Statistical construction learning: Does a Zipfian problem space ensure robust language learning?

Nick C. Ellis and Matthew Brook O'Donnell

1. Introductory Overview

One of the key mysteries of language development is that each of us as learners has had different language experiences and yet somehow we have converged on broadly the same language system. From diverse, noisy samples, we end up with similar competence. How so? Some views hold that there are constraints in the learner's estimation of how language works, as expectations of linguistic universals pre-programmed in some innate language acquisition device. Others hold that the constraints are in the dynamics of language itself – that language form, language meaning, and language usage come together to promote robust induction by means of statistical learning over limited samples. The research described here explores this question with regard English verbs, their grammatical form, semantics, and patterns of usage.

As a child, you engaged your parents and friends talking about things of shared interest using words and phrases that came to mind, and all the while you learned language. We were privy to none of this. Yet somehow we have converged upon a similar-enough 'English' to be able to communicate here. Our experience allows us similar interpretations of novel utterances like "the ball mandoolz across the ground" or "the teacher spugged the boy the book." You know that *mandool* is a verb of motion and have some idea of how *mandooling* works – its action semantics. You know that *spugging* involves some sort of transfer, that the teacher is the donor, the boy the recipient, and that the book is the transferred object. How is this possible, given that you have never heard these verbs before? Each word of the construction contributes individual meaning, and the verb meanings in these Verb-Argument Constructions (VACs) is usually at the core. But the larger configuration of words has come to carry meaning as a whole too. The VAC as a category has inherited its schematic meaning from all of the examples you have heard. *Mandool* inherits its interpretation from the echoes of the verbs that occupy this VAC – words like *come*, *walk*, *move*, . . . , *scud*, *skitter and flit* – in just the same way that you can conjure up an idea of the first author's dog Phoebe, who you have never met either, from the conspiracy of your memories of dogs.

Knowledge of language is based on these types of inference, and verbs are the cornerstone of the syntax-semantics interface. To appreciate your idea of Phoebe, we would need a record of your relevant evidence (all of the dogs you have experienced, in their various forms and frequencies) and an understanding of the cognitive mechanisms that underpin categorization and abstraction. In the same way, if we want a scientific understanding of language knowledge, we need to know the evidence upon which such psycholinguistic inferences are based, and the relevant psychology of learning. These are the goals of our research. To describe the evidence, we take here a sample of VACs based upon English form, function, and usage distribution. The relevant psychology of learning, as we will explain, suggests that learnability will be optimized for constructions that are (1) Zipfian in their type-token distributions in usage (the most frequent word occurring approximately twice as often as the second most frequent word, which occurs twice as often as the fourth most frequent word, etc.), (2) selective in their verb form occupancy, and (3) coherent in their semantics. We assess whether these factors hold for our sample of VACs.

In summary, our methods are as follows; we will return to explain each step in detail. We search a tagged and dependency-parsed version of the British National Corpus (BNC 2007), a representative 100-million word corpus of English, for 23 example VACs previously identified in the *Grammar Patterns* volumes (Francis, Hunston, and Manning 1996; Hunston and Francis 1996) resulting from the COBUILD corpus-based dictionary project (Sinclair 1987). For each VAC, such as the pattern *V*(erb) *across N*(oun phrase), we generate (1) a list of verb types that occupy each construction (e.g. *walk*, *move*, *skitter*). We tally the frequencies of these verbs to produce (2) a frequency ranked type-token profile for these verbs, and we determine the degree to which this is Zipfian (e.g. *come* 474 . . . *spread* 146 . . . *throw* 17 . . . *stagger* 5; see Fig. 1 below). Because some verbs are faithful to one construction while others are more promiscuous, we next produce (3) a contingency-weighted list which reflects their statistical association (e.g. *scud*, *skitter*, *sprawl*, *flit* have the strongest association with *V across N*). Because verbs are highly polysemous, we apply word sense disambiguation algorithms to assign (4) senses to these verbs in the sentences where they are present, according to WordNet (Miller 2009). We use techniques for identifying clustering and degrees of separation in networks to determine (5) the degree to which there is semantic cohesion of the verbs occupying each construction (e.g., semantic fields TRAVEL and MOVE are most frequent for *V across N*), and whether they follow a prototype/radial category structure. In order to gauge the degree to which each VAC is more coherent than expected by chance in terms of the association of its grammatical form and semantics we generate a distributionally-yoked control (a 'control ersatz construction', CEC), matched for type-token distribution but otherwise randomly selected to be grammatically and semantically uninformed. Through the comparison of VACs and CECS of these various measures, and following what is known of the psychology of learning, we assess the consequences for acquisition.

This work is a preliminary interdisciplinary test, across significantly large language usage and learning corpora, of the generalizability of construction grammar theories of language learning informed by cognitive linguistics, learning theory, categorization, statistical learning, usage-based child language acquisition, and complex systems theory.

2. Construction Grammar and Usage

Constructions are form-meaning mappings, conventionalized in the speech community, and entrenched as language knowledge in the learner's mind. They are the symbolic units of language relating the defining properties of their morphological, lexical, and syntactic form with particular semantic, pragmatic, and discourse functions (Goldberg 2006, 1995). Verbs are central in this: their semantic behavior is strongly intertwined with the syntagmatic constraints governing their distributions. Construction Grammar argues that all grammatical phenomena can be understood as learned pairings of form (from morphemes, words, idioms, to partially lexically filled and fully general phrasal patterns) and their associated semantic or discourse functions: "the network of constructions captures our grammatical knowledge *in toto*, i.e. it's constructions all the way down" (Goldberg, 2006, p. 18). Such beliefs, increasingly influential in the study of child language acquisition, emphasize data-driven, emergent accounts of linguistic systematicities (e.g., Tomasello 2003; Clark and Kelly 2006).

Frequency, learning, and language come together in usage-based approaches which hold that we learn linguistic constructions while engaging in communication (Bybee 2010). The last 50 years of psycholinguistic research provides the evidence of usage-based acquisition in its demonstrations that language processing is exquisitely sensitive to usage frequency at all levels of language representation from phonology, through lexis and syntax, to sentence processing (Ellis 2002). That language users are sensitive to the input frequencies of these patterns entails that they must have registered their occurrence in processing. These frequency effects are thus compelling evidence for usage-based models of language acquisition which emphasize the role of input. Language knowledge involves statistical knowledge, so humans learn more easily and process more fluently high frequency forms and 'regular' patterns which are exemplified by many types and which have few competitors (e.g., MacWhinney 2001). Psycholinguistic perspectives thus hold that language learning is the associative learning of representations that reflect the probabilities of occurrence of form-function mappings.

If constructions as form-function mappings are the units of language, then language acquisition involves inducing these associations from experience of language usage. Constructionist accounts of language acquisition thus involve the distributional analysis of the language stream and the parallel analysis of contingent perceptuo-motor activity, with abstract constructions being learned as categories from the conspiracy of concrete exemplars of usage following statistical learning mechanisms (Christiansen and Chater 2001; Jurafsky and Martin 2000; Bybee and Hopper 2001; Bod, Hay, and Jannedy 2003; Ellis 2002; Perruchet and Pacton 2006) relating input and learner cognition.

3. Determinants of Construction Learning

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: (1) input frequency (type-token frequency, Zipfian distribution), (2) form (salience and perception), (3) function (prototypicality of meaning), and (4) interactions between these (contingency of form-function mapping) (Ellis and Cadierno 2009). We will briefly consider each in turn, along with studies demonstrating their applicability:

3.1. Input Frequency

3.1.1. Construction Frequency

Frequency of exposure promotes learning and entrenchment (e.g., Anderson 2000; Ebbinghaus 1885; Bartlett [1932] 1967). Learning, memory and perception are all affected by frequency of usage: the more times we experience something, the stronger our memory for it, and the more fluently it is accessed. The more recently we have experienced something, the stronger our memory for it, and the more fluently it is accessed [hence your reading this sentence more fluently than the preceding one]. The more times we experience conjunctions of features, the more they become associated in our minds and the more these subsequently affect perception and categorization; so a stimulus becomes associated to a context and we become more likely to perceive it in that context.

Frequency of exposure also underpins statistical learning of categories (Mintz 2002; Hunt and Aslin 2010; Lakoff 1987; Taylor 1998; Harnad 1987). Human categorization ability provides the most persuasive testament to our incessant unconscious figuring or 'tallying'. We know that natural categories are fuzzy rather than monothetic. Wittgenstein's (1953) consideration of the concept game showed that no set of features that we can list covers all the things that we call games, ranging as the exemplars variously do from soccer, through chess, bridge, and poker, to solitaire. Instead, what organizes these exemplars into the game category is a set of family resemblances among these members – son may be like mother, and mother like sister, but in a very different way. And we learn about these families, like our own, from experience. Exemplars are similar if they have many features in common and few distinctive attributes (features belonging to one but not the other); the more similar are two objects on these quantitative grounds, the faster are people at judging them to be similar (Tversky 1977). 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. The operationalization of this criterion predicts the speed of human categorization performance – people more quickly classify as *dogs* Labradors (or other typically sized, typically colored, typically tailed, typically featured specimens) than they do dogs with less common features or feature combinations like Shar Peis or Neapolitan Mastiffs. Prototypes are judged faster and more accurately, even if they themselves have never been seen before – someone who has never seen a Labrador, yet who has experienced the rest of the run of the canine mill, will still be fast and accurate in judging it to be a dog (Posner and Keele 1970). Such effects make it very clear that although people don't go around consciously counting features, they nevertheless have very accurate knowledge of the underlying frequency distributions and their central tendencies.

(V9 5/4/12 14:41) WDG-LCB (155mm×230mm) TimesNRMT 1382 Rebuschat pp. 265–304 1382 Rebuschat_09_pisoni (p. 269)

3.1.2. Type and Token Frequency

Token frequency counts how often a particular form appears in the input. Type frequency, on the other hand, refers to the number of distinct lexical items that can be substituted in a given slot in a construction, whether it is a word-level construction for inflection or a syntactic construction specifying the relation among words. For example, the "regular" English past tense *-ed* has a very high type frequency because it applies to thousands of different types of verbs, whereas the vowel change exemplified in *swam* and *rang* has much lower type frequency. The productivity of phonological, morphological, and syntactic patterns is a function of type rather than token frequency (Bybee and Hopper 2001). This is because: (a) the more lexical items that are heard in a certain position in a construction, the less likely it is that the construction is associated with a particular lexical item and the more likely it is that a general category is formed over the items that occur in that position; (b) the more items the category must cover, the more general are its criterial features and the more likely it is to extend to new items; and (c) high type frequency ensures that a construction is used frequently, thus strengthening its representational schema and making it more accessible for further use with new items (Bybee and Thompson 2000). In contrast, high token frequency promotes the entrenchment or conservation of irregular forms and idioms; the irregular forms only survive because they are high frequency. There is related evidence for type-token matters in statistical learning research (Gómez 2002; Onnis et al. 2004). These findings support language's place at the center of cognitive research into human categorization, which also emphasizes the importance of type frequency in classification.

