

Optimizing the Input:
Frequency and Sampling in Usage-based and Form-focussed Learning

Nick C. Ellis

Chapter for

Michael Long & Cathy Doughty (Eds.)

The Handbook of Second and Foreign Language Teaching

Blackwell Handbooks in Linguistics

Draft of July 17, 2007

Estimating how language works:

From tokens to types to systemⁱ

Learners' understanding of language and of how it works is based upon their experience of language. They have to estimate the system from a sample. This chapter considers the effects of input sample, construction frequency, and processing orientation on learning. It draws out implications for usage-based acquisition and form-focussed instruction for second (L2) and foreign (FL) language learners.

A language is not a fixed system. It varies in usage over speakers, places, and time. Yet despite the fact that no two speakers own an identical language, communication is possible to the degree that they share constructions (form-meaning correspondences) relevant to their discourseⁱⁱ. Language learners have to acquire these constructions from usage, and beginners don't have much to go on in building the foundations for basic interpersonal communication. They have to induce the types of construction from experience of a limited number of tokens. Their very limited exposure poses them the task of *estimating* how linguistic constructions work from an input sample that is incomplete, uncertain, and noisy. How do they achieve this, and what types of experience can best support the process?

Native-like fluency, idiomaticity, and selection are another level of difficulty again. For a good fit, every utterance has to be chosen, from a wide range of possible expressions, to be appropriate for that idea, for that speaker, for that place, and for that time. And again, learners can only *estimate* this from their finite experience. What are the best usage histories to support these abilities?

Language, a moving target, can neither be described nor experienced comprehensively, and so, in essence, *language learning is estimation from sample*. Like other estimation problems, successful determination of the population characteristics is a matter of statistical sampling, description, and inference. For language learning the estimations include: What is the range of constructions in the language? What are their major types? Which are the really useful ones? What is their relative frequency distribution? How do they map function and form, and how reliably so? How can this information best be organized to allow its appropriate and fluent access in recognition and production? Are there variable ways of expressing similar meanings? How are they distributed across different contexts? And so on. Etcetera. And so forth. Like.

Frequency of usage, in various guises, determines acquisition (Ellis, 2002a, 2002b). There are three fundamental aspects of this conception of language learning as statistical sampling and estimation.

- The first and foremost concerns *sample size*: As in all surveys, the bigger the sample, the more accurate the estimates, but also the greater the costs. Native speakers estimate their language over a lifespan of usage. L2 and FL learners just don't have that much time or resource. Thus, both of these *additional language* (AL) learner groups are faced with a task of optimizing their estimates of language from a limited sample of exposure. Broadly, power analysis dictates that attaining native-like fluency and idiomaticity requires much larger usage samples than does basic interpersonal communicative competence in predictable contexts. But for the particulars, what sort of sample is needed adequately to assess the workings of constructions of, respectively, high, medium, and low base

occurrence rates, of more categorical vs. more fuzzy patterns, of regular vs. irregular systems, of simple vs. complex ‘rules’, of dense vs. sparse neighbourhoods, etcetera?

- The second concerns *sample selection*: Principles of survey design dictate that a sample must properly represent the strata of the population of greatest concern. Thus, *Needs Analysis* (chapter 18) is relevant to all AL learners. Thus, too, the truism that FL learners, who have much more limited access to the authentic natural source language than L2 learners, are going to have greater problems of adequate description. But what about learning particular constructions? What is the best sample of experience to support this? How many examples do we need? In what proportion of types and tokens? Are there better sequences of experience to optimize estimation? What learning increment comes from each experience? Is this a constant or does it diminish over time as dictated by the power law of practice? And so forth.
- A final implication of language acquisition as estimation concerns *sampling history*: How does knowledge of some cues and constructions affect estimation of the function of others? What is the best sequence of language to promote learning new constructions? And what is the best processing orientation to make this sample of language the appropriate sample of usage? Like.

This chapter first describes the units of language acquisition – linguistic constructions – and then considers how sample size and sample selection affect the development of constructions (their consolidation, generalization, and probabilistic

tuning) from naturalistic input. There are established effects of input token frequency, type frequency, Zipfian frequency distributionⁱⁱⁱ of the construction-family, and neighbourhood homogeneity.

Next, it describes how sample size and sample selection affect usage-based language acquisition across the board -- native and AL both. It reviews how learners' models of language broadly reflect the constructions in their sample of experience and how they unconsciously tally and collate a rich knowledge of the relative frequencies of these constructions in their input history. Because language learning is less an issue of the collection of linguistic constructions than of their cataloguing, organization, and marshalling for efficient appropriate use, this *implicit* knowledge is essential to fluent processing. In order for the estimation procedures rationally to produce a model of the language that optimizes the probabilistic knowledge of constructions and their mappings, learners must be exposed to a representative sample of authentic input that is appropriate to their needs. It also considers the implications of modularity and transfer-appropriate processing for tuning the full range of necessary representative modalities and functions of usage.

Finally it nods at analyses of transfer in AL acquisition, how prior estimation of L1 biases the usage-based estimation of an AL, and why form-focussed instruction may be necessary to reset some counters to tally the L2 more appropriately.

The Units of Language Acquisition

Construction Grammar (Goldberg, 1995, 2003, 2006; Tomasello, 2003) and other *Cognitive Linguistic* theories of first (Croft & Cruise, 2004; Langacker, 1987; Taylor, 2002; Tomasello, 1998) and second language (Robinson & Ellis, 2007b) acquisition hold

that the basic units of language representation are *constructions*. These are form-meaning mappings, conventionalized in the speech community, and entrenched as language knowledge in the learner's mind. Constructions vary in specificity and in complexity, including morphemes (*anti-*, *-ing*, *N-s*), words (*aardvark*, *and*), complex words (*antediluvian*, *multimorphemic*), idioms (*hit the jackpot*), semi-productive patterns (*Good <time of day>*), and syntactic patterns [Subj [V Obj1 Obj2]]; [Subj *be-* Tns V *-en by* Obj]. Hence morphology, lexicon, and syntax are uniformly represented in Construction Grammar. Constructions are symbolic, in that their defining properties of morphological, lexical, and syntactic form are associated with particular semantic, pragmatic, and discourse functions. Constructions form a structured inventory of a speaker's knowledge of the conventions of their language, where schematic constructions can be abstracted over the less schematic ones, which are inferred inductively by the speaker in acquisition. A construction may provide a partial specification of the structure of an utterance; hence, an utterance's structure is specified by a number of distinct constructions. Constructions are independently represented units in a speaker's mind. Any construction with unique, idiosyncratic formal or functional properties must be represented independently in order to capture a speaker's knowledge of their language. However, absence of any unique property of a construction does not entail that it is not represented independently and simply derived from other, more general or schematic constructions. Frequency of occurrence may lead to independent representation of even "regular" constructional patterns.

