7 Usage-Based Language Learning

NICK C. ELLIS, MATTHEW BROOK O'DONNELL, AND UTE RÖMER

1. Introduction

As a child, you talked with your significant others about things of shared interest using words and phrases that came to mind, and all the while you learned language. We authors of this paper were privy to none of your language experience. Nor did you have access to ours. Yet somehow we have all converged upon a similar enough "English" to be able to communicate here. How so?

Quine argued that the robustness of language lies in the commonalities of language usage, E pluribus unum:

> Each of us learns his [or her] language from other people, through the observable mouthings of words under conspicuously intersubjective circumstances. Linguistically, and hence conceptually, the things in sharpest focus are the things that are public enough to be talked of publicly, common and conspicuous enough to be talked of often, and near enough to sense to be quickly identified and learned by name; it is to these that words apply first and foremost. (Quine, 1960: 1)
>
> The uniformity that unites us in communication and belief is a uniformity of resultant patterns overlying a chaotic subjective diversity of connections between words and experience. Uniformity comes where it matters socially. (Quine, 1960: 8)

Hence his metaphor of language as a topiary:

> Different persons growing up in the same language are like different bushes trimmed and trained to take the shape of identical elephants. The anatomical detail of twigs and branches will fulfill the elephantine form differently from bush to bush, but the overall outward results are alike. (Quine, 1960: 8)

Brown (1958) analyzed these regularities in the ways in which adults name things for children and their preference for one of many possible names. Adults give a thing the name it has most commonly been given in their experience. This most common name

The Handbook of Language Emergence, First Edition. Edited by Brian MacWhinney and William O'Grady.

categorizes things as they need to be categorized for the communities' non-linguistic purposes, so the most common name is at the level of usual utility:

> The occasion for a name is ordinarily some particular thing. In the naming it is categorized. … The preference among possible names seems to go to the one that is most commonly applied to the referent in question. That name will ordinarily categorize the referent so as to observe the equivalences and differences that figure in its usual utilization. (Brown, 1958: 18)

Rosch, Mervis, Gray, Johnson, and Boyes-Braem (1976) showed how basic categories, those that carry the most information in clustering the things of the world, are those whose members possess significant numbers of attributes in common, are visually imageable with similar shapes, and have associated motor programs which are similar to one another. Basic categories are also those that are the most codable (naming is faster), most coded, and are most frequently utilized. Children acquire basic-category terms like *dog*, *hammer*, *apple* earlier than they do their superordinates *animal*, *tool*, *fruit*, or subordinates *collie*, *ball-peen hammer*, *Granny Smith*. Arguably, it is the reliable coming together of visual and motor-perceptual experience along with frequent and highly contingent labels which makes these nouns reliably and robustly learnable.

E. Clark (chapter 15, this volume) analyzes the ways in which adults and children establish the "common ground" that is necessary for the intersubjective circumstances of communication. She argues that children's social experience of their surroundings and highly repetitive routines underpins knowledge of perceptual and conceptual categories, and that their participation in reciprocal games, object-exchanges, and proto-turn-taking establishes conceptual and social knowledge combined with a general attentiveness to the other in interaction. Establishing common ground requires joint attention, physical co-presence, and conversational co-presence. This, plus adult feedback when children express something in a non-conventional fashion, shapes the language:

> Adult feedback consistently provides conventional forms, whether phonological or syntactic, morphological or lexical. These are the forms that children need in order to understand the intentions of others, and to convey their own intentions and be understood. Mastery of these conventions plays a central role for common ground: knowledge of a language and its use offers extensive communal common ground with other users of that language and so allows for more extensive and detailed communication of both needs and interests. Finally, adult reformulations of child errors also attest to the importance of interaction for the acquisition of language. It is in conversation that children master the conventions and so also learn how to use common ground. (Clark, chapter 15, this volume: 338)

These ideas are key to our understanding of how, as Brown had it, a thing shall be called. And they show commonalities of emphasis over the last fifty years of thinking on these matters. But what about actions and verbs? How shall doing be called?

Cognitive linguistics, particularly construction grammar, has since extended these ideas to language as a whole. It is not just that nouns typically relate to the things of the world, but, because language has emerged to describe our experiences of the world, so whole sentences are used to describe the doing of the referents of nouns in our world of experiences. Linguistic constructions which correspond to basic sentence types encode as their prototypical senses event types that are basic to human experience: those of something moving, something being in a state, someone causing something, someone possessing something, something causing a change of state or location, someone causing a change of possession, something undergoing a change of state or location, something having an effect on someone, etc. (Croft, 2001, 2012; Goldberg, 1995; Levin, 1993).

The prominence of particular aspects of the scene and the perspective of the internal observer (i.e. the attentional focus of the speaker and the intended attentional focus of the listener) are key elements in determining regularities of association between elements of visuo-spatial experience and elements of phonological form. In language comprehension, abstract linguistic constructions (like locatives, datives, and passives) guide the listener's attention to a particular perspective on a scene while backgrounding other aspects (Langacker, 1987; MacWhinney, 1998, 1999; Talmy, 2000; Taylor, 2002). In child language, research on the learning of verbs and constructions emphasizes the importance of item-based patterns and their perceptual groundings in acquisition (Goldberg, 2006; MacWhinney, 1999; Tomasello, 1992).