3.1.3. Zipfian Distribution

In natural language, Zipf's law (Zipf 1935) describes how the highest frequency words account for the most linguistic tokens. Zipf's law states that the frequency of words decreases as a power function of their rank in the frequency table. If p_f is the proportion of words whose frequency in a given language sample is *f*, then $p_f \sim f^{-\gamma}$ with $\gamma \approx 1$. Zipf showed this scaling law holds across a wide variety of language samples. Subsequent research provides support for this law as a linguistic universal. Many language events across scales of analysis follow his power law: phoneme and letter strings (Kello and Beltz 2009), words (Evert 2005), grammatical constructs (Ninio 2006; O'Donnell and Ellis 2010), formulaic phrases (O'Donnell and Ellis 2009) etc. Scale-free laws also pervade language structures, such as scale-free networks in collocation (Solé et al. 2005; Bannard and Lieven 2009), in morphosyntactic productivity (Baayen 2008), in grammatical dependencies (Ferrer i Cancho & Solé, 2001, 2003; Ferrer i Cancho, Solé, & Köhler, 2004), and in networks of speakers, and language dynamics such as in speech perception and production, in language processing, in language acquisition, and in language change (Ninio 2006; Ellis 2008). Zipfian covering, where, as concepts need to be refined for clear communication, they are split, then split again hierarchically, determines basic categorization, the structure of semantic classes, and the language form-semantic structure interface (Steyvers and Tennenbaum 2005; Manin 2008). Scale-free laws pervade both language structure and usage. And not just language structure and use. Power law behavior like this has since been shown to apply to a wide variety of structures, networks, and dynamic processes in physical, biological, technological, social, cognitive, and psychological systems of various kinds (e.g. magnitudes of earthquakes, sizes of meteor craters, populations of cities, citations of scientific papers, number of hits received by web sites, perceptual psychophysics, memory, categorization, etc.) (Newman 2005; Kello et al. 2010). It has become a hallmark of Complex Systems theory. Zipfian scale-free laws are universal. Complexity theorists suspect them to be fundamental, and are beginning to investigate how they might underlie language processing, learnability, acquisition, usage and change (Beckner, et al., 2009; Ellis & Larsen-Freeman, 2009b; Ferrer i Cancho & Solé, 2001, 2003; Ferrer i Cancho, et al., 2004; Solé, et al., 2005) Various usage-based/functionalist/ cognitive linguists (e.g., Boyd & Goldberg, 2009; Bybee, 2008, 2010; Ellis, 2008a; Goldberg, 2006; Goldberg, Casenhiser, & Sethuraman, 2004; Lieven & Tomasello, 2008; Ninio, 1999, 2006) argue that it is the coming together of these distributions across linguistic form and linguistic function that makes language robustly learnable despite learners' idiosyncratic experience and the 'poverty of the stimulus'.

In first language acquisition, Goldberg, Casenhiser & Sethuraman (2004) demonstrated that there is a strong tendency for VACs to be occupied by one single verb with very high frequency in comparison to other verbs used, a profile which closely mirrors that of the mothers' speech to these children. They argue that this promotes language acquisition: In the early stages of learning categories from exemplars, acquisition is optimized by the introduction of an initial, low-variance sample centered upon prototypical exemplars. This low variance sample allows learners to get a fix on what will account for most of the category members, with the bounds of the category being defined later by experience of the full breadth of exemplar types.

In naturalistic second language (L2) acquisition, Ellis and Ferreira-Junior (2009) investigated type/token distributions in the items comprising the linguistic form of English VACs (VL verb locative, VOL verb object locative, VOO ditransitive) and showed that VAC verb type/token distribution in the input is Zipfian and that learners first acquire the most frequent, prototypical and generic exemplar (e.g. *put* in VOL, *give* in VOO, etc.).

3.2. Function (Prototypicality of Meaning)

Categories have graded structure, with some members being better exemplars than others. In the prototype theory of concepts (Rosch and 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.

Ellis & Ferreira-Junior (2009) show that the verbs that L2 learners first used in particular VACs are prototypical and generic in function (*go* for VL, *put* for VOL, and *give* for VOO). 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 (1999) argues 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, "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 proportionately less frequent as children acquire more words for specific categories of actions" (p. 53).

3.3. Interactions between these (Contingency of Form-Function Mapping)

Psychological research into associative learning has long recognized that while frequency of form is important, so too is contingency of mapping (Shanks 1995). 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 2008; MacWhinney 1987; Ellis 2006, 2006; Gries and Wulff 2005), with the most developed account for L2 acquisition being that of the Competition model (MacWhinney 1987, 1997, 2001).

Ellis and Ferreira-Junior (2009) use a variety of metrics to show that VAC acquisition is determined by their contingency of form-function mapping. They show that the one-way dependency statistic ΔP (Allan 1980) that is commonly used in the associative learning literature (Shanks 1995), as well as collostructional analysis measures current in corpus linguistics (Gries and Stefanowitsch 2004; Stefanowitsch and Gries 2003) predict effects of form-function contingency upon L2 VAC acquisition. Other researchers use conditional probabilities to investigate contingency effects in VAC acquisition. This is still an active area of inquiry, and more research is required before we know which statistical measures of form-function contingency are more predictive of acquisition and processing.

Ellis and Larsen-Freeman (2009) provided computational (Emergent connectionist) serial-recurrent network simulations of these various factors as they play out in the emergence of constructions as generalized linguistic schema from their frequency distributions in the input. This fundamental claim that Zipfian distributional properties of language usage helps to make language learnable has thus begun to be explored for these three VACs, at least. But three VACs is a pitifully small sample of English grammar. It remains an important research agenda to explore its generality across the wide range of the verb constructicon.

(V9 5/4/12 14:41) WDG-LCB (155mm×230mm) TimesNRMT 1382 Rebuschat pp. 265–304 1382 Rebuschat_09_pisoni (p. 273)

The primary motivation of construction grammar is that we must bring together linguistic form, learner cognition, and usage. An important consequence is that constructions cannot be defined purely on the basis of linguistic form, *or* semantics, *or* frequency of usage *alone*. All three factors are necessary in their operationalization and measurement. Psychology theory relating to the statistical learning of categories suggests that constructions are robustly learnable when they are (1) Zipfian in their type-token distributions in usage, (2) selective in their verb form occupancy, and (3) coherent in their semantics. Our research aims to assess this for a sample of the verbal grammar of English, analyzing the way VACs map form and meaning, and providing an inventory of the verbs that exemplify these constructions and their frequency.

4. Method

As a starting point, we considered several of the major theories and datasets of construction grammar such as FrameNet (Fillmore, Johnson, and Petruck 2003). However, because our research aims to empirically determine the semantic associations of particular linguistic forms, it is important that such forms are initially defined by bottom-up means that are semantics-free. There is no one in corpus linguistics who 'trusts the text' more than Sinclair (2004) in his operationalizations of linguistic constructions on the basis of repeated patterns of words in collocation, colligation, and phrases. Therefore we chose the definitions of VACs presented in the *Verb Grammar Patterns* (Hunston and Francis 1996) that arose out of the COBUILD project (Sinclair 1987) for our first analyses. We focus on a convenience sample of 23 constructions for our initial explorations here. Most of these follow the verb – preposition – noun phrase structure, such as *V into N*, *V after N*, *V as N* (Goldberg 2006), but we also include other classic examples such as the ditransitive, and the *way* construction (Jackendoff 1997).

4.1. Step 1 Construction Inventory: COBUILD Verb Patterns

The form-based patterns described in the COBUILD Verb Patterns volume (Francis, Hunston, and Manning 1996) take the form of word class and lexis combinations, such as *V across N*, *V into N* and *V N N*. For each of these patterns the resource provides information as to the structural configurations and meaning groups found around these patterns through detailed concordance analysis of the Bank of English corpus during the construction of the COBUILD dictionary. For instance, the following is provided for the *V across N* pattern (Francis, Hunston, and Manning 1996):

> The verb is followed by a prepositional phrase which consists of *across* and a noun group.
>
> This pattern has one structure:
>
> * Verb with Adjunct.
> *I cut across the field.*

Further example sentences drawn from the corpus are provided and a list of verbs found in the pattern and that are semantically typical are given. For the *V across N* pattern these are: *brush*, *cut*, *fall*, *flicker*, *flit*, *plane*, *skim*, *sweep*. No indication is given as to how frequent each of these types are or how comprehensive the list of types is. Further structural (syntactical) characteristics of the pattern are sometimes provided, such as the fact that for *V across N* the prepositional phrase is an adjunct and that the verb is never passive. There are over 700 patterns of varying complexity in the *Grammar Patterns* volume. In subsequent work we hope to analyze them all in the same ways we describe here for our sample of 23.

4.2. Step 2 Corpus: BNC XML Parsed Corpora

To get a representative sample of usage, the analysis of verb type-token distribution in the kinds of construction patterns described in *Step 1* should be done across corpora in the magnitude of the tens or hundreds of millions of words. Searching for the pattern as specified requires that the corpora be part-of-speech tagged, and some kind of partial parsing and chunking is necessary to apply the necessary structural constraints (see Mason and Hunston 2004 for exploratory methodology). For this initial work, we chose to use the 100 million word BNC (2007) on account of its size, the breadth of text types it contains and the consistent lemmatization and part-of-speech tagging. Andersen et al. (2008) parsed the XML version of the BNC using the RASP parser (Briscoe, Carroll, and Watson 2006). RASP is a statistical feature-based parser that produces a probabilistically ordered set of parse trees for a given sentence and additionally a set of grammatical relations that capture "those aspects of predicate-argument structure that the system is able to recover and is the most stable and grammar independent representation available" (Briscoe, Carroll, and

Watson 2006). For each VAC, we translate the formal specifications from the COBUILD patterns into queries to retrieve instances of the pattern from the parsed corpus.

4.3. Step 3 Searching Construction Patterns

Using a combination of part-of-speech, lemma and dependency constraints we construct queries for each of the construction patterns. For example, the *V across N* pattern is identified by looking for sentences that have a verb form within 3 words of an instance of *across* as a preposition, where there is an indirect object relation holding between *across* and the verb and the verb does not have any other object or complement relations to following words in the sentence. Table 1 shows our 23 constructions, the number of verb types that occupy them, the total number of tokens found, and the type-token ratio.

