Preprint of

Ellis, N. C. & Ogden, D. C. (2015). Language Cognition: comments on 'The Ubiquity of Frequency Effects in First Language.' *Journal of Child Language, 42*, 182-186

Nick C. Ellis & Dave C. Ogden

ncellis@umich.edu ogdend@umich.edu

University of Michigan

Ambridge, Kidd, Rowland and Theakston (this issue) review child language research and show how frequency effects pervade the acquisition of lexis, morphology, and syntax. They argue therefore that any successful account of language acquisition, from whatever theoretical standpoint, must be frequency sensitive to the extent that it can explain these effects.

The same conclusion follows from 60 years of psycholinguistic research into the fluent language processing that culminates from acquisition: Language processing is exquisitely sensitive to usage frequency at all levels of representation: phonology and phonotactics, reading, spelling, lexis, morphosyntax, formulaic language, language comprehension, grammaticality, sentence production, and syntax (Ellis, 2002). If language processing is sensitive to frequency of experience, it follows that experiences of language processing each individually affect the system. Acquisition is usage-based.

The ubiquity of frequency effects across all domains of child language leads Ambridge et al. to favor accounts of language acquisition which have as their core learning mechanism the storage and re-use of strings from the input. Exemplar-based cognitive theories of this type are common in explaining schematization and categorization across cognitive domains, and they generate the observed frequency effects for free.

Thus the evidence of frequency effects links language to general cognition. Research in language acquisition has long grappled with whether language is special, with its own cognitive underpinnings apart from other domains, or subject to broadly the same learning mechanisms and cognitive constraints as the rest of our knowledge and behavior. The answer may lie somewhere between, and may be approached from either direction. We take the view that it is parsimonious to begin with the latter and determine exactly what might be learnable of language by general cognitive means, and frequency effects have been integral to research in this vein.

But let us not leave it there. As Ambridge et al. make clear, many other factors interact with frequency in language learning. There are many hallmarks of general mechanisms of learning that show themselves across cognitive contents and domains. Let us broaden the enquiry to look out for these other 'usual suspects' too. The more we observe, the stronger the inference that language is cut of the same cloth as the rest of cognition, and the more inter-disciplinary our cognitive scientific understanding of how language emerges from usage.

In this brief commentary we provide an I-Spy guide to the most likely candidates to look for in language acquisition. We start with the fundamentals of cognitive and associative learning: *frequency*, *recency*, and *context*. Theoretical analyses show that these underpin *rational cognition*, the optimal extraction of regularities from experience. Additionally we emphasize *contingency learning* (the statistical mapping of language forms to language functions) and the factors that drive *categorization*. We consider the statistical measurement of these phenomena. Finally, the many interacting factors in language usage, cognition, and acquisition lead us to advocate complex systems approaches involving computational modeling.

An I-Spy for Language Acquisition: The General Forces of Cognition

1.1. Frequency, Recency, and Context

From its beginnings, psychological research recognized, frequency, recency, and context of usage as the driving forces of cognition (Ebbinghaus, 1885). Learning, memory and perception are all affected by frequency of usage: The more times we experience something, the stronger our memory for it, and the more fluently it is accessed. The more recently we have experienced something, the stronger our memory for it. The more times we experience conjunctions of features, the more they become associated in our minds and the more these subsequently affect perception and categorization; so a stimulus becomes associated to a context and we become more likely to perceive it in that context. The *power law of learning* describes the relationship between practice and performance in the acquisition of a wide range of cognitive skills – the greater the practice, the greater the performance, with effects of practice larger at early stages of leaning. The power function relating probability of recall and recency is the *forgetting curve*.

Learners aren't experiencing the linguistic system in isolation, but in connection with socially meaningful communication. Constructions, from fixed words and expressions to abstract morphology and syntax, are symbols, mappings of linguistic form and their interpretation. Psychological research into associative learning has long recognized that while frequency of form is important, more so is *contingency*. Cues with multiple interpretations are ambiguous and so hard to resolve, whereas cue-outcome associations of high contingency are reliable and readily processed. Consider how, in the learning of the category of birds, while eyes and wings are equally frequently experienced features in the exemplars, it is wings which are distinctive in differentiating birds from other animals. Contingency of cue-outcome mapping is a driving force of all associative learning, and is central in the Competition Model (Bates & MacWhinney, 1987) and other models of the rational learning of form-function constructions.

Every stimulus is ambiguous: "*Perception is of definite and probable things*" (William James, 1890). A percept is a complex state of consciousness in which antecedent sensation is supplemented by consequent ideas which are associated with it. If a certain sensation is strongly associated with the attributes of a certain thing, that thing is perceived when we get that sensation. But where the sensation is associated with more than one reality, unconscious processes weigh the odds, and we perceive the most probable thing: "all brain-processes are such as give rise to what we may call FIGURED consciousness" (James, 1890, p. 82). Accurate and fluent perception thus rests on the perceiver having acquired the appropriately weighted range of associations for sensory inputs. Learners have to figure stimuli out: to learn the probability distribution P(*interpretation|cue, context*), the probability of an interpretation given a stimulus cue. This figuring is achieved, and cognition optimized, by implicit tallying of the frequency, recency, and context of stimuli in our experience.

