*SSLA*, **18**, 361–368. Printed in the United States of America.

# ANALYZING LANGUAGE SEQUENCE IN THE SEQUENCE OF LANGUAGE ACQUISITION

## *Some Comments on Major and Ioup*

Nick C. Ellis
*University of Wales, Bangor*

Apparent complexity may come more from the problem than from the system that learns to solve it. Simon (1969) illustrated this by describing the path of an ant making its homeward journey on a pebbled beach. The path seems complicated. The ant probes, doubles back, circumnavigates, and zigzags. But these actions are not deep and mysterious manifestations of intellectual power. Closer scrutiny reveals that the control decisions are both simple and few in number. An environment-driven problem-solver often produces behavior that is complex only because a complex environment drives it.

Language learners have to solve the extremely complex problem of language. Thus, theories of language acquisition, like explanations of Simon's ant, might easily overestimate the degree of control sophistication and innate neurological predisposition that is required. Both Major and Ioup correctly characterize associative (connectionist, constructivist) learning approaches to language acquisition as being investigations into what representations can result when simple learning mechanisms for distributional analysis are exposed to complex language evidence. Occam's Razor is influential in the associationists' attributions of learning mechanisms:

> Implicit knowledge of language may be stored in connections among simple processing units organized in networks. While the behavior of such networks may be describable (at least approximately) as conforming to some system of rules, we suggest that an account of the fine structure of the phenomena of

This paper was assisted by grant R000236099 from the Economic and Social Research Council (UK). Thanks are due to Ivan Uemlianin, Marilyn Vihman, Ginny Gathercole, Gordon Brown, Kim Ellis, Gabe Ellis, Cathair O'Dochartaigh, Arthur Reber, Ernest Lee, and Fergus Lowe and other colleagues at UWB for providing the fertile environment that encouraged the growth of this paper.

Address reprint requests and correspondence to Nick Ellis, School of Psychology, University of Wales Bangor, Gwynedd, Wales, LL57 2DG, UK; e-mail: n.ellis@bangor.ac.uk.

> language use can best be formulated in models that make reference to the characteristics of the underlying networks. (Rumelhart & McClelland, 1987, p. 196)

Connectionists test such conceptualizations by evaluating the effectiveness of their implementations. Computational practicalities dictate that, for the moment at least, this can only be done in a piecemeal fashion: Many separate models now address the acquisition of morphology, phonology, novel word repetition, prosody, semantic structure, and syntactic structure (Levy, Bairaktaris, Bullinaria, & Cairns, 1995; MacWhinney, 1987; MacWhinney & Leinbach, 1991; Rumelhart & McClelland, 1986). Even these simple "test-tube" demonstrations put the lie to Major's claim that behaviorism and connectionism "cannot explain creative aspects of language" (p. 351); rather, they repeatedly show that connectionist models can extract the regularities in each of these domains of language and then operate in a rule-like (but not rule-governed) way. The current connectionist enterprise is an effort toward giving these simple learning mechanisms access to the true complexity of the language evidence, (a) by expanding the models within each domain, improving the low-level representations to give the learning mechanisms better access to the true richness that is there in the speech stream, and (b) by combining different sources of evidence (prosodic, semantic, lexical distributional, etc.) to allow interaction between these domains.

Constructivists believe that

> many universal or at least high-probability outcomes are so inevitable given a certain "problem-space" that extensive genetic underwriting is unnecessary.... Just as the conceptual components of language may derive from cognitive content, so might the computational facts about language stem from nonlinguistic processing, that is, from the multitude of competing and converging constraints imposed by perception, production, and memory for linear forms in real time. (Bates, 1984, pp. 188–190)

At this point in her argument, Bates refers to Slobin's (1973) operating principles for the learner processing strategies that result in this extraction of structure. But, constructivist to the core, she elsewhere inquires, Where do these strategies come from? They come from the environment, from the cues in language itself (Bates & MacWhinney, 1981). Language is a complex, hierarchical, behavioral structure. Its acquisition is rich in sequential dependencies: syllables and formulaic phrases before phonemes and features, holophrases before words, words before simple sentences, simple sentences before lexical categories, and lexical categories before complex sentences.

