significant at $p < .05$. Table 4 shows that certain subjects are more significantly associated with certain VACs, for example *it* and *I* for VOO, and that VOL is more often used with the subject implied_you in the imperative. Nevertheless, comparison of the data in Tables 4 and 2 shows that verbs are generally much more distinctively associated with these VACs than Subjs both in terms of Collocation Strength, and ΔP measures. Thus while the occupants of Subj do follow a Zipfian distribution lead by pronouns, and thus could indeed signal the beginning of a VAC parse, they tend not to be associated with any particular VAC.

The same analysis is presented for the Prep islands in Table 5 which shows that the prepositions are much more like the verbs in their selectivity: *to*, *back*, *in* and

Table 5. Top ten Prep island types for the three different VACs VL, VOL, VOO ordered by (1) by frequency in the NNS learners' speech, (2) frequency in the NS speech, (3) collexeme strength p_{log} , (4) contingency (ΔP Construction \rightarrow Word), (5) contingency (ΔP Word \rightarrow Construction)

	Frequency in that VAC in the NNS Learners		Frequency in that VAC in the NS input	Collocation Strength Fisher Yates Exact p_{log}		Delta P Construction \rightarrow Word		Delta P Word \rightarrow Construction	
VL		to	226	to	171.30	to	0.74	to	0.88
	93	in	100	back	41.52	in	0.28	down	0.79
	90	out	73	at	35.83	back	0.23	left	0.78
	31	on	62	in	33.85	at	0.20	right	0.78
	30	down	56	out	25.04	out	0.17	away	0.78
	26	there	52	into	23.63	there	0.16	straight	0.78
	20	inside	48	from	21.94	from	0.15	across	0.78
	17	up	41	there	20.99	into	0.13	along	0.78
	15	here	35	here	17.98	here	0.11	anywhere	0.78
	10	outside	20	down	13.14	down	0.07	near	0.78
	6								
VOL		in	52	on	18.98	on	0.14	on	0.58
	35	on	45	off	6.12	in	0.08	off	0.53
	28	there	17	up	5.05	off	0.05	up	0.50
	14	the_table	15	in	3.48	up	0.04	under	0.44
	12	up	14	around	1.85	around	0.02	around	0.28
	7	from	13	round	1.28	round	0.01	round	0.28
	6	the_bag	11	inside	1.08	inside	0.01	inside	0.28
	6	down	11	under	0.89	there	0.01	outside	0.20
	5	inside	8	outside	0.71	outside	0.01	over	0.18
	5		7	over	0.50	under	0.01	in	0.13
	4								
VOO				to	-12.44	to	-0.20	to	-0.13
				in	-7.68	in	-0.13	in	-0.12
				back	-3.92	back	-0.07	back	-0.11
				out	-3.42	out	-0.06	at	-0.11
				there	-3.37	there	-0.06	there	-0.11
				at	-3.27	at	-0.06	out	-0.11
				from	-2.87	from	-0.05	on	-0.11
				on	-2.82	on	-0.05	from	-0.11
				into	-2.12	into	-0.04	here	-0.11
				here	-1.93	here	-0.03	into	-0.11
	(REPULSION)								

out are distinctively associated with VL both in terms of Collocation Strength (all top 10 prepositions associated with VL have Fisher Yates p_{log} scores in excess of 10) and ΔP Construction \rightarrow Word; *on*, *off*, and *up* are strongly selective of VOL; and all of these prepositions are repulsed by VOO.

Finally, this analysis is shown for the Obj1 islands in Table 6 where any Obj1 repulses VL, in VOL *it*, *money*, *them* and *that* are very significantly distinctive in terms of their Collocation strength and their ΔP Word \rightarrow Construction, and the object pronouns *you*, *me*, *him* and *her* are distinctive recipients in VOO, again with strong selectivity in terms of Collocation strength and ΔP , particularly Word \rightarrow Construction.

Together, these analyses demonstrate that, while the verb island is most distinctive, the constituency of the other islands is by no means negligible in determining VAC identity. In particular, VL and VOL are highly selective in terms of their Prep occupancy, and Obj1 types clearly select between VOO, VOL and VL.

