

Construction learning as category learning: A cognitive analysis

Nick C. Ellis
Matthew Brook O'Donnell

1. Introduction

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 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 carries 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*. Knowledge of language is based on these types of inference, and verbs are the cornerstone of the syntax-semantics interface.

This chapter reviews psychological theory that relates to the learning of constructions as categories. It then analyses the frequency, form and function of a sample of 23 VACs in 100 million words of usage to show how language form, language meaning, and language usage come together to promote robust induction by means of statistical learning over limited samples.

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 1995, 2006). 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: 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 their 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-meaning/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 ab-

stract constructions being learned as categories from the conspiracy of concrete exemplars of usage following statistical learning mechanisms¹ relating input and learner cognition.

3. Determinants of construction learning

Psychological analyses of the learning of constructions as form-meaning pairs are 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) meaning and 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

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

1. Cf. Christiansen and Chater (2001), Jurafsky and Martin (2000), Bybee and Hopper (2001), Bod, Hay, and Jannedy (2003), Ellis (2002), Perruchet and Pacton (2006).

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 do not go around consciously counting features, they nevertheless have very accurate knowledge of the underlying frequency distributions and their central tendencies.

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.

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 and Solé 2001, 2003, Ferrer i Cancho, Solé, and 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 2008a). Zipfian covering, where, as concepts need to be refined for clear communication, they are split, then split again hierarchically (e.g. animal, canine, dog, retriever, Labrador ...), determines basic categorization, the structure of semantic classes, and the language form-semantic structure interface (Steyvers and Tenenbaum 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 (cf. Ellis and Larsen-Freeman 2009b, Beckner et al. 2009, Ferrer i Cancho and Solé 2001, 2003, Ferrer i Cancho, Solé, and Köhler 2004, Solé et al. 2005). Various usage-based/functionalist/cognitive linguists (Goldberg 2006, Goldberg, Casenhiser, and Sethuraman 2004, Lieven and Tomasello 2008, Bybee 2008, e.g. Boyd and Goldberg 2009, Ellis 2008c, Ninio 1999, 2006, Bybee 2010) 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 and 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 (2009a) 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 Form (salience and perception)

The general perceived strength of stimuli is commonly referred to as their salience. Low salience cues tend to be less readily learned. Ellis (2006a, 2006b) summarized the associative learning research demonstrating that selective attention, salience, expectation, and surprise are key elements in the analysis of all learning, animal and human alike. As the Rescorla-Wagner (1972) model encapsulates, the amount of learning induced from an experience of a cue-outcome association depends crucially upon the salience of the cue and the importance of the outcome.

Many grammatical meaning-form relationships, particularly those that are notoriously difficult for second language learners like grammatical particles and inflections such as the third person singular *-s* of English, are of low salience in the language stream. For example, some forms are more salient: *'today'* is a stronger psychophysical form in the input than is the morpheme *'-s'* marking 3rd person singular present tense, thus while both provide cues to present time, *today* is much 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 2006b, 2008a, Goldschneider and DeKeyser 2001). As for the cue, so for the interpretation – the meaning of *'today'* is more salient, tangible, and less abstract than that of *'-s'*.

3.3 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 and Ferreira-Junior (2009a) 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 transitiv-

ity 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 (1978: 53) 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”.

3.4 Interactions between form and function (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 2008b, MacWhinney 1987a, Ellis 2006a, Ellis 2006b, Gries and Wulff 2005), with the most developed account for L2 acquisition being that of the Competition model (MacWhinney 1987b, 1997, 2001).

Ellis and Ferreira-Junior (2009b) 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 collocation 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 (2009a) provided computational (Emergent connectionist) serial-recurrent network simulations of these various factors as they play out in the emergence of VACs (VL, VOL, VOO] as generalized linguistic schemas from their frequency distributions in the input. This fundamental claim that Zipfian distributional properties of language usage help 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 construction.

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 here aims to assess this for a larger 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

Our research aims to empirically determine the semantic associations of particular linguistic forms, therefore 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. There are over 700 patterns of varying complexity in this volume. In subsequent work we hope to analyze them all in the same ways. Here we focus on a convenience sample of 23 constructions for our initial explorations. 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).

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 the *V across N* pattern:

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.

Step 2 Corpus: BNC XML parsed corpora

To get a representative sample of usage, the verb type-token distribution of these VACs was determined in the 100 million word British National Corpus BNC (2007) parsed using the XML version of the BNC using the RASP parser (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.