Acquiring Constructions

Usage-based theories of naturalistic language acquisition hold that we learn language through using language. Creative linguistic competence emerges from learners' piecemeal acquisition of the many thousands of constructions experienced in communication, and from their frequency-biased abstraction of the regularities in this history of usage. Competence and performance both *emerge* from the conspiracy of memorized exemplars of construction usage, with competence being the integrated sum of prior usage and performance its dynamic contextualized activation (Ellis, 1998, 2003, 2006a, 2007; Ellis & Larsen Freeman, 2006).

Many of the constructions we know are quite specific, formulaic utterances based on particular lexical items, ranging, for example, from a simple 'Wonderful!' to increasingly complex phrases like 'One, two, three', 'Once upon a time', or 'Won the battle, lost the war'. These sequential patterns of sound, like words, are acquired as a result of chunking from repeated usage (Ellis, 1996; Pawley & Syder, 1983; Wray, 2002). In building up these sequences, learners bind together the chunks that they already know, with high frequency sequences being more strongly bound than lower frequency ones (Ellis, 2002a). In analyzing these sequences, the highest frequency chunks stand out as the most likely constituents of the parse. The constructions already acquired by the learner constitute the sample of evidence from which they implicitly and explicitly identify regularities, so generalizing their knowledge by inducing unconscious schema and prototypes that map meaning and form, and by abducing conscious metalinguistic hypotheses about language, too. These are the foundations, then, of new expressions and new understandings.

Constructionist approaches to language acquisition (Bybee & Hopper, 2001; Goldberg, 2003; Robinson & Ellis, 2007b; Tomasello, 2003, 1998) thus emphasize piecemeal learning from concrete exemplars. A high proportion of children's early multi-word speech is produced from a developing set of slot-and-frame patterns. These patterns are often based around chunks of one or two words or phrases, and they have 'slots' into which the child can place a variety of words, for instance subgroups of nouns or verbs (e.g., *I can't + VERB*; *where's + NOUN + gone?*). Children are very productive with these patterns, and both the number of patterns and their structure develop over time. But initially, they are lexically specific. For example, if a child has two patterns, *I can't + X* and *I don't + X*, the verbs used in these two X slots typically show little or no overlap, suggesting (1) that the patterns are not yet related through an underlying grammar (the child doesn't 'know' that *can't* and *don't* are both auxiliaries or that the words that appear in the patterns all belong to a category of Verb), and (2) that learners are picking up frequent patterns from what they hear around them and only slowly making more abstract generalizations as the database of related utterances grows (Pine & Lieven, 1993; Pine, Lieven, & Rowland, 1998; Tomasello, 1992). Tomasello's (1992) *Verb Island* hypothesis holds that it is verbs and relational terms that are the individual islands of organization in young children's otherwise unorganized grammatical system: the child initially learns about arguments and syntactic markings on a verb-by-verb basis, and ordering patterns and morphological markers learned for one verb do not immediately generalize to other verbs. Positional analysis of each verb island requires memories of the verb's usage, the exemplars of its collocations and the constructions it commonly

inhabits. Over experience, syntagmatic categories emerge from the regularities in this dataset, the learner's sample of language.

The chapters in Robinson and Ellis (2007b) extend these cognitive linguistic / construction grammar theories of child language acquisition to the naturalistic acquisition of ALs in adulthood, so developing a Usage-based approach to SLA. Some of the key features are as follows.

Frequency and the roles of input

AL Learners' knowledge of a linguistic construction depends, too, on their experience of its use, the sample of its manifestations of usage. Different frequencies of exemplification, and different types of repetition of a linguistic pattern, have different effects upon acquisition – the consolidation, generalization, and productive use of constructions. A key separation is between type and token frequency.

Type and Token Frequency

The token frequency of a construction is how often in the input that particular word or specific phrase appears; we can count in a sample corpus the token frequency of any specific form (e.g., the syllable [ka], the trigram *aze*, the word *frog*, the phrase *on the whole*, the sentence *I love you*). Type frequency, on the other hand, is the calculation of how many different lexical items a certain pattern, paradigm, or construction applies to, i.e., 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. Similarly the prepositional transfer construction [Subj [V ObjDir *to* ObjInd]] has a high type frequency (*give, read, pass, donate, display, explain...*) because many different verbs can be used in this way, whereas the ditransitive alternative [Subj [V ObjInd ObjDir]] is only used with a small set of verbs like *give, read, and pass* and not others (**donate, *display, *explain*).

Consolidating a particular formulaic construction: The role of Token frequency

Like other concrete constructions, a word, *can* be sketchily learned from a single exposure, as a *fast mapping* (Carey & Bartlett, 1978), a relation between an approximation of its sound and its likely meaning, forged as an explicit episodic memory relating its form and the perception of its likely referent (Ellis, 2005). The hippocampus and limbic structures in the brain allow us such unitary bindings from single experiences, rapid explicit memory, one-off learning, the establishment of new conjunctions of arbitrarily different elements (Ellis, 2002b; Squire, 1992), the learning of separate discrete episodes – what you saw across the field as your friend said *gavagai* or the particular colour of tray that accompanied hearing *chromium* for the first time. There is benefit in being able to keep such episodic records distinct. But fast mappings are rough, ready, fragile, and, without reiteration, often transient. Repetition strengthens memories (Ebbinghaus, 1885), and there are clearly defined effects of frequency, spacing, and distribution of practice in the consolidation, elaboration, and explicit learning of foreign-language vocabulary, both naturalistically and from flash-cards, CALL programs, and the like (Ellis, 1995).