> It is a general rule of both phylogeny and ontogeny that complex structures evolve by differentiation of smaller structures from larger. Accordingly, we do not expect children to build words from phonemes, as adults do; rather, we should expect phonemes to emerge from words... similar principles must apply to the development of word classes and syntactic structures, a fact not generally recognized in developmental psycholinguistics. (Studdert-Kennedy, 1991, p. 16)

Nor, I would add, is it sufficiently recognized in SLA. Sequence analysis underpins the acquisition of both phonology and syntax.

## PHONOLOGICAL ACQUISITION

I agree with both Major and Ioup that we acquire abstract systems of phonology. Their various examples show that (a) there are perceptual constraints on phonology (the /f/ and /θ/ contrast is notoriously difficult for both adults and children [Miller & Nicely, 1955; Velleman, 1988]), (b) there are motoric constraints on phonology (open syllables are indeed produced before closed ones, but do we need to postulate more specifically linguistic universals to describe the primitive infant vocal tract?), and (c) these at times moderate the direct influence of frequency of input. But, surely, we do not disagree concerning the all-pervading influence of frequency in the acquisition of phonology, whereby successful language learners acquire the phonological system to which they are exposed? Vihman (e.g., 1996, chapter 6) presents a bootstrapping account of the acquisition of phonological systematicity where frequency and salience in input interact with the learner's productive repertoire—the available vocal motor schemes serve as an "articulatory filter," which makes salient particular words in the input stream whose phonetic pattern falls within the range of the learner's productive repertoire.

I particularly agree with Ioup's emphasis on the serial nature of the speech stream where "the same acoustic features that characterize a sound in one environment will be used to identify a totally different sound when the environment changes" (p. 356). The acknowledgment of the sequence analysis aspects of speech perception, and the interaction of acoustic cues, preceding and following context, and even lexical, semantic, and syntactic context, served as the starting point for the first (McClelland & Elman, 1986) connectionist model of speech perception. There are more recent statistical models of speech showing how low-level phonotactics could be used by the neonate as a cue for initially breaking up the continuous stream of speech and how a metrical segmentation strategy could arise without recourse to positing sensitivity to metrical information as part of a genetic endowment (Cairns, Shillcock, Chater, & Levy, 1995).

## BOTTOM-UP OR TOP-DOWN PROCESSING

Major is mistaken when he says that my account involves only bottom-up processing and that I eschew top-down models (p. 352). Ellis (1996) lists the range of LTM influences on STM, describes cyclical reciprocal interactions as learners bootstrap their way to knowledge of L2 structure, and discusses the role of salience and attentional processes in the acquisition of grammar. This account is entirely within the spirit of Neisser's (1976) "cycle of perception." Experience of our environment leads to modification of our schema, our schema direct our exploration of the environment, our exploration samples the available information in the environment, and thus the cycle continues. The same systems that perceive language represent language. Thus, the "cycle of perception" is also the "cycle of learning"; bottom-up and top-down processes are in constant interaction.

## THE "HOMOPHONY PROBLEM" IN PDP MODELS

Ioup quotes the Pinker and Prince (1988) criticism of the Rumelhart and McClelland (1986) past-tense model: The original model should assign the same past-tense morphology to homophonic verb pairs like *ring/wring* and *right/write* because they contain the same sound sequences. Pinker and Prince (1988, p. 174) claim that this problem arises because parallel distributed processing (PDP) models are incapable of representing lexical identity. This is not true. By definition, homophones are words of the same pronunciation but different meaning. To distinguish between them, the model must have access to some level of semantic representation. MacWhinney and Leinbach (1991) report simulations using a revised verb-learning model where the phonological architecture is supplemented with additional nodes for semantic features. These simulations rapidly learned that *rang* was the past tense of *ring* $_1$ [action, auditory-result, high-pitch, object-thing, sharp-onset], whereas *ringed* was the past tense of *ring* $_2$ [action, circle, completive, object-thing, positional-change, surround]. The revised implementation also answered a wide range of the other Pinker and Prince (1988) and Lachter and Bever (1988) criticisms of the original past-tense model.

