Robust Language Acquisition – an Emergent Consequence of Language as a Complex Adaptive System

Nick C. Ellis (NCELLIS@Umich.Edu)
Department of Psychology and English Language Institute, University of Michigan
1085 South University Ave., Ann Arbor, MI 48109-1107 USA

Matthew Brook O'Donnell (MBOD@Umich.Edu)
English Language Institute, University of Michigan
1085 South University Ave., Ann Arbor, MI 48109-1107 USA

Abstract

Each of us as learners had different language experiences, yet 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. Analyses of a 100-million-word corpus show how Zipfian scale-free distributions of usage ensure robust learning of linguistic constructions as categories: constructions are (1) Zipfian in their type-token distributions in usage, (2) selective in their verb form occupancy, and (3) coherent in their semantics. Parallel psycholinguistic experiments demonstrate the psychological reality of these constructions in language users.

Keywords: Language as a Complex Adaptive System, Zipf's law, Verb Argument Constructions; Syntax-semantics interface; Usage-based models.

Verb Argument Constructions in Usage, Acquisition, and Mind

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* - in just the same way that you can conjure up an idea of my 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 a large corpus of usage. Parallel psycholinguistic experiments demonstrate the psychological reality of these constructions in language users.

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). Verbs are central in this: their semantic behavior is strongly intertwined with the syntagmatic constraints. 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. Such beliefs, increasingly influential in the study of child language acquisition, emphasize data-driven, emergent accounts of linguistic systematicities (e.g., Tomasello, 2003).

Frequency, learning, and language come together in usage-based approaches which hold that we learn linguistic constructions while engaging in communication (Bybee, 2010). Fifty years of psycholinguistic research substantiates usage-based acquisition through its demonstrations of language processing being exquisitely sensitive to usage frequency at all levels, from phonology, through lexis and syntax, to sentence processing (Ellis, 2002). Frequency is a key determinant of acquisition because 'rules' of language emerge as categories from the conspiracy of concrete exemplars of usage following statistical learning mechanisms relating input and learner cognition.

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 & Cadierno, 2009).

Determinants of construction learning

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 $p_f \sim f^{-b}$, with $b \approx 1$. Zipf showed this scaling relation holds across a wide variety of language samples. Subsequent research generalises this law as a linguistic universal: it holds acrioss many language events (e.g., frequencies of phoneme and letter strings, of words, of grammatical constructs, of formulaic phrases, etc.) across scales of analysis (Solé, Murtra, Valverde, & Steels, 2005).

Goldberg, Casenhiser & Sethuraman (2004) demonstrated that in samples of child language acquisition, for a variety of verb-argument constructions (VACs), there is a strong tendency for one single verb to occur with very high frequency in comparison to other verbs used, a profile which closely mirrors that of the mothers' speech to these children. They argue that this promotes acquisition since the pathbreaking verb which accounts for the lion's share of instances of each argument frame is the one with the prototypical meaning from which the construction is derived. 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. The bounds of the category are defined later by experience of the full breadth of exemplar types.

Ellis and Ferreira-Junior (2009) investigate effects upon naturalistic second language acquisition of type/token distributions in the islands comprising the linguistic form of three English verb-argument constructions (VL verb locative, VOL verb object locative, VOO ditransitive). They show 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.). Their work further illustrates how acquisition is affected by the frequency and frequency distribution of exemplars within each island of the construction (e.g. [Subj V Obj $\text{Obl}_{\text{path/loc}}$]), by their prototypicality, and, using a variety of psychological and corpus linguistic association metrics, by their contingency of form-function mapping. The fundamental claim that Zipfian distributional properties of language usage helps to make language learnable has thus been explored for these three VACs, at least. It remains important to explore its generality across a wider range of the constructicon. We do this here for a sample of 23 constructions.