Repeated processing of a particular construction facilitates its fluency of subsequent processing, too, and these effects occur whether the learner is conscious of this processing or not. Your reading of the various occurrences of the word *chunk* in this chapter so far has primed the subsequent reading of this word and contributed to your lifetime usage practice of it, despite the fact that you cannot remember where in the text these occurrences fell. Although you are conscious of words in your visual focus, you definitely did not just now consciously label the word *focus* as a noun. On reading it, you were surely unaware of its nine alternative meanings, though in a different sentence you would instantly have brought a different meaning to mind. What happens to the other meanings? Psycholinguistic evidence demonstrates that some of them exist unconsciously for a few tenths of a second before your brain decides on the right one. Most words (over 80% in English) have multiple meanings, but only one of these can become conscious at a time. So your reading of *focus* has primed subsequent reading of that letter string (whatever its interpretation), and your interpretation of *focus* as a noun has primed that particular subsequent interpretation of it. In this way, particular constructions (e.g., [ba], *ave, kept, man, dead boring, on the whole, I love you, [wʌn] = one*) with high token frequency are remembered better, recognized faster, produced more readily and otherwise processed with greater facility than low token frequency constructions (e.g., [za], *aze, leapt, artichoke, sublimely boring, on the organelle, I venerate you, [wʌn] = won*) (see Ellis, 2002a for review). Each token of use thus strengthens the memory traces of a construction, priming its subsequent use and accessibility following the power law of practice relationship, whereby the increase in strength afforded by early increments of experience are greater than those from later

additional practice. In these ways, language learning involves considerable unconscious ‘tallying’ (Ellis, 2002a) of construction frequencies, and language use requires exploitation of this implicit statistical knowledge (Bod, Hay, & Jannedy, 2003; Bybee & Hopper, 2001; Chater & Manning, 2006).

High token repetition is said to *entrench* constructions (Langacker, 1987), protecting them from change. Thus it is that it is the high frequency past tenses in English that are irregular (*went, was, kept*), their ready accessibility holding off the forces of regularization from the default paradigm (**goed, *beed, *keeped*), whereas neighbours of lower frequency eventually succumb (with *leaped* starting to rival *leapt* in usage). Bybee (2008) calls this the *conserving* function of high token frequency. High token frequency also leads to *autonomy*, whereby creative constructions learned by rote may never be analyzed into their constituent units, e.g., learners may never have considered that *gimme* consists of *give + me*, nor the literal roots of a *dicey situation*. Finally, considerable practice with a particular token also results in automaticity of production and processes of reduction, assimilation, and lenition involving loss and overlap of gestures. A maxim of Bybee (2003, p. 112), on a variant of Hebb’s “Cells that fire together wire together,” is that “Items that are used together fuse together”. The phenomenon is entirely graded -- the degree of reduction is a continuous function of the frequency of the target word and the conditional probability of the target given the previous word and that of the target given the next word (Bybee & Hopper, 2001; Ellis, 2002a; Jurafsky, Bell, Gregory, & Raymond, 2001). Such changes underpin grammaticalization in language change (Bybee, 2000; Croft, 2000).

In sum, although a particular construction *can* be roughly learned from a single exposure, multiple repetitions of that same token in different contexts are needed to enmesh and elaborate it into the meaning system -- to turn it from a fast-mapped tentative working hypothesis to a more complete rich representation of the full connotations of a the word (Carey & Bartlett, 1978). For example, it has been estimated that between eight and 12 encounters are needed of a novel word in text before its meaning will be adequately comprehended from inference and its form and meaning retained (Horst, Cobb, & Meara, 1998; Saragi, Nation, & Meister, 1978). Multiple repetitions are also necessary for entrenched representation, ready accessibility, automatized processing, idiomatic autonomy, and fast, fluent, and phonetically reduced production.

Generalizing a construction from formula to limited scope pattern to productive abstract schema: The role of Type frequency

The productivity of phonological, morphological, and syntactic patterns is a function of their type rather than token frequency (Bybee, 1995; Bybee & Hopper, 2001). Type frequency determines productivity because: (1) 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. As novel exemplars are added in memory, they affect the category too, their features resonate with the whole population, adding their weight to the prototype, and stretching the bounds slightly in their direction. (2) 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. (3) 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 & Thompson, 2000).

When a construction is variously experienced with different items occupying a position, it allows the parsing of its schematic structure. Having an initial formulaic exemplar of the Caused-Motion construction [Subj V Obj Prep Obl_{path/loc}], perhaps *she pushed it down the road*, subsequent experience of *she pushed it ((up) the hill)*, *she pushed it ((to) the service station)*, *she pushed it ((to) the gas pump)* allows identification of the common components, their structural commonalities, and their regularities of reference. Common items (pronouns like *she*, *he*, *I*, rather than complex noun phrases *Mrs. Struthers*, *the miraculous moose*, *the distressed driver*, etc.; high frequency prepositions like *to*, *up*, *down*, etc., rather than complex locatives *Alabama-way*, *paralleling the path of flight*, etc.) repeat more in these slots and thus help to bring out the commonalities of the adjacent slot-fillers. Braine (1987) showed in experiments involving the learning of artificial languages that it was relatively easy to learn ‘categories’ and rules for combining them providing the ‘words’ exemplifying these categories were either preceded or followed by a fixed item. Otherwise, the categories were difficult or impossible to learn. In natural language, it is the grammatical words that often serve as anchors like this. It is the closed class ‘little words’, the grammatical functors, that have both the highest frequency in the language and the highest connectivity or *degree*. When the sequential co-occurrences of words in discourse are described in terms of graphs of word connections, mapping the interactions like social networks, the world wide web, or other complex systems, these graphs show so-called small-world properties of being highly clustered and richly interconnected (Ferrer i

Cancho & Solé, 2001; Ferrer i Cancho, Solé, & Köhler, 2004). Despite having many thousands of nodes (the >450,000 words populating a language), the average number of jumps in the path needed to get from any word to any other in this graph is remarkably small, at less than three. A small number of highly connected words allows these properties. And it is the function words, the prepositions, pronouns, determiners, etc., that do this, having *both* high token frequency and high degree of connectivity^{iv}.