Ioup concludes this section: "The child must . . . determine which verbs have rule-governed past tenses and which are to be stored as irregular forms" (p. 356). This is the classic false equation of rule-like equals rule-governed. Connectionist models demonstrate how one distributed associative learning system can mimic the human learning characteristics of regular and irregular inflectional morphology both in the L1 (Plunkett & Marchman, 1993) and in the L2 (Ellis & Schmidt, 1996).

## SYNTACTIC ACQUISITION

Ioup offers the ability of 3-year-old Elias to distinguish between embedded interrogatives and free relative clauses as evidence against sequential order accounts. I am reluctant to comment without seeing a representative corpus of Elias's utterances in context, but my interpretation is exactly the reverse. Elias is showing that he knows something about the word order of questions. Furthermore, he is demonstrating knowledge that words like *ask* and *know* behave differently from words like *sat* and *found.* Elias has implicit knowledge about subcategories of verbs. In the terms of UG, these words have different c-selection properties that would be represented in the lexicon in terms of different subcategorization frames. Within UG, the differing subcategorization frames then affect interpretation or production of a sentence according to the Projection Principle (Chomsky, 1981, p. 29). The lexical items drive the process. The Projection Principle marks a milestone in the evolution, over the last 20 years, of theories of grammar increasingly putting more syntax into the lexicon and correspondingly less into rules; the Minimalist Program (Chomsky, 1993) marks the current extreme of this progression. So, we must ask, whence does the learner acquire these subcategorization frames? No one, however nativist, holds that this lexical information is innate. Everyone agrees that it is learned. The relevant sources of evidence are (a) the distributional properties of words and (b) semantics—

the s-selection properties of lexical items, determining the particular number of arguments bearing particular theta-roles, are a remarkably good predictor of c-selection properties. Chomsky (1986, pp. 88–90) suggests that c-selection may eventually be entirely reducible to s-selection (Cook & Newson, 1996).

But putting semantics aside for the moment, there is a tremendous amount of information about the subcategorization requirements of words to be had from simple distributional analysis, as long as the analysis is done on a suitably large collection of language (rather more than the handful of example utterances that commonly appear preceded by number labels, either with or without a leading *, in linguistics texts). In my original article (Ellis, 1996, pp. 111–112), I referred to computational work on Markovian analysis of English word class. In particular, Finch and Chater (1994) report the results of applying a standard statistical clustering algorithm to data concerning the distributional frequencies of words in a very noisy corpus of 40 million items from the USENET newsgroups. The distributional frequency was simply counted over four words of context, that is, the two preceding and two following lexical items. The cluster analysis produces a clear separation of different word classes; for example, verbs cluster separately from nouns, conjunctions are a bit like *wh*-words that resemble prepositions, which in turn are very different from determiners (p. 301). But further, we can look within each of these classes, for example, to see how the verb *-ing* forms behave. Then we see that *ask* and *tell* behave very similarly, and they are quite like *show* and *hear* and very different from *be*, and *use*, and *put*, and *find* (p. 303). Finch and Chater only report the data on 28 verb *-ing* forms, and I fear that Elias's *know* isn't there. But these distributional analyses, totally devoid of either grammatical or semantic input and based on simple counts, can produce very interesting descriptions of linguistic structure with not only clusters suggestive of word class information but also subclusters of words taking different subcategorization frames.