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. The types of sentence captured are diverse, of course, as can be seen from these two example results:

- (1) *She **walked** across the yard to check Shine On.*
- (2) *'The intellectual and rational conception of life has given way to a more creative interpretation', wrote the British Surrealist Eileen Agar in 1931, 'and artistic life is under the sway of womb-magic'; and Agar give expression to this 'womb-magic' in the foetal and embryonic forms which play a central part in paintings such as 'Family Trio' or 'The Autobiography of an Embryo' where fluid shapes **float** across the picture plane to be captured in a net of geometric planes.*

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.

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

Construction	Types	Tokens	TTR	Lead verb type
<i>V about N</i>	365	3519	10.37	<i>talk</i>
<i>V across N</i>	799	4889	16.34	<i>come</i>
<i>V after N</i>	1168	7528	15.52	<i>look</i>
<i>V among pl-N</i>	417	1228	33.96	<i>find</i>
<i>V around N</i>	761	3801	20.02	<i>look</i>
<i>V as adj</i>	235	1012	23.22	<i>know</i>
<i>V as N</i>	1702	34383	4.95	<i>know</i>
<i>V at N</i>	1302	9700	13.42	<i>look</i>
<i>V between pl-N</i>	669	3572	18.73	<i>distinguish</i>
<i>V for N</i>	2779	79894	3.48	<i>look</i>
<i>V in N</i>	2671	37766	7.07	<i>find</i>
<i>V into N</i>	1873	46488	4.03	<i>go</i>
<i>V like N</i>	548	1972	27.79	<i>look</i>
<i>V N N</i>	663	9183	7.22	<i>give</i>
<i>V off N</i>	299	1032	28.97	<i>take</i>
<i>V of N</i>	1222	25155	4.86	<i>think</i>
<i>V over N</i>	1312	9269	14.15	<i>go</i>
<i>V through N</i>	842	4936	17.06	<i>go</i>

<i>V to N</i>	707	7823	9.04	<i>go</i>
<i>V towards N</i>	190	732	25.96	<i>move</i>
<i>V under N</i>	1243	8514	14.6	<i>come</i>
<i>V way prep</i>	365	2896	12.6	<i>make</i>
<i>V with N</i>	1942	24932	7.79	<i>deal</i>

Full details of the search methods, example sentences, and our methods for estimating precision and recall of the searches can be seen in Römer, O'Donnell, and Ellis (2013). Mean precision for the 23 constructions was 0.78, mean recall 0.53.

Step 4 A frequency ranked type-token VAC profile

The sentences extracted using this procedure outlined for each of the 23 construction patterns produced verb type distributions like the following one for the *V across N* VAC pattern:

come	483				
walk	203				
cut	199	...			
run	175	veer	4		
...		slice	4	...	
		...		navigate	1
				scythe	1
				scroll	1

These distributions appear to be Zipfian, exhibiting the characteristic long right tail in a plot of rank against frequency. We generated logarithmic plots and linear regression to examine the extent of this trend using logarithmic binning of frequency against log cumulative frequency. Figure 1 shows such a plot for verb type frequency of the *V across N* construction, Figure 2 shows the same type of plot for verb type frequency of the ditransitive *V N N* construction. Both distributions produce a good fit of Zipfian type-token frequency with $R^2 > 0.97$ and slope (γ) around 1. Inspection of the construction verb types, from most frequent down, also demonstrates that the lead member is prototypical of the construction and generic in its action semantics. You may notice that some of these items, such as 'come across N', are phrasal verbs and have metaphorically extended their mean-

ing (e.g. “I came across an interesting book”) while others like *walk* and *run* are more literal. We allowed our searches to capture this mixed bag because our goal was to analyse just what meanings were associated with the verb types that appeared in VACs that were simply operationalised at this first stage on the basis of their linguistic form.