So, these highest frequency components and chunks are the recurrent constituents of the construction that anchor its parse: as sub-unit constructions with high token frequency, they are recognized faster, produced more readily and otherwise processed with greater facility than low token frequency constructions, and, thus, they outline and bracket the schematic structure of the construction more readily. In 11-month-old infants, it is these frequently occurring functor forms that serve as a framework against which potential candidates for vocabulary membership may be identified and extracted from the speech stream (Shi, Cutler, Werker, & Cruikshank, 2006). In these ways, although verb islands predominate in seeding generalizations, patterns based on other high frequency lexical types, such as bound morphemes, auxiliary verbs and case-marking pronouns ('pronoun islands'), are also important in the parsing and identification of the schematic structure of constructions (Childers & Tomasello, 2001; McClure, Lieven, & Pine, in press; Pine, Lieven, & Rowland, 1998; Wilson, 2003)^v. In growth, too, these are the high-degree nodes of the kernel lexicon of the language network, to which new sub-unit constructions are preferentially attached, allowing scale-free growth distribution according to the so-called Barabási-Albert model (Barabási & Albert, 1999; Ferrer i Cancho & Solé, 2001).

Chunking is a ubiquitous feature of human learning and memory. Chunking affords the ability to build up structures recursively, with the embedding of small chunks within larger ones leading to a hierarchical organization in nature (Simon, 1962, see particularly his parable of the two watchmakers, Hora and Tempus), in memory (Newell, 1990), and in the hierarchies and tree structures of grammar (Bybee, 2003; Ellis, 1996, 2003). In these ways, constituent structure is emergent, with constructions as grammatical schemas at all levels of specificity (from very specific (*my chapter*), through limited scope (*my* + NOUN), more general (POSSESSIVE + NOUN), to fully general (DETERMINER + NOUN)) emerging from the conspiracy of component constructions whose commonalities, in turn, are defined by their inclusion in the networks of other constructions (Bybee, 2003, 2008).

Functional motivations

Constructions are useful because of the symbolic functions that they serve. It is their communicative functions, semantic, pragmatic, or discursive, that motivate their learning. Goldberg (1995) claims that verb-centred constructions are more likely to be salient in the input because they relate to certain fundamental perceptual primitives, and, thus, that this construction of grammar involves in parallel the distributional analysis of the language stream and the analysis of contingent perceptual activity. It has been argued that basic level categories (e.g., hammer, dog) are acquired earlier and are more frequently used than superordinate (tools, canines) or subordinate (ball peen hammer, weimaraner) terms because, besides their frequency of use, this is the level at which the world is optimally split for function, the level where objects within the class share the same broad visual shape and motoric function, and, thus, where the categories of

language most directly map onto perceptual form and motoric function (Lakoff, 1987; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Rosch, Varela, & Thompson, 1991). Goldberg extends this notion to argument structure more generally:

“Constructions which correspond to basic sentence types encode as their central senses event types that are basic to human experience... that of someone causing something, something moving, something being in a state, someone possessing something, something causing a change of state or location, something undergoing a change of state or location, and something having an effect on someone.” (Goldberg, 1995, p. 39).

Ninio (1999) and Goldberg, Casenhiser and Sethuraman (2004) show for child language acquisition that individual “pathbreaking” semantically prototypic verbs form the seeds of verb-centered argument-structure patterns, with generalizations of the verb-centered instances emerging gradually as the verb-centered categories themselves are analyzed into more abstract argument structure constructions. The verb is a better predictor of sentence meaning than any other word in the sentence and plays the central role in determining the syntactic structure of a sentence. Since the same functional concerns motivate AL and L1 both, we should expect the same pattern for L2 and FL acquisition.

Learning Categories and Prototypes: From Tokens to Types

Because constructions are linguistic categories, we need to consider the psychology of concept and category learning (Ashby & Maddox, 2005; Cohen & Lefebvre, 2005): Humans can readily induce a category from experience of exemplars. Categories have graded structures (Rosch & Mervis, 1975). Rather than all instances of a

category being “equal,” certain instances are better exemplars than others. The prototype is the best example among the members of a category and serves as the benchmark against which the surrounding “poorer,” more borderline instances are categorized; it combines the most representative attributes of that category in the conspiracy of its memorized exemplars. People have memory for the tokens they have seen before – previously experienced patterns are better judged than novel ones of equal distortion from the prototype. Although we don’t go around consciously counting types and tokens, we nevertheless have very accurate implicit knowledge of the underlying distributions and their most usual settings. Similarity and frequency are, thus, important determinants of learning and generalization:

The more *similar* an instance is to the other members of its category and the less similar it is to members of contrast categories, the easier it is to classify (e.g., we better classify sparrows [or other average-sized, average-coloured, average-beaked, average-featured specimens] as birds than we do birds with less common features or feature combinations, like geese or albatrosses)(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 of the category (e.g., sparrows are rated as highly typical birds because they are frequently experienced examples of the category birds). The unmarked form of linguistic oppositions are more frequent than their marked form (Greenberg, 1966). Token frequency is particularly important in this way in early and intermediate levels of learning, less so as learning approaches asymptote (Homa, Dunbar, & Nohre, 1991; Nosofsky, 1988).

There are important effects of presentation order in the implicit tallying that underlies category formation. In learning, the greater the variability of exemplars, the lower the rate of acquisition but the more robust the categorization / the less variability of distortion, the faster the category is learned (Posner & Keele, 1968, 1970). But it looks like there's an optimal balance to be had here. When people try to teach a category to someone else *explicitly*, there is high agreement on the teaching sequences that are naturally adopted: The typical sequence starts with several ideal positive cases, followed by an ideal negative case and then borderline cases (Avrahami et al., 1997). Avrahami et al. tested to see whether this is indeed an optimal instruction sequence by comparing it with other orders that emphasized the full breadth of category from the outset.