We have still to carry out a systematic precision-recall analysis, but the strict constraints using the dependency relations provides us with a good precision and the size of the corpus results in a reasonable number of tokens to carry out distributional analysis. In future, we plan to use a number of different parsers [e.g. Stanford (Klein and Manning 2003), Pro3Gres (Schneider, Rinaldi, and Dowdall 2004), MALT (Nivre, Hall, and Nilsson 2004), and Link (Grinberg, Lafferty, and Sleator 1995)] over the same corpora and use a consensus-based selection method where sentences will be counted if two or more parsers agree (according to queries particular to their parsing output) that it is an instance of a particular construction pattern. Further we will select samples of certain VAC distributions for manual evaluation.

4.4. Step 4 A Frequency Ranked Type-Token VAC Profile

The sentences extracted using this procedure outlined for each of the construction patterns are stored in a document database. This database can then be queried to produce verb type distributions such as those in Table 2 for the *V across N* VAC pattern. These distributions appear to be Zipfian, exhibiting the characteristic long-tailed distribution in a plot of rank against frequency. We have developed scripts in R (R Development Core Team 2008) to generate logarithmic plots and linear regression to examine the extent of this trend. Dorogovstev & Mendes (2003) outline the use of logarithmic binning of frequency against log cumulative frequency methods for measuring distributions of this type. Linear regression can be applied to the resulting plots and goodness of fit (R^2) and the slope (γ)

Table 1. Type-Token data for 23 VACs drawn from COBUILD Verb Patterns retrieved from the BNC

Construction	Types	Tokens	TTR	Lead verb type
V about N	365	3519	10.37	*talk*
V across N	799	4889	16.34	*come*
V after N	1168	7528	15.52	*look*
V among pl-*N*	417	1228	33.96	*find*
V around N	761	3801	20.02	*look*
V as adj	235	1012	23.22	*know*
V as N	1702	34383	4.95	*know*
V at N	1302	9700	13.42	*look*
V between pl-*N*	669	3572	18.73	*distinguish*
V for N	2779	79894	3.48	*look*
V in N	2671	37766	7.07	*find*
V into N	1873	46488	4.03	*go*
V like N	548	1972	27.79	*look*
V N N	663	9183	7.22	*give*
V off N	299	1032	28.97	*take*
V of N	1222	25155	4.86	*think*
V over N	1312	9269	14.15	*go*
V through N	842	4936	17.06	*go*
V to N	707	7823	9.04	*go*
V towards N	190	732	25.96	*move*
V under N	1243	8514	14.6	*come*
V way prep	365	2896	12.6	*make*
V with N	1942	24932	7.79	*deal*

recorded. Figure 1 shows such a plot for verb type frequency of the *V across N* construction pattern extracted from the BNC grouping types into 20 logarithmic bins according to their frequency. Each point represents one bin and a verb from each group is randomly selected to label the point with its token frequency in parentheses. For example, the type *look* occurs 102 times in the *V across N* pattern and is placed into the 15th bin with the types *go*, *lie* and *lean*. Points towards the lower right of the plot indicate high-frequency low-type groupings and those towards the top left low-frequency high-type groupings, that is the fat- or long-tail of the distribution.

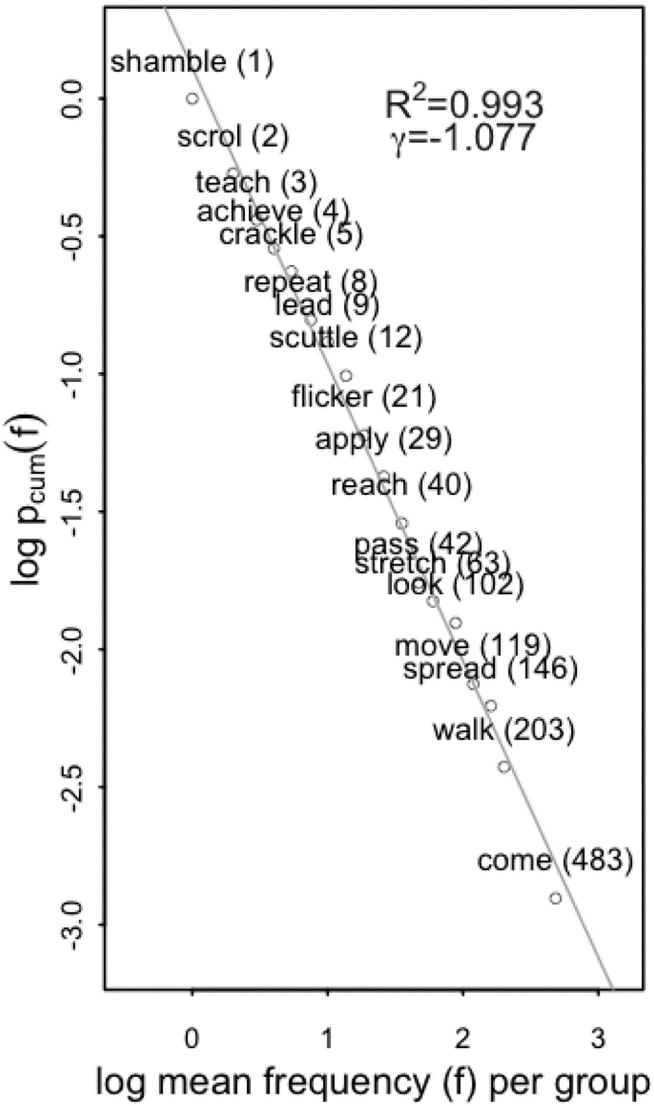

Figure 1. Verb type distribution for *V across N*

Figure 2 shows such the same type of plot for verb type frequency of the ditransitive *V N N* construction pattern extracted and binned in the same way. Both distributions produce a good fit ($R^2 > 0.99$) with a straight regression line, indicating a Zipfian type-token frequency distributions

(V9 5/4/12 14:41) WDG-LCB (155mm×230mm) TimesNRMT 1382 Rebuschat pp. 265–304 1382 Rebuschat_09_pisoni (p. 278)

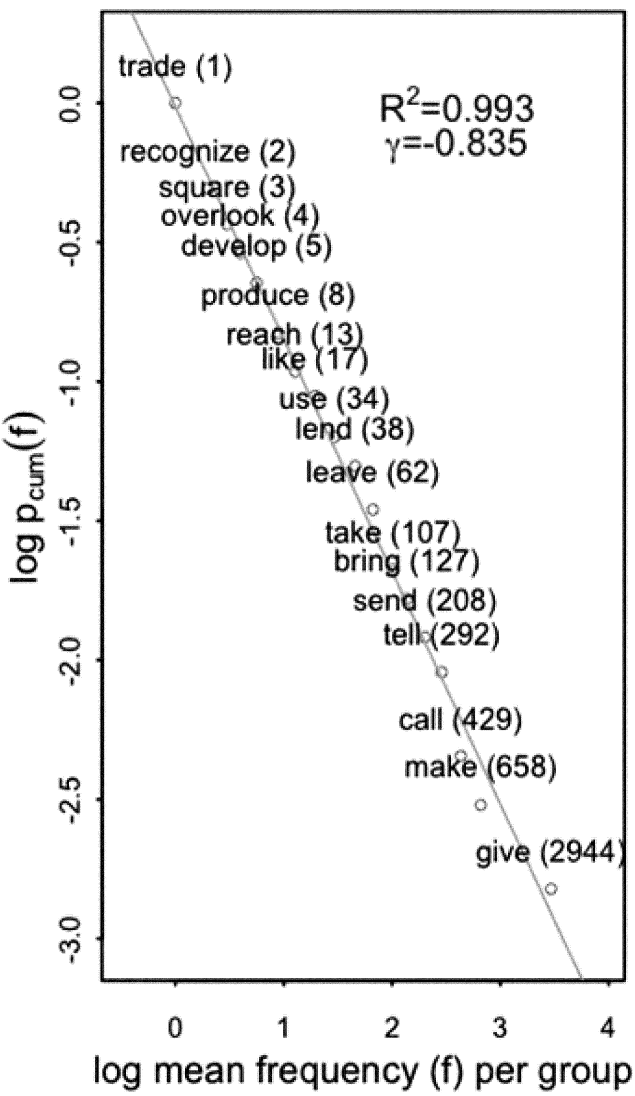

Figure 2. Verb type distribution for *V N N*

for these constructions. Inspection of the construction verb types, from most frequent down, also demonstrates that, as in prior research (Ellis & Ferreira-Junior, 2009b; Goldberg, et al., 2004; Ninio, 1999, 2006), the lead member is prototypical of the construction and generic in its action semantics.

If Zipf's law applies across language, then any sample of language will be Zipfian-distributed, rendering such findings potentially trivial (we elaborate on this in *Step 7*). But they become much more interesting if the company of verb forms occupying a construction is selective, i.e. if the frequencies of the particular VAC verb members cannot be predicted from their frequencies in language as a whole. We measure the degree to which VACs are selective like this using a variety of measures including a chi-square goodness-of-fit test, and the statistic '1-tau' where Kendall's tau measures the correlation between the rank verb frequencies in the construction and in language as a whole. Higher scores on both of these metrics indicate greater VAC selectivity. Another useful measure is Shannon entropy for the distribution. Entropy is a measure of the uncertainty associated with a random variable – it is affected by the number of types in the system and the distribution of the tokens of the types. If there is just one type, then the system is far from random, and entropy is low. If there are ten types of equal probability, the system is quite random, but if 99% of the tokens are of just one type, it is far less random, and so on. The lower the entropy the more coherent the VAC verb family. Construction scores on all these measures are given later in Table 4.