Could these associations of form and function allow, by processes of syntactic and semantic bootstrapping, linguistic constructions to be learnable, exemplar by exemplar, with abstract schematic patterns being induced from the conspiracy of particular usage patterns and their interpretations? Researching this question requires interdisciplinary collaborations: The investigation of form requires structuralist, corpus-linguistic, and computational-linguistic approaches. The investigation of function requires functionalist, cognitive-linguistic, and psycholinguistic analyses, the study of embodied force dynamics, an understanding of semantic organization, and more. Their association requires quantitative linguistics for the statistical tallying of form and function as well as an understanding of the psychology of learning. The result of these collaborations will not be a dictionary, nor a grammar, nor a frequency list. Rather it should be a systemic network integrating the syntactic constructions of a language, the lexis they select, their meanings, and the distributions and mappings of these forms and functions.

In what follows we sketch how we believe this work might progress, illustrating it with some preliminary findings of ongoing investigations of our own where we focus upon a large corpus-linguistic analysis of English usage of verb–argument constructions (VACs), including verb locative (VL), e.g. "V *across* n" as in "she walked across the street," verb object locative (VOL), e.g. "VO *in* n" as in "he put it in the fridge," and double object ditransitive (VOO) as in "she gave him a present." These initial studies convince us that the evidence of language usage is rich in latent structure. Learners' explorations of this problem-space are grounded and contextualized. There is much latent structure to scaffold development in the frequency distributions of exemplars of linguistic constructions and in the network structure of the corresponding semantic space.

Our shared language understanding suggests that, just as for nouns, there is a basic variety of VACs each with its basic-level verb core: Despite the fact that we have not heard the same input, our experience allows us similar interpretations of novel utterances like "it mandools 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 previously heard these verbs? Each word of the construction contributes individual meaning, and the verb meanings in these VACs are usually at the heart. But the larger configuration of words as a whole carries meaning too. The VAC as a category has inherited its schematic meaning from the conspiracy of 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*.

As you read these utterances, you parse them and identify their syntagmatic form: "it mandools across the ground" as a verb locative (VL) construction, "the teacher spugged the boy the book" as a double object (VOO) construction. Then the paradigmatic associations of the types of verb that fill these slots are awakened: for the VL "V *across* n" pattern *come*, *walk*, *move*, … , *scud*, *skitter*, and *flit*, for VOO *give*, *send*, *pass*, … , *read*, *loan*, and *fax*. Knowledge of language is based on these types of inference of syntactic and semantic bootstrapping, and verbs are the cornerstone of the syntax–semantics interface, which is why we focus upon VACs in our research.

In the rest of this chapter, first we will consider the nature of VACs and the psychology of their learning. Then we will turn to a complex systems analysis of the dynamic structure of language usage and how it might support robust language learning.

2. Constructions and Their Acquisition

2.1 *Construction grammar*

We take the Saussurean (1916) view that the units of language are constructions; 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 (Bates and MacWhinney, 1981; Goldberg, 1995, 2006). 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., Ambridge and Lieven, 2011; Ellis, 2011; Tomasello, 2003).

2.2 *The psychology of learning*

Usage-based approaches hold that we learn linguistic constructions while engaging in communication (Bybee, 2010). 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 that 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., Ellis, 2006a; MacWhinney, 2001).

Constructionist accounts of language learning 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 integrated experience of concrete exemplars of usage following statistical learning mechanisms (Bybee and Hopper, 2001; Christiansen and Chater, 2001; Ellis, 2006b; Jurafsky and Martin, 2009) that relate input and learner cognition.

2.3 Determinants of construction learning

Psychological analysis 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) form frequency in the input (type−token frequency, Zipfian distribution), (2) function (prototypicality of meaning), and (3) interactions between these (contingency of form−function mapping) (Ellis and Cadierno, 2009).

2.3.1 Construction frequency Frequency of exposure promotes learning and entrenchment (e.g., Anderson, 2000; Ebbinghaus, 1913). 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 times we experience conjunctions of features, the more they become associated in our minds and the more these subsequently affect perception and categorization (Harnad, 1987; Lakoff, 1987; Taylor, 1998).

2.3.2 Type and token frequency Token frequency counts how often a particular form appears in the input. 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. 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 constructions is a function of type rather than token frequency (Bybee and Hopper, 2001).

2.3.3 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. Thus in English, the most frequent word (*the* with a token frequency of ~60,000/million words) occurs approximately twice as often as the second most frequent word, three times as often as the third most frequent word, etc. 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: words (Evert, 2005), collocations (Bannard and Lieven, 2009; Solé, Murtra, Valverde, and Steels, 2005), formulaic phrases (O'Donnell and Ellis, 2009), morphosyntactic productivity (Baayen, 2008), grammatical constructs (Ninio, 2006; O'Donnell and Ellis, 2010), and grammatical dependencies (Ferrer i Cancho and Solé, 2001, 2003; Ferrer i Cancho, Solé, and Köhler, 2004). 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 (Manin, 2008; Steyvers and Tenenbaum, 2005). Scale-free laws pervade language structure and usage.

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, populations of cities, citations of scientific papers, number of hits received by web sites, sizes of airline hubs, perceptual psychophysics, memory, categorization) (Kello, Brown, Ferrer-i-Cancho et al., 2010; Newman, 2005). It has become a hallmark of Complex Systems theory. It is tempting to think of Zipfian scale-free laws as universals. Complexity theorists suspect them to be fundamental, and are beginning to investigate how they might underlie language processing, learnability, acquisition, usage, and change (Beckner, Blythe, Bybee et al., 2009; Ellis and Larsen-Freeman, 2009b; Ferrer i Cancho and Solé, 2001, 2003; Ferrer i Cancho, Solé, and Köhler, 2004; Solé et al., 2005).