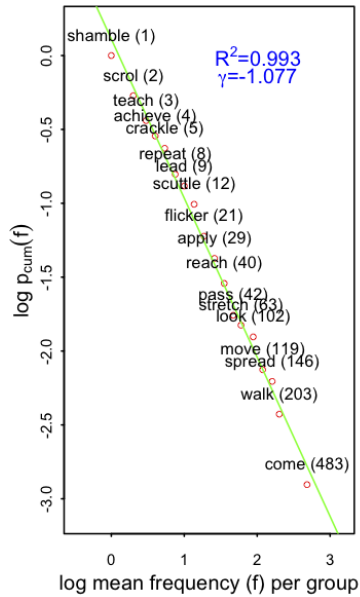


Figure 1. Verb type distribution for *V across N*

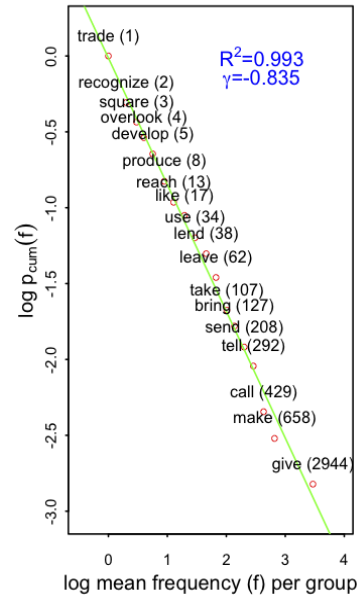


Figure 2. Verb type distribution for *VNN*

Since Zipf’s law applies across language, the Zipfian nature of these distributions is potentially trivial. But they are 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 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. The lower the entropy, the more coherent the VAC verb family. Scores on all these metrics are given for all VACs later in Table 3.

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). 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 should be more readily acquired. The measures of contingency adopted here are (1) faithfulness (also known as “reliance”, Schmid and Küchenhoff 2013) – the proportion of tokens of total verb usage that appear in this particular construction (e.g. the faithfulness of *give* to the ditransitive is approximately 0.40; that of *leave* is 0.01), and (2) 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 2003). Table 2 lists these contingency measures for the verbs occupying the *V across N VAC* pattern.

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

Verb	Constr. Freq.	Corpus Freq.	Faith.	MI Word \rightarrow Construction	MI Construction \rightarrow Word
<i>come</i>	474	122107	0.0039	15.369	10.726
<i>walk</i>	203	17820	0.0114	16.922	15.056
<i>cut</i>	197	16200	0.0122	17.016	15.288
<i>run</i>	175	36163	0.0048	15.687	12.800
<i>spread</i>	146	5503	0.0265	18.142	17.971
<i>move</i>	114	34774	0.0033	15.125	12.295
<i>look</i>	102	93727	0.0011	13.534	9.273
<i>go</i>	93	175298	0.0005	12.498	7.333
<i>lie</i>	80	18468	0.0043	15.527	13.610
<i>lean</i>	75	4320	0.0174	17.530	17.708
<i>stretch</i>	62	4307	0.0144	17.260	17.442
<i>fall</i>	57	24656	0.0023	14.621	12.287
<i>get</i>	52	146096	0.0004	11.922	7.020
<i>pass</i>	42	18592	0.0023	14.588	12.661
<i>reach</i>	40	21645	0.0018	14.298	12.152
<i>travel</i>	39	8176	0.0048	15.666	14.924

<i>fly</i>	38	8250	0.0046	15.616	14.861
<i>stride</i>	38	1022	0.0372	18.629	20.887
<i>scatter</i>	35	1499	0.0233	17.957	19.663
<i>sweep</i>	34	2883	0.0118	16.972	17.734

Step 6 Identifying the meaning of verb types occupying the constructions

Our semantic analyses use 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*.

The values on the metrics we have described so far are illustrated for the 23 VACs in Table 3. It can be seen that for all of the VACs, the type-token distribution is Zipfian (mean $R^2 = 0.98$) and that there is contingency between verbs and VACs (mean $MI_{\text{word-construction}} = 14.16$) – particular verbs select particular constructions, and vice versa.

Table 3. Values for our 23 Verb Argument Constructions on metrics of Zipfian distribution, verb form selectivity, and semantic coherence

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

Step 7 Generating distributionally-matched, control ersatz constructions (CECs)

Because so much of language distribution is Zipfian, for each of the 23 VACs we analyze, we generate a distributionally-yoked control (a ‘control ersatz construction’ [CEC]), which is matched for type-token distribution but otherwise randomly selected to be grammatically and semantically uninformed. 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 *walk* 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 to 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. 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 the VACs and their yoked CECs in Table 4.

Step 8 Evaluating semantic cohesion in the VAC distributions

The VAC type-token list shows that the tokens list captures the most general and prototypical senses (*come, walk, move* etc. for V *across* N and *give, make, tell*, 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 WordNet, where each synset can be traced back to a root or top-level synset, we compared the semantic cohesion of the top 20 verbs, using their disambiguated WordNet senses, from a given VAC to its matching CEC. For example, in V *across* N, the top level hypernym synset *travel.v.01* accounts for 15% of tokens, whereas the most frequent root synset for the matching CEC, *pronounce.v.1*, accounts for just 4% of the tokens. The VAC has a more compact semantic distribution in that the 3 top-level synsets account for 25% of the tokens compared to just 11% for the CEC.

