Language Learning 48:4, December 1998, pp. 631–664

Emergentism, Connectionism and Language Learning

Nick C. Ellis
University of Wales

This review summarizes a range of theoretical approaches to language acquisition. It argues that language representations emerge from interactions at all levels from brain to society. Simple learning mechanisms, operating in and across the human systems for perception, motor-action and cognition as they are exposed to language data as part of a social environment, suffice to drive the emergence of complex language representations. Connectionism provides a set of computational tools for exploring the conditions under which emergent properties arise. I present various simulations of emergence of linguistic regularity for illustration.

> If it be asked: What is it you claim to be emergent?—the brief reply is Some new kind of relation. Consider the atom, the molecule, the thing (e.g., a crystal), the organism, the person. At each ascending step there is a new entity in virtue of some new kind of relation, or set of relations, within it . . . It may still be asked in what distinctive sense the relations are new. The reply is that their specific nature could not be predicted before they appear in the evidence, or prior to their occurrence. (Lloyd Morgan, 1925, pp. 64–65)

Correspondence concerning this article may be addressed to the author at School of Psychology, University of Wales, Bangor, Gwynedd, Wales, LL57 2DG, United Kingdom: Internet: N.Ellis@bangor.ac.uk

The -*ists* and -*istics* of Language Learning Research[1]

Generative linguistics analyzes the relations in language. It has been the dominant linguistic paradigm for studying language during most of the period celebrated by this Jubilee issue. A generative grammar of a language is the set of rules that defines the unlimited number of sentences of the language, and associates each with an appropriate grammatical description. Such descriptions come from formal linguistic models of often elegantly abstract mathematical structure. The Government/Binding (GB) theory of syntax (Chomsky, 1981) describes knowledge of language as consisting of universal language principles (e.g., structure-dependency, binding, subjacency, etc.) unbroken in any language, and of parameters (e.g., the head parameter, the pro-drop parameter, the opacity parameter, etc.) with a limited set of values. The theory posits that principles and parameters are part of the structure of mind; accordingly, GB research tries to determine the essential patterns and relations in language, in languages, in all languages.[2]

Following Chomsky (1965, 1981, 1986), generative research has been guided by the following assumptions:

1. *Modularity*: language is a separate faculty of mind;

2. *Grammar as a System of Symbol-Manipulating Rules*: knowledge about language is a grammar, a complex set of rules and constraints that allows people to distinguish grammatical from ungrammatical sentences;

3. *Competence*: research should investigate grammatical competence as an idealized hygienic abstraction rather than language use which is sullied by performance factors;

4. *Poverty of the Stimulus*: because learners converge on the same grammar in broadly similar patterns of acquisition even though language input is degenerate, variable and lacking in reliable negative evidence, learnability arguments suggest that there must be strong constraints on the possible forms of

grammars, the determination of which constitutes the enterprise of Universal Grammar (UG);

5. *Language Instinct*: the essential constraints of UG are innately represented in the brain, language is an instinct, linguistic universals are inherited, the language faculty is modular by design;

6. *Acquisition as Parameter Setting*: language acquisition is therefore the acquisition of the lexical items of a particular language and the appropriate setting of parameters for that language.

The Jubilee of *Language Learning* falls at an exciting time in the history of language research when doubt is being cast on these UG assumptions (Bates, Thal, & Marchman, 1991; Elman, et al., 1996; MacWhinney, 1998; Quartz & Sejnowski, in press; Seidenberg, 1997; Tomasello, 1995). Note that it is the assumptions of UG that are under attack, not the generative grammar descriptions of the relations between the linguistic units. These hold: To the extent to which they are systematic observations, they are as valid as the universals derived from typological classification research concerning the analysis of congruent facts gleaned from wide cross-linguistic samples and subsequently categorised according to absolute universals, universal tendencies and implicational universals (Greenberg, 1963). When language is dissected in isolation, there appear many complex and fascinating structural systematicities. A complete theory of language learning must necessarily include these in all their rich sophistication. But critics are concerned that GB has raised the systematicities of syntax from explanandum to explanans. Not the extensive descriptions of linguistic relations are being questioned, but the origins and implications of these relations.