Exemplifying category breadth from the outset, borderline cases and central cases all, produced slower and less accurate explicit learning. For *implicit* learning of categories from exemplars, so, too, acquisition is optimized by the introduction of an initial, low-variance sample centered upon prototypical exemplars (Elio & Anderson, 1981, 1984). This low variance sample allows learners to get a 'fix' on what will account for most of the category members, Then the bounds of the category can later be defined by experience of the full breadth of exemplars.

Form, function, and frequency: Zipfian family profiles

Goldberg, Casenhiser & Sethuraman (2004) tested the applicability of these generalizations to the particular case of children acquiring constructions. Phrasal form-meaning correspondences (e.g., X causes Y to move Z_{path/loc} [Subj V Obj Obl_{path/loc}]) do exist independently of particular verbs, but there is a close relationship between the types of verb that appear therein (in this case *put, get, take, push*, etc.). Furthermore, in natural

language, the frequency profile of the verbs in the family follows a Zipfian profile (Zipf, 1935) whereby the highest frequency words accounted for the most linguistic tokens. Goldberg et al. demonstrated that in samples of child language acquisition, for a variety of constructions, there is a strong tendency for one single verb to occur with very high frequency in comparison to other verbs used (e.g., the [Subj V Obj Obl_{path/loc}]) construction is exemplified in the children's speech by *put* 31% of the time, *get* 16%, *take* 10%, and *do/pick* 6%). This profile closely mirrored that of the mothers' speech to these children (with, e.g., *put* appearing 38% of the time in this construction that was otherwise exemplified by 43 different verbs). Ellis and Ferreira Junior (Ellis, Ferreira Junior, & Ke, in preparation) have replicated the Zipfian family profiles of these same constructions for the speech of naturalistic adult learners of English as a second language in the ESF project (Perdue, 1993).

The same can be seen in the constructions for compliments. Manes and Wolfson (1989) examined a corpus of seven hundred examples of compliments uttered in day-to-day interactions. Just three constructions accounted for 85% of these: [NP <*is / looks*> (*really*) ADJ] (53%), [I (*really*) <*like / love*> NP] (16%), and [PRO *is (really) (a)* ADJ NP] (15%). Eighty per cent of these depended on an adjective to carry the positive semantic load. While the number of positive adjectives that could be used is virtually unlimited, in fact two-thirds of all adjectival compliments in the corpus used only five adjectives: *nice* (23%), *good* (20%), *pretty* (9%), *beautiful* (9%), and *great* (6%). Non-adjectival compliments were focussed on a handful of semantically positive verbs, with *like* and *love* accounting for 86%.

Thus, it appears that in natural language, at least for the constructions considered in this way so far, tokens of one particular verb account for the lion's share of instances of argument frames, and that the pathbreaking verb for each is the one with the prototypical meaning from which that construction is derived. How about that? As Morales & Taylor (in press) put it: "Language is exquisitely adaptive to the learning capabilities of its users." The natural structure of natural language seems to provide exactly the familial type:token frequency distribution to ensure optimized acquisition of linguistic constructions as categories.

Optimizing instruction samples for construction learning

What are the implications for instruction using curriculum-driven input samples? What we know about category formation suggests that these type:token frequency considerations should apply here too. *Optimal acquisition should occur when the central members of the category are presented early and often.*

For syntactic constructions, Goldberg, Casenhiser & Sethuraman (2004) tested whether, when training novel patterns (a construction of the form [Subj Obj V-*o*] signalling the appearance of the subject in a particular location, for example, *the king the ball moopo-ed*) exemplified by 5 different novel verbs, it is better to train with relatively balanced token frequencies (4-4-4-2-2) or with a family frequency profile where one exemplar had a particularly high token frequency (8-2-2-2-2). Undergraduate native speakers of English learned this novel construction from 3 minutes of training using videos. They were then tested for the generalization of the semantics of this construction to novel verbs and new scenes. Learners in the high token frequency condition showed

significantly better learning than those in the balanced condition, a finding Goldberg (Goldberg, 2006; , 2007) has now observed in studies of child acquisition too.

For morphological constructions, Bybee (2008) analyzed the ways that natural frequency skewing affects the acquisition of verbal inflexions. The most frequent forms of a paradigm (3rd person /1st person singular) either have no affix or a short affix, and the other forms of the paradigm can typically be derived from them. Thus, she argues, the high token frequency forms of the paradigm are the anchoring points of the other forms. Lower frequency forms are analysed and learned in terms of these more robust forms creating a relationship of dependency.

Frequency variation is ubiquitous across natural languages. Morales and Taylor (in press) present connectionist simulations evidencing how learning can be enhanced through frequency variation: training samples where there were variable numbers of tokens per type produced more accurate and more economical learning than did training with more uniform frequency profiles.

There is clearly need to extend these initial studies to explore more thoroughly the sampling of exemplars of a wide range of second language constructions for optimal acquisition, but in the interim, the best informed practice is to introduce a new construction using an initial, low-variance sample centered upon prototypical exemplars to allow learners to get a 'fix' on the central tendency that will account for most of the category members. Tokens that are more frequent have stronger representations in memory and serve as the analogical basis for forming novel instances of the category.

Corpus and Cognitive Linguistic analyses are essential to the determination of which constructions of differing degrees of schematicity are worthy of instruction, their

relative frequency, and their best (= central and most frequent) examples for instruction and assessment (Biber, Conrad, & Reppen, 1998; Biber, Johansson, Leech, Conrad, & Finegan, 1999). Gries (2007) describes how the three basic methods of corpus linguistics (frequency lists, concordances, and collocations) inform the instruction of second language constructions. Achard (2008), Tyler (2008), Robinson and Ellis (2008a) and other readings in Robinson and Ellis (2008b) show how an understanding of the item-based nature of construction learning inspires the creation and evaluation of instructional tasks, materials, and syllabi, and how cognitive linguistic analyses can be used to inform learners how constructions are conventionalized ways of matching certain expressions to specific situations and to guide instructors in precisely isolating and clearly presenting the various conditions that motivate speaker choice.