4.5. Step 5 Determining the Contingency between Verbs and VACs

Some verbs are closely tied to a particular construction (for example, *give* is highly indicative of the ditransitive construction, whereas *leave*, although it can form a ditransitive, is more often associated with other constructions such as the simple transitive or intransitive). As we described above, the more reliable the contingency between a cue and an outcome, the more readily an association between them can be learned (Shanks 1995), so constructions with more faithful verb members are more transparent and thus should be more readily acquired (Ellis 2006). The measures of contingency that we adopt here are (1) faithfulness – the proportion of tokens of total verb usage that appear this particular construction (e.g., the faithfulness of *give* to the ditransitive is approximately 0.40; that of *leave* is 0.01, (2) directional one-way associations, contingency (ΔP Construction $\rightarrow$ Word: *give* 0.314, *leave* 0.003) and (ΔP Word $\rightarrow$ Construction: *give* 0.025, *leave* 0.001) (e.g. Ellis & Ferreira-Junior, 2009), and (3) directional mutual information (MI Word $\rightarrow$ Construction: *give* 16.26, *leave* 11.73 and MI Construction $\rightarrow$ Word: *give* 12.61 *leave* 9.11), an information science statistic that has been shown to predict language processing fluency (e.g., Ellis, Simpson-Vlach, and Maynard 2008; Jurafsky

(V9 5/4/12 14:41) WDG-LCB (155mm×230mm) TimesNRMT 1382 Rebuschat pp. 265–304 1382 Rebuschat_09_pisoni (p. 280)

Table 2. Top 20 verbs found in the *V across N* construction pattern in the BNC

Verb	Constr. Freq.	Corpus Freq.	Faith.	Token* Faith	MI word → constr	MI constr → word	ΔP word → constr	ΔP constr → word
come	474	122107	0.0039	1.840	15.369	10.726	0.004	0.089
walk	203	17820	0.0114	2.313	16.922	15.056	0.011	0.040
cut	197	16200	0.0122	2.396	17.016	15.288	0.012	0.039
run	175	36163	0.0048	0.847	15.687	12.800	0.005	0.034
spread	146	5503	0.0265	3.874	18.142	17.971	0.026	0.030
move	114	34774	0.0033	0.374	15.125	12.295	0.003	0.021
look	102	93727	0.0011	0.111	13.534	9.273	0.001	0.015
go	93	175298	0.0005	0.049	12.498	7.333	0.000	0.008
lie	80	18468	0.0043	0.347	15.527	13.610	0.004	0.015
lean	75	4320	0.0174	1.302	17.530	17.708	0.017	0.015
stretch	62	4307	0.0144	0.893	17.260	17.442	0.014	0.012
fall	57	24656	0.0023	0.132	14.621	12.287	0.002	0.010
get	52	146096	0.0004	0.019	11.922	7.020	0.000	0.002
pass	42	18592	0.0023	0.095	14.588	12.661	0.002	0.007
reach	40	21645	0.0018	0.074	14.298	12.152	0.002	0.007
travel	39	8176	0.0048	0.186	15.666	14.924	0.004	0.007
fly	38	8250	0.0046	0.175	15.616	14.861	0.004	0.007
stride	38	1022	0.0372	1.413	18.629	20.887	0.037	0.008
scatter	35	1499	0.0233	0.817	17.957	19.663	0.023	0.007
sweep	34	2883	0.0118	0.401	16.972	17.734	0.011	0.007

2002). Table 2 lists some of these contingency measures for the verbs occupying the *V across N* VAC pattern.

It is instructive to reorder the distribution according to these measures and consider the top items in terms of how characteristic of the VAC semantics they are (this is a standard option for each VAC listed on the website we are developing to allow open-access to our analyses). For the *V across N* VAC pattern, more generic movement verbs *come*, *walk*, *cut*, *run*, *spread* and *move* top the list ordered by token frequency. But when ordered according to verb to construction faithfulness, the items are much more specific in their meaning, though of low frequency: *scud*, *skitter*, *sprawl*, *flit*, *emblazon* and *slant*. The average faithfulness, MI and ΔP scores across the members of the construction are also important metrics, illustrating the degree to which VACs are selective in their membership. We show examples later in Table 4.

4.6. Step 6 Identifying the Meaning of Verb Types Occupying the Constructions

We are investigating several ways of analyzing verb semantics. Because our research aims to empirically determine the semantic associations of particular linguistic forms, ideally the semantic classes we employ should be defined in a way that is free of linguistic distributional information, otherwise we would be building in circularity. Therefore distributional semantic methods such as Latent Semantic Analysis (LSA, Landauer et al. 2007) are not our first choice here. Instead, here we utilize WordNet, a distribution-free semantic database based upon psycholinguistic theory which has been in development since 1985 (Miller 2009). WordNet places words into a hierarchical network. At the top level, the hierarchy of verbs is organized into 559 distinct root synonym sets ('synsets' such as *move1* expressing translational movement, *move2* movement without displacement, etc.) which then split into over 13,700 verb synsets. Verbs are linked in the hierarchy according to relations such as hypernym [verb Y is a hypernym of the verb X if the activity X is a (kind of) Y (to *perceive* is an hypernym of to *listen*], and hyponym [verb Y is a hyponym of the verb X if the activity Y is doing X in some manner (to *lisp* is a hyponym of to *talk*)]. Various algorithms to determine the semantic similarity between WordNet synsets have been developed which consider the distance between the conceptual categories of words, as well as considering the hierarchical structure of the WordNet (Pedersen, Patwardhan, and Michelizzi 2004).

Polysemy is a significant issue of working with lexical resources such as WordNet, particularly when analyzing verb semantics. For example, in WordNet the lemma forms *move*, *run* and *give* used as verbs are found in 16, 41 and 44 different synsets respectively. To address this we have applied word sense disambiguation tools specifically designed to work with WordNet (Pedersen and Kolhatkar 2009) to the sentences retrieved at *Step 3*.

4.7. Step 7 Generating Distributionally-Matched, Control Ersatz Constructions (CECs)

Miller (1965) in his preface to the MIT Press edition of Zipf's (1935) *Psychobiology of Language* claimed that Zipfian type-token frequency distributions are essentially uninteresting artifacts of language in use rather than important factors in acquisition. His "monkey at the typewriter" (1957) word generation model produces random words of arbitrary average length as follows: With a probability *s*, a word separator is generated at each step, with probability $(1 - s)/N$, a letter from an alphabet of size N is generated, each letter having the same probability. That the monkey at the typewriter model produces gibberish that is Zipfian well-distributed thence rendered Zipf's law uninteresting for linguistics for several decades (see also Manning and Schütze 1999). Li (1992) reawakened the issue with further demonstrations that random texts exhibit Zipf's law-like word frequency distributions. Ferrer-i-Cancho and Solé (2002) responded by showing that random texts lose the Zipfian shape in the frequency versus rank plot when words are restricted to a certain length, which is not the case in real texts. As they conclude: "By assuming that Zipf's law is a trivial statistical regularity, some authors have declined to include it as part of the features of language origin. Instead, it has been used as a given statistical fact with no need for explanation. Our observations do not give support to this view." Nevertheless Yang (2010) claims that item/usage-based approaches to language acquisition, which typically make use of the notion of constructions, have failed to amass sufficient empirical evidence and to apply the necessary statistical analysis to support their conclusions. He asserts that it is the Zipfian nature of language itself ('the sparse data problem') that gives rise to *apparent* item-specific patterns. In response to these possibilities, for every VAC we analyze, we generate a distributionally-yoked control which is matched for type-token distribution but otherwise randomly selected to be grammatically and semantically uninformed. We refer to these distributions as 'control ersatz constructions' (CECs). We then assess, using paired-sample tests, the degree to which VACs are more coherent than expected by chance in terms of the association of their grammatical form and semantics. We show such comparisons for illustration VACs and their yoked CECs later in Tables 4, 5 and 6.

The goal in generating CECs is to produce a distribution with the same number of types and tokens as the VAC. To do this we use the following method. For each type in a distribution derived from a VAC pattern (e.g. *walk* in *V across N* occurs 203 times), ascertain its corpus frequency (*walk* occurs 17820 times in the BNC) and randomly select a replacement type from the list of all verb types in the corpus found within the same frequency band (e.g. from *learn*, *increase*, *explain*, *watch*, *stay*, etc. which occur with similar frequencies to *give* in the BNC). This results in a matching number of types that reflect the same general frequency profile as those from the VAC. Then, using this list of replacement types, sample the same number of tokens (along with their sentence contexts) as in the VAC distribution (e.g. 4889 for *V across N*) following the probability distribution of the replacement types in the whole corpus (e.g. *walk*, with a corpus frequency of 17820, will be sampled roughly twice as often as *extend*, which occurs 9290 times). The resulting distribution has an identical number of types and tokens its matching VAC, although, if the VAC does attract particular verbs, the lead members of the CEC distribution will have a token frequency somewhat lower than those in the VAC.

4.8. Step 8 Evaluating Semantic Cohesion in the VAC Distributions

We have suggested that an intuitive reading of VAC type-token lists such as in Table 2 shows that the tokens list captures the most general and prototypical senses (*come*, *walk*, *move* etc. for *V across N* and *give*, *make*, *tell*, *offer* for *V N N*), while the list ordered by faithfulness highlights some quite construction specific (and low frequency) items, such as *scud*, *flit* and *flicker* for *V across N*. Using the structure of the verb component of the WordNet dictionary, where each synset can be traced back to a root or top-level synset, we are able to compare the semantic cohesion of the top 20 verbs, using their disambiguated WordNet senses, from a given VAC to its matching CEC. So for each verb in a VAC or CEC we query the database for the disambiguated WordNet senses for the verb in the instance sentences. For example, in *V across N*, the verb type *move* occurs 114 times across 5 synsets: move.v.1 (86x), move.v.2 (18x), move.v.3 (5x), move.v.7 (1x) and move.v.8 (4x). Each of these synsets can be traced back to a top or root level synset or may itself be that synset: move.v.1 → travel.v.1, move.v.2 → move.v.2, move.v.3 → move.v.3, move.v.7 → change.v.3, move.v.8 → act.v.1. Table 3 shows this for the *V across N* VAC pattern, where the synsets come.v.1, walk.v.1, run.v.1, move.v.1, go.v.1, fall.v.2, pass.v.1, travel.v.1, stride.v.1, stride.v.2 account for 744 of the 4889 (15%) tokens, and share the top level hypernym synset travel.v.01. In comparison, the most frequent root synset for the matching CEC, pronounce.v.1, accounts for just 4% of the tokens. The VAC has a much more compact semantic distribution, in that 5 top level synsets account for a third of the tokens compared to the 21 required to account for the same proportion for the CEC

We have explored two methods of evaluating the differences between the semantic sense distributions, such as the one in Table 3, for each VAC-CEC pair. First, we can measure the amount of variation in the distribution (i.e. its compactness) using Shannon entropy as we did in Step 4. For these semantic distributions this can be done according to (1) number of sense types per root (*V across N* VAC: 2.75 CEC: 3.37) (so ignoring the token frequency column in Table 3) and (2) the token fre-

Table 3. Disambiguated WordNet senses for the top 20 verbs found in the *V across N* VAC and yoked CEC distributions from the BNC and the root verb synsets to which they belong (Top 12 root synsets shown for VAC and CEC).