Many of the criticisms address generative linguistics' taking the uniquely human faculty of language and then studying it in isolation, divorced from semantics, the functions of language, and the other social, biological, experiential and cognitive aspects of humankind. This autism has two consequences. First, it concentrates the study of language on grammar, ignoring such areas as lexis, fluency, idiomaticity, pragmatics and discourse. Second, it dramatically restricts the potential findings of the study of grammar: If the investigation never looks outside of language, it can never identify any external influences on language.

Cognitive linguistics (Ungerer & Schmid, 1996) counters that, in order to understand language, we must understand its grounding in our experience and our embodiment, which represents the world in a very particular way. The meaning of the words of a given language, and how they can be used in combination, depend on the perception and categorization of the real world. Because people constantly observe and play an active role in this world, they know a great deal about the entities of which it consists, and this experience and familiarity is reflected in the nature of language. People have expectations of the world that are represented as complex packets of related information (schemata, scripts, or *frames*), varying in complexity and generality from, for example, "how the parts of a chair inter-relate," up to "trips to the dentist" or "buying and selling." Ultimately, everything humans know is organized and related to other knowledge in some meaningful way or other, and everything they perceive is affected by their perceptual apparatus and perceptual history. Language reflects this embodiment and this experience. The different degrees of salience or prominence of elements involved in situations people wish to describe affect the selection of subject, object, adverbials and other clause arrangement. Figure/ground segregation, which originated from Gestalt psychological analyses of visual perception, and perspective-taking, again very much in the domains of vision and attention, are mirrored in language and have systematic relations with syntactic structure. In production, what people express reflects which parts of an event attract attention; depending on how people direct their attention, they can select and highlight different aspects of the frame, thus arriving at different linguistic expressions. In comprehension, abstract linguistic constructions (like simple transitives, locatives, datives, resultatives and passives) serve as a "zoom lens" for the listener, guiding attention to a particular perspective on a scene while backgrounding other aspects (Fisher, Gleitmann & Gleitmann, 1991; Goldberg, 1995; Tomasello & Brooks, in press).

All of these concerns—the experiential grounding of language, humans' embodiment that represents the world in a very particular way, the relations between perceptual and imagery representations and the language used to describe them, perspective and attentional focus—are central to Cognitive Linguistic analyses of language learning (Goldberg, 1995; Lakoff, 1987; Lakoff & Johnson, 1980; Langacker, 1987, 1991; Talmy, 1988).

Corpus linguistics (McEnery & Wilson, 1996) argues that the proper object of study is language as it is used, and that when looking at this evidence it becomes clear (a) that it is impossible to describe syntax and lexis independently; and (b) that meaning and usage have a profound and systematic effect on each other, or, in other words, that syntax is inextricable from semantics and function (Gross, 1975). Thus, for example, the Collins Cobuild (1996) analysis of the verbs in the 250 million words of the British National Corpus 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: (a) the *increase* and *decrease* group [*climb, decline, decrease,* etc.], (b) the *win* and *lose* group [*lead, lose, win*], (c) the *overrun* group [*overrun, overspend*]). Any Comp pattern is describable *only* in terms of its lexis. Such patterns went largely unnoticed until analysis was done on a corpus of language data large enough for their reliability to be apparent and with techniques rigorous enough to capture them.

The same enterprise has also made it clear that language use is far more idiomatic and less open-class than previously believed. 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 analyzable 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 idiom 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 text is interpretable by this principle.