Table 6. Top ten Obj1 island types for the three different VACs VL, VOL, VOO ordered by (1) by frequency in the NNS learners' speech, (2) frequency in the NS speech, (3) collexeme strength p_{log} , (4) contingency (ΔP Construction \rightarrow Word), (5) contingency (ΔP Word \rightarrow Construction)

	Frequency in that VAC in the NNS Learners		Frequency in that VAC in the NS input		Collocation Strength Fisher Yates Exact p_{log}		Delta P Construction > Word		Delta P Word->Construction	
VL (REPULSION)				you	-37.62	you	-0.17	you	-0.71	
				it	-34.94	it	-0.16	it	-0.71	
				me	-18.30	me	-0.08	me	-0.69	
				money	-8.29	money	-0.04	money	-0.68	
				him	-7.80	him	-0.04	him	-0.68	
				them	-7.80	them	-0.04	them	-0.68	
				that	-4.85	that	-0.02	that	-0.68	
				her	-4.37	her	-0.02	her	-0.68	
				bag	-4.37	bag	-0.02	bag	-0.68	
				string	-4.37	string	-0.02	string	-0.68	
VOL	in	61	it	63	it	35.54	money	0.06	it	0.73
	it	42	money	17	money	11.14	that	0.03	money	0.78
	on	7	them	14	them	7.27	bag	0.03	them	0.66
	him	6	that	10	that	6.51	string	0.03	that	0.78
	the_bag	6	bag	9	bag	5.86	something	0.01	bag	0.78
	er	5	string	9	string	5.86	it	0.20	string	0.78
	right	5	you	7	something	2.59	them	0.04	me	-0.23
	left	4	something	4	us	0.40	us	0.00	you	-0.14
	that	4	her	2	microphone	0.00	her	0.00	something	0.78
	a_look	3	us	1	me	-4.18	you	-0.04	him	-0.23
VOO	me	20	you	67	you	64.56	you	0.48	me	0.92
	you	6	me	37	me	38.51	me	0.27	him	0.91
	him	4	him	16	him	16.11	him	0.12	you	0.85
	#	1	her	7	her	5.48	her	0.05	her	0.68
	her	1	it	6	us	0.71	us	0.01	us	0.40
	money	1	them	2	them	0.30	them	0.00	them	0.02
	my_daught	1	us	1	microphone	0.00	microphone	0.00	it	-0.02
	us	1	money	0	money	-0.81	something	0.00	something	-0.10
			that	0	that	-0.48	bag	-0.01	bag	-0.10
			bag	0	bag	-0.43	string	-0.01	string	-0.10

Conclusions

In terms of our original specific hypotheses we can conclude as follows:

Verb islands

- H1. The frequency distribution for the types occupying the verb island of each VAC is Zipfian.
- H2. The first-learned verbs in each VAC are those which appear more frequently in that construction in the input. Correlations between learner uptake and frequency of lemma use in the input are in excess of $r > 0.89$.
- H3. The pathbreaking verb for each VAC is much more frequent than the other members.
- H4. The first-learned verbs in each VAC is prototypical of that construction's action semantics but also generic and thus widely applicable. Other verbs which fit the VAC prototype well, but which have additional specifications of manner which restrict their usage, tend to be acquired later.

H5. The first-learned verbs in each construction are very distinctively associated with that construction in the input. Correlations between learner uptake and contingency are near perfect: $r > 0.96$ for collexeme strength (Fisher-Yates) and $r > 0.89$ for ΔP (Construction- \rightarrow Word).

The other islands in the VAC archipelago

Many of these findings are also true for the other islands in each VAC.

H6. The frequency distribution for the types occupying all of the VAC islands is Zipfian, both in the NS input and in the learners' uptake. Each of these slots can serve as a distinctive attractor because frequency of experience cuts a distinctive groove for each element of the construction, like the key in Figure 4.

H7. The first-learned types in each VAC island appear more frequently in that construction island in the input. For each island there is a high correspondence between the types that occupy them in the NS input and those first picked up by the learners.

H8. The first-learned types for each VAC island are more frequent but, unlike the verb islands, there is not one unique slot-filler that initially dominates each. Instead, for each construction island, there is high overlap between the top 5–10 occupant types in NNS uptake and those in NS use.

H9. The first-learned types in each VAC island, as was true for verbs, are prototypical and generic.

H10. The first-learned types in each VAC island do tend to be more distinctively associated with that construction island in the input, but to varying degrees. Verb island occupancy certainly plays the major role in this respect. But other island occupants, particularly Prep and Obj1, make their contributions too in distinguishing between these VACs. It is less the case for the Subj slot which is occupied by high frequency pronouns which very usefully mark the beginning of a clause but not which type of VAC.

There is good evidence that these factors first play out in learning to comprehend the L2. The analyses of NNS here are done irrespective of total accuracy of form in production. While 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; Purdue, 1993). This typically involves a pragmatic topic-comment word ordering, where old information goes first and new information follows. Examples for the VOL include:

yeah this television put it up the # book #
 this bag <he put him> [/?] put in the st [/?] er floor # <bag> [>1]
 a horse # put in there <> [\$ laughs]
 you know which block put down
 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”.