Actual *V across N* VAC distribution				**Random *V across N* CEC distribution**			
Root verb synset	Specific WordNet senses	Freq.	Cum. %	Root verb synset	Specific WordNet senses	Freq.	Cum. %
travel.v.01	come.v.1, walk.v.1, run.v.1, move.v.1, go.v.1, fall.v.2, pass.v.1, travel.v.1, stride.v.1, stride.v.2	744	15	pronounce.v.01	say.v.6	193	04
be.v.03	come.v.9, run.v.3, go.v.7, lie.v.1, stretch.v.1, pass.v.6, reach.v.6, sweep.v.5, sweep.v.8	259	21	be.v.01	make.v.31, go.v.10, go.v.6, take.v.38, come.v.14, look.v.2, need.v.2, work.v.14, seem.v.1	183	08
be.v.01	come.v.12, come.v.14, cut.v.25, run.v.12, look.v.2, lie.v.4, lean.v.3, fall.v.16, fall.v.4, get.v.33	233	25	travel.v.01	go.v.1, come.v.1	154	11
move.v.02	cut.v.1, run.v.26, move.v.2, lean.v.2, fly.v.4	210	30	make.v.03	make.v.3, make.v.5, see.v.4, give.v.13, think.v.5, work.v.11	123	13

(V9 5/4/12 14:41) WDG-LCB (155mm×230mm) TimesNRMT 1382 Rebuschat pp. 265–304 1382 Rebuschat_09_pisoni (p. 285)

Actual *V across N* VAC distribution				**Random *V across N* CEC distribution**			
change.v.02	come.v.4, cut.v.39, run.v.38, run.v.39, spread.v.4, go.v.4, lean.v.1, stretch.v.3, stretch.v.9, fall.v.26, fall.v.3, get.v.12, get.v.2, pass.v.18, fly.v.7	198	34	think.v.03	see.v.5, know.v.6, give.v.10, think.v.1, think.v.2, think.v.3, try.v.2	100	15
spread.v.01	spread.v.1, scatter.v.3	172	37	move.v.02	say.v.5, set.v.1, put.v.1	84	17
move.v.03	cut.v.14, run.v.6, spread.v.2, move.v.3, stretch.v.11, reach.v.3, sweep.v.2	106	39	transfer.v.05	give.v.17, give.v.3	84	19
get.v.01	run.v.36, get.v.1	41	40	understand.v.01	see.v.24, take.v.6, work.v.24	83	21
touch.v.01	fly.v.3	30	41	know.v.01	know.v.1	73	22
reach.v.01	reach.v.1	26	41	use.v.01	give.v.18, use.v.1, work.v.23, put.v.4	73	24
guide.v.05	sweep.v.3	14	42	remove.v.01	take.v.17	72	25
happen.v.01	come.v.19, come.v.3, pass.v.8	14	42	change.v.02	make.v.30, go.v.17, go.v.30, go.v.4, see.v.21, see.v.3, know.v.5, take.v.5, come.v.4, give.v.26, find.v.12, leave.v.8	67	26

quency per root (*V across N* VAC: 2.08 CEC: 3.08), the lower the entropy the more coherent the VAC verb semantics. These figures are calculated for all 23 VACs and CECs and shown in Tables 4 and 5 as (1) Type entropy per root synset and (2) Token entropy per root synset. Secondly, we can develop the observation for the distribution in Table 3 that the top three root synsets, in the VAC account for 25% (1236) of the tokens compared to 11% (530) for the CEC. Third, we quantify the semantic coherence or 'clumpiness' of the disambiguated senses for the top 20 verb forms in the VAC and CEC distributions using measures of semantic similarity from WordNet and Roget's. Pedersen et al (2004) outline six measures in their Perl WordNet::Similarity package, three (*path, lch* and *wup*) based on the path length between concepts in WordNet Synsets and three (*res*, *jcn* and *lin*) that incorporate a measure called 'information content' related to concept specificity. For instance, using the *res* similarity measure (Resnik 1995) the top 20 verbs in *V across N* VAC distribution have a mean similarity score of 0.353 compared to 0.174 for the matching CEC.

5. Results

Our core research questions concern the degree to which VAC form, function, and usage promote robust learning. As we explained in the theoretical background, the psychology of learning as it relates to these psycholinguistic matters suggests, in essence, that learnability will be optimized for constructions that are (1) Zipfian in their type-token distributions in usage, (2) selective in their verb form occupancy, (3) coherent in their semantics. Their values on the metrics we have described so far are illustrated for the 23 VACs in Table 4 along with those for their yoked CECs in Table 5.

Table 6 contrasts between the VACs and the CECs on these measures as the results of paired-sample t-tests.

The results demonstrate:

5.1. Type-token Usage Distributions

All of the VACs are Zipfian in their type-token distributions in usage (VACs: $M\gamma = -1.00$, $MR^2 = 0.98$). So too are their matched CECs ($M\gamma = -1.12$, $MR^2 = 0.96$). The fit is slightly better for the VACs than the CECs because the yoked-matching algorithm tends to make the topmost types of the CEC somewhat less extreme in frequency than is found in the real VACs (because particular verbs are attracted to particular

Table 4. Values for our 23 Verb Argument Constructions on metrics of Zipfian distribution, verb form selectivity, and semantic coherence.

VAC Pattern	R^2	γ	Entropy	$\chi 2$	$1\text{-}\tau$	Mean $MI_{w\text{-}c}$	Mean $\Delta P_{c\text{-}w}$	Type entropy per root synset	Token entropy per root synset	Proportion of tokens covered by top 3 synsets	*lch*	*res*
V about N	0.98	−0.80	3.79	29919	0.74	15.55	0.011	3.17	2.42	0.45	0.162	0.271
V across N	0.99	−1.08	5.30	23324	0.77	15.49	0.003	2.75	2.08	0.25	0.194	0.353
V after N	0.99	−1.04	5.04	48065	0.69	12.87	0.002	3.33	2.12	0.31	0.103	0.184
V among pl-*N*	0.99	−1.43	5.36	9196	0.77	17.51	0.009	2.93	2.79	0.11	0.096	0.174
V around N	0.97	−1.17	5.51	40241	0.77	15.96	0.004	2.80	2.43	0.19	0.155	0.284
V as adj	0.96	−0.98	4.05	8993	0.76	17.88	0.020	3.20	2.48	0.34	0.078	0.141
V as N	0.99	−0.80	4.84	184085	0.87	10.36	0.003	3.55	2.56	0.25	0.079	0.146
V at N	0.97	−1.02	4.94	66633	0.79	12.51	0.003	3.23	1.72	0.36	0.099	0.185
V between pl-*N*	0.98	−1.08	5.17	47503	0.80	15.18	0.005	3.11	2.61	0.21	0.078	0.149
V for N	0.97	−0.79	5.58	212342	0.73	9.54	0.002	3.38	2.70	0.16	0.117	0.198
V in N	0.96	−0.96	6.22	61215	0.72	10.48	0.002	3.56	2.90	0.10	0.079	0.138
V into N	0.98	−0.82	5.22	82396	0.71	11.44	0.003	3.21	2.39	0.26	0.168	0.289
V like N	0.98	−1.08	4.80	12141	0.66	15.84	0.009	2.99	1.92	0.34	0.121	0.216
V N N	0.99	−0.84	3.79	51652	0.66	11.52	0.004	3.21	2.38	0.41	0.139	0.236
V off N	0.98	−1.29	4.89	10101	0.60	17.84	0.011	2.64	2.46	0.21	0.198	0.358
V of N	0.97	−0.76	4.26	319284	0.88	11.15	0.003	3.31	2.56	0.33	0.11	0.189
V over N	0.98	−1.08	5.95	77407	0.87	13.72	0.002	2.87	2.33	0.17	0.237	0.404
V through N	0.99	−1.11	5.37	29525	0.83	14.84	0.003	3.05	2.10	0.26	0.147	0.266
V to N	0.95	−0.92	5.02	25729	0.72	13.50	0.003	2.88	2.59	0.19	0.189	0.325
V towards N	0.98	−1.16	4.36	15127	0.78	19.59	0.017	2.68	2.35	0.31	0.149	0.274
V under N	0.97	−1.10	5.74	19244	0.70	13.13	0.002	3.07	2.54	0.16	0.14	0.248
V way prep	0.99	−0.83	3.61	29827	0.81	17.26	0.013	3.27	2.46	0.39	0.105	0.194
V with N	0.98	−0.96	5.59	192521	0.81	12.56	0.003	3.16	2.50	0.18	0.136	0.231
Mean	0.98	−1.00	4.97	69412	0.76	14.16	0.006	3.10	2.41	0.26	0.134	0.237

Table 5. Values for our 23 Control Ersatz Constructions on metrics of Zipfian distribution, verb form selectivity, and semantic coherence.

VAC Pattern	R^2	γ	Entropy	$\chi 2$	1-τ	Mean $MI_{w\text{-}c}$	Mean $\Delta P_{c\text{-}w}$	Type entropy per root synset	Token entropy per root synset	Proportion of tokens covered by top 3 synsets	*lch*	*res*
V about N	0.94	–1.04	4.80	441	0.17	14.02	0.004	3.52	3.07	0.15	0.084	0.152
V across N	0.96	–1.14	5.55	232	0.19	13.29	0.003	3.37	3.08	0.11	0.098	0.174
V after N	0.97	–1.21	5.95	222	0.22	12.05	0.002	3.65	3.11	0.09	0.083	0.146
V among pl-*N*	0.99	–1.45	5.39	867	0.35	15.25	0.006	3.42	3.10	0.10	0.068	0.123
V around N	0.96	–1.17	5.57	366	0.22	13.71	0.003	3.51	3.14	0.10	0.093	0.16
V as adj	0.96	–1.26	4.74	1232	0.31	15.97	0.010	2.90	2.84	0.16	0.165	0.286
V as N	0.94	–0.99	5.98	203	0.14	15.97	0.010	3.64	3.07	0.09	0.088	0.154
V at N	0.96	–1.18	6.01	248	0.21	11.67	0.002	3.53	3.11	0.08	0.083	0.151
V between pl-*N*	0.97	–1.19	5.46	329	0.26	13.52	0.003	3.33	3.04	0.11	0.092	0.149
V for N	0.93	–0.94	6.22	205	0.10	8.42	0.001	3.83	3.12	0.08	0.075	0.198
V in N	0.96	–1.05	6.33	228	0.15	9.47	0.001	3.70	3.07	0.09	0.082	0.138
V into N	0.94	–0.86	5.81	225	0.11	9.62	0.002	3.71	3.12	0.10	0.088	0.289
V like N	0.95	–1.32	5.46	678	0.24	14.57	0.004	3.38	3.10	0.12	0.083	0.216
V N N	0.94	–1.05	5.48	226	0.20	11.98	0.002	3.53	3.10	0.10	0.093	0.236
V off N	0.96	–1.23	4.89	3853	0.28	16.22	0.010	3.42	3.20	0.13	0.072	0.358
V of N	0.93	–0.97	5.70	193	0.13	10.38	0.002	3.72	3.08	0.10	0.078	0.189
V over N	0.97	–1.18	5.99	264	0.22	11.79	0.002	3.60	3.07	0.09	0.081	0.404
V through N	0.97	–1.19	5.66	293	0.24	12.98	0.002	3.51	3.21	0.09	0.099	0.266
V to N	0.93	–1.05	5.46	253	0.18	12.48	0.003	3.52	3.11	0.10	0.091	0.325
V towards N	0.98	–1.14	4.52	3414	0.25	17.08	0.015	3.15	2.84	0.20	0.085	0.274
V under N	0.97	–1.18	5.99	237	0.24	11.82	0.002	3.59	3.22	0.08	0.14	0.248
V way prep	0.95	–0.90	4.32	1628	0.21	11.82	0.002	3.57	2.96	0.22	0.105	0.194
V with N	0.95	–1.05	6.07	220	0.16	10.22	0.002	3.57	3.13	0.09	0.136	0.231
Mean	0.96	–1.12	5.54	698	0.21	12.80	0.004	3.51	3.08	0.11	0.094	0.22