## ALM AND OTHER METHODOLOGIES

Ioup claims that the audiolingual method (ALM) was prominent in education for about 20 years but that, as a learning technique, it "was a dismal failure" (p. 358). By this evidence, then, it's fared just about as well as Grammar-translation, the Cognitive Code Method, Total Physical Response, the Natural Approach, and other competitors. The histories of second and foreign language teaching demonstrate radical swings in favored methodology (see, e.g., Kelly, 1969; Stern, 1983). All too often a method is adopted for the decade or two necessary for students and their parents to become dissatisfied, and it is then replaced by a fundamentally different alternative. This holds sway for 10–15 years and the cycle begins again. There is a serious lack of empirical evaluation of the outcomes of different techniques. One can only hope that moves toward a scientist-practitioner model of SLA will redress this balance.

In my original article (p. 97), I mentioned ALM in the context of a range of other teaching methods. It is far too simplistic to believe that a theoretical account of language acquisition that focuses on surface form as a sequence learning problem necessarily implies ALM as a pedagogical method. I stress again that I believe in a bootstrapping model where bottom-up and top-down processes are in constant interaction. Some parts of our environment can be made more salient (e.g., by "grammatical consciousness raising" or "input enhancement" or "focus on form"), and learners are more likely to learn about the *parts of the environment* to which they selectively *attend.* I am deeply concerned with the issue of consciousness in language learning (Ellis, 1994). I take a functional perspective on language acquisition that holds that the primary goal of the learner is to communicate meanings. Thus, my position (Ellis & Laporte, in press) concerning the relationships between the theory and practice of SLA is somewhat more sophisticated than that implied by Ioup.

## CONCLUSIONS

Ioup concludes her response by saying that "Ellis acknowledges that there are innate grammatical resources at work in language acquisition" (p. 359). I reject this characterization and request a more careful reading (Ellis, 1996, pp. 101, 116–120) for my views on linguistic universals. Although I do acknowledge that work within the UG framework has provided the most complete description of language competence to date, like Studdert-Kennedy (1991), I suspect that UG is neither a prescription nor a program for development, but rather it is a partial and a posteriori description of the phenotypic product of the developmental system. In this view, UG is a consequence, not a condition, of development.

Meteorology has its rules and principles of the phenomena of the atmosphere that allow the prediction of weather. Geology has its rules and principles to describe and summarize the successive changes in the earth's crust. But these "rules" are the descriptions and heuristics of science. They play no causal role in shifting even a grain of sand or a molecule of water. It is the interaction of water and rocks that smoothes the irregularites and grinds the pebbles and sand. UG is like the other -ologies with its principles and parameters to describe language, and the rules of UG have a similar causal status.

The evidence of language grinds on our perceptuomotor and cognitive apparatus. The proper study of language acquisition is to chart the course by which perceptual, motoric, and cognitive functions induce structure, from undifferentiated novice performance to that remarkably differentiated native-like competence that can be described by theories of linguistics such as UG. For me, a more relevant universal concerns process and learning rather than content: It is to be found in efforts to rationalize intelligence in terms of models of optimal (Bayesian) inference in the presence of uncertainty. Nativists and empiricists will continue to argue about their essential and principled differences, but in practice their enterprises are highly complementary. SLA research is so much fun because there is both language and learning to be understood.

(*Received 3 March 1996*)

## REFERENCES


Bates, E. (1984). Bioprograms and the innateness hypothesis. *Behavioral and Brain Sciences, 7,* 188–190.

Bates, E., & MacWhinney, B. (1981). Second language acquisition from a functionalist perspective. In H. Winitz (Ed.), *Native language and foreign language acquisition. Annals of the New York Academy of Sciences, 379,* 190–214.

Cairns, P., Shillcock, R., Chater, N., & Levy, J. P. (1995). Bottom-up connectionist modeling of speech. In J. P. Levy, D. Bairaktaris, J. A. Bullinaria, & P. Cairns (Eds.), *Connectionist models of memory and language* (pp. 289–310). London: UCL Press.