Tuning the System: Frequency and the Attainment of Nativelike Fluency and Selection

Language is fundamentally probabilistic: every piece is ambiguous. Each of these example formulas ('One, two, three', 'Once upon a time', 'Wonderful!', 'Won the battle, lost the war') begins with the sound 'wʌn'. At this point, what should the appropriate interpretation be? A general property of human perception is that when a sensation is associated with more than one reality, unconscious processes weigh the odds, and we perceive the most probable thing. *Psycholinguistic* analyses demonstrate that fluent language users are sensitive to the relative probabilities of occurrence of different constructions in the speech stream (Bod, Hay, & Jannedy, 2003; Bybee & Hopper, 2001; Chater & Manning, 2006; Ellis, 2002a, 2002b; Jurafsky & Martin, 2000). Since learners have experienced many more tokens of 'one' than they have 'won', in the absence of any

further information, they typically favour the unitary interpretation over that involving gain or advantage. But they need to be able to suppress this interpretation in a context of ‘Alice in wλn...’ Learners have to figure language out: their task is, in essence, to learn the probability distribution $P(\textit{interpretation}|\textit{cue}, \textit{context})$, the probability of an interpretation given a formal cue, a mapping from form to meaning conditioned by context. This figuring is achieved, and communication optimized, by implicit tallying of the *frequency*, *recency*, and *context* of constructions.

This incidental learning from usage allows language users to be *Rational* in the sense that their mental models of the way language works are optimal given their linguistic experience to date (Ellis, 2006b). The words that they are likely to hear next, the most likely senses of these words, the linguistic constructions they are most likely to utter next, the syllables they are likely to hear next, the graphemes they are likely to read next, the interpretations that are most relevant, and the rest of what’s coming next across all levels of language representation, are made more readily available to fluent speakers by their language processing systems. Their unconscious language representations are adaptively probability-tuned to predict the linguistic constructions that are most likely to be relevant in the ongoing discourse context, optimally preparing them for comprehension and production. With practice comes modularization too, the development of autonomous specialist systems for different aspects of language processing. These ‘zombie agents’ are independent – experience of reading a word facilitates subsequent reading of that word, experience of speaking a word facilitates subsequent speaking of that word, but cross-modal priming effects are null or slight in fluent speakers. So reading practice tallies the reading system, speaking practice tunes the speaking system, etc. Fluency in each

separate module requires its own usage practice (see Gatbonton & Segalowitz, 2005 for communicative approaches designed to engender this). This specificity of practice gain from different forms of processing underlies many failures of leaning and generalization as summarized in the Transfer-Appropriate Processing (TAP) framework (Morris, Bransford, & Franks, 1977). Lightbown (in press) reviews the implications of TAP for L2 instruction, how there is a need to increase the number of settings and processing types in which learners encounter the material they need to learn.

Just as extensive sampling is required for nativelylike fluency, so it is, too, for nativelylike selection. Many of the forms required for idiomatic use are, nevertheless, of relatively low frequency, and the learner thus needs a large input sample just to encounter them. More usage still is required to allow the tunings underpinning nativelylike use of collocation – something which even advanced learners have particular difficulty with. Hence the emphasis on the representative samples necessary for EAP and ESP (e.g., Swales, 1990). Linguists interested in the description of language (e.g., British National Corpus, 2006) have come to realize that really large corpora are necessary to describe it adequately – 100 million words is just a start, and each genre, dialect, and type requires its own properly targeted sampling. Child language researchers have also begun the relevant power analyses to explore the relations between construction frequency and sample size for accurate description, reaching the conclusion that for many constructions of interest, *dense* corpora are an absolute necessity (Tomasello & Stahl, 2004). So, too, in learners' attainment of fluent language processing, whether in L1 or AL, there is no substitute for usage, lots of appropriate usage.

Becoming fluent requires a sufficient sample of needs-relevant authentic input for the necessary implicit tunings to take place. The ‘two puzzles for linguistic theory’, nativelike selection and nativelike fluency (Pawley & Syder, 1983), are less perplexing when considered in these terms of frequency and probability. There’s a lot of tallying to be done here. The necessary sample is certainly to be counted in terms of thousands of hours on task.

The Language Calculator has no ‘Clear’ button

A final implication of language acquisition as estimation relates again to *sampling history*, this time in terms of the difference between L1A and ALA. AL learners are distinguished from infant L1 acquirers by the fact that they have previously devoted considerable resources to the estimation of the characteristics of another language -- the native tongue in which they have considerable fluency (and any others subsequently acquired). Since they are using the same apparatus to survey their additional language too, their computations and induction are often affected by transfer, with L1-tuned expectations and selective attention (Ellis, 2006c) blinding the computational system to aspects of the AL sample, thus rendering biased estimates from naturalistic usage and the limited endstate typical of L2A. These effects have been explored within the traditions of Contrastive Analysis (James, 1980), Language Transfer (Odlin, 1989), and more recently within Cognitive Linguistics (Robinson & Ellis, 2007b). From our L1 we learn how language frames the world and how to use it to describe action therein, focussing our listeners’ attention appropriately. Cognitive Linguistics is the analysis of these mechanisms and processes that underpin what Slobin (1996) called ‘thinking for

speaking.’ But learning an AL requires ‘rethinking for speaking’ (Robinson & Ellis, 2007a). In order to counteract the L1 biases to allow estimation procedures to optimize induction, all of the AL input needs to be made to count (as it does in L1A), not just the restricted sample typical of the biased intake of L2A. Certain types of form-focused instruction can help to achieve this by recruiting learners’ explicit, conscious processing to allow them to consolidate unitized form-function bindings of novel AL constructions (Ellis, 2005). Once a construction has been represented in this way, so its use in subsequent processing can update the statistical tallying of its frequency of usage and probabilities of form-function mapping.