Various usage-based linguists (e.g., Boyd and Goldberg, 2009; Bybee, 2008, 2010; Ellis 2009; Goldberg, 2006; Goldberg, Casenhiser, and Sethuraman, 2004; Lieven and Tomasello, 2008; Ninio, 1999, 2006) suspect that it is the Zipfian coming together of linguistic form and function that makes language robustly learnable despite learners' idiosyncratic experience. For example, in first language acquisition, Goldberg, Casenhiser and Sethuraman (2004) demonstrated that there is a strong tendency for VL, verb object locative (VOL), and double object ditransitive (VOO) 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 because the 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.

2.3.4 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 the central ideal is the center of the category, appropriately summarizing the most representative attributes of a category. In some types of category, there might indeed be one particular exemplar that is typical of the instances of a category, serving as the benchmark against which surrounding, less representative instances are classified. In others, there might be no one prototype, but as in network structures such as airline connectivity or the world wide web, there might be a well-connected group of high-degree hubs at the core.

In child language acquisition, a small group of semantically general verbs (e.g., *go*, *do*, *make*, *come*) are learned early (Clark, 1978; Goldberg, 2006; Ninio, 1999; Pinker, 1989). Ellis and Ferreira-Junior (2009a) show the same is true of the second language acquisition of VL, VOL, and VOO constructions. These first verbs are prototypical and generic in function (*go* for VL, *put* for VOL, and *give* for VOO). In the early stages of learning categories from exemplars, acquisition might thus be optimized by the introduction of an initial, low-variance sample centered upon prototypical exemplars.

You will have noticed how our conclusions here concerning function converge upon those concerning frequency of form in 2.3.3.

2.3.5 Contingency of form–function mapping Psychological research into associative learning has long recognized that while frequency of form is important, contingency of mapping is more so (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 that are distinctive in differentiating birds from other animals. Wings are important features for learning the category of birds, because they are reliably associated with class membership, whereas eyes are not. Raw frequency of occurrence is less important than the contingency between cue and interpretation (Rescorla, 1968). Contingency, reliability of form–function mapping and associated aspects of predictive value, information gain, and statistical association are driving forces of learning. They are central in psycholinguistic theories of language acquisition too (Ellis, 2006a, 2006b, 2008; Gries and Wulff, 2005; MacWhinney, 1987).

2.4 *Usage-based acquisition*

The primary motivation of construction grammar is that we must bring together linguistic form, learner cognition, and usage. 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, and have (4) a high contingency between form and function.

A necessary research program, then, is to measure whether language usage provides experience of this type. If so, then VACs as linguistic constructions should be robustly learnable. Is language, shaped by the human brain (Christiansen and Chater, 2008), consequently shaped *for* the human brain in that the structures latent in language usage make language robustly learnable?

3. Language Usage as a Complex Adaptive System

This fundamental claim that Zipfian distributional properties of language usage help to make language learnable has just begun to be explored for a small number of VACs in first (Goldberg, 2006; Goldberg et al., 2004; Ninio, 2006, 2011) and second (Ellis and Ferreira-Junior, 2009a, 2009b; Ellis and Larsen-Freeman, 2009a) language acquisition. It remains a priority to explore its generality across the wide range of English verbal grammar. We need (1) an integrative analysis of the VACs of a language, the lexis they select, their meanings, and the distributions and mappings of these forms and functions, (2) to determine if the latent structures therein are of the type that would promote robust learning, and (3) to measure corpora of learner language (L1 and L2) to see if learning is shaped by the input. This is no small task. Here we describe some of our pilot work and suggestions for further research.

3.1 A usage-based grammar of English verbs and verb–argument constructions

3.1.1 A catalogue of VACs In order to avoid circularity, the determination of the semantic associations of particular linguistic forms should start from structuralist definitions of VACs 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. We therefore use the patterns presented in *Grammar Patterns: Verbs* (Francis, Hunston, and Manning, 1996) that arose out of the COBUILD project (Sinclair, 1987) for our initial analyses. There are over 700 form-based patterns of varying complexity in this volume. These 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.*

Our initial research (for further details see Ellis and O'Donnell, 2012; Römer, O'Donnell, and Ellis, 2014) describes the methods and findings for an initial 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.

3.1.2 A large corpus of English To get a representative sample of usage, one needs a large corpus. We investigate the verb type–token distribution of these VACs in the 100-million-word British National Corpus (2007), parsed using the XML version and 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. Using a combination of part-of-speech, lemma and dependency constraints we formulate queries for each of the construction patterns. For example, the "V *across* n" pattern is identified by looking for sentences in which (1) a verb form occurs within three words of an instance of *across* as a preposition, (2) a selectional relationship holds between *across* and the verb, and (3) the verb does not have any other object or complement relations to following words in the sentence (Römer, O'Donnell, and Ellis, 2014).