Psycholinguistics demonstrates that language skill intimately reflects prior language use, in that it is tuned to lifespan practice effects and the learner's relative frequencies of experience of language and the world. For example, lexical recognition processes (both for speech perception and reading) and lexical production processes (articulation and writing) are independently governed by the *power law of practice* whereby performance, typically speed, improves with practice according to a power law relationship in which the amount of improvement decreases as a function of increasing practice or frequency (Kirsner, 1994). Anderson (1982) showed that this function applies to a wide range of skills including cigar rolling, syllogistic reasoning, book writing, industrial production, reading inverted text and lexical decision. The power law seems to be ubiquitous throughout language: it is certainly not restricted to lexis. Larsen-Freeman (1976) was the first to propose that the common acquisition order of English morphemes to which learners of English as a second language (ESL) adhere, despite their different ages and language backgrounds, is a function of the frequency of ocurrence of these morphemes in adult native-speaker (NS) speech. More recently, Ellis and Schmidt (1997) and DeKeyser (1997) have shown that the power law applies to the acquisition of morphosyntax and that it is this acquisition function which underlies interactions of regularity and frequency in this domain. The Competition Model (MacWhinney, 1987, 1997) sees all of language acquisition as the process of acquiring from language input the particular cues which relate phonological forms and conceptual meanings or communicative intentions, and the determination of the frequencies, reliabilities and validities of these cues. This information then serves as the knowledge base for sentence production and comprehension in lexicalist constraint-satisfaction theories, which hold that sentence processing is the simultaneous satisfaction of the multiple probabilistic constraints afforded by the cues present in each particular sentence (MacDonald, Pearlmutter & Seidenberg, 1994; MacWhinney & Bates, 1989).

Psycholinguistic analyses thus suggest that language is cut of the same cloth as other cognitive processes, and that language acquisition, like other skills, can be understood in terms of models of optimal (Bayesian) inference in the presence of uncertainty.

Sociolinguistics (Preston, 1989; Tarone, 1988, 1997) emphasizes that language learners are social beings who acquire language in social contexts. Thus, no account of their cognitive development can be complete without a description of the learners' sociolinguistic environmental history. In particular, sociolinguists are concerned with the way that social factors affect learners' language input and their interpretation and assimilation of this input. In first language (L1) research these issues initially arose with interest in motherese or caregiver speech, in second language (L2) with native speaker/nonnative speaker (NS/NNS) interactions and with instructor/learner interactions in the classroom. The nature of these interactions affects a cascade of factors, starting with the complexity of the language input and the clarity of its reference and comprehensibility and running thence to the determination of whether there is provision of interactional modification: repetition or clarification, or negative evidence, focus on form, or recasts if the learner's utterances are in error (Long, 1983;

Lyster & Ranta, 1997; Tarone, 1997; Tarone & Swain, 1995). The social environment may tune the learners' input to something far more optimally scaffolding than the malevolent tutor of Gold's learnability analysis (Ellison, 1997; Gold, 1967).

Connectionists are concerned that although language behavior can be described as being rule-like, this does not imply that language behavior is rule-governed. Instead, they investigate how simple learning mechanisms in artificial neural networks are able to acquire the associations between, for example, forms and meanings, along with their respective reliabilities and validities, and then use these associations to produce novel responses by "on-line" generalization. Connectionist models demonstrate how subsymbolic associative systems, where there are neither given nor identifiable rules, nevertheless simulate rule-like grammatical behavior (Levy, Bairaktaris, Bullinaria & Cairns, 1995; McClelland, Rumelhart, & PDP Group, 1986; Miikkulainen, 1993).