Table 6. Comparisons of values for our 23 VACs and CECs on metrics of Zipfian distribution, verb form selectivity, and semantic coherence.

Pattern	R^2	γ	Entropy	χ^2	1-τ	Mean $MI_{w\text{-}c}$	Mean $\Delta P_{c\text{-}w}$	Type entropy per root synset	Token entropy per root synset	Proportion of tokens covered by top 3 synsets	*lch*	*res*
Mean VACs	0.98	−1.00	4.97	69412	0.76	14.16	0.006	3.10	2.41	0.26	0.134	0.237
Mean CECs	0.96	−1.12	5.54	698	0.21	12.80	0.004	3.51	3.08	0.11	0.094	0.22
p value for paired t-test (d.f. 22)	1.6 e-06 ***	4.4 e-06 ***	4.89 e-04 ***	5.5 e-18 ***	1.9 e-03 ***	1.1 e-02 ***	5.1 e-05 ***	1.7 e-08 ***	1.2 e-10 ***	3.2 e-08 ***	2.0 e-04 ***	1.6 e-06 ***

VACs), and so the fit line is not pulled out into so extreme a tail. Inspection of the graphs for each of the 23 VACs shows that the highest frequency items take the lion's share of the distribution and, as in prior research (Ellis & Ferreira-Junior, 2009b; Goldberg, et al., 2004; Ninio, 1999, 2006), the lead member is prototypical of the construction and generic in its action semantics (see the rightmost column in Table 1).

5.2. Family Membership and Type Occupancy

VACs are selective in their verb form family occupancy. There is much less entropy in the VACs than the CECs, with fewer forms of a less evenly-distributed nature (*M* distribution Entropy VAC 4.97, CEC 5.54, $p < .0001$). The distribution deviation from verb frequency in the language as a whole is much greater in the VACs than the CECs ($M\,\chi^2$ VAC 69411, CEC 698, $p < .0001$). The lack of overall correlation between VAC verb frequency and overall verb frequency in the language is much greater in the VACs ($M\ 1 - \tau$ VAC 0.76, CEC 0.21, $p < .002$).

Individual verbs select particular constructions (M $\mathrm{MI}_{\text{w-c}}$ VAC 14.16, CEC 12.80, $p < .01$) and particular constructions select particular words (M $\Delta P_{\text{c-w}}$ VAC 0.006, CEC 0.004, $p < .0001$). Overall then, there is greater contingency between verb types and constructions.

5.3. Semantic Coherence

VACS are coherent in their semantics with lower type (*M* VAC 3.10, CEC 3.51, $p < .0001$) and token (*M* VAC 2.41, CEC 3.08, $p < .0001$) sense entropy. Figure 3 shows distributions of the root synsets for the top 20 types of each of the VAC-CEC pairs through plots of logarithmic token frequency against rank – in each pair, fewer senses cover more of the VAC uses than the CEC. Figure 3 also shows the proportion of tokens accounted for by the top three root synsets (e.g. for *V across N*: VAC 0.25 CEC 0.11). The proportion of the total tokens covered by their three most frequent WordNet roots is much higher in the VACs (*M* VAC 0.26, CEC 0.11, $p < .0001$). Finally, the VAC distributions are higher on the Pedersen semantic similarity measures (*M lch* VAC 0.13, CEC 0.09, $p < .0002$) (M *res* VAC 0.24, CEC 0.22, $p < .0001$).

6. Discussion

Twenty-three constructions is a better sample of constructions than three, and the 16,141,058 tokens of verb usage analyzed here is a lot more representative than the 14, 474 analyzed in Ellis & Ferreira-Junior (2009a,b).

(V9 5/4/12 14:41) WDG-LCB (155mm×230mm) TimesNRMT 1382 Rebuschat pp. 265–304 1382 Rebuschat_09_pisoni (p. 291)

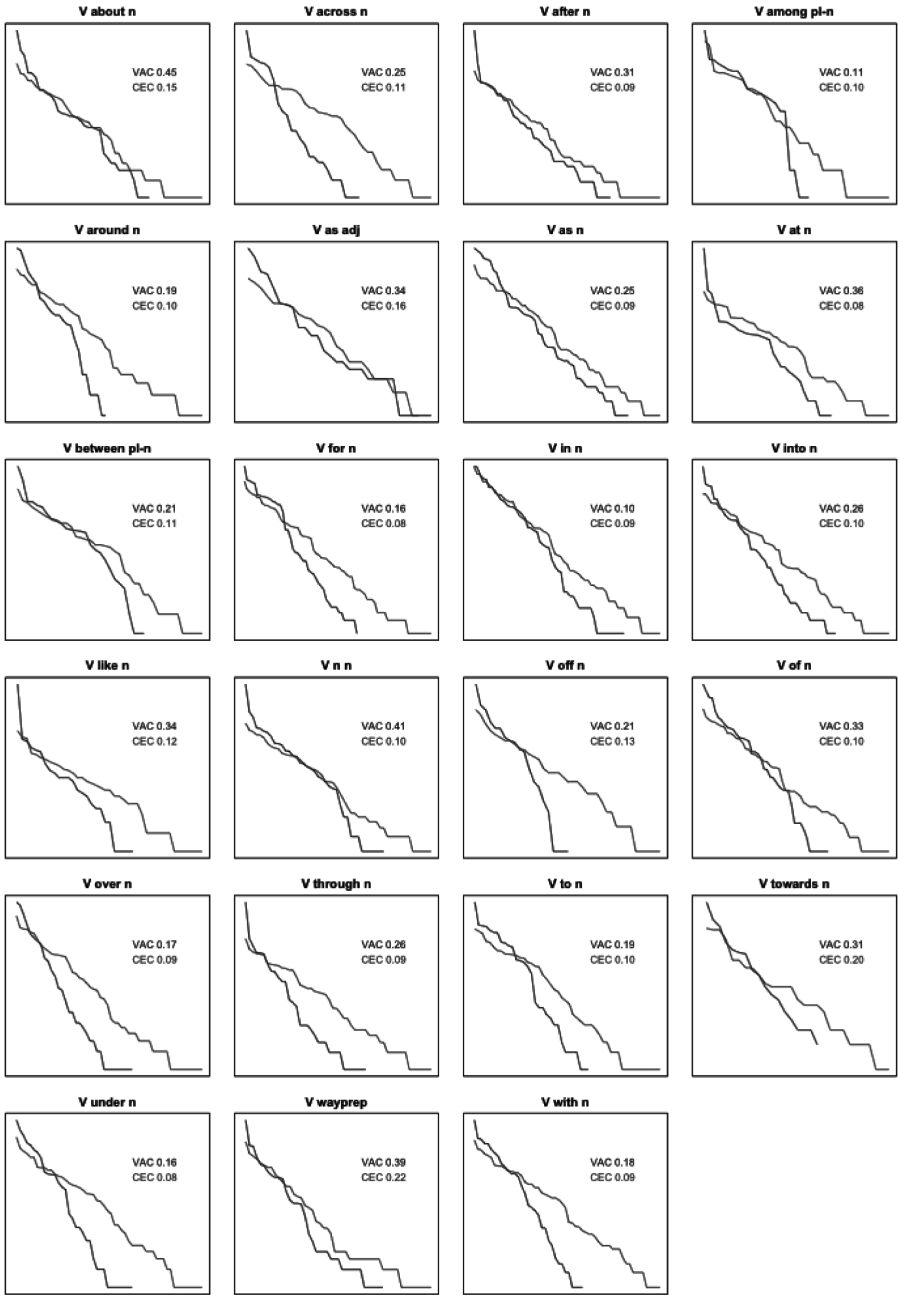

Figure 3. Distribution of WordNet root verb synsets for VACs and CECs

Nevertheless, the conclusions from those earlier studies seem to generalize. We have shown:

- The frequency distribution for the types occupying the verb island of each VAC are Zipfian.
- The most frequent verb for each VAC is much more frequent than the other members, taking the lion's share of the distribution.
- The most frequent verb in each VAC is prototypical of that construction's functional interpretation, albeit generic in its action semantics.
- VACs are selective in their verb form family occupancy:
  - Individual verbs select particular constructions.
  - Particular constructions select particular verbs.
  - There is greater contingency between verb types and constructions.
- VACS are coherent in their semantics.

Psychology theory relating to the statistical learning of categories suggests that these are the factors which make concepts robustly learnable. We suggest, therefore, that these are the mechanisms which make linguistic constructions robustly learnable too, and that they are learned by similar means.

7. Future Work

7.1. An Exhaustive Inventory of English VACs

This is still a small sample from which to generalize. In subsequent work we intend to analyze the 700+ patterns of *Verb Pattern Grammar* volume as found in the 100 million words of the BNC. Other theories of construction grammar start from different motivations, some more semantic [e.g. Framenet (Fillmore, Johnson, and Petruck 2003) and VerbNet (Kipper et al. 2008; Palmer 2010; Levin 1993)], some alternatively syntactic [e.g. the Erlangen Valency Patternbank (Herbst and Uhrig 2010; Herbst et al. 2004)], and so present different, complementary descriptions of English verb grammar. Given time, we hope to analyze usage patterns from these descriptions too. We are particularly interested in whether these inventories represent optimal partitioning of verb semantics, starting with basic categories of action semantics and proceeding to greater specificity via Zipfian mapping.

