The Epigenesis of Language:

Acquisition as a Sequence Learning Problem

Nick Ellis

School of Psychology,

University of Wales,

Bangor,

Gwynedd,

Wales

LL57 2DG

UK

N.Ellis@bangor.ac.uk

Paper submitted to the Proceedings of the BAAL Annual Meeting 1996,

Evolving Models of Language

First draft: 29 September, 1996

"Precisely because evolution produced an animal capable of tackling whatever challenge the environment might offer, the answer must be that very few behavioral patterns are rigidly built into the human brain."

(Leakey & Lewin, 1977, p. 254).

Constructivist Approaches to Language Acquisition

Human beings are not present in the embryo as miniature complete forms. Rather, the development of an embryo consists of the gradual production and organisation of parts - this is the theory of epigenesis: "the additament of parts budding one out of another" (Harvey, 1653, p. 272). Human language is no more preformed in the embryo than is human neuroanatomy or human society. This paper advocates a constructivist research agenda which views language acquisition as a problem of *sequence learning* and which denies innate linguistic universals and rules of language as having any causal influence in language acquisition.

Our ability as humans to be awed by complexity is rarely matched by our talent for scientific understanding. Wonderment makes us posit predeterminism. The theologian William Paley (1828), awed by the intricacy of the human eye and its fittingness for purpose, took its complexity as evidence of God as an artificer who formed it for its purpose. Yet when the focus of study turned towards the development process, it soon

became apparent how simple processes of evolution, operating in complex environments, gradually resulted in this impressive phylogenetic feat (Darwin, 1859; Dawkins, 1986). Chomsky (1965), arguably knowing more about the intricacies and universalities of human language than anybody else at the time, took the complexity of these competencies as evidence of an evolutionary process culminating in the provision of linguistic universals represented in a language acquisition device as a general human inheritance. The Chomskian legacy of Universal Grammar continues to predominate linguistics and second language acquisition.

Yet the last 20 years has seen a revival of approaches to language acquisition, traceable as far back as Locke (1690), that all ideas originate in experience, and that there are no innate ideas. Language is *learned*. These constructivist views of language acquisition, including connectionist approaches (McClelland & Rumelhart, 1986; Levy, Bairaktaris, Bullinaria & Cairns, 1995), functional linguistics (Bates & MacWhinney, 1981; MacWhinney & Bates, 1989), and cognitive linguistics (Lakoff, 1987; Langacker, 1987), believe that Chomsky's (1965) assumption is as suspect as that of Paley. As the study of language turns to consider ontogenetic acquisition processes, it favours a conclusion whereby the complexity of the final result stems from simple learning processes applied, over extended periods of practice in the learner's lifespan, to the rich and complex problem space of language evidence.

Apparent complexity may come more from the problem than from the system which 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 behaviour that is complex only because a complex environment drives it. Language learners have to solve the problem of language. Thus in this case, like that of Simon's ant, it is all too easy to overestimate the degree of control sophistication and innate neurological predisposition required in its solution.

The complexity is in the language, not the learning process: "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, commenting on Bickerton's Language Bioprogram). Bates refers to Slobin's (1973) operating principles as being good candidates for the learner's processing strategies which 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 (see Bates & MacWhinney, 1981, The Competition Model).

Constructivists are unhappy with nativist explanations simply because they may not be necessary - why posit predeterminism, like magic, when simpler explanations might suffice (Tomasello, 1995; Sampson, 1980)? They are additionally unhappy because the innateness hypothesis has no process explanation - our current theories of brain function, process and development do not readily allow for the inheritance of structures which might serve as principles or parameters of UG. Without such a process explanation, innatist theories are left with a 'and here a miracle occurs' step in their argumentation. Incompleteness of explanation is not a fatal flaw - Mendel was correct long before Crick and Watson provided a process explanation - but we do expect the gaps

to be filled eventually, and current neuroscience makes implausible any assumptions about inherited parameter 'switches'.