Corpus analyses of 23 VACs in 100-million words of usage

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. Therefore we use the definition of VACs presented in the *Verb Grammar Patterns* that arose out of the Cobuild project (Hunston & Francis, 1996). 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 take a convenience sample of 23 VACs, most of which 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 V n n ditransitive, and the *way* construction.

Method

Step 1 Construction inventory: Cobuild Verb Patterns The VACs described in *Verb Grammar Patterns* take the form of word class and lexis combinations, such as V *across* n:

> 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 <u>cut</u> 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. For each VAC, we translated 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 constructed queries for each of the construction patterns. For example, the V *across* n pattern was 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.

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-tailed 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 such the same type of plot for verb type frequency of the ditransitive V *of* n construction. Both distributions produce a good fit of Zipfian type-token frequency. 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.

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.

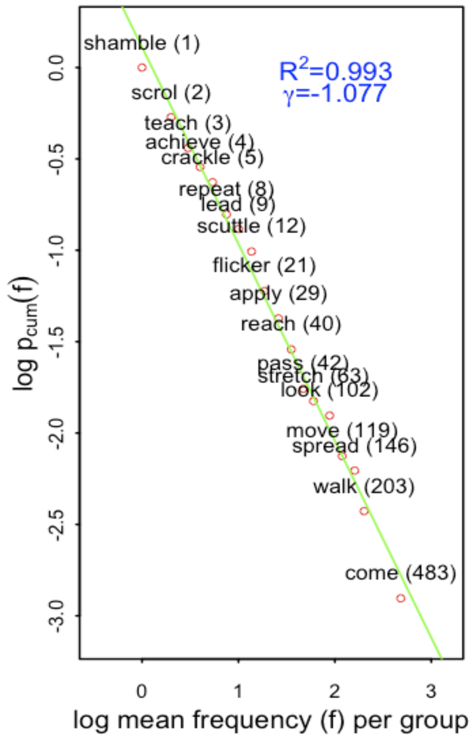


Figure 1 Type-token distribution for V *across* n

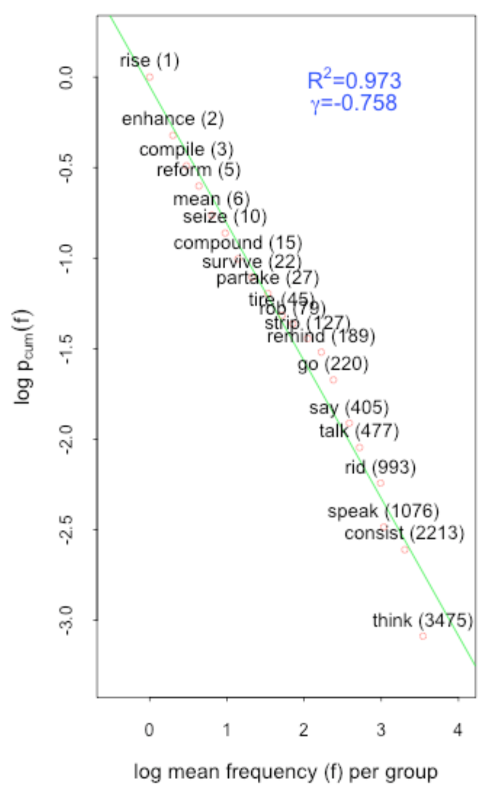


Figure 2 Type-token distribution for V *of* n

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 – 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 → Word: *give* 0.314, *leave* 0.003) and (ΔP Word → Construction: *give* 0.025, *leave* 0.001), and (3) directional mutual information (MI Word → Construction: *give* 16.26, *leave* 11.73 and MI Construction → Word: *give* 12.61 *leave* 9.11), an information science statistic that has been shown to predict language processing fluency.

Step 6 Identifying the meaning of verb types occupying the constructions

Our semantic analyses use WordNet (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 and hyponym. 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, & Michelizzi, 2004). Polysemy is a significant issue when analyzing verb semantics. For example, in WordNet the lemma forms *move*, *run* and *give* are found in 16, 41 and 44 different synsets respectively. To address this we applied word sense disambiguation tools specifically designed to work with WordNet (Pedersen & Kolhatkar, 2009) to the sentences retrieved at Step 3.