3.1.3 The lexical constituency of verbs in VACs The sentences extracted using this procedure produced verb type-frequency distributions like the following one for the "V *across* n" VAC:

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 tail in a plot of rank against frequency. We generated logarithmic plots and linear regressions to examine the extent of this trend using logarithmic binning of frequency against log cumulative frequency following Adamic and Huberman (2002). Figure 7.1 shows such a plot for verb type frequency of the "V *across* n" construction alongside the same plot for verb type frequency of the ditransitive "V n n" construction. In these graphs we randomly select one verb from each frequency bin for illustration. Both distributions

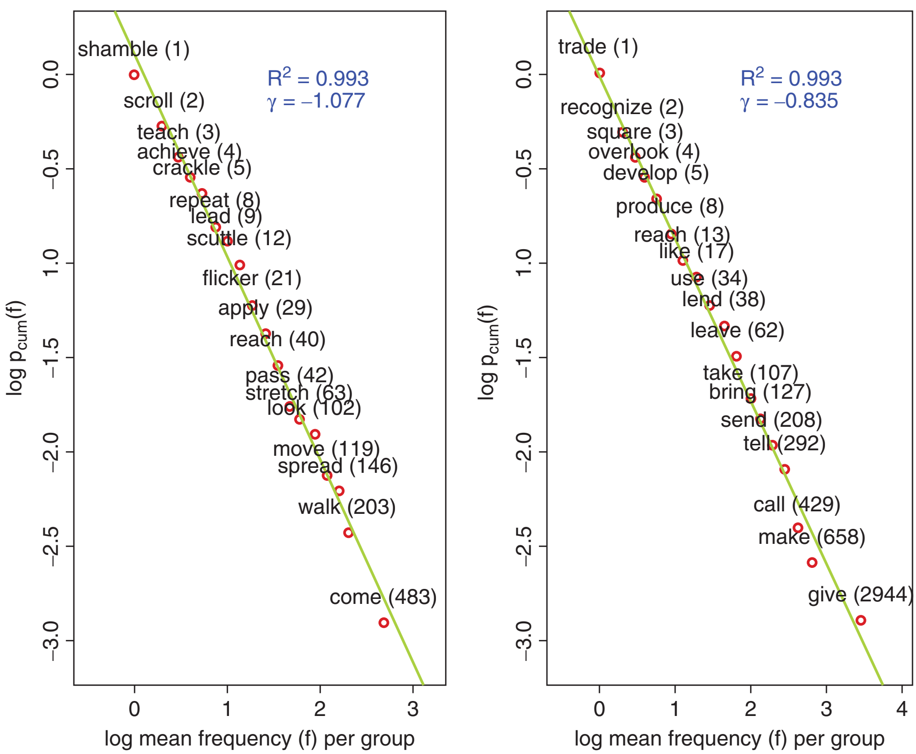

Figure 7.1. BNC verb type distribution for "V across n" (left) and for "V n n" (right) (Ellis, O'Donnell, and Römer, 2013)

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.

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 measures of association 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. For the VACs studied so far, 1−tau is typically about 0.7, showing that the rankings of verbs in particular VACs differ markedly from the rankings of verbs in the language as a whole. VACs are selective in their verb constituency.

3.1.4 Verb−VAC contingency Another way of looking at this is to assess verb−VAC contingency. 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, and higher-contingency verbs should be learned first in a VAC and should first come to mind when processing that VAC. The measures of contingency we adopt are (1) faithfulness – 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, (2) directional mutual information (MI Word → Construction: *give* 16.26, *leave* 11.73, and MI Construction → Word: *give* 12.61, *leave* 9.11), and (3) the one-way dependency statistic ΔP (Allan, 1980) used in the associative learning literature (Shanks, 1995) and in other studies of form−function contingency in construction usage, knowledge, and processing (Ellis, 2006a; Ellis and Ferreira-Junior, 2009b). Our analyses for the 23 VACs studied so far show a general pattern whereby individual verbs tend to select particular constructions (MI_{wc}, ΔP_{wc}) and particular constructions select particular verbs (MI_{cw}, ΔP_{cw}) (for details see Ellis and O'Donnell, 2012; Ellis, O'Donnell, and Römer, 2013).

3.1.5 VAC meanings and coherence 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 a 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). We take the lists of verbs that occupy each VAC according to the methods described in 3.1.3 and compare the verbs pairwise on these metrics. We then apply the graph-based algorithms of networks science (de Nooy, Mrvar, and Batagelj, 2010) to build semantic networks in which the nodes represent verb types and the edges strong semantic similarity for each VAC. Standard measures of network density, average clustering, degree centrality, transitivity, etc. are used to assess the cohesion of these semantic networks, and we also apply algorithms for the detection of communities within the networks representing different semantic sets (Clauset, Newman, and Moore, 2004; Danon, Díaz-Guilera, Duch, and Arenas, 2005). The network for "V *across* n" is shown as an example in Figure 7.2. The network is fairly dense. The hubs, shown here as larger nodes, are those that are most connected, i.e., have the highest degree. They are *go*, *move*, and *travel* – the prototypical "V *across* n" senses. However, there are also subcommunities, shown in different colors in the on-line version: for example, one