However, constructivists have an onus beyond criticising other views - they must additionally demonstrate the power of their own explanations. Thus, with the rise of connectionism, much recent constructivist research has concerned computational 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 behaviour 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 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; Ellis & Schmidt, in press; Rumelhart & McClelland, 1986). Yet even these simple "test-tube" demonstrations demonstrate 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 towards 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 provide 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. We are, as yet, many generations away from a connectionist system that might pass the Turing test for nativelike language use, but the demonstrations to date are encouragingly persuasive that some of the interesting systematicities of language development that can be described as if they are the product of rule-governed systems, can indeed result from very simple learning mechanisms which statistically abstract the regularities that are present in the evidence of language.

Thus the constructivist view is that language learning results from general processes of human inductive reasoning being applied to the specific problem of language. There is no language acquisition device specifiable in terms of linguistic universals, principles and parameters, or language-specific learning mechanisms. Rather,

language is cut of the same cloth as other cognitive *processes*, but it is special in terms of its cognitive *content*. Learners' language comes not directly from their genes, but rather from the structure of adult language, from the structure of their cognitive and social cognitive skills, and from the constraints on communication inherent in expressing non-linear cognition into the *linear* channel provided by the human vocal-auditory apparatus (Bates, Thal, & Marchman, 1991).

This leads us to the characterization of language as a *sequence* learning problem. "(L)anguage, as a complex, hierarchical, behavioral structure with a lengthy course of development ... 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, lexical categories before complex sentences, and so on." (Studdert-Kennedy, 1991, p. 10). This assumption that each of the subsystems, phonology and syntax, evolved hierarchically by repeated cycles of differentiation and integration, promises an understanding of language development in functional terms. "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" (Studdert-Kennedy, 1991, p. 16). Studdert-Kennedy goes on to point out that "similar principles must apply to the development of word classes and syntactic structures, a fact not generally recognized in developmental psycholinguistics." Nor is it sufficiently recognized in the study of second- or foreign-language acquisition.

The remainder of this paper will concentrate on the sequence learning aspects of the acquisition of language. It will demonstrate (i) individual learners' verbal short-term memory ability is a good predictor of their language learning aptitude, (ii) much of fluent language use is memory based, following the idiom principle rather than open-class models, (iii) simple sequence analysis of large collections of language allows the automatic induction of grammatical word-class, (iv) analysis of learners' speech, both in L1 and L2, demonstrates how statistical learning processes of chunking and sequence analysis characterise their progression to fluency. I have gathered evidence for some of these claims before (Ellis, 1996). Rather than repeat this here, I will try in the main to bolster the arguments with additional material.

Language Learning as Sequence Analysis

One possible description of language learning is as follows. *Language learning is the learning and analysis of sequences:* (a) the learner must acquire sound sequences in words, (b) the learner must acquire word sequences in phrases, and (c) these sequences form the database for the abstraction of grammar. (a) Learning word structure involves identifying the categorical units of speech perception, their particular sequences in particular words, and their general sequential probabilities in the language. (b) Learning discourse structure largely involves learning particular sequences of words in stock phrases and collocations. The idiom principle underlies much of fluent language use and language learners need to acquire particular sequences of words in particular phrases, and

the general sequential probabilities of words in the language. (c) Learning the grammatical word-class of a particular word, and learning grammatical structures more generally, involves in large part the automatic implicit analysis of the word's sequential position relative to other words in the learner's stock of known phrases which contain it.

This is not an explanation, it is a description which, while hardly contentious, clearly demonstrates how sequence analysis pervades all levels of language processing.

Verbal sequencing ability predicts language aptitude

One advantage of this focus on sequence analysis is that it allows insight into individual differences in language aptitude. Individuals differ in their ability to repeat phonological sequences (this is known as phonological short-term memory [STM] span). This ability to repeat verbal sequences (for example, new phone numbers or nonwords like 'sloppendash') immediately after hearing them, is a good predictor of learner's facility to acquire vocabulary and syntax in first, second, and foreign language learning. Ellis (1996) reviews a wide range of evidence for this: (i) phonological STM span predicts vocabulary acquisition in L1 and L2, (ii) interfering with phonological STM by means of articulatory suppression disrupts vocabulary learning, (iii) repetition and productive rehearsal of novel words promotes their long-term consolidation and retention, (iv) phonological STM predicts syntax acquisition in L1 and L2, (v) phonological rehearsal of L2 utterances results in superior performance in receptive skills in terms of learning to comprehend and translate L2 words and phrases, explicit metalinguistic knowledge of the detailed content of grammatical regularities, acquisition of the L2 forms of words and phrases, accuracy in L2 pronunciation, and grammatical fluency and accuracy (Ellis & Sinclair, 1996).