Every linguistic form is ambiguous too. We should ask therefore whether the same probability distribution holds - P(*interpretation|form, context*) and whether this figuring is achieved, and language cognition optimized, by implicit tallying of the frequency, recency, and context of linguistic constructions in usage.

Rational analysis (Anderson, 1990) aims to answer *why* human cognition is the way it is. Its guiding principle is that the cognitive system optimizes the adaptation of the behavior of the organism to its environment in the sense that the behavior of the mechanism is as efficient as it conceivably could be given the structure of the problem space or input-output mapping it must solve. For the case of memory, the criterial factor is the optimal estimation of an item's need probability. Rational analysis considers the way that human memory corresponds to this needs function. Consider the relative availability of items in the mental lexicon. Three factors determine *information need*: frequency, recency, and context. *Frequency*: the probability of a word occurring in a particular source is predicted by its past probability of occurrence in that source. It works for all sorts of information. Whether organizing a library, a mental lexicon, or an instrument tray, you should put the most-used items nearer to hand. *Recency*: there is a power function relating probability of a word occurring on day *n* to how long it has been since the word previously occurred. The forgetting curve is rational in that it follows this trend. *Context*: a particular word is more likely to occur when other words that have historically co-occurred with it are present. See how these three aspects of *information need* in the problem space are satisfied by their *cognitive* counterparts in learning, memory and perception summarized at the beginning of this I-Spy.

1.2. Categorization

Ambridge et al. correctly emphasize that frequency effects are observed at different levels, including that of abstract categories. Investigating linguistic categories, their nature, their utility in communication, and how they come to be, should draw on domain-general categorization research. Categories have graded structure, with some members being better exemplars than others. Research shows:

For exemplar learning:

A frequency effect: Exemplars presented with higher frequency are classified more accurately and are seen to be more typical of the category.

A family resemblance / prototypicality effect: Exemplars that are more similar to other members of the category, and less similar to members of contrasting categories, are learned faster, classified more accurately and quickly, and considered to be more typical of the category.

An in-crowd effect: Exemplars that are more similar to the higher frequency items are classified more accurately and are seen to be more typical.

For category learning:

A variability effect: Less variable categories, where exemplars are close to the prototype, are acquired easier than categories comprised of diverse, more variable exemplars. Low variance samples are particularly conducive to forming strong category generalizations. In contrast, diversity increases overgeneralization.

An order of exposure effect: Learners acquire more knowledge about a category if they are shown typical items at the beginning of learning followed by more atypical items.

Measurement

These mechanisms of general cognition are ripe for continued investigation in language acquisition research, and there are useful statistical operationalizations for their measurement in linguistic corpora (see Gries & Ellis, 2015, for a review).

Language as a Complex Adaptive System

Constructions map categories of language form to categories of our experience. Constructionist accounts of acquisition thus involve the distributional analysis of the language stream and the parallel analysis of contingent perceptual activity, with abstract constructions being learned from the conspiracy of concrete exemplars of usage following statistical learning mechanisms relating input and learner cognition.

The statistical patterns gathered here are as telling in the emergence of language as is the fossil record in the evolution of species. As in the fossil record there are gaps. Ambridge et al. emphasize the complex interactions between numerous variables whose potencies vary with proficiency. Constructions are nested and overlap at various levels (morphology within lexis within grammar; hierarchical semantic organizations, etc.). Sequential elements can be multiply memorized as wholes at these different levels. So there is no one direction of growth, but rather continuing interplay between modalities, between top-down and bottom-up processes, and between memorized structures and more open constructions. Constructions develop hierarchically by repeated cycles of differentiation and integration. This is why we need to go beyond univariate statistics, beyond multivariate statistics still, towards computational modeling (richly informed by corpus data), and why there is sense in viewing language as a complex adaptive system (Beckner et al., 2009).

Nevertheless, as we explore the field of language, these patterns are everywhere (e.g., Ellis, O'Donnell, & Römer, 2014; Ogden & Ellis, 2014). Following Ambridge et al., let us identify them, and recognize their theoretical significance.

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (ed.), *Mechanisms of Language Acquisition,* 157-193. Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Ebbinghaus, H. (1885). *Memory: A contribution to experimental psychology* (H. A. R. C. E. B. (1913), Trans.). New York: Teachers College, Columbia.

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

Gries, S. Th., & Ellis, N.C. (in press, 2015). Statistical measures for Usage-based linguistics. *Currents in Language* Learning, ***2***.

James, W. (1890). *The principles of psychology* (Vol. 1). New York: Holt.

Ogden, D. C. & Ellis, N. C. (2014). *Child Acquisition of Distributional and Semantic Properties of Verb-Argument Constructions and Its Lead in Adult Usage*. Paper presented at the 5th UK Cognitive Linguistics Conference.