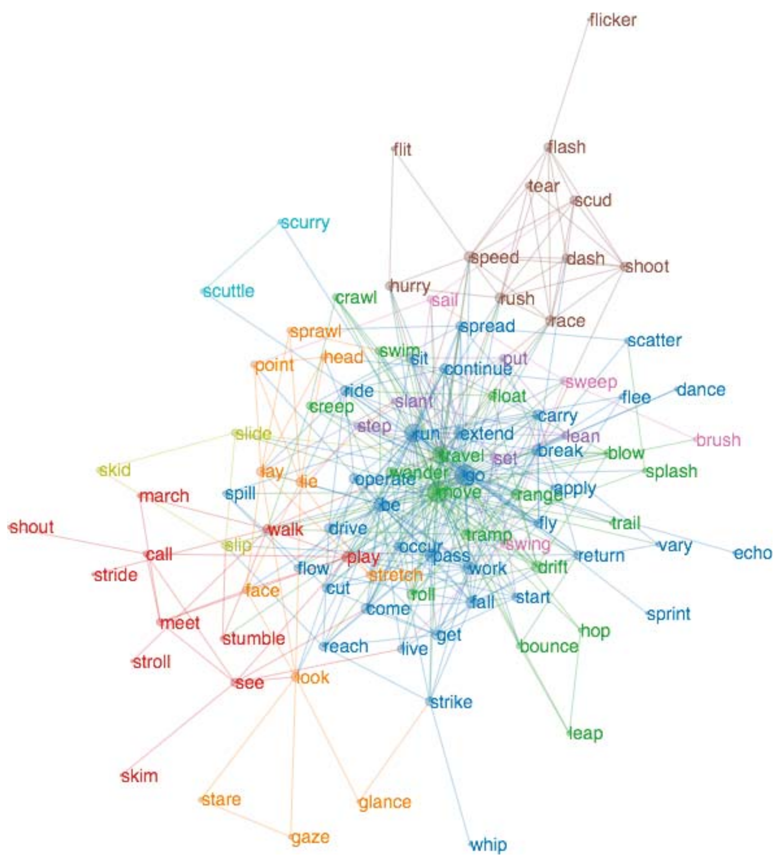

Figure 7.2. A semantic network for "V across n" from the BNC using WordNet as a base (Ellis, O'Donnell, and Römer, 2013)

relating to vision that includes *look*, *stare*, *gaze*, *face*, another to high-speed movement with unspecified action semantics: *shoot*, *scud*, *race*, *rush*, etc., and another emphasizing flat contact, *lay*, *lie*, *sprawl*, etc. Note that degree in the network is unrelated to token frequency in the corpus; it simply reflects verb type connectivity within the network. So too is "betweenness centrality," a measure developed to quantify the control of a human over the communication between other humans in a social network. Betweenness centrality is a measure of a node's centrality in a network equal to the number of shortest paths from all vertices to all others that pass through that node. In semantic networks, central nodes are those which are prototypical of the network as a whole.

Across the VACs we have investigated to date (O'Donnell, Ellis, and Corden, 2012), the semantic networks are coherent, with short average path-lengths between the nodes and having degree distributions which approximate a Zipfian power function. Satisfaction of these two criteria classifies them as "small-world networks" (Barabási, 2002). Steyvers and Tenenbaum (2005) show that, more generally across semantic networks such as Roget's thesaurus and WordNet, the distributions of the number of connections as a whole follow power laws that indicate a scale-free pattern of connectivity, with most nodes having relatively few connections joined together through a small number of hubs with many connections. Small-world networks are interesting because they are more robust to perturbations than other network architectures (Barabási, 2002). They achieve this through the high connectivity of their hubs.

3.2 Is the latent structure of VACs of the type that would promote robust acquisition following the learning principles proposed in section 2?

Our core research questions concern the degree to which VAC form, function, and usage might promote robust learning. As we explained in 2.3, 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, and (3) coherent in their semantics. The findings we summarize in 3.1.3−3.1.5 confirm that these factors apply.

4. Further Directions and Conclusions

These initial investigations make it clear that usage is intricately structured in ways typical of complex adaptive systems – there are scale-free distributions in verb usage frequency within constructions and scale-free connectivity patterns within semantic networks – latent structure that could potentially scaffold robust development.

Note that these results are preliminary, being based on an analysis of only about 20 constructions to date. It remains to explore a more complete verbal grammar of English. There are over 700 patterns of varying complexity in the *Grammar Patterns: Verbs* (Francis, Hunston, and Manning, 1996) – a lot to be done.

There is a considerable amount of additional statistical analysis and modeling to be done, as well as analyses of corpora of language acquisition (and the matching child-directed speech or native speaker interlocutor language) to test out the predictions of learning, and experimental psycholinguistic research to test the psychological validity of VACs.

We identify the following priorities:

- Learning theory, as applied to input structured in these ways, makes predictions about language acquisition and processing. VACs that have higher frequency in the language, with more verb types, and greater semantic cohesion, should be acquired earlier and accessed more readily in speakers' minds. Similarly, verbs which are more frequent constituents of a VAC, more faithful to that VAC, and closer to the VAC semantic prototype should be acquired earlier in that VAC and accessed more readily in speakers' minds when they consider that VAC schema. We are currently assessing these predictions in the UK and US English child language corpora available through CHILDES (MacWhinney, 2000).
- This work needs to be done for second language acquisition too. We have made a start (Ellis and Ferreira-Junior, 2009a, 2009b; Römer, Roberson, O'Donnell, and Ellis, 2013), but there is an imperative for larger L2 corpora. Second language acquisition is special too in that while, like L1A, it is driven by target-language usage, learners' interpretation of that usage is made through systems tuned by knowledge of a prior language, and so it is affected by transfer and learned attention (Ellis, 2008; Römer, O'Donnell, and Ellis, in press).
- The psychological reality of VACs needs further work. In various studies we use 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 responses of adult native and fluent L2 learners are highly predicted by corpus frequencies, showing that users have implicit knowledge of the occupancy and frequencies of verbs in VACs. But, further, we have identified separable independent influences of (1) the frequency of the verb in the language, (2) VAC–verb contingency (ΔP_{cw}), and (3) the prototypicality of the verb in the semantic network (as indexed by its betweenness centrality) (Ellis, O'Donnell, and Römer, 2014, in press).
- The semantic analyses here are crude; other distributional measures could well be applied alongside techniques for investigating network growth (O'Donnell et al., 2012).
- The acquisition data here are basically correlational. There need to be experimental studies comparing the relative learnability of Zipfian skewed input compared to constructions with flatter frequency profiles. Casenhiser and Goldberg (2005) and Goldberg et al. (2004) have made important steps in doing this in children and in adults, but they investigate the learning of just one construction from a small number of trials, and there is need for larger causal studies of the effects of combined Zipfian frequency distributions and Zipfian semantic connectivity upon more complete approximations to natural language.
- To better understand the processes of how these latent structures of usage affect robust acquisition and stable usage, there is need for modeling, both connectionist (Ellis and Larsen-Freeman, 2009a) and agent-based (Beckner et al., 2009).
- These processes are not going to hold for all grammatical constructions. There are more arbitrary subcategorizations (e.g., German nominal gender), as well as general facts of typology (an SOV language is almost certainly going to have postpositions