Language is in its dynamic usage. It ever changes. For learners and linguists alike, its sum can only ever be estimated from limited samples of experience. Understanding the units and the processes of their estimation helps guide theory and application, learning and instruction.

References

- Achard, M. (2008). Cognitive pedagogical grammar. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition*. London: Routledge.
- Ashby, E. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology*, 56, 149-178.
- Avrahami, J., Kareev, Y., Bogot, Y., Caspi, R., Dunaevsky, S., & Lerner, S. (1997). Teaching by Examples: Implications for the Process of Category Acquisition. *The Quarterly Journal of Experimental Psychology: Section A*, 50(3), 586 - 606.
- Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286, 509-511.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. New York: Cambridge University Press.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow, UK: Pearson Education.
- Bod, R., Hay, J., & Jannedy, S. (Eds.). (2003). *Probabilistic linguistics*. Cambridge, MA: MIT Press.

- Braine, M. D. (1987). What is learned in acquiring word classes - a step towards acquisition theory. In B. MacWhinney (Ed.), *Mechanisms of language acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- British National Corpus. (2006). from <http://www.natcorp.ox.ac.uk/>
- Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, 10, 425-455.
- Bybee, J. (2000). *Mechanisms of change in grammaticalization: The role of frequency*. Unpublished manuscript.
- Bybee, J. (2003). Sequentiality as the basis of constituent structure. In T. Givón & B. F. Malle (Eds.), *The evolution of language out of pre-language* (pp. 109-132). Amsterdam: John Benjamins.
- Bybee, J. (2008). Usage-based grammar and second language acquisition. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition*. London: Routledge.
- Bybee, J., & Hopper, P. (Eds.). (2001). *Frequency and the emergence of linguistic structure*. Amsterdam: Benjamins.
- Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Proceedings of the Stanford Child Language Conference / Papers and Reports on Child Language Development*, 15, 17-29.
- Chater, N., & Manning, C. (2006). Probabilistic models of language processing and acquisition. *Trends in Cognitive Science*, 10, 335-344.
- Childers, J. B., & Tomasello, M. (2001). The role of pronouns in young children's acquisition of the English transitive construction. *Developmental Psychology*, 37, 739-748.
- Cohen, H., & Lefebvre, C. (Eds.). (2005). *Handbook of categorization in cognitive science*. Mahwah, NJ: Elsevier.
- Croft, W. (2000). *Explaining language change: An evolutionary approach*. London: Longman.
- Croft, W., & Cruise, A. (2004). *Cognitive linguistics*. Cambridge: Cambridge University Press.
- Ebbinghaus, H. (1885). *Memory: A contribution to experimental psychology* (H. A. R. C. E. B. (1913), Trans.). New York: Teachers College, Columbia.
- Elio, R., & Anderson, J. R. (1981). The effects of category generalizations and instance similarity on schema abstraction. *Journal of Experimental Psychology: Human Learning & Memory*, 7(6), 397-417.
- Elio, R., & Anderson, J. R. (1984). The effects of information order and learning mode on schema abstraction. *Memory & Cognition*, 12(1), 20-30.
- Ellis, N. C. (1995). The psychology of foreign language vocabulary acquisition: Implications for CALL. *Computer Assisted Language Learning*, 8, 2-3.
- Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, 18(1), 91-126.
- Ellis, N. C. (1998). Emergentism, connectionism and language learning. *Language Learning*, 48(4), 631-664.
- Ellis, N. C. (2002a). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 143-188.

- Ellis, N. C. (2002b). Reflections on frequency effects in language processing. *Studies in Second Language Acquisition*, 24(2), 297-339.
- Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition*. Oxford: Blackwell.
- Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, 27, 305-352.
- Ellis, N. C. (2006a). Cognitive perspectives on SLA: The Associative Cognitive CREED. *AILA Review*, 19, 100-121.
- Ellis, N. C. (2006b). Language acquisition as rational contingency learning. *Applied Linguistics*, 27(1), 1-24.
- Ellis, N. C. (2006c). Selective attention and transfer phenomena in SLA: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics*, 27(2), 1-31.
- Ellis, N. C. (2007). Dynamic Systems and SLA: The wood and the trees. *Bilingualism: Language & Cognition*, 10, 23-25.
- Ellis, N. C. (2008). The dynamics of language use, language change, and first and second language acquisition. *Modern Language Journal*, xx, xxx-xxx.
- Ellis, N. C., Ferreira Junior, F., & Ke, J.-Y. (in preparation). Form, function, and frequency: Zipfian family construction profiles in SLA.
- Ellis, N. C., & Larsen Freeman, D. (2006). Language emergence: Implications for Applied Linguistics (Introduction to the Special Issue). *Applied Linguistics*, 27(4), 558-589.
- Ferrer i Cancho, R., & Solé, R. V. (2001). The small world of human language. *Proceedings of the Royal Society of London, B.*, 268, 2261-2265.
- Ferrer i Cancho, R., Solé, R. V., & Köhler, R. (2004). Patterns in syntactic dependency networks. *Physical Review*, E69, 0519151-0519158.
- Gatbonton, E., & Segalowitz, N. (2005). Rethinking communicative language teaching: A focus on ACCESS to fluency. *Canadian Modern Language Review*, 61, 325-353.
- Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- Goldberg, A. E. (2003). Constructions: a new theoretical approach to language. *Trends in Cognitive Science*, 7, 219-224.
- Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- Goldberg, A. E. (2007). The language of constructions. In P. Robinson & N. C. Ellis (Eds.), *A handbook of cognitive linguistics and SLA*. Mahwah, NJ: Erlbaum.
- Goldberg, A. E., Casenhiser, D. M., & Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive Linguistics*, 15, 289-316.
- Greenberg, J. (1966). *Language Universals*. The Hague: Mouton.
- Gries, S. T. (2007). Corpus-based methods in analyses of SLA data. In P. Robinson & N. C. Ellis (Eds.), *A handbook of cognitive linguistics and SLA*. Mahwah, NJ: Erlbaum.