Neurobiologists and *Emergentists* are concerned that the innateness assumption of the language instinct hypothesis lacks any plausible process explanation (Elman et al., 1996; Quartz & Sejnowski, in press). 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. In the Emergentist perspective (Elman et al., 1996; MacWhinney, 1998), interactions occurring at all levels, from genes to environment, give rise to emergent forms and behavior. These outcomes may be highly constrained and universal, but they are not themselves directly contained in the genes in any domain-specific way. Information theory analyses suggest that humans are more than 20 orders of magnitude short of being mosaic organisms, where development is completely prespecified in the genes. Instead, human growth is under regulatory control, where precise pathways to adulthood reflect numerous interactions at the cellular level occurring throughout development. The human cortex is plastic; its architecture reflects experience to a remarkable degree. Heterotopic transplant studies have shown that, for example, an area of visual cortex, when transplanted into the somatosensory area that normally forms whisker barrels (local cortical areas, each specialized to sensing input from a particular whisker on a rodent's cheek) will, given normal thalamic afferents, develop apparently normal barrel fields (Schlaggar & O'Leary, 1991). Auditory cortex, in situ, can come to see. Rewiring the thalamic inputs to what would normally become auditory cortex can cause these areas to specialize for vision (Sur, Pallas & Roe, 1990). The form of representational map is not an intrinsic property of the cortex. A cortical area can come to support different types of maps depending on its early experience. Given this plasticity and enslavement to the periphery, it is hard to see how genetic information might prescribe rigid representation of UG in the developing cortex. This is not to deny Fodorian modular faculties in adulthood (Fodor, 1983), or areas of cortex specialized in function. However, (a) the attainment of modularity and (but not necessarily in 1:1 correspondence) cortical specialization may be more the result of learning and the development of automaticity than the cause (Elman et al., 1996); (b) studies of neural imaging of the human brain are resulting in a proliferation of language areas, including Broca's area and Wernicke's area, on the right side as well as the left, in parts of the cerebellum, in a number of subcortical structures, and in high frontal and parietal areas (Damasio & Damasio, 1992; Posner & Raichle, 1994)); and (c) none of these regions is uniquely active for language but is involved in other forms of processing as well; all collaborate in language processing. Ask someone to do a language task as simple as choosing the action verb that goes with a noun (hammer–hit) and one can observe more than one of these areas "light up" (Peterson, Fox, Posner, Mintun & Raichle, 1988). Although individual subtraction studies might show partially non-overlapping patterns of activation for linguistically interesting contrasts (e.g., nouns vs. verbs, content vs. function words, syntactic vs. semantic violations), the same is true for contrasts that do not relate to linguistic theory (e.g., long vs. short words, high vs. low frequency words): "Not every difference is a difference in kind" (Bates, in press).

Innate specification of synaptic connectivity in the cortex is unlikely. On these grounds, linguistic representational nativism seems untenable. Theories of language must reflect this; they must be biologically, developmentally and ecologically plausible.

Neural constructivists (Quartz & Sejnowski, in press) refute the arguments of learnability and the poverty of the stimulus. Learnability analyses must make some assumptions about the nature of the learning model. Classic learnability theory is premised on the assumption that a system's learning properties can be deduced from a particular model of selective induction running on a fixed computational architecture (Gold, 1967). But if the developmental mechanism is a dynamic interaction between the informational structure of the environment and neuronal growth mechanisms, the representational properties of the cortex are constructed by the nature of the problem domain confronting it. This results in a uniquely powerful and general learning strategy. It minimizes the need for prespecification yet allows for the hypothesis space to be constructed as it learns; the inductive bias required to make search tractable need not be prespecified in the genes, it can result from incremental exposure to the problem space. The representations are optimally tuned to the problem during learning, and the neural ability to slowly add representational capacity as performance on the problem demands undermines the fixed-capacity assumptions of classic learnability theory. The question of learnability becomes relaxed from that of what is learnable from some particular representational class to that of what is learnable from any representational class. Baum (1988, 1989) demonstrated that networks with the power to add structure as a function of learning are complete representations, capable of learning in polynomial time any learning problem that can be solved in polynomial time by any algorithm.

Functional linguistics (Bates & MacWhinney, 1981; MacWhinney & Bates, 1989), *Emergentistism* (Elman et al., 1996; MacWhinney, 1998) and *Constructivist* child language researchers (Tomasello 1992; Tomasello & Brooks, in press) believe that language study's consideration of ontogenetic acquisition processes favors 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. 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 language competencies.