So what is the involvement of phonological memory in language learning? Ellis (1996) presents an account which is based on the basic learning mechanism of 'chunking'. This term was coined by George Miller in his classical review of short-term memory (Miller, 1956). It is the development of permanent sets of associative connections in long-term storage and is the process which underlies the attainment of automaticity and fluency in language. "A chunk is a unit of memory organization, formed by bringing together a set of already formed chunks in memory and welding them together into a larger unit. Chunking implies the ability to build up such structures recursively, thus leading to a hierarchical organization of memory. Chunking appears to be a ubiquitous feature of human memory. Conceivably, it could form the basis for an equally ubiquitous law of practice."(Newell, 1990, p. 7). Chunking allows us to bootstrap our way into language: Repetition of sequences in phonological STM allows their consolidation in phonological long-term memory (LTM). Perception of frequent sequences, and the more frequent subsequences within them, allows their chunking in

phonological LTM. The same cognitive system which does the LTM for phonological sequences does the perception of phonological sequences. Thus the tuning of phonological LTM to regular sequences allows more ready perception of input which contains regular sequences. Regular sequences are thus perceived as chunks and, as a result, L2-experienced individuals' phonological STM for regular sequences is greater than for irregular ones. Such influences of LTM on STM make the relationship between these systems truly reciprocal and underlie the development of automaticity (LaBerge & Samuels, 1974; McLaughlin, 1987). This is an epigenetic view of language whereby acquisition is the additament of chunks budding one out of another.

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 which perceive language represent language. Thus the "cycle of perception" (Neisser, 1976) is also the "cycle of learning" - bottom-up and top-down processes are in constant interaction.

Chunking and the Idiom Principle

It is easy to see this cycle of learning in vocabulary acquisition: As learners' L2 vocabulary extends, as they practise hearing and producing L2 words, so they automatically and implicitly acquire knowledge of the statistical frequencies and sequential probabilities of the phonotactics of the L2. Their input and output modules for L2 processing begin to abstract knowledge of L2 regularities, thus to become more proficient at short-term repetition of novel L2 words. And so L2 vocabulary learning lifts itself up by its bootstraps. But the same processes operate at all levels of language. Language reception and production are mediated by learner's representations of chunks of language: "Suppose that, instead of shaping discourse according to rules, one really pulls old language from memory (particularly old language, with all its words in and everything,) and then reshapes it to the current context: "context shaping', as Bateson puts it, 'is just another term for grammar'" (Becker, 1983, p. 218). This is why elicited

imitation tests serve so well as measures of second-language competence (Lado, 1965; Bley-Vroman & Chaudron, 1994).

As we analyse language performance, so the underlying chunks become readily apparent. Sinclair (1991), as a result of his experience directing the Cobuild project, the largest lexicographic analysis of the English language to date, proposed *the principle of idiom* - "a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments. To some extent this may reflect the recurrence of similar situations in human affairs; it may illustrate a natural tendency to economy of effort; or it may be motivated in part by the exigencies of real-time conversation. However it arises, it has been relegated to an inferior position in most current linguistics, because it does not fit the open-choice model" (Sinclair, 1991, p. 110). Rather than its being a rather minor feature, compared with grammar, Sinclair suggests that for normal texts, the first mode of analysis to be applied is the idiom principle, since most of text is interpretable by this principle. Comparisons of written and spoken corpora demonstrate that collocations are even more frequent in spoken language (Butler, 1995). Collocations and stock phrases are viewed with just the same importance in FL research where they are known as holophrases (Corder, 1973), prefabricated routines and patterns (Hakuta, 1974), formulaic speech (Wong-Fillmore, 1976), memorised sentences and lexicalized stems (Pawley & Syder, 1983), lexical phrases (Nattinger & DeCarrico, 1992), or formulas (R. Ellis, 1994). An important index of nativelike competence is that the learner uses idioms fluently. So language learning involves learning sequences of words (frequent collocations, phrases, and idioms) as well as sequences within words.