7.2. Learner Language

We are also interested in extending these approaches to learner language to investigate whether first and L2 learners' acquisition follows the same construction distributional profiles. We have done some initial pilot work to test the viability of our methods by extracting 18 of the same VAC patterns from American English and British English child language acquisition corpora in CHILDES (MacWhinney 2000, 2000) transcripts. Child directed speech (CDS, over 6.8 million words) was separated from the speech of the target child (over 3.6 million words) for the UK and USA components of the database where dependency parsing of each utterance is available (Sagae et al. 2007). The same analysis steps described here are equally viable with learner language. In our initial explorations (O'Donnell and Ellis submitted) we build on the types of analysis carried out in Goldberg, Casenhiser & Sethuraman (2004) that demonstrate how the frequency profiles of CDS are reproduced in child language. For example, for the *V across N* VAC pattern *go* is the most frequent type in both CDS and child speech. Likewise, for *V over N* we found *go* and *jump* as the first types in both samples. For *V with N* the top 4 types, *play*, *go*, *do*, *come*, are shared, as they are for *V under N*: *go*, *look*, *get*, *hide* and the top two for *V like N*: *look* and *go*. The nature of CDS with respect to more general English can also be examined. Applying the various contingency and semantic measures discussed above we found the 10 most faithful types to the VAC pattern *V like N* were: 1) from the BNC: *glitter*, *behave*, *gleam*, *bulge*, *shape*, *flutter*, *glow*, *shine*, *sound*, *sway* (with a *wup* similarity score of 0.3559) and 2) for CDS: *sound*, *act*, *shape*, *smell*, *taste*, *look*, *yell*, *feel*, *talk*, *fit* (*wup* 0.4564). This initial analysis points both to the more frequent use of generic verbs (e.g. *go* and *do*) in CDS and a tighter semantic coherence in the items most associated with specific VACs. These steps need next to be done for the complete inventory of VACs so that a comparison can be made of general usage (BNC), CDS, and child language acquisition at different stages.

7.3. Determinants of Learning

Once we have these parallel datasets of sufficient scale, we can undertake a principled empirical analysis of the degree to which the psychological factors outlined really do determine acquisition. For each VAC in the input we will have the data relating to frequency, distributional, contingency, and semantic factors which learning theory considers important in acquisition. With the staged child language acquisition analyzed in the same way, we can test out these predictions and explore how the different factors conspire in the emergence of language.

7.4. Modeling Acquisition

As we have argued in an upcoming review of statistical corpus linguistics and language cognition (Ellis in press), the field as a whole needs to work on how to combine the various corpus metrics that contribute to learnability into a model of acquisition rather than a series of piecemeal univariate snapshots. We have developed some connectionist methods for looking at this and trialed them with just the three VACs VL, VOL, and VOO (Ellis and Larsen-Freeman 2009), but that enterprise and the current one are of hugely different scales. We need models of acquisition that relate such VAC measures as applied to the BNC and CDS to longitudinal patterns of child language and L2 acquisition.

8. Conclusion

This research shows some promise towards an English verb grammar operationalized as an inventory of VACs, their verb membership and their type-token frequency distributions, their contingency of mapping, and their semantic motivations. Our initial analyses show that constructions are (1) Zipfian in their type-token distributions in usage, (2) selective in their verb form occupancy, and (3) coherent in their semantics. Psychology theory relating to the statistical learning of categories suggests that these are the factors which make concepts robustly learnable. We suggest therefore, that these are the mechanisms which make linguistic constructions robustly learnable too, and that they are learned by similar means.

9. Epilogue

Phoebe was a black and brindle collie-cross (Figure 4). She was 12 years old when we brought her to (VOL_{to}) the US. It was Michigan, February, blue skies over 12″ of snow. We collected her, dehydrated, from (VOL_{from}) DTW, left the airport, and pulled onto (VL_{onto}) the nearest safe verge to let her out (VOL_{out}) of her travel-kennel. It had been a long flight and we were somewhat concerned, but after a typically warm reunion, she looked at (VL_{at}) the strange whiteness, and then, like a wolf pouncing on (VL_{on}) a mouse, she ponked into (VL_{into}) the snow.

(V9 5/4/12 14:41) WDG-LCB (155mm×230mm) TimesNRMT 1382 Rebuschat pp. 265–304 1382 Rebuschat_09_pisoni (p. 295)

Figure 4. Phoebe. One particular dog. How was your estimate?

References

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

Andersen, Øistein E., Julien Nioche, Ted Briscoe, and John Carroll
2008 The BNC Parsed with RASP4UIMA. *Proceedings of the Sixth International Language Resources and Evaluation (LREC08)*: 28–30.

Anderson, John Robert
2000 *Cognitive psychology and its implications (5th ed.)*. New York: W.H. Freeman.

Baayen, Harald
2008 Corpus linguistics in morphology: morphological productivity. In *Corpus Linguistics. An international handbook*, edited by A. Ludeling and M. Kyto. Berlin: Mouton De Gruyter.

Bannard, Colin and Elena Lieven
2009 Repetition and reuse in child language learning. In *Formulaic Language: Volume 2 Acquisition, loss, psychological reality, and functional explanations*, edited by R. Corrigan, E. A. Moravcsik, H. Ouali and K. M. Wheatley. Amsterdam: John Benjamins.

Bartlett, Frederick Charles
[1932] 1967 *Remembering: A Study in Experimental and Social Psychology*. Cambridge: Cambridge University Press
BNC
2007 BNC XML Edition http://www.natcorp.ox.ac.uk/corpus/.
Bod, Rens, Jennifer Hay, and Stefanie Jannedy, eds.
2003 *Probabilistic linguistics*. Cambridge, MA: MIT Press.
Briscoe, Ted, John Carroll, and Rebecca Watson
2006 The Second Release of the RASP System. Paper read at Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions, at Sydney, Australia.
Bybee, Joan
2010 *Language, usage, and cognition*. Cambridge: Cambridge University Press.
Bybee, Joan and Paul Hopper, eds.
2001 *Frequency and the emergence of linguistic structure*. Amsterdam: Benjamins.
Bybee, Joan and Sandra Thompson
2000 Three frequency effects in syntax. *Berkeley Linguistic Society* 23: 65–85.
Christiansen, Morten H., and Nick Chater, eds.
2001 *Connectionist psycholinguistics*. Westport, CO: Ablex.
Clark, Eve V.
1978 Discovering what words can do. In *Papers from the parasession on the lexicon, Chicago Linguistics Society April 14–15, 1978*, edited by D. Farkas, W. M. Jacobsen and K. W. Todrys. Chicago: Chicago Linguistics Society.
Clark, Eve V. and Barb Kelly, eds.
2006 *Constructions in acquisition*. Chicago: University of Chicago Press.
Dorogovstev, Sergey N. and José F. F. Mendes
2003 *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford: Oxford University Press.
Ebbinghaus, Hermann
1885 *Memory: A contribution to experimental psychology*. Translated by H. A. R. C. E. B. (1913). New York: Teachers College, Columbia.
Ellis, Nick C.
2002 Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition* 24(2): 143–188.
Ellis, Nick C.
2006 Language acquisition as rational contingency learning. *Applied Linguistics* 27(1): 1–24.

Ellis, Nick C.
2006 Selective attention and transfer phenomena in SLA: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics* 27(2): 1–31.

Ellis, Nick C.
2008 The dynamics of second language emergence: Cycles of language use, language change, and language acquisition. *Modern Language Journal* 41(3): 232–249.

Ellis, Nick C.
2008 Usage-based and form-focused language acquisition: The associative learning of constructions, learned-attention, and the limited L2 endstate. In *Handbook of cognitive linguistics and second language acquisition*, edited by P. Robinson and N. C. Ellis. London: Routledge.

Ellis, Nick C.
in press What can we count in language, and what counts in language acquisition, cognition, and use? In *Frequency effects in cognitive linguistics (Vol. 1): Statistical effects in learnability, processing and change*, edited by S. T. Gries and D. S. Divjak. Berlin: Mouton de Gruyter.

Ellis, Nick C. and Teresa Cadierno
2009 Constructing a second language. *Annual Review of Cognitive Linguistics* 7 (Special section): 111–290.

Ellis, Nick C. and Fernando Ferreira-Junior
2009 Construction Learning as a function of Frequency, Frequency Distribution, and Function *Modern Language Journal* 93: 370–386.

Ellis, Nick C. and Fernando Ferreira-Junior
2009 Constructions and their acquisition: Islands and the distinctiveness of their occupancy. *Annual Review of Cognitive Linguistics*: 111–139.

Ellis, Nick C. and Diane Larsen-Freeman
2009 Constructing a second language: Analyses and computational simulations of the emergence of linguistic constructions from usage. *Language Learning* 59 (Supplement 1): 93–128.

Ellis, Nick C., Rita Simpson-Vlach, and Carson Maynard
2008 Formulaic Language in Native and Second-Language Speakers: Psycholinguistics, Corpus Linguistics, and TESOL. *TESOL Quarterly* 42(3): 375–396.

Evert, Stefan
2005 The Statistics of Word Cooccurrences: Word Pairs and Collocations, University of Stuttgart, Stuttgart.

Ferrer-i-Cancho, Ramon and Ricard V. Sole
2002 Zipf's law and random texts. *Advances in Complex Systems* 5(1): 1–6.

Fillmore, Charles J., Christopher R. Johnson, and Miriam R. L. Petruck
2003 Background to Framenet. *International Journal of Lexicography* 16: 235–250.

Francis, Gill, Susan Hunston, and Elizabeth Manning, eds.
1996 *Grammar Patterns 1: Verbs. The COBUILD Series*. London: Harper Collins.

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

Goldberg, Adele E.
2006 *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Goldberg, Adele E., Devin M. Casenhiser, and Nitya Sethuraman
2004 Learning argument structure generalizations. *Cognitive Linguistics* 15: 289–316.

Gómez, Rebecca
2002 Variability and detection of invariant structure. *Psychological Science* 13: 431–436.

Gries, Stefan Th. and Anatol Stefanowitsch
2004 Extending collostructional analysis: a corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics* 9: 97–129.

Gries, Stefan Th., and Stefanie Wulff
2005 Do foreign language learners also have constructions? Evidence from priming, sorting, and corpora. *Annual Review of Cognitive Linguistics* 3: 182–200.

Grinberg, Dennis, John Lafferty, and Danniel Sleator
1995 A robust parsing algorithm for link grammars. Carnegie Mellon University Computer Science technical report CMU-CS-95–125.