Tomasello (1992) used Wittgenstein to illustrate the epigenesis of language: "Language games are the forms of language with which a child begins to make use of words . . . 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, 1969, p. 17). But one doesn't see the emergence of the new forms without very detailed longitudinal records of individual child language development; hence, researchers in these -isms have strongly supported the CHILDES project, which allows cataloguing and analysis of such data and its exchange among empirical researchers (MacWhinney, 1995). Tomasello's own (1992) study involved 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 saw that the new forms that Travis daily added to her language were not simply resultant, but instead were genuinely emergent:

> It is not until the child has produced or comprehended a number of sentences with a particular verb [cut] 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, pp. 273–274; see also Tomasello & Brooks, in press)

In these views, language is learned, and syntax is an emergent phenomenon, not a condition of development.

Interactions and Emergence

The times are exciting and the disputes are impassioned. Theories of language acquisition are changing, and the wheel's still in spin. One very clear trend, however, is that towards interdisciplinarity. A complete understanding of language is not going to come from one discipline alone. As Cook and Seidlhofer usefully summarize, language can be viewed as:

> a genetic inheritance, a mathematical system, a social fact, the expression of individual identity, the expression of cultural identity, the outcome of a dialogic interaction, a social semiotic, the intuitions of native speakers, the sum of attested data, a collection of memorized chunks, a rule-governed discrete combinatory system, or electrical activation in a distributed network . . . We do not have to choose. Language can be all of these things at once. (Cook & Seidlhofer, 1995, p. 4)[3]

Emergentists and chaos/complexity scientists (Larsen-Freeman, 1997) recognize that, because language is all of these things at once, language at any one of these levels is in fact a result of interactions between language at all of these levels: The sum is a dynamic, complex, non-linear system where the timing of events can have dramatic influence on the developmental course and outcomes (Elman et al., 1996; MacWhinney, in press).

For example, the complexity of a solution emerges from the interaction of problem and solver. Apparent complexity may come more from the problem than from the system that learns to solve it. H.A. 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 problem of language. 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 solver.

Again, rule-like reality can emerge from apparently unregulated behavior. Perhaps, in your mind's eyes, the ant's path seems rather too haphazard to constitute a compelling example. If so, consider a case of emergent systematicity: the growth of queues at traffic lights on a multi-lane highway. The longer the lights have been red, the longer the queues. The greater the volume of traffic, the longer the queues. So far, so obvious. But, more interesting, typically the lengths of the queues in the various lanes are roughly equal. There is no prescription to this effect in the Highway Code. Instead, the "rule" that equalizes the number of cars in the carriageways emerges from satisfying the constraints of the more basic goals and behaviors of drivers, traffic planners, cars and time.

MacWhinney (1998, in press) has pointed out that nature is replete with examples of this type of emergence—the form of beaches and mountain ridges, the geometry of snowflakes and crystals, the movement of the Gulf Stream; biology is too—the hexagonal shape of cells in a honeycomb, the pattern of stripes on a tiger, the structuring of catalytic enzymes, fingerprints. Science investigates all of these phenomena and tries to form useful descriptions. 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 summarize the successive changes in the earth's crust. But these play no causal role in shifting even a grain of sand or a molecule of water; the interaction of water and rocks smoothes the irregularities and grinds the pebbles and sand (Ellis, 1996b). For emergentists, the rules of UG have a similar status—the regularities of generative grammar provide well-researched patterns in need of explanations.

Language is like the majority of complex systems which exist in nature and which empirically exhibit hierarchical structure (H.A. Simon, 1962). And as with these other systems, emergentists believe that the complexity of language emerges from relatively simple developmental processes being exposed to a massive and complex environment. Thus emergentists substitute a process description for a state description, study development rather than the final state, and focus on the language acquisition process (LAP) rather than language acquisition device (LAD).

> 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)

Emergentists believe that the universals of language have emerged, just as the universals of human transport solutions have emerged. As the other Simon says: "Cars are cars all over the world" (P. Simon, 1983). Yet, their universal properties have not come from some grand preordained design; rather they have arisen from the constraints imposed by human transport goals, society, physics, ergonomics and the availability of natural resources.

Humans have evolved systems for perceiving and representing different sources of information—vision, space, audition, touch, motor-action, emotion, and so forth. Simple learning mechanisms, operating in and across these systems as they are exposed to language data as part of a communicatively-rich human social environment by an organism eager to exploit the functionality of language, suffice to drive the emergence of complex language representations.