- Homa, D., Dunbar, S., & Nohre, L. (1991). Instance frequency, categorization, and the modulating effect of experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 444-458.
- Horst, M., Cobb, T., & Meara, P. (1998). Beyond A Clockwork Orange: Acquiring second-language vocabulary through reading. *Reading in a Foreign Language*, *11*, 207-223.
- James, C. (1980). *Contrastive analysis*. London: Longman.
- Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229-254). Amsterdam: Benjamins.
- Jurafsky, D., & Martin, J. H. (2000). *Speech and language processing: An introduction to natural language processing, speech recognition, and computational linguistics*. Englewood Cliffs, NJ: Prentice-Hall.
- Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
- Langacker, R. W. (1987). *Foundations of cognitive grammar: Vol. 1. Theoretical prerequisites*. Stanford, CA: Stanford University Press.
- Lightbown, P. M. (in press). Transfer appropriate processing in classroom second language acquisition. In Z.-H. Han (Ed.), *Fossilization*.
- Manes, J., & Wolfson, N. (1989). The compliment formula. In F. Coulmas (Ed.), *Conversational routine*. The Hague: Mouton.
- McClure, C., Lieven, E., & Pine, J. M. (in press). Investigating the abstractness of children's early knowledge of argument structure. *Journal of Child Language*.
- Morales, F., & Taylor, J. R. (in press). Learning and relative frequency. In C. Zelinski (Ed.), *Memory and language*.
- Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning & Verbal Behavior*, *16*, 519-533.
- Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
- Ninio, A. (1999). Pathbreaking verbs in syntactic development and the question of prototypical transitivity. *Journal of Child Language*, *26*, 619-653.
- Nosofsky, R. M. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 54-65.
- Odlin, T. (1989). *Language transfer*. New York: Cambridge University Press.
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191-225). London: Longman.
- Perdue, C. (Ed.). (1993). *Adult language acquisition: Crosslinguistic perspectives*. Cambridge: Cambridge University Press.
- Pine, J. M., & Lieven, E. V. (1993). Reanalyzing rote-learned phrases: Individual differences in the transition to multi-word speech. *Journal of Child Language*, *20*, 551-571.
- Pine, J. M., Lieven, E. V., & Rowland, C. F. (1998). Comparing different models of the development of the English verb category. *Linguistics*, *36*, 807-830.

- Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, 77, 353-363.
- Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology*, 83, 304-308.
- Robinson, P., & Ellis, N. C. (2007a). Final synthesis chapter *****. In P. Robinson & N. C. Ellis (Eds.), *A handbook of cognitive linguistics and SLA*. Mahwah, NJ: Lawrence Erlbaum.
- Robinson, P., & Ellis, N. C. (2008a). Conclusion: Cognitive Linguistics, Second Language Acquisition and L2 Instruction—Issues for Research. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition*. London: Routledge.
- Robinson, P., & Ellis, N. C. (Eds.). (2007b). *A handbook of cognitive linguistics and SLA*. Mahwah, NJ: Lawrence Erlbaum.
- Robinson, P., & Ellis, N. C. (Eds.). (2008b). *A handbook of cognitive linguistics and second language acquisition*. London: Routledge.
- Rosch, E., & Mervis, C. B. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104, 192-233.
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8, 382-439.
- Rosch, E., Varela, F., & Thompson, E. (1991). *The embodied mind*. Boston, MA: MIT Press.
- Saragi, T., Nation, I. S. P., & Meister, G. F. (1978). Vocabulary learning and reading. *System*, 6, 72-128.
- Shi, R., Cutler, A., Werker, J. E., & Cruikshank, M. (2006). Frequency and form as determinants of functor sensitivity in English-acquiring infants. *Journal of the Acoustic Society of America*, 119, EL61-EL67.
- Simon, H. A. (1962). *The sciences of the artificial*. Cambridge, MA: MIT Press.
- Slobin, D. I. (1996). From "thought and language" to "thinking for speaking." In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity*. Cambridge: Cambridge University Press.
- Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, 99, 195-231.
- Swales, J. M. (1990). *Genre analysis : English in academic and research settings*. Cambridge: Cambridge University Press.
- 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*. Boston, MA: Harvard University Press.
- Tomasello, M. (Ed.). (1998). *The new psychology of language: Cognitive and functional approaches to language structure*. Mahwah, NJ: Erlbaum.
- Tomasello, M., & Stahl, D. (2004). Sampling children's spontaneous speech: How much is enough? *Journal of Child Language*, 31, 101-121.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327-352.
- Tyler, A. (2008). Cognitive linguistics and second language instruction. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition*. London: Routledge.

- Wilson, S. (2003). Lexically specific constructions in the acquisition of inflection in English. *Journal of Child Language*, 30, 75-115.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- Zipf, G. K. (1935). *The psycho-biology of language: An introduction to dynamic philology*. Cambridge, MA: The M.I.T. Press.

Signature

Nick C. Ellis

Biographic Information

Nick Ellis is Research Scientist at the English Language Institute and Professor of Psychology at the University of Michigan. His research interests include psycholinguistic, cognitive, cognitive linguistic, corpus linguistic, emergentist, and behavioral genetic aspects of adult language acquisition.

Contact address

ncellis@umich.edu

ⁱ I thank Patsy Lightbown for constructive comments on a previous draft.

ⁱⁱ Depending as well, of course, upon degree of shared context, embodiment, attention, cultural understandings, communicative intent, etc.

ⁱⁱⁱ Whereby the frequency of the tokens of verbs seeding a construction type decays as a power function of their rank (Zipf, 1935)

^{iv} The high token frequency of these items, though, means that in the course of language use, they have become eroded. These items lack perceptual salience and are consequently difficult to perceive from bottom-up, data-driven sources alone, a factor which makes their second language acquisition difficult (Ellis, 2006c, 2008). They are also semantically light, abstract, and often homonymous, factors also making them difficult to acquire (Ellis, 2008). So it is the semantically rich and basic verbs which seed the constructions, these other grammatical functors making their contribution by marking the commonalities of the parse pattern.

^v Again, emphasizing the proviso concerning their low salience, low contingency, and abstractness.