We use various methods of evaluating the differences between the semantic sense distributions for each VAC-CEC pair. First, we measure the amount of variation in the distribution using Shannon entropy according to (1) number of sense types per root (V *across* N VAC: 2.75 CEC: 3.37) and (2) the token frequency per root (V *across* N VAC: 2.08 CEC: 3.08), the lower the entropy the more coherent the VAC verb semantics. Second, we assess the coverage of the top three root synsets in the VAC and its corresponding CEC. Third, we quantify the semantic coherence of the disambiguated senses of the top 20 verb forms in the VAC and CEC distributions using two measures of semantic similarity from Pedersen, Patwardhan and Michelizzi's (2004) Perl WordNet::Similarity package, *lch* based on the path length between concepts in WordNet Synsets and *res* that additionally incorporates a measure called 'information content' related to concept specificity. For instance, using the *res* similarity measure the top 20 verbs in V *across* n VAC distribution have a mean similarity score of 0.35 compared to 0.17 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. We show comparisons for the VACs and their yoked CECs on these aspects in Table 4.

Table 4. Comparisons of values for our 23 VACs and CECs on metrics of Zipfian distribution, verb form selectivity, and semantic coherence

Criterion dimension	Metric	Mean VACs	Mean CECs	t value for paired t-test (d.f. 22) ***=p<.001
Zipfian distribution	R^2	0.98	0.96	6.49 ***
	γ	-1.00	-1.12	6.04 ***
	χ^2	69412	698	4.09 ***
	$1-\tau$	0.76	0.21	25.94 ***
	Entropy	4.97	5.54	5.76 ***
Verb form selectivity	Faithfulness	0.016	0.002	5.13 ***
	Mean MIw-c	14.16	12.80	3.53 ***
	Mean MIc-w	14.11	10.86	10.79 ***
Semantic Coherence	Type entropy per root synset	3.1	3.51	5.01 ***
	Token entropy per root synset	2.41	3.08	5.51 ***
	Proportion of tokens covered by top 3 synsets	0.26	0.11	5.23 ***
	lch	0.134	0.094	4.30 ***
	res	0.237	0.22	4.45 ***

The results demonstrate:

(1) Type-token usage distributions: All of the VACs are Zipfian in their type-token distributions in usage (VACs: Mean $\gamma = -1.00$, Mean $R^2 = 0.98$). So too are their matched CECs (Mean $\gamma = -1.12$, Mean $R^2 = 0.96$). 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, the lead member is prototypical of the construction and generic in its action semantics.

(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. The distribution deviation (χ^2) from verb frequency in the language as a whole is much greater in the VACs than the CECs. The lack of overall correlation ($1-\tau$) between VAC verb frequency and overall verb frequency in the language is much greater in the VACs. Verbs are more faithful to VACs than to CECs. Individual verbs select particular constructions (Mean MIw-c) and particular constructions select particular words (Mean MIC-w). Overall then, there is greater contingency between verb types and constructions.

(3) Semantic coherence: VACs are coherent in their semantics with lower type and token sense entropy. The proportion of the total tokens covered by their three most frequent WordNet roots is much higher in the VACs. Finally, the VAC distributions are higher on the Pedersen semantic similarity measures (lch and res).

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 and Ferreira-Junior (2009a,b). Nevertheless, the conclusions from those earlier studies seem to generalize. These 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. Complex systems are characterized by their robustness to different

kinds of perturbations, by their scale-free properties, and by their structures emerging from the interactions of agents and components at many levels (Page 2009). We believe that the robustness of language emerges as a consequence of its dynamics as a complex adaptive system (Beckner et al. 2009).

Acknowledgement

Nick Ellis is grateful for the support of Freiburg Institute for Advanced Studies (FRIAS), where he was based as an External Senior Fellow while revising this paper.