With the benefit of hindsight, it comes as no surprise that language is acquired in this way. The formation of chunks, as stable intermediate structures, is the mechanism underlying the evolution and organization of many complex hierarchical systems in biology, society, and physics (Simon, 1962; Dawkins, 1976).

Working out how words work - The distributional analysis of memorized collocations

As we analyse word sequence chunks, so we discover that they have characteristic structures all of their own. Linguists call these regularities grammar. And if we take a bottom-up approach, and simply describe the distributional properties of words in chunks, so we discover that something very close to traditional grammatical word-class information emerges. Kiss (1973) provided the first computational model of the acquisition of grammatical word class from accumulating evidence of word distributions. An associative learning program was exposed to an input corpus of 15,000 words gathered from tape recordings of seven Scottish middle class mothers talking to their children who were between one and three years of age. The program read the corpus and established associative links between the words and their contexts (here defined as their *immediate successor*). Thus, for example, the program counted that *the* was followed by *house* 4.1% of the time, by *horse* 3.4%, by *same* 1%, by *put* never, etc., that *a* was connected to *horse* 4.2%, to *house* 2.9%, to *put* never, etc., etc. For computational reasons (this work was done in the days of punched cards) such 'right-context' distributional vectors were only computed for 31 frequent words of the corpus. These

vectors constituted a level of associative representation which was a network of transitions. Next a classification learning program analyzed this information to produce connections between word representations which had strengths determined by the degree of similarity between the words in terms of the degree to which they tended to occur together after a common predecessor (i.e. the degree of similarity based on their 'left-contexts'). This information formed a level of representation which was a network of word similarities. Finally the classification program analyzed this similarity information to produce a third network which clustered them into groups of similar words. The clusters that arose were as follows: (*hen sheep pig farmer cow house horse*) (*can are do think see*) (*little big nice*) (*this he that it*) (*a the*) (*you I*). It seemed that these processes discovered word classes which were nounlike, verblike, adjectivelike, articlelike, pronounlike, etc. Thus the third level of representation, which arises from simple analysis of word distributional properties, can be said to be that of word-class. Kiss argues that in this way language learners can bootstrap their way into discovering word classes. More recent, and much larger, demonstrations show that such bootstrapping results from a variety of analysis methods including statistical, recurrent neural network, or self-organizing map models (Sampson, 1987; Charniak, 1993; Finch & Chater, 1994; Honkela, Pulkki & Kohonen, 1995).

The Cobuild project represents the largest descriptive enterprise of this type for English, and three key conclusions of this research are (i) that is impossible to describe syntax and lexis independently, (ii) that syntax and semantics are inextricable, and (iii) that language is best described as being collocational streams where patterns flow into each other (often going over the clause boundary). "Through the reliability and

objectivity of the computer evidence, verbs can be divided according to the pattern, and pattern can be seen to correlate with meaning - that is to say, verbs with similar patterns have similar meanings... We can now see that this relation between meaning and pattern is inevitable - that meaning and usage have a profound and systematic effect on one another." (Sinclair, forward to Cobuild Grammar Patterns: Verbs, 1996, p. iv). Thus the Collins Cobuild (1996) analysis of English verbs shows that there are perhaps 100 major patterns of English verbs (of the type, for example, V *by* amount: the verb is followed by a prepositional phrase which consists of the preposition *by* and a noun group indicating an amount as in 'Their incomes have dropped by 30 per cent', 'The Reds were leading by two runs', etc.). Verbs with the same Comp pattern share meanings (the above-illustrated pattern is used by three meaning groups: (i) the 'increase' and 'decrease' group (inc. 'climb', 'decline', 'decrease', 'depreciate', etc.), (ii) the 'win' and 'lose' group (inc. 'lead', 'lose', and 'win'), (iii) the' overrun' group (inc. 'overrun', 'overspend'). Any Comp pattern is describable *only* in terms of its lexis.