Connectionist Explorations of Emergence

However plausible these claims, there's clearly too little explanation of the processes involved. Perhaps this is not surprising: When one doesn't properly understand any of the individual domains of the preceding paragraph, how can one hope to understand the emergent products of their interactions? Furthermore, the interactions are going to be so complicated that, as Lloyd Morgan (1925) said, their specific nature cannot be predicted before they appear in the evidence. For these reasons, emergentists look to connectionism because it provides a set of computational tools for exploring the conditions under which emergent properties arise.

Connectionism's advantages for this purpose include: neural inspiration; distributed representation and control; data-driven processing with prototypical representations emerging rather than being innately pre-specified; graceful degradation; emphasis on acquisition rather than static description; slow, incremental, non-linear, content- and structure-sensitive learning; blurring of the representation/learning distinction; graded, distributed and non-static representations; generalization and transfer as natural products of learning; and, since the models must actually run, less scope for hand-waving (see Churchland & Sejnowski, 1992; Elman et al., 1996; McClelland et al., 1986; Plunkett & Elman, 1997).

Connectionist approaches to language acquisition investigate the representations that can result when simple learning mechanisms are exposed to complex language evidence. Lloyd Morgan's canon (1925: In no case may we interpret an action as the outcome of a higher psychical faculty if it can be interpreted as the outcome of one which stands lower in the psychological scale) is influential in connectionists' 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 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 their hypotheses about the emergence of representation by evaluating the effectiveness of their implementations as computer models consisting of many artificial neurons connected in parallel. Each neuron has an associated activation value, often between 0 and 1, roughly analogous to the firing rate of a real neuron. Psychologically meaningful objects can then be represented as patterns of this activity across the set of artificial neurons. For example, in a model of vocabulary acquisition, one subpopulation of the units in the network might be used to represent picture detectors and another to set the corresponding word forms. The units in the artificial network are typically multiply interconnected by connections with variable strengths or weights. These connections permit the level of activity in any one unit to influence the level of activity in all the units to which it is connected. The connection strengths are then adjusted by a suitable learning algorithm, in such a way that when a particular pattern of activation appears across one population it can lead to a desired pattern of activity arising in another set of units. If the learning algorithm has set the connection strengths appropriately, then units representing the detection of particular pictures cause the units that represent the appropriate lexical labels for that stimulus to become activated. Thus, the network could be said to have learned the appropriate verbal output for that picture stimulus.

There are various standard architectures of model, each suited to particular types of classification. The most common has three layers: the input layer of units, the output layer, and an intervening layer of hidden units (so-called because they are hidden from direct contact with the input or the output). The presence of these hidden units enables more difficult input and output mappings to be learned than would be possible if the input units were directly connected to the output units (Elman et al., 1996; Rumelhart & McClelland, 1986). The most common learning algorithm is back propagation, in which, on each learning trial, the network compares its output with the target output, and propagates any difference or error back to the hidden unit weights, and in turn to the input weights, in a way that reduces the error.

There are now many separate connectionist simulations of the acquisition of morphology, phonological rules, novel word repetition, prosody, semantic structure, syntactic structure, etc. (e.g., Levy, et al., 1995; MacWhinney & Leinbach, 1991; Rumelhart & McClelland, 1986). Yet these simple "test-tube" demonstrations 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 past 10 years of connectionist research has produced enough substantive demonstrations of emergent language representations to qualify emergentism as something more than the mere rhetoric of a rallying-call. In the remainder of this paper I will sketch out where connectionism is looking for processes of emergence of linguistic phenomena (see Ellis, in press, for more detail).