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.
- Anderson, John R.
2000 *Cognitive Psychology and its Implications* (5th ed.). New York: W. H. Freeman.
- Baayen, R. Harald
2008 Corpus linguistics in morphology: Morphological productivity. In *Corpus Linguistics: An International Handbook*, Anke Ludeling, and Merja Kyto (eds.), 899–919. Berlin: Mouton de Gruyter.
- Bannard, Colin, and Elena Lieven
2009 Repetition and reuse in child language learning. In *Formulaic Language: Vol. 2 Acquisition, Loss, Psychological Reality, and Functional Explanations*, Roberta Corrigan, Edith A. Moravcsik, Hamid Ouali, and Kathleen M. Wheatley (eds.), 297–322. Amsterdam: John Benjamins.
- Bartlett, Frederic Charles
1967 *Remembering: A Study in Experimental and Social Psychology*. First published in 1932. Cambridge: Cambridge University Press
- Beckner, Clay, Richard Blythe, Joan Bybee, Morten H. Christiansen, William Croft, Nick C. Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman, and Tom Schoenemann
2009 Language is a complex adaptive system. Position paper. *Language Learning* 59 Suppl. 1: 1–26.
- 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.
- Boyd, Jeremy K., and Adele E. Goldberg
2009 Input effects within a constructionist framework. *Modern Language Journal* 93 (2): 418–429.
- 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
2008 Usage-based grammar and second language acquisition. In *Handbook of Cognitive Linguistics and Second Language Acquisition*, Peter Robinson, and Nick C. Ellis (eds.), 216–236. London: Routledge.
- 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, Donka Farkas, Wesley M. Jacobsen, and Karol W. Todrys (eds.), 34–57. Chicago: Chicago Linguistics Society.
- Clark, Eve V., and Barb Kelly (eds.)
2006 *Constructions in Acquisition*. Chicago: University of Chicago Press.
- Ebbinghaus, Hermann
1885 *Memory: A Contribution to Experimental Psychology*, translated by Clara E. Bussenius and Henry A. Ruger (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.
2006a Language acquisition as rational contingency learning. *Applied Linguistics* 27 (1): 1–24.

- Ellis, Nick C.
 2006b 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.
 2008a 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.
 2008b Optimizing the input: Frequency and sampling in usage-based and form-focussed learning. In *Handbook of Second and Foreign Language Teaching*, Michael H. Long, and Catherine Doughty (eds.), 139–158. Oxford: Blackwell.
- Ellis, Nick C.
 2008c Usage-based and form-focused language acquisition: The associative learning of constructions, learned-attention, and the limited L2 end-state. In *Handbook of Cognitive Linguistics and Second Language Acquisition*, Peter Robinson, and Nick C. Ellis (eds.), 372–405. London: Routledge.
- Ellis, Nick C., and Teresa Cadierno (eds.)
 2009 Constructing a second language. *Annual Review of Cognitive Linguistics* 7 (Special section):111–290.
- Ellis, Nick C., and Fernando Ferreira-Junior
 2009a Construction learning as a function of frequency, frequency distribution, and function. *Modern Language Journal* 93: 370–386.
- Ellis, Nick C., and Fernando Ferreira-Junior
 2009b Constructions and their acquisition: Islands and the distinctiveness of their occupancy. *Annual Review of Cognitive Linguistics* 7: 111–139.
- Ellis, Nick C., and Diane Larsen-Freeman
 2009a 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., and Diane Larsen-Freeman (eds.)
 2009b Language as a complex adaptive system (special issue). *Language Learning* 59, Supplement 1.
- 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*, Dissertation. Stuttgart: University of Stuttgart.

- Ferrer i Cancho, Ramon, and Ricard V. Solé
2001 The small world of human language. *Proceedings of the Royal Society of London*, B. 268: 2261–2265.
- Ferrer i Cancho, Ramon, and Ricard V. Solé
2003 Least effort and the origins of scaling in human language. *PNAS* 100: 788–791.
- Ferrer i Cancho, Ramon, Ricard V. Solé, and Reinhard Köhler
2004 Patterns in syntactic dependency networks. *Physical Review* E69: 0519151–0519158.
- 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.
- Goldschneider, Jennifer M., and Robert DeKeyser
2001 Explaining the “natural order of L2 morpheme acquisition” in English: A meta-analysis of multiple determinants. *Language Learning* 51: 1–50.
- Gómez, Rebecca
2002 Variability and detection of invariant structure. *Psychological Science* 13: 431–436.
- Gries, Stefan Th., and Anatol Stefanowitsch
2004 Extending collocation 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.
- Harnad, Stevan (ed.)
1987 *Categorical Perception: The Groundwork of Cognition*. New York: Cambridge University Press.
- 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
 2003 Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. In *Probabilistic Linguistics*, Rens Bod, Jennifer Hay, and Stefanie Jannedy (eds.), 39–95. 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, Christopher T., and Brandon C. Beltz
 2009 Scale-free networks in phonological and orthographic wordform lexicons. In *Approaches to Phonological Complexity*, Ioana Chitoran, Christophe Coupé, Egidio Marsico, and François Pellegrino (eds.), 178–197. Berlin: Mouton de Gruyter.
- Kello, Christopher 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.
- Lakoff, George
 1987 *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.
- Lieven, Elena, and Michael Tomasello.
 2008 Children's first language acquisition from a usage-based perspective. In *Handbook of Cognitive Linguistics and Second Language Acquisition*, Peter Robinson, and Nick C. Ellis (eds.), 168–196. New York/London: Routledge.
- MacWhinney, Brian
 1987a Applying the competition model to bilingualism. *Applied Psycholinguistics* 8 (4): 315–327.
- MacWhinney, Brian
 1987b The competition model. In *Mechanisms of Language Acquisition*, Brian MacWhinney (ed.), 249–308. Hillsdale, NJ: Erlbaum.
- MacWhinney, Brian
 1997 Second language acquisition and the competition model. In *Tutorials in Bilingualism: Psycholinguistic Perspectives*, Annette M. B. de Groot, and Judith F. Kroll (eds.), 113–144. Mahwah, NJ: Lawrence Erlbaum Associates.