Perhaps surprisingly, Chomsky's recent accounts of syntax in the Minimalist Program for Linguistic Theory (MPLT) (Chomsky, 1992, 1995) shares this emphasis on lexis and sequence analysis. (Chomsky, 1989, emphasis added) stated: "there is only one human language apart from the lexicon, and *language acquisition is in essence a matter of determining lexical idiosyncrasies*". Within the MPLT, "*differences between languages are attributed to differences between the features of lexical items in the languages* and specifically between the features of lexical items belonging to the functional categories AGR and Tense... Vs and Ns are taken from the lexicon fully inflected with inflectional affixes... specific bundles of these features of the category AGR and T are lexical items

and *differences between the sets of bundles available in the lexicon account for cross-linguistic syntactic differences between languages*." (Marantz, 1995, p. 366)

Thus this corpus linguistic approach and the MPLT are alike in focusing on lexis as being at the centre of syntax. In both accounts, syntax acquisition reduces to vocabulary acquisition - the analysis of how words work in sequence. As Singleton (1996) quipped, this is the linguistic analogue of the economic maxim: "Look after the pennies and the pounds will look after themselves".

Sequences in learner talk

We have seen that as powerful computers are used for distributional analysis of large language corpora, so they demonstrate the underlying chunks of language and the ways in which lexical items, with their particular valences and subcategorization requirements, operate in these patterns. The final part of the argument for viewing language learning as sequence learning is a demonstration from analyses of collections of *learner's* language that they acquire collocations in their path to fluency, and that their analyses of these chunks gives them the information about lexical idiosyncrasies that allows later more open-class productions. This evidence is there for both L1 and L2.

L1

Tomasello (1992), from whom I have taken the notion of language epigenesis, begins his book, *First Verbs: A Case Study of Early Grammatical Development*, with the following observation from Wittgenstein: "Language games are the forms of language with which a child begins to make use of words... When we look at the simple forms of language the mental mist which seems to enshroud our ordinary use of language

disappears. We see activities, reactions, which are clear-cut and transparent. On the other hand we recognize in these simple processes forms of language not separated by a break from our more complicated ones. We see that we can build up the more complicated forms from the primitive ones by gradually adding new forms." (Wittgenstein, The Blue Book).

Tomasello (1992) kept a detailed diary of his daughter Travis' language between 1 and 2 years old. On the basis of a fine-grained analysis of this corpus he proposed the Verb Island hypothesis: Young children's early verbs and relational terms are individual islands of organization in an otherwise unorganized grammatical system. In the early stages the child learns about arguments and syntactic marking on a verb-by-verb basis, and ordering patterns and morphological markers learned for one verb do not immediately generalize to other verbs. The reason for this is that the nascent language learners do not have any adultlike syntactic categories or rules, nor do they have any kind of word class of verb that would support generalizations across verbs.

Particular summary observations supporting this claim were as follows:

> "There is individuality and contextedness everywhere, signs of broad-based rules nowhere. T did bring order and systematicity to her language during her 2nd year of life, but it was a gradual, constructive process. It did not resemble in any way the instantaneous and irrevocable setting of parameters...
>
> T's earliest three-or-more-word sentences (18-21 months) were almost all structured by verbs. The vast majority of these involved straight-forward coordinations of already produced word combinations

(93%), preserving in almost all cases the established ordering patterns of the constituents (99%).

T began marking the syntagmatic relations in these three-or-more-word sentences through the use of contrastive word order and prepositions. She did this, however, on a verb-by-verb basis. By far the best predictor of the arguments and argument markings that T used with a particular verb at a particular time was previous usage of that verb, not same time usage of other verbs." (Tomasello, 1992, pp. 264-266)

He states his conclusion for epigenesis as follows: "It is not until the child has produced or comprehended a number of sentences with a particular verb that she can construct a syntagmatic category of 'cutter', for example. Not until she has done this with a number of verbs can she construct the more general syntagmatic category of agent or actor. Not until the child has constructed a number of sentences in which various words serve as various types of arguments for various predicates can she construct word classes such as noun or verb. Not until the child has constructed sentences with these more general categories can certain types of complex sentences be produced." (Tomasello, 1992, p. 273-274).