rather than prepositions, a language that allows indirect objects to be relativized is going to allow subjects and direct objects to be relativized), and so on, where the semantic motivations are less clear-cut (Diessel, 2004) and where, as discussed throughout this volume, a wider range of factors play out in their emergence over longer timescales.

However, we can at least say that for the VL, VOL, and VOO constructions studied here, the input from which learners acquire language is far from unstructured or unhelpful. The evidence of language usage is rich in latent structure. There are two different Zipfian distributions described in 3.1.3 and 3.1.5. The first relates to type–token frequency distribution in the language. The second relates to node connectivity (degree distribution and betweenness centrality) in the semantic network, and it has no regard of corpus frequencies. Ellis et al. (2014) and (Ellis et al., in press) show that these factors separately affect the association of verbs and VACs and hence their availability in processing. Nevertheless, the high-degree items in the semantic distribution tend also to be the high-token-frequency items in VAC usage. It is this coming together of the two Zipfian distributions, we believe, that makes language robustly learnable. The VAC pattern is seeded by a high-token-frequency exemplar that is also prototypical in the action-dynamic construal of events to which it relates. Thereafter the forms and functions of verbs added to the VACs resonate with the category itself. "[T]his process is actually the organization of exemplars of utterances and of verb-specific constructions into clusters of greater or lesser size, with greater or lesser syntactic or semantic coherence" (Croft, 2012: 393).

Roger Brown concluded his "naming of things" paper, "Adults give a thing the name it is most commonly given. We have now come full circle and are left with the question, 'Why is one name for a thing more common than another?' It seems likely that things are first named to categorize them in a maximally useful way" (Brown, 1958: 20). The convergence of the two Zipfian distributions in our statistical analyses of VACs in usage here shows these principles to hold for the naming of doing too.

Acknowledgments

We are grateful to the University of Michigan "LSA Scholarship/Research" funding opportunity for supporting the project "Piloting the development of an inventory of usage of English verb grammar," and to Brian MacWhinney and William O'Grady for helpful comments on earlier versions.

REFERENCES

Adamic, L. A. and B. A. Huberman. 2002. Zipf's law and the Internet. *Glottometrics* 3: 143–150.

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

Ambridge, B. and E. Lieven. 2011. *Child Language Acquisition: Contrasting Theoretical Approaches*. Cambridge: Cambridge University Press.

Anderson, J. R. 2000. *Cognitive Psychology and its Implications*. 5th ed. New York: W.H. Freeman.

Baayen, R. H. 2008. Corpus linguistics in morphology: Morphological productivity. In A. Lüdeling and M. Kytö (eds.), *Corpus Linguistics: An International Handbook*, pp. 899–919. Berlin: Mouton de Gruyter.

Bannard, C. and E. Lieven. 2009. Repetition and reuse in child language learning. In R. Corrigan, E. A. Moravcsik, H. Ouali, and K. M. Wheatley (eds.), *Formulaic Language. Volume 2: Acquisition, Loss, Psychological Reality, and Functional Explanations*, pp. 299–321. Amsterdam: John Benjamins.

Barabási, A.-L. 2002. *Linked: The New Science of Networks*. Cambridge, MA: Perseus Books.

Bates, E. and B. MacWhinney. 1981. Second-language acquisition from a functionalist perspective: Pragmatic, semantic, and perceptual strategies. *Annals of the New York Academy of Sciences* 379: 190–214.

Beckner, C., R. Blythe, J. Bybee, M. H. Christiansen, W. Croft, N. C. Ellis, J. Holland, Jinyun Ke, Diane Larsen-Freeman and T. Schoenemann. 2009. Language is a complex adaptive system: Position paper. *Language Learning* 59, Supplement 1: 1–26.

Boyd, J. K. and A. E. Goldberg. 2009. Input effects within a constructionist framework. *Modern Language Journal* 93(2): 418–429.

Briscoe, E., J. Carroll, and R. Watson. 2006. The second release of the RASP system. *COLING-ACL '06: Proceedings of the COLING/ACL on Interactive Presentation Sessions*, pp. 77–80. Stroudsburg, PA: Association for Computational Linguistics.

British National Corpus. 2007. *BNC XML Edition*. http://www.natcorp.ox.ac.uk/corpus/.

Brown, R. 1958. How shall a thing be called? *Psychological Review* 65: 14–21.

Bybee, J. 2008. Usage-based grammar and second language acquisition. In P. Robinson and N. C. Ellis (eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition*, pp. 216–236. London: Routledge.