- MacWhinney, Brian
2001 The competition model: The input, the context, and the brain. In *Cognition and Second Language Instruction*, Peter Robinson (ed.), 69–90. New York: Cambridge University Press.
- Manin, Dimitri Y.
2008 Zipf's law and avoidance of excessive synonymy. *Cognitive Science* 32: 1075–1098.
- 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, M. E. J.
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.
- O'Donnell, Matthew, 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, 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, June 1–6, 2010, 9–16. 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*, 1047–1052. Mahwah, NJ: Lawrence Erlbaum.
- Page, Scott E.
2009 *Understanding Complexity* [DVD-ROM]. Chantilly, VA: The Teaching Company.
- 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 Technol-

- ogy 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 Sébastien 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 Steven W. Keele
 1970 Retention of abstract ideas. *Journal of Experimental Psychology* 83: 304–308.
- 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.
- Rescorla, Robert A., and Allan R. Wagner
 1972 A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In *Classical Conditioning II: Current Theory and Research*, Abraham H. Black, and William F. Prokasy (eds.), 64–99. New York: Appleton-Century-Crofts.
- Römer, Ute, Matthew Brook O'Donnell, and Nick C. Ellis
 2013 Using COBUILD grammar patterns for a large-scale analysis of verb-argument constructions: Exploring corpus data and speaker knowledge. In *Corpora, Grammar, Text and Discourse: In Honour of Susan Hunston*, Maggie Charles, Nick Groom and Suganthi John (eds.). Amsterdam: John Benjamins (in press).
- 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.

- Schmid, Hans-Jörg, and Helmut Küchenhoff
2013 Collostructional analysis and other ways of measuring lexicogram-
matical attraction: Theoretical premises, practical problems and cog-
nitive underpinnings. *Cognitive Linguistics* 24 (3): 531–577.
- Shanks, David R.
1995 *The Psychology of Associative Learning*. New York: Cambridge
University Press.
- Sinclair, John (ed.)
1987 *Looking Up: An Account of The COBUILD Project in Lexical Com-
puting*. London: Collins ELT.
- Sinclair, John
2004 *Trust the Text: Language, Corpus and Discourse*. London:
Routledge.
- Solé, Ricard V., Bernat Corominas Murtra, Sergi Valverde, and Luc Steels
2005 Language networks: Their structure, function and evolution. *Com-
plexity* 15 (6): 20–26.
- Stefanowitsch, Anatol, and Stefan Th. Gries.
2003 Collostructions: Investigating the interaction between words and
constructions. *International Journal of Corpus Linguistics* 8: 209–
243.
- Steyvers, Mark, and Joshua Tenenbaum
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 Psy-
chology of Language: Cognitive and Functional Approaches to Lan-
guage Structure*, Michael Tomasello (ed.), 177–202. Mahwah, NJ:
Erlbaum.
- Tomasello, Michael
2003 *Constructing a Language*. Boston, MA: Harvard University Press.
- Tversky, Amos
1977 Features of similarity. *Psychological Review* 84: 327–352.
- Wittgenstein, Ludwig
1953 *Philosophical investigations*. Translated by Gertrude E. M.
Anscombe. Oxford: Blackwell.
- Zipf, Georg Kingsley
1935 *The Psycho-Biology of Language: An Introduction to Dynamic Phi-
lology*. Cambridge, MA: M.I.T. Press.