Other analyses of child language corpora point to similar conclusions. For example, Lieven, Pine and Dresner Barnes (1992) show formulae to be both frequent (children's first 100 words typically contain about 20 formulae) and productive (in providing templates which, following analysis, are converted into lexically based patterns).

L2

No observation is entirely theory-free. Yet we are fortunate to have some descriptions of stages of L2 proficiency which were drawn up in as atheoretical way as possible by the American Council on the Teaching of Foreign Languages (ACTFL) (Higgs, 1984). The ACTFL (1986) Oral Proficiency Guidelines include the following descriptions of novice and intermediate levels:

"Novice Low:

Oral production consists of *isolated words and perhaps a few high-frequency phrases*. Essentially no functional communicative ability.

Novice Mid:

Oral production continues to consist of *isolated words and learned phrases* within very predictable areas of need, although quantity is increased. Vocabulary is sufficient only for handling simple, elementary needs and expressing basic courtesies. Utterances rarely consist of more than two or three words and show frequent long pauses and repetition of interlocutor's words.

Novice High:

Able to satisfy partially the requirements of basic communicative exchanges *by relying heavily on learned utterances but occasionally expanding these through simple recombinations of their elements.* Can ask questions or make statements *involving learned material.* Shows signs of spontaneity, although this fails short of real autonomy of expression. *Speech continues to consist of learned utterances* rather than of personalized, situationally adapted ones. Vocabulary centers on areas such as basic objects, places, and most common kinship terms. *Pronunciation may still be strongly influenced by first language*.

Intermediate:

The intermediate level is characterized by an ability to - create with the language *by combining and recombining learned elements*, though primarily in a reactive mode; - initiate, minimally sustain, and close in a simple way basic communicative tasks; and - ask and answer questions.

Intermediate Low:

Able to handle successfully a limited number of interactive, task-oriented social situations. Can ask and answer questions, initiate and respond to simple statements, and maintain face-to-face conversation, *although in a highly restricted manner* and with much linguistic inaccuracy. Within these limitations, can perform such tasks as introducing self, ordering a meal, asking directions, and making purchases. Vocabulary is adequate to express only the most elementary needs. *Strong interference from native language may occur.*

Intermediate-Mid:

Able to handle successfully a variety of uncomplicated, basic communicative tasks and social situations. Can talk simply about self and family members. Can ask and answer questions and participate in simple conversations on topics beyond the most immediate needs; e.g., personal history and leisure-time activities. *Utterance length increases slightly, but speech may continue to be characterized by frequent long pauses, since the smooth incorporation of even basic conversational strategies is often hindered as the speaker struggles to create appropriate language forms. Pronunciation may continue to be strongly influenced by first language* and fluency may still be strained." (ACTFL, 1986, emphases added).

It is clear that L2A, like L1A, is characterised by the acquisition of collocations and chunks of language which are slowly analysed on a word-by-word basis to allow the

determination of grammatical word class and c-selection. L2A additionally demonstrates significant transfer from L1, as is predicted by constructivist accounts which emphasise sequential analysis (Ellis, 1996) or the competition of multiple cues (MacWhinney, in press).

Conclusions

Fluent language users have had tens of thousands of hours on task. They have processed many millions of utterances involving tens of thousands of types presented as innumerable tokens. The evidence of language has ground on their perceptuo-motor and cognitive apparatus to result in complex competencies which can be described by formal theories of linguistics such as UG. It is more than a 'simplifying assumption' that language learning 'can be conceptualized as an instantaneous process' (Chomsky, 1976, pp. 14-15). It is an error which compounds into the fallacy of predeterminism.