Bybee, J. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.

Bybee, J. and P. Hopper (eds.). 2001. *Frequency and the Emergence of Linguistic Structure*. Amsterdam: Benjamins.

Casenhiser, D. and A. E. Goldberg. 2005. Fast mapping between a phrasal form and meaning. *Developmental Science* 8: 500–508.

Christiansen, M. H. and N. Chater (eds.). 2001. *Connectionist Psycholinguistics*. Westport, CO: Ablex.

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

Clark, E. V. 1978. Discovering what words can do. In D. Farkas, W. M. Jacobsen, and K. W. Todrys (eds.), *Papers from the Parasession on the Lexicon, Chicago Linguistic Society April 14–15, 1978*, pp. 34–57. Chicago: Chicago Linguistic Society.

Clauset, A., M. E. J. Newman, and C. Moore. 2004. Finding community structure in very large networks. *Physical Review E* 70: 066111.

Croft, W. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

Croft, W. 2012. *Verbs: Aspect and Causal Structure*. Oxford: Oxford University Press.

Danon, L., A. Díaz-Guilera, J. Duch, and A. Arenas. 2005. Comparing community structure identification methods. *Journal of Statistical Mechanics* 29: P09008.

de Nooy, W., A. Mrvar, and V. Batagelj. 2010. *Exploratory Social Network Analysis with Pajek*. Cambridge: Cambridge University Press.

Diessel, H. 2004. *The Acquisition of Complex Sentences in English*. Cambridge: Cambridge University Press.

Ebbinghaus, H. 1913. *Memory: A Contribution to Experimental Psychology* (trans. H. A. Ruger and C. E. Bussenius). New York: Teachers College, Columbia University. Originally published as *Über das Gedächtnis*, 1885.

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. 2006a. Language acquisition as rational contingency learning. *Applied Linguistics* 27(1): 1–24.

Ellis, N. 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, N. C. 2008. Usage-based and form-focused language acquisition: The associative learning of constructions, learned-attention, and the limited L2 endstate. In P. Robinson and N. C. Ellis (eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition*, pp. 372–405. London: Routledge.

Ellis, N. C. 2009. Optimizing the input: Frequency and sampling in usage-based and form-focused learning. In M. H. Long and C. Doughty (eds.), *The Handbook of Language Teaching*, pp. 139–158. Chichester: Wiley-Blackwell.

Ellis, N. C. 2011. The emergence of language as a complex adaptive system. In J. Simpson (ed.), *The Routledge Handbook of Applied Linguistics*, pp. 654–667. London: Routledge.

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

Ellis, N. C. and F. Ferreira-Junior. 2009a. Construction learning as a function of frequency, frequency distribution, and function. *Modern Language Journal* 93: 370–386.

Ellis, N. C. and F. Ferreira-Junior. 2009b. Constructions and their acquisition: Islands and the distinctiveness of their occupancy. *Annual Review of Cognitive Linguistics* 7: 111–139.

Ellis, N. C. and D. 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, N. C. and D. Larsen-Freeman (eds.). 2009b. *Language as a Complex Adaptive System. Language Learning* 59: Supplement 1.

Ellis, N. C. and M. B. O'Donnell. 2012. Statistical construction learning: Does a Zipfian problem space ensure robust language learning? In J. Rebuschat and J. Williams (eds.), *Statistical Learning and Language Acquisition*, pp. 265–304. Berlin: Mouton de Gruyter.

Ellis, N. C., M. B. O'Donnell, and U. Römer. 2013. Usage-based language: Investigating the latent structures that underpin acquisition. *Language Learning* 63, Supplement 1: 25–51.

Ellis, N. C., M. B. O'Donnell, and U. Römer. 2014. The processing of verb-argument constructions is sensitive to form, function, frequency, contingency, and prototypicality. *Cognitive Linguistics* 25(1): 55–98.

Ellis, N. C., M. B. O'Donnell, and U. Römer. In press. Second language verb–argument constructions are sensitive to form, function, frequency, contingency, and prototypicality. *Linguistic Approaches to Bilingualism* 4(4).

Evert, S. 2005. The statistics of word cooccurrences: Word pairs and collocations. Doctoral thesis, University of Stuttgart.

Ferrer i Cancho, R. and R. V. Solé. 2001. The small world of human language. *Proceedings of the Royal Society of London, B* 268: 2261–2265.

Ferrer i Cancho, R. and R. V. Solé. 2003. Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences* 100: 788–791.

Ferrer i Cancho, R., R. V. Solé, and R. Köhler. 2004. Patterns in syntactic dependency networks. *Physical Review E* 69: 051915-1–0519158.

Francis, G., S. Hunston, and E. Manning (eds.). 1996. *Grammar Patterns. 1: Verbs*. COBUILD Series. London: HarperCollins.

Goldberg, A. E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

Goldberg, A. E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

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

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

Harnad, S. (ed.). 1987. *Categorical Perception: The Groundwork of Cognition*. New York: Cambridge University Press.

Hunston, S. and G. Francis. 1996. *Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English*. Amsterdam: Benjamins.

Jurafsky, D. and J. H. Martin. 2009. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 2nd ed. Englewood Cliffs, NJ: Prentice-Hall.

Kello, C. T., G. D. A. Brown, R. Ferrer-i-Cancho, J. G. Holden, K. Linkenkaer-Hansen, T. Rhodes, and G. C. Van Orden. 2010. Scaling laws in cognitive sciences. *Trends in Cognitive Science* 14: 223–232.