Connectionist theories are data-rich and process-light: massively parallel systems of artificial neurons use simple learning processes to statistically abstract information from masses of input data. What evidence is there in the input stream from which simple learning mechanisms might abstract generalizations? The Saussurean linguistic sign as a set of mappings between phonological forms and conceptual meanings or communicative intentions gives a starting point. Learning to understand a language involves parsing the speech stream into chunks which reliably mark meaning. The learner doesn't care about theoretical analyses of language. From a functional perspective, the role of language is to communicate meanings, and the learner wants to acquire the label-meaning relations. This task is made more tractable by the patterns of language. Learners' attention to the evidence to which they are exposed soon demonstrates that there are recurring chunks of language. Thus, in the first instance, important aspects of language learning must concern the learning of phonological forms and the analysis of phonological sequences: the categorical units of speech perception, their particular sequences in particular words and their general sequential probabilities in the language, particular sequences of words in stock phrases and collocations and the general sequential probabilities of words in the language (Ellis, 1996a, 1996b, 1997, in press). In this view, phonology, lexis and syntax develop hierarchically by repeated cycles of differentiation and integration of chunks of sequences. This process seems likely because 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 (Dawkins, 1976; H.A. Simon, 1962).

Phonological Sequences

Elman (1990) used a simple recurrent network to investigate the temporal properties of phonologically sequential inputs of language. In simple recurrent networks, the input to the network is the current letter in a language stream, and the output represents the network's best guess as to the next letter. The difference between the predicted state and the correct subsequent state (the target output) is used by the learning algorithm to adjust the weights in the network at every time step. In this way the network improves its accuracy with experience. A context layer is a special subset of inputs that receive no external input but which feed the result of the previous processing back into the internal representations. Thus, at Time 2 the hidden layer processes both the input of Time 2 and, from the context layer, the results of processing at Time 1. And so on, recursively. Thus, simple recurrent networks capture the sequential nature of temporal inputs. Such networks are not given any explicit information about the structure of language.

Elman (1990) fed the network one letter (or phoneme) at a time; it had to predict the next letter in the sequence. It was trained on

200 sentences varying in length from 4 to 9 words. There was no word or sentence boundary information; thus, part of the stream was: *Manyyearsagoaboyandgirllivedbytheseatheyplayedhappily* . . .

The error patterns for a network trained on this task demonstrate that it can abstract a lot of information about the structure of English. If the error is high, it means that the network has trouble predicting this letter. Errors tend to be high at the beginning of a word and decrease until the word boundary is reached, thus demonstrating that the model has extracted orthographic sequential probabilities. Before it is exposed to the first letter in a word, the network is unsure what is to follow. But the identity of the first two phonemes is usually sufficient to enable the network to predict with a high degree of confidence the subsequent phonemes in the word. In Elman's (1990) experiment, the time course of this process was as predicted by cohort models of word recognition (e.g., Marslen-Wilson, 1993). Once the input string reached the end of the word, the network could not be sure which word was to follow, so the error increased. The resultant saw-tooth shaped error function, with the teeth appearing after unit boundaries, demonstrated that the model had learned the common recurring units (the morphemes and words) and some word sequence information too (Elman, 1990).

At times, when the network could not predict the actual next phoneme, it nonetheless predicted the correct category of phoneme: vowel/consonant, and so on (Elman, 1990; see also Elman & Zipser, 1988, who trained networks on a large corpus of unsegmented continuous raw speech without labels). Thus, the network moved from processing mere surface regularities to representing something more abstract, but without this being built in as a pre-specified constraint; linguistically useful generalizations emerged. Simple sequence learning processes learned regular chunks like words, bound morphemes, collocations and idioms; they learned regularities of transition between these surface chunks; and they acquired abstract generalizations from the patterns in these data (Elman, 1990).

Such chunks are potential labels, but what about reference? The more any word or formula, be it L1 or L2, is repeated in phonological working memory, the more its regularities and chunks are abstracted, and the more accurately and readily these can be called to working memory, either for accurate pronunciation as articulatory output or as labels for association with other representations. From these potential associations with other representations other interesting properties of language emerge.