Language is like the majority of complex systems which exist in nature and which empirically exhibit hierarchical structure (Simon, 1962). And like these other systems, its complexity emerges from simple developmental processes being exposed to a massive and complex environment. We are enlightened when we substitute a process description for a state description - when we describe development rather than the final state.

Meteorology has its rules and principles of the phenomena of the atmosphere which allow the prediction of weather. Geology has its rules and principles to describe and summarise 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 which smoothes the irregularities 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[1]. 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 nativelike competence. There is a more relevant Universal which 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. Language acquisition researchers would do well to mirror language learners in investigating the conditional probabilities of words in sequence.

[1] This should not be taken to deny any role of pedagogical rules in language learning. Unlike rocks or clouds, humans are reactive to verbal instruction. Thus some parts of their 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* (Ellis, 1994; 1995).

References

ACTFL. (1986). *The ACTFL proficiency guidelines.* Yonkers, NY: The American Council on the Teaching of Foreign Languages (ACTFL).

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.

Bates, E., Thal, D., & Marchman, V. (1991). Symbols and syntax: A Darwinian approach to language development. In N. A. Krasnegor, D. M. Rumbaugh, R. L. Schiefelbusch, & M. Studdert-Kennedy (Eds.), *Biological and behavioral determinants of language development* (pp. 29-66). Hillsdale, NJ: Lawrence Erlbaum Associates.

Becker, A.L. (1983). Toward a post-structuralist view of language learning: A short essay. *Language Learning, 33*, 217-220.

Bley-Vroman, R., & Chaudron, C. (1994). Elicited imitation as a measure of second-language competence. In E. Tarone, S.M. Gass, & A.D. Cohen (Eds.), *Research methodology in second-language acquisition.* Hillsdale, N.J. : L. Erlbaum.

Butler, C. (1995). Between lexis and grammar: Repeated word sequences and collocational frameworks in Spanish. Paper presented to the 5th Dyffryn Conference on Vocabulary and Lexis, Cardiff, March 31-April 2, 1995.

Charniak, E. (1993). *Statistical language learning*. Cambridge, MA: MIT Press.

Chomsky, N. (1965). *Aspects of a theory of syntax*. Cambridge, Mass.: MIT Press.

Chomsky, N. (1976). *Reflections on language.* London: Temple Smith.

Chomsky, N. (1989). Some notes on economy of derivation and representation. *MIT Working Papers in Linguistics*, *10*, 43-74.

Chomsky, N. (1992). A minimalist program for linguistic theory. *MIT Occasional Papers in Linguistics*, *1*. (Published (1993), 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.)

Chomsky, N. (1995). Bare phrase structure. In G. Webelhuth (Ed.), *Government and binding theory and the minimalist program* (pp. 383-420). Oxford: Blackwell.

Collins Cobuild. (1996). *Grammar patterns 1: Verbs*. London: HarperCollins.

Corder, S.P. (1973). *Introducing applied linguistics*. Harmondsworth: Penguin.

Darwin, C. (1859). *The origin of the species*. Reprinted. London: Penguin.

Dawkins, R. (1976). Hierarchical organisation: A candidate principle for ethology. In P.P.G. Bateson & R.A. Hinde (Eds.), *Growing points in ethology* (pp. 7-54). Cambridge: Cambridge University Press.

Dawkins, R. (1986). *The blind watchmaker*. Harlow: Longman.

Ellis, N.C. (1994). Implicit and explicit processes in language acquisition: An introduction. In N. Ellis (Ed.) *Implicit and explicit learning of languages* (pp. 1-32). London: Academic Press.

Ellis, N.C. (1995). Consciousness in Second Language Acquisition: A Review of Field Studies and Laboratory Experiments. *Language Awareness,* **4***, 123-146.*

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., & Schmidt, R. (in press, 1997). Morphology and Longer-Distance Dependencies: Laboratory Research Illuminating the **A** in SL**A** *Studies in Second Language Acquisition*. **19:2**

Ellis, N. C., & Sinclair, S. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *Quarterly Journal of Experimental Psychology, 49A*, 234-250.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

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.