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 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 the VACs and their yoked CECs later in Table 1.

Step 8 Evaluating semantic cohesion in the VAC distributions

The VAC type-token lists 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 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 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 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 measures of semantic similarity using Pedersen et al's (2004) 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 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.

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 mean values on the metrics we have described so far are contrasted for the 23 VACs and their yoked CECs in Table 1.

Table 1: A comparison of 23 VACs and CECs for distribution, contingency, and semantic cohesion

Pattern	Mean VACs	Mean CECs	p value for paired t-test (d.f. 22)
R^2	0.98	0.96	1.6 e-06 ***
γ	-1.00	-1.12	4.4 e-06 ***
Entropy	4.97	5.54	4.9 e-04 ***
χ^2	69412	698	5.5 e-18 ***
1-τ	0.76	0.21	1.9 e-03 ***
Mean MIw-c	14.16	12.8	1.1 e-02 ***
Mean ΔPc-w	0.006	0.004	5.1 e-05 ***
Type entropy per root synset	3.1	3.51	1.7 e-08 ***
Token entropy per root synset	2.41	3.08	1.2 e-10 ***
Proportion of tokens covered by top 3 synsets	0.26	0.11	3.2 e-08 ***
lch	0.134	0.094	2.0 e-04 ***
res	0.237	0.22	1.6 e-06 ***

These results demonstrate:

(1) Type-token usage distributions All of the VACs are Zipfian in their type-token distributions in usage (VACs: M γ = -1.00, M R^2 = 0.98). So too are their matched CECs (M γ = -1.12, M 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-τ) between VAC verb frequency and overall verb frequency in the language is much greater in the VACs. Individual verbs select particular constructions (M MIw-c) and particular constructions select particular words (M ΔPc-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).

Discussion

We have shown for these 23 constructions:

- 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.

Assessing Psychological Validity of these VAC structures

We have shown these structural properties of VACs in usage. But are these also the structural properties of VAC representations in the minds of language users? Are these structural properties psychologically valid? We used free association tasks to have people think of the first word that comes to mind to fill the V slot in a particular VAC frame. The range of the verbs that they generate, and their speed of access, inform us about the representation of these VACs in the human mind.

Method

A convenience sample of 274 native English speakers volunteered for a free-association task over the internet. They were asked to type the first verb that came to mind to fill frames for 20 VACS given as pronoun_v-slot_determiner frames such as he __ *across* the... , it __ *across* the... , he __ *of* the... , it __ *of* the..., etc. Their responses were collated across VACs and the distributions assessed for the degree to which they accorded the usage statistics detemined in the previous corpus analyses.

Results

There were strong correspondences between people's free associations to particular VAC frames and the frequencies of verb exemplars in natural usage. We illustrate this in Figures 3 and 4 with the data for the V *across* n and V *of* n VACs. The fact that frames even as apparently abstract as v *of* n generate clusters of appropriate mentation verbs such as *think*, *know*, perception verbs such as *speak, hear, tell*, and perception verbs such as *smell, reek* make it clear that there are strong psychological associations between particular verbs semantics and particular VAC syntagmatics, i.e., that VACs are psychologically real.