Harnad, Steven, ed.
1987 *Categorical perception: The groundwork of cognition*. New York: Cambridge University Press.

Herbst, Thomas, David Heath, Ian Roe, and Dieter Götz
2004 *Valency Dictionary of English*. Berlin/New York: Mouton de Gruyter.

Herbst, Thomas and Peter Uhrig. *Erlangen Valency Patternbank*
2010 Available from http://www.patternbank.uni-erlangen.de/cgi-bin/patternbank.cgi?do=help.

Hunston, Susan and Gill Francis
1996 *Pattern grammar: A corpus driven approach to the lexical grammar of English*. Amsterdam: Benjamins.

Hunt, Ruskin H. and Richard N. Aslin
2010 Category induction via distributional analysis: Evidence from a serial reaction time task. *Journal of Memory and Language* 62: 98–112.

Jackendoff, Ray
1997 Twistin' the night away. *Language* 73: 543–559.

Jurafsky, Daniel
2002 Probabilistic Modeling in Psycholinguistics: Linguistic Comprehension and Production. In *Probabilistic Linguistics*, edited by R. Bod, J. Hay and S. Jannedy. Harvard, MA: MIT Press.

Jurafsky, Daniel and James H. Martin
2000 *Speech and language processing: An introduction to natural language processing, speech recognition, and computational linguistics.* Englewood Cliffs, NJ: Prentice-Hall.

Kello, Chris T. and Brandon C. Beltz
2009 Scale-free networks in phonological and orthographic wordform lexicons. In *Approaches to Phonological Complexity*, edited by I. Chitoran, C. Coupé, E. Marsico and F. Pellegrino. Berlin: Mouton de Gruyter.

Kello, Chris T., Gordon D. A. Brown, Ramon Ferrer-i-Cancho, John G. Holden, Klaus Linkenkaer-Hansen, Theo Rhodes, and Guy C. Van Orden
2010 Scaling laws in cognitive sciences. *Trends in Cognitive Science* 14: 223–232.

Kipper, Karin, Anna Korhonen, Neville Ryant, and Martha Palmer
2008 A large-scale classification of English verbs. *Language Resources and Evaluation* 41: 21–40.

Klein, Dan and Christopher D. Manning
2003 Fast Exact Inference with a Factored Model for Natural Language Parsing. In *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press.

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

Landauer, Thomas K., Danielle S. McNamara, Simon Dennis, and Walter Kintsch, eds.
2007 *Handbook of Latent Semantic Analysis*. Mahwah, NJ: Lawrence Erlbaum.

Levin, Beth
1993 *English verb classes and alternations: A preliminary analysis*. Chicago: Chicago University Press.

Li, Wentian
1992 Random Texts Exhibit Zipf's Law-Like Word Frequency Distribution. *IEEE Transactions on Information Theory* 38.6: 1842–1845.

MacWhinney, Brian
1987 Applying the Competition Model to bilingualism. *Applied Psycholinguistics* 8(4): 315–327.

MacWhinney, Brian
1987 The Competition Model. In *Mechanisms of language acquisition*, edited by B. MacWhinney. Hillsdale, NJ: Erlbaum.

(V9 5/4/12 14:41) WDG-LCB (155mm×230mm) TimesNRMT 1382 Rebuschat pp. 265–304 1382 Rebuschat_09_pisoni (p. 300)

MacWhinney, Brian
1997 Second language acquisition and the Competition Model. In *Tutorials in bilingualism: Psycholinguistic perspectives*, edited by A. M. B. De Groot and J. F. Kroll. Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney, Brian
2000 *The CHILDES project: Tools for analyzing talk, Vol 1: Transcription format and programs (3rd ed.)*: (2000). xi, 366pp.

MacWhinney, Brian
2000 *The CHILDES Project: Tools for analyzing talk, Vol 2: The database (3rd ed.)*: (2000). viii, 418pp.

MacWhinney, Brian
2001 The competition model: The input, the context, and the brain. In *Cognition and second language instruction*, edited by P. Robinson. New York: Cambridge University Press.

Manin, Dmitrii Y.
2008 Zipf's Law and Avoidance of Excessive Synonymy *Cognitive Science* 32: 1075–1098.

Manning, Chris D. and Hinrich Schütze
1999 *Foundations of statistical natural language processing*. Cambridge, MA: The MIT Press.

Mason, Oliver and Susan Hunston
2004 The automatic recognition of verb patterns: A feasibility study. *International Journal of Corpus Linguistics* 9: 253–270.

Miller, George A.
1957 Some effects of intermittent silence. *American Journal of Psychology* 70: 311–314.

Miller, George A.
1965 *Preface to the MIT press publication of G. H. K. Zipf (1935) The psychobiology of language: An introduction to dynamic philology.* Boston MA: MIT Press.

Miller, George A.
2009 WordNet – About us. Princeton University.

Mintz, Toben
2002 Category induction from distributional cues in an artificial language. *Memory & Cognition* 30: 678–686.

Newman, Mark
2005 Power laws, Pareto distributions and Zipf's law. *Contemporary Physics* 46: 323–351

Ninio, Anat
1999 Pathbreaking verbs in syntactic development and the question of prototypical transitivity. *Journal of Child Language* 26:619–653.

Ninio, Anat
2006 *Language and the learning curve: A new theory of syntactic development*. Oxford: Oxford University Press.

(V9 5/4/12 14:41) WDG-LCB (155mm×230mm) TimesNRMT 1382 Rebuschat pp. 265–304 1382 Rebuschat_09_pisoni (p. 301)

Nivre, Joakim, Johan Hall, and Jens Nilsson
2004 Memory-Based Dependency Parsing. Paper read at Proceedings of the Eighth Conference on Computational Natural Language Learning (CoNLL).

O'Donnell, Matthew B. and Nick C. Ellis
submitted 'Zipfing across the input' Analyzing the distribution and semantics of verbs in verb argument patterns in child directed and general English. In *BUCLD 35*. Boston.

O'Donnell, Matthew B. and Nick C. Ellis
2009 Measuring formulaic language in corpora from the perspective of Language as a Complex System. Paper read at 5th Corpus Linguistics Conference, 20–23 July, 2009, at University of Liverpool.

O'Donnell, Matthew B. and Nick C. Ellis
2010 Towards an Inventory of English Verb Argument Constructions. *Proceedings of the 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics, Los Angeles.*

Onnis, Luca, Padraic Monaghan, Morten H. Christiansen, and Nick Chater
2004 Variability is the spice of learning, and a crucial ingredient for detecting and generalising nonadjacent dependencies. *Proceedings of the 26th Annual Conference of the Cognitive Science Society*.

Palmer, Martha
[redacted] *VerbNet: A class based verb lexicon* 2010. Available from http://verbs.colorado.edu/~mpalmer/projects/verbnet.html.

Pedersen, Ted and Varada Kolhatkar
2009 WordNet::SenseRelate::AllWords – A Broad Coverage Word Sense Tagger that Maximizes Semantic Relatedness Paper read at Proceedings of the Demonstration Session of the Human Language Technology Conference and the Tenth Annual Meeting of the North American Chapter of the Association for Computational Linguistics, at Boulder, Colorado.

Pedersen, Ted, Siddharth Patwardhan, and Jason Michelizzi
2004 WordNet::Similarity – Measuring the Relatedness of Concepts. Paper read at Proceedings of Fifth Annual Meeting of the North American Chapter of the Association of Computational Linguistics (NAACL 2004).

Perruchet, Pierre and Sébastian Pacton
2006 Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in Cognitive Sciences* 10: 233–238.

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

Posner, Michael I., and Stephen W. Keele
1970 Retention of abstract ideas. *Journal of Experimental Psychology* 83: 304–308.

(V9 5/4/12 14:41) WDG-LCB (155mm×230mm) TimesNRMT 1382 Rebuschat pp. 265–304 1382 Rebuschat_09_pisoni (p. 302)

R:
[redacted] A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rescorla, Robert A.
1968 Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative and Physiological Psychology* 66: 1–5.

Resnik, Philip
1995 Using Information Content to Evaluate Semantic Similarity in a Taxonomy. *IJCAI*: 448–453.

Rosch, Eleanor and Carolyn B. Mervis
1975 Cognitive representations of semantic categories. *Journal of Experimental Psychology: General* 104: 192–233.

Rosch, Eleanor, Carolyn B. Mervis, Wayne D. Gray, David M. Johnson, and Penny Boyes-Braem
1976 Basic objects in natural categories. *Cognitive Psychology* 8: 382–439.

Sagae, Kenji, Eric Davis, Alon Lavie, Brain MacWhinney, and Shuly Wintner
2007 High-accuracy annotation and parsing of CHILDES transcripts. In *Proceedings of the ACL-2007 Workshop on Cognitive Aspects of Computational Language Acquisition*. Prague: Czech Republic.

Schneider, Gerold, Fabio Rinaldi, and James Dowdall
2004 Fast, deep-linguistic statistical dependency parsing. Paper read at Recent advances in DG workshop, Coling 2004, at Geneva.

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

Sinclair, John
2004 *Trust the text: Language, corpus and discourse*. London: Routledge.

Sinclair, John, ed.
1987 *Looking Up: An Account of The COBUILD Project in Lexical Computing*. London: Collins ELT.

Solé, Ricard V., Bernat Murtra, Sergi Valverde, and Luc Steels
2005 Language Networks: their structure, function and evolution. *Trends in Cognitive Sciences* 12.

Stefanowitsch, Anatol and Stefan Th. Gries
2003 Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics* 8: 209–43.

Steyvers, Mark and Josh Tennenbaum
2005 The large-scale structure of semantic networks: statistical analyses and a model of semantic growth. *Cognitive Science* 29: 41–78.

Taylor, John R.
1998 Syntactic constructions as prototype categories. In *The new psychology of language: Cognitive and functional approaches to language structure*, edited by M. Tomasello. Mahwah, NJ: Erlbaum.

Tomasello Michael
2003 *Constructing a language*. Boston, MA: Harvard University Press.

Tversky, Alan
1977 Features of similarity. *Psychological Review* 84: 327–352.

Wittgenstein, Ludwig
1953 *Philosophical investigations*. Translated by G. E. M. Anscombe. Oxford: Blackwell.

Yang, Charles
2010 Who's afraid of George Kingsley Zipf? http://www.ling.upenn.edu/~ycharles/papers/zipfnew.pdf.

Zipf, George K.
1935 *The psycho-biology of language: An introduction to dynamic philology*. Cambridge, MA: The M.I.T. Press.