Hakuta, K. (1974). Prefabricated patterns and the emergence of structure in second language acquisition. *Language Learning*, *24*, 287-298.

Harvey, W. (1653). *Anatomical exercises*. Translated by Richard Lowdnes, London.

Higgs, T.V. (1984). (Ed.) *Teaching for proficiency, the organizing principle*. Lincolnwood, Ill: National Textbook Company.

Honkela, T., Pulkki, V., & Kohonen, T. (1995). Contextual relations of words in Grimm Tales, Analyzed by self-organizing maps. In F. Fogelman-Soulie & P. Gallinari (Eds.), *Proceedings of International Conference on Artificial Neural Networks: ICANN-95.* EC2 et Cie, Paris, pp. 3-7.

Kiss, G.R. (1973). Grammatical word classes: A learning process and its simulation. In G.H. Bower (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory.* Vol. *7*. New York: Academic Press.

LaBerge, D., & Samuels, S.J. (1974). Towards a theory of automatic information processing in reading. *Cognitive Psychology*, *6*, 292-323.

Lado, R. (1965). Memory span as a factor in second language learning. *IRAL, III*, 123-129.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind.* Chicago: University of Chicago Press.

Langacker, R. (1987). *Foundations of cognitive grammar.* Stanford, CA: Stanford University Press.

Leakey, R.E., & Lewin, R. (1977). *Origins: What new discoveries reveal about the emergence of our species and its possible future*. London: Futura.

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

Lieven, E.V.M., Pine, J.M., & Dresner Barnes, H. (1992). Individual differences in early vocabulary development: Redefining the referential-expressive dimension. *Journal of Child Language*, *19*, 287-310.

Locke, J. (1690). *An essay concerning human understanding.* Edition of P.H. Nidditch, Oxford, 1975.

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

MacWhinney, B. (in press). Second language acquisition and the competition model. In A. M. B. de Groot and J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives*. Hillsdale, N.J.: Lawrence Erlbaum.

MacWhinney, B., & Bates, E. (1989). (Eds.) *The crosslinguistic study of sentence processing*. New York: Cambridge University Press.

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

Marantz, A. (1995). The minimalist program. In G. Webelhuth (Ed.), *Government and binding theory and the minimalist program* (pp. 349-382). Oxford: Blackwell.

McClelland, J.L., Rumelhart, D.E., & the PDP Research Group (Eds.) (1986). *Parallel distributed processing: Explorations in the microstructure of cognition, Volume 2: Psychological and biological models*. Cambridge, MA: MIT Press.

McLaughlin, B (1987). *Theories of second language acquisition.* London: Edward Arnold.

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81-97.

Nattinger, J.R., & DeCarrico, J.S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.

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

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA.: Harvard University Press.

Paley, W. (1828). *Natural theology*. 2nd edn. Oxford: J. Vincent.

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.* London: Longman.

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tense of English verbs. In J. L. McClelland, D. E. Rumelhart and the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition, Volume 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*. Hillsdale, NJ: Erlbaum.

Sampson, G.R. (1980). *Making sense*. Oxford: Oxford University Press.

Sampson, G. (1987). Probabilistic models of analysis. In R. Garside, G. Leech & G. Sampson (Eds.), *The computational analysis of english*. Harlow, Essex: Longman.

Simon, H.A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society, 106*, 467-482. Reprinted in Simon, H. A. (1969), *The sciences of the artificial*. Cambridge, Mass.: MIT Press.

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

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Singleton, D. (1996). Overview of the lexical processing symposium. Paper presented at the 11th World Congress of Applied Linguistics, Jyväskylä, Finland, 4-9 August, 1996.

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.

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: Lawrence Erlbaum Associates.

Tomasello, M. (1992). *First verbs: A case study of early grammatical development.* Cambridge: Cambridge University Press.

Tomasello, M. (1995). Language is not an instinct. *Cognitive Development, 10*, 131-156.

Wittgenstein, L. (1969). *Preliminary studies for the philosophical investigations generally known as the blue and brown books by Ludwig Wittgenstein*. Oxford: Blackwell.