Chomsky, N. (1981). *Lectures on government and binding.* Dordrecht: Foris.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin and use.* New York: Praeger.

Chomsky, N. (1993). A minimalist program for linguistic theory. In K. Hale & S. J. Keyser (Eds.), *The view from Building 20: Essays in linguistics in honor of Sylvian Bromberger* (pp. 1-52). Cambridge, MA: MIT Press.

Cook, V. J., & Newson, M. (1996). *Chomsky's Universal Grammar: An introduction,* 2nd ed. Oxford: Blackwell.

Ellis, N. C. (Ed.). (1994). *Implicit and explicit learning of languages.* London: Academic Press.

Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition, 18,* 91-126.

Ellis, N. C., & Laporte, N. (in press). Contexts of acquisition: Effects of formal instruction and naturalistic exposure on SLA. In A. de Groot & J. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives.* Hillsdale, NJ: Erlbaum.

Ellis, N. C., & Schmidt, R. (1996). Rules or associations in the acquisition of second language morphology? Human learning of regular and irregular morphosyntax and connectionist simulations of the frequency by regularity interaction. Manuscript in preparation.

Finch, S., & Chater, N. (1994). Learning syntactic categories: A statistical approach. In M. Oaksford & G. D. A. Brown (Eds.), *Neurodynamics and psychology* (pp. 295–321). London: Academic Press.

Kelly, L. G. (1969). *Twenty-five centuries of language teaching.* Rowley, MA: Newbury House.

Lachter, J., & Bever, T. (1988). The relation between linguistic structure and associative theories of language learning: A constructive critique of some connectionist learning models. *Cognition, 28,* 195–247.

Levy, J. P., Bairaktaris, D., Bullinaria, J. A., & Cairns, P. (1995). *Connectionist models of memory and language.* London: UCL Press.

MacWhinney, B. (Ed.). (1987). *Mechanisms of language acquisition.* Hillsdale, NJ: Erlbaum.

MacWhinney, B., & Leinbach, J. (1991). Implementations are not conceptualisations: Revising the verb learning model. *Cognition, 40,* 121–157.

Marchman, V. A., & Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language, 21,* 339–366.

McClelland, J. L., & Elman, J. L. (1986). Interactive processes in speech perception: The TRACE Model. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 2. Psychological and biological models* (pp. 58–121). Cambridge, MA: MIT Press.

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustic Society of America, 27,* 338–352.

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology.* San Francisco: Freeman.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition, 29,* 195–247.

Plunkett, K., & Marchmann, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition, 48,* 73–193.

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tense of English verbs. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models* (pp. 216–271). Cambridge, MA: MIT Press.

Rumelhart, D. E., & McClelland, J. L. (1987). Learning the past tense of English verbs: Implicit rules or parallel distributed processes? In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 195–248). Hillsdale, NJ: Erlbaum.

Simon, H. A. (1969). *The sciences of the artificial.* Cambridge, MA: MIT Press.

Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In C. A. Ferguson & D. I. Slobin (Eds.), *Studies of child language development* (pp. 175–208). New York: Holt Rinehart Winston.

Stern, H. H. (1983). *Fundamental concepts of language teaching.* Oxford: Oxford University Press.

Studdert-Kennedy, M. (1991). Language development from an evolutionary perspective. In N. A. Krasnegor, D. M. Rumbaugh, R. L. Schiefelbusch, & M. Studdert-Kennedy (Eds.), *Biological and behavioral determinants of language development* (pp. 5–28). Hillsdale, NJ: Erlbaum.

Velleman, S. L. (1988). The role of linguistic perception in later phonological development. *Applied Psycholinguistics, 9*, 221–236.

Vihman, M. M. (1996). *Phonological development: The origins of language in the child.* Oxford: Blackwell.