Lakoff, G. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.

Langacker, R. W. 1987. *Foundations of Cognitive Grammar. Volume 1: Theoretical Prerequisites*. Stanford, CA: Stanford University Press.

Levin, B. 1993. *English Verb Classes and Alternations: A Preliminary Analysis*. Chicago: Chicago University Press.

Lieven, E. and M. Tomasello. 2008. Children's first language acquisition from a usage-based perspective. In P. Robinson and N. C. Ellis (eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition*, pp. 168–196. New York and London: Routledge.

MacWhinney, B. 1987. The competition model. In B. MacWhinney (ed.), *Mechanisms of Language Acquisition*, pp. 249–308. Hillsdale, NJ: Erlbaum.

MacWhinney, B. 1998. Models of the emergence of language. *Annual Review of Psychology* 49: 199–227.

MacWhinney, B. 1999. The emergence of language from embodiment. In B. MacWhinney (ed.), *The Emergence of Language*, pp. 213–256. Mahwah, NJ: Lawrence Erlbaum.

MacWhinney, B. 2000. *The CHILDES Project: Tools for Analyzing Talk. Volume 2: The Database*. 3rd ed. Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney, B. 2001. The competition model: The input, the context, and the brain. In P. Robinson (ed.), *Cognition and Second Language Instruction*, pp. 69–90. New York: Cambridge University Press.

Manin, D. Y. 2008. Zipf's Law and avoidance of excessive synonymy. *Cognitive Science* 32: 1075–1098.

Miller, G. A. 2009. WordNet – About us. http://wordnet.princeton.edu. Retrieved March 1, 2010.

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

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

Ninio, A. 2006. *Language and the Learning Curve: A New Theory of Syntactic Development*. Oxford: Oxford University Press.

Ninio, A. 2011. *Syntactic Development, its Input and Output*. Oxford: Oxford University Press.

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

O'Donnell, M. B. and N. C. Ellis. 2010. Towards an inventory of English verb argument constructions. In *Proceedings of the NAACL HLT Workshop on Extracting and Using Constructions in Computational Linguistics*, Los Angeles, California, June 2010, pp. 9–16. Stroudsburg, PA: Association for Computational Linguistics.

O'Donnell, M. B., N. C. Ellis, and G. Corden. 2012. Exploring semantics in verb argument constructions using community identification algorithms. Paper presented at the Language and Network Science symposium, International Conference on Network Science NETSCI 2012, June 18–22, Northwestern University.

Pedersen, T., S. Patwardhan, and J. Michelizzi, 2004. WordNet::Similarity – Measuring the relatedness of concepts. In *Proceedings of the Fifth Annual Meeting of the North American Chapter of the Association of Computational Linguistics (NAACL 2004)*. Stroudsburg, PA: Association for Computational Linguistics.

Pinker, S. 1989. *Learnability and Cognition: The Acquisition of Argument Structure*. Cambridge, MA: Bradford Books.

Quine, W. V. O. 1960. *Word and Object*. Cambridge, MA: MIT Press.

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

Römer, U., M. O'Donnell, and N. C. Ellis. 2014. Using COBUILD grammar patterns for a large-scale analysis of verb-argument constructions: Exploring corpus data and speaker knowledge. In M. Charles, N. Groom, and L. Suganthi (eds.), *Corpora, Grammar, Text and Discourse: In Honour of Susan Hunston*. Amsterdam: John Benjamins.

Römer, U., M. B. O'Donnell, and N. C. Ellis. In press. Second language learner knowledge of verb–argument constructions: Effects of language transfer and typology. *Modern Language Journal* 98(4).

Römer, U., A. Roberson, M. B. O'Donnell, and N. C. Ellis. 2013. Linking learner corpus and experimental data in studying second language learners' knowledge of verb-argument constructions. Paper presented at the 34th conference of the International

Computer Archive of Modern and Medieval English (ICAME), Santiago de Compostela, Spain.

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

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

Saussure, F. de., 1916. *Cours de linguistique générale*. Paris: Payot. English trans. *Course in General Linguistics* (trans. Roy Harris). 1983. London: Duckworth.

Shanks, D. R. 1995. *The Psychology of Associative Learning*. New York: Cambridge University Press.

Sinclair, J. M. (ed.). 1987. *Looking Up: An Account of the COBUILD Project in Lexical Computing*. London: Collins ELT.

Sinclair, J. M. 2004. *Trust the Text: Language, Corpus and Discourse*. London: Routledge.

Solé, R. V., B. Murtra, S. Valverde, and L. Steels. 2005. Language networks: Their structure, function and evolution. *Complexity* 15(6): 20–26.

Steyvers, M. and J. Tenenbaum. 2005. The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science* 29: 41–78.

Talmy, L. 2000. *Toward a Cognitive Semantics: Concept Structuring Systems*. Cambridge, MA: MIT Press.

Taylor, J. R. 1998. Syntactic constructions as prototype categories. In M. Tomasello (ed.), *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure*, pp. 177–202. Mahwah, NJ: Erlbaum.

Taylor, J. R. 2002. *Cognitive Grammar*. Oxford: Oxford University Press.

Tomasello, M. 1992. *First Verbs: A Case Study of Early Grammatical Development*. New York: Cambridge University Press.

Tomasello, M. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.

Zipf, G. K. 1935. *The Psycho-biology of Language: An Introduction to Dynamic Philology*. Cambridge, MA: MIT Press.