In sum, these findings suggest that the acquisition of linguistic constructions is affected by a wide range of factors. For each island there is:

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,
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 the ways that language as a complex adaptive system has evolved to be learnable. Ellis & Larsen-Freeman (2009) presents various connectionist simulations of the emergence of these VACs.

Constructivists view language acquisition as repeated cycles of differentiation and integration (Studdert-Kennedy, 1991). In Figure 4 we illustrated the putative default naturalistic sequence of naturalistic acquisition from high utility generic formulaic phrase, to limited scope formula, to analyzed schematic construction (Ellis, 2002a, 2002b; Lieven & Tomasello, 2008; Tomasello, 2000, 2003). In our investigation of the lead occupants of each island in Table 3, we saw how selecting either of the top two alternatives and moving left to right generated such VL sequences as “come in” or “I went to the shop”, such VOL sequences as “you put it in it”, “take them in there”, or “put it in the bag”, and such VOO sequences as “they wrote me a letter”, and “I’ll give you money”. We have come full cycle. The analysis of the islands of the schematic construction, and their predominant occupants in usage, identifies high frequency, prototypical, generic, and often distinctive occupants of each, which, in combination, generate high utility generic phrases.

Each island in each VAC archipelago thus makes a significant contribution to its identification and interpretation, and our general conclusion is that the acquisition of these linguistic constructions can be understood in terms of the cognitive

science of concept formation. It follows the general associative principles of the induction of categories from the experience of the features of their exemplars. In natural language, the type-token frequency distributions of the occupants of each of these VAC islands, their prototypicality and generality of function in these roles, and their reliability of mappings between these, together conspire to optimize learning.

Acknowledgements

This research was supported in part by a doctoral grant from the Brazilian government through the CAPES Foundation, grant BEX # 0043060.

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. & B. MacWhinney. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of Language Acquisition* (pp. 157–193). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bencini, G. & A.E. Goldberg. (2000). The contribution of argument structure constructions to sentence meaning. *Journal of Memory and Language*, 43, 640–651.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating language structure and use*. New York: Cambridge University Press.
- Bybee, J. (2005). *From Usage to Grammar: The mind's response to repetition*. Paper presented at the Linguistic Society of America, Oakland.
- Childers, J.B. & M. Tomasello. (2001). The role of pronouns in young children's acquisition of the English transitive construction. *Developmental Psychology*, 37, 739–748.
- Cohen, H. & C. Lefebvre. (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: John Benjamins.
- Elio, R. & J.R. Anderson. (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. & J.R. Anderson. (1984). The effects of information order and learning mode on schema abstraction. *Memory & Cognition*, 12(1), 20–30.
- Ellis, N.C. (1998). Emergentism, connectionism and language learning. *Language Learning*, 48(4), 631–664.
- Ellis, N.C. (2002a). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 143–188.
- Ellis, N.C. (2002b). Reflections on frequency effects in language processing. *Studies in Second Language Acquisition*, 24(2), 297–339.

- 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: 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), 1–24.
- Ellis, N.C. (2006c). Selective attention and transfer phenomena in SLA: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics*, 27(2), 1–31.
- Ellis, N.C. (2008a). The dynamics of language use, language change, and first and second language acquisition. *Modern Language Journal*, 41(3), 232–249.
- Ellis, N.C. (2008b). Usage-based and form-focused language acquisition: The associative learning of constructions, learned-attention, and the limited L2 endstate. In P. Robinson & N.C. Ellis (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition*. London: Routledge.
- Ellis, N.C. & F. Ferreira-Junior. (2009). Construction Learning as a function of Frequency, Frequency Distribution, and Function. *Modern Language Journal*, 93(3), 370–385.
- Ellis, N.C. & D. Larsen-Freeman. (2009). Emergent models of the acquisition of Verb Argument Constructions. *Language Learning*, 59, Supplement 1.
- Feldweg, H. (1991). *The European Science Foundation Second Language Database*. Nijmegen: Max-Planck-Institute for Psycholinguistics.
- Gleitman, L. (1990). The structural source of verb meaning. *Language Acquisition*, 1, 3–55.
- Goldberg, A.E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- 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.
- Gries, S.T. (2007). Coll.analysis 3.2. A program for R for Windows 2.x.
- Gries, S.T. & A. Stefanowitsch. (2004). Extending collocation analysis: a corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics*, 9, 97–129.
- Gries, S.T. & S. Wulff. (2005). Do foreign language learners also have constructions? Evidence from priming, sorting, and corpora. *Annual Review of Cognitive Linguistics*, 3, 182–200.
- Klein, W. & C. Purdue. (1992). *Utterance Structure: Developing grammars again*. Amsterdam: John 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.
- Langacker, R.W. (2000). A dynamic usage-based model. In M. Barlow & S. Kemmer (Eds.), *Usage-based Models of Language* (pp. 1–63). Stanford, CA: CSLI Publications.
- Lieven, E. & M. Tomasello. (2008). Children's first language acquisition from a usage-based perspective. In P. Robinson & N.C. Ellis (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition*. New York and London: Routledge.