Figure 4: V across n, English L1

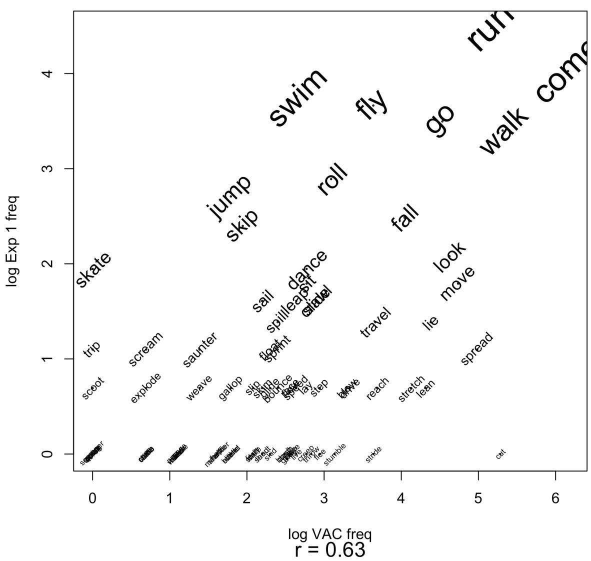
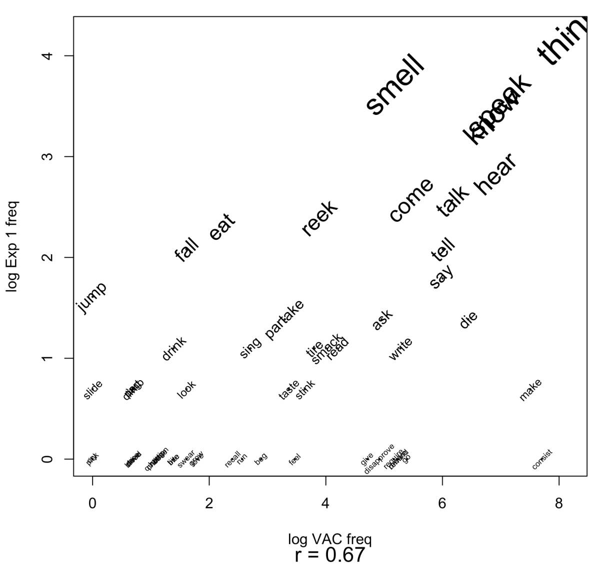


Figure 4: V of n, English L1



Conclusions – Robustness in Language and Other Complex Adaptive Systems

We have shown that Zipfian scale-free type-token distributions in language focus-forge together characteristic semantic functions and characteristic syntactic frames, both in language usage and in language cognition. Complex systems are characterised 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).

References

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., et al. (2009). Language is a complex adaptive system. Position paper. *Language Learning, 59 Supplement 1*, 1-26.

BNC (2007). BNC XML Edition http://www.natcorp.ox.ac.uk/corpus/.

Bybee, J. (2010). *Language, usage, and cognition*. Cambridge: Cambridge University Press.

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24*(2), 143-188.

Ellis, N. C., & Cadierno, T. (2009). Constructing a second language. *Annual Review of Cognitive Linguistics, 7* (Special section), 111-290.

Ellis, N. C., & Ferreira-Junior, F. (2009). Constructions and their acquisition: Islands and the distinctiveness of their occupancy. *Annual Review of Cognitive Linguistics*, 111-139.

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

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

Hunston, S., & Francis, G. (1996). *Pattern grammar: A corpus driven approach to the lexical grammar of English*. Amsterdam: Benjamins.

Miller, G. A. (2009). *WordNet - About us*. Retrieved March 1, 2010, from Princeton University: http://wordnet.princeton.edu

Page, S. E. (2009). *Understanding Complexity* [DVD-ROM]. Chantilly, VA: The Teaching Company.

Pedersen, T., & Kolhatkar, V. (2009). *WordNet:: SenseRelate:: AllWords*. Paper presented at the Tenth Annual Meeting of the North American Chapter of the Association for Computational Linguistics, Boulder, Colorado.

Pedersen, T., Patwardhan, S., & Michelizzi, J. (2004). *WordNet::Similarity – Measuring the Relatedness of Concepts*. Paper presented at the Fifth Annual Meeting of the North American Chapter of the Association of Computational Linguistics (NAACL 2004).

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

Solé, R. V., Murtra, B., Valverde, S., & Steels, L. (2005). Language Networks: their structure, function and evolution. *Trends in Cognitive Sciences, 12*.

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

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