- MacWhinney, B. (1987). The Competition Model. In B. MacWhinney (Ed.), *Mechanisms of Language Acquisition* (pp. 249–308). Hillsdale, NJ: Erlbaum.
- MacWhinney, B. (2000a). *The CHILDES Project: Tools for analyzing talk, Vol 1: Transcription format and programs (3rd ed.)*: (2000). xi, 366pp.
- MacWhinney, B. (2000b). *The CHILDES Project: Tools for analyzing talk, Vol 2: The database (3rd ed.)*: (2000). viii, 418pp.
- Murphy, G.L. (2003). *The Big Book of Concepts*. Boston, MA: 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.
- Perdue, C. (Ed.). (1993). *Adult Language Acquisition: Crosslinguistic perspectives*. Cambridge: Cambridge University Press.
- Pinker, S. (1989). *Learnability and Cognition: The acquisition of argument structure*. Cambridge, MA: Bradford Books.
- Posner, M.I. & S.W. Keele. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, 77, 353–363.
- Posner, M.I. & S.W. Keele. (1970). Retention of abstract ideas. *Journal of Experimental Psychology*, 83, 304–308.
- Rescorla, R.A. (1968). Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative and Physiological Psychology*, 66, 1–5.
- Robinson, P. & N.C. Ellis. (Eds.). (2008). *A Handbook of Cognitive Linguistics and Second Language Acquisition*. London: Routledge.
- Shanks, D.R. (1995). *The Psychology of Associative Learning*. New York: Cambridge University Press.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, J. (2004). *Trust the Text: Language, corpus and discourse*. London: Routledge.
- Stefanowitsch, A. & S.T. Gries. (2003). Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics*, 8, 209–243.
- Studdert-Kennedy, M. (1991). Language development from an evolutionary perspective. In N.A. Krasnegor, D. M. Rumbaugh, R. L. Schiefelbusch & M. Studdert-Kennedy (Eds.), *Biological and Behavioral Determinants of Language Development* (pp. 5–28). Mahwah, NJ: Erlbaum.
- Tomasello, M. (1992). *First Verbs: A case study of early grammatical development*. New York: Cambridge University Press.
- Tomasello, M. (2000). The item based nature of children's early syntactic development. *Trends in Cognitive Sciences*, 4, 156–163.
- Tomasello, M. (2003). *Constructing a Language*. Boston, MA: Harvard University Press.
- Wilson, S. (2003). Lexically specific constructions in the acquisition of inflection in English. *Journal of Child Language*, 30, 75–115.
- Wulff, S., Ellis, N.C., Römer, U., Bardovi-Harlig, K., & LeBlanc, C. (2009). The acquisition of tense-aspect: Converging evidence from corpora, cognition, and learner constructions. *Modern Language Journal*, 93(3), 370–385.
- Zipf, G.K. (1935). *The Psycho-biology of Language: An introduction to dynamic philology*. Cambridge, MA: The M.I.T. Press.

Author's address

Nick C. Ellis
University of Michigan
500, E. Washington Street
Ann Arbor
MI 48104
ncellis@umich.edu

About the authors

Nick Ellis is a Professor of Psychology and a Research Scientist at the English Language Institute of the University of Michigan. His research interests include psycholinguistic, cognitive linguistic, applied linguistic, corpus linguistic, and emergentist approaches to second and foreign language acquisition. He serves as General Editor of the journal *Language Learning*.

Fernando G. Ferreira-Junior did his PhD in Linguistics at the Federal University of Minas Gerais, Brazil. He was a one-year Visiting Scholar at the English Language Institute, University of Michigan, and has been a Senior Lecturer in English at IFET Ouro Preto since 1999. His main research interests are the cognitive and psycholinguistic processes involved in language learning within a constructionist and connectionist framework.