Lexical Syntactic Information

Learning the grammatical word-class of a particular word, and learning grammatical structures more generally, involves 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. Elman (1990) trained a recurrent network on sequences of words following a simple grammar, the network having to learn to predict the next word in the sequence. At the end of training, Elman cluster-analyzed the representations that the model had formed across its hidden unit activations for each word+context vector. This showed that the network had discovered several major categories of words—large categories of verbs and nouns, smaller categories of inanimates or animates nouns, smaller still categories of human and nonhuman animals, and so on (e.g., "dragon" occurred as a pattern in activation space in the region corresponding to the category "animals," and also in the larger region shared by animates, and finally in the area reserved for nouns). The category structure was hierarchical, soft and implicit.

The network moved from processing mere surface regularities to representing something more abstract, but without this being built in as a pre-specified syntactic or other linguistic constraint and without provision of semantics or real world grounding. Relatively general architectural constraints gave rise to language-specific representational constraints as a product of processing the input strings. These linguistically-relevant representations are an *emergent* property of the network's functioning (see Finch & Chater, 1994; and Redington & Chater, 1998, for larger analyses of this type on corpora of natural language). Learning the grammatical categories and requirements of words and word groups reduces to the analysis of the sequence in which words work in chunks.

Lexical Semantics

Landauer and Dumais (1997) presented a Latent Semantic Analysis (LSA) model which simulated L1 and L2 learners' acquisition of vocabulary from text. The model simply treated words as being alike if they tended to co-occur with the same neighboring words in text passages. By inducing global knowledge indirectly from local co-occurrence data in a large body of representative text, LSA acquired knowledge about the full vocabulary of English at a rate comparable to that for school-children. After the model had been trained by exposing it to text samples from over 30,000 articles from Grolier's *Academic American Encyclopedia*, it achieved a score of 64% on the synonym portion of the Test of English as a Foreign Language (a level expected of a good ESL learner). The performance of LSA was surprisingly good for a model that had no prior linguistic or grammatical knowledge and that could neither see nor hear, thus being unable to use phonology, morphology or real-world perceptual knowledge. In this account, lexical semantic acquisition emerged from the analysis of word co-occurrence.

Morphosyntax

The processes of phonological sequencing (see above) generate words, fuzzy word-class clusters and letter sequences which are fairly reliable morphological markers (e.g., -*s*, -*ing*, -*ed*, etc. in English). If particular combinations of these are reliably associated with particular temporal perspectives (for tense and aspect) or number of referents (for noun plural marking) for example, then one has the information necessary for the beginnings of a system which can generate inflectional morphology. There have been a number of compelling connectionist models of the acquisition of morphology. The pioneers, Rumelhart and McClelland (1986), showed that a simple learning model reproduced, to a remarkable degree, the characteristics of young children learning the morphology of the past tense in English. The model generated the so-called U-shaped learning curve for irregular forms; it exhibited a tendency to overgeneralize, and, in the model as in children, different past-tense forms for the same word could co-exist at the same time. Yet there was no "rule"; "it is possible to imagine that the system simply stores a set of rote-associations between base and past-tense forms with novel responses generated by 'on-line' generalizations from the stored exemplars" (Rumelhart & McClelland, 1986, p. 267). This original past-tense model was very influential. It laid the foundations for the connectionist approach to language research; it generated a large number of criticisms (Lachter & Bever, 1988; Pinker & Prince, 1988), some of them undeniably valid; and, in turn, it spawned a number of revised and improved connectionist models of different aspects of the acquisition of the English past tense.

These recent models have successfully captured the regularities that are present (a) in associating phonological form of lemma with phonological form of inflected form (Daugherty & Seidenberg, 1994; MacWhinney & Leinbach, 1991; Marchman, 1993; Plunkett & Marchman, 1991), and (b) between referents (+past tense or +plural) and associated inflected perfect or plural forms (Cottrell & Plunkett, 1994; Ellis & Schmidt, 1997); in the process they have closely simulated the error patterns, profiles of acquisition, differential difficulties, false-friends effects, reaction times for production, and interactions of regularity and frequency that are found in human learners (both L1 and L2), as well as acquiring default case allowing generalization on "wug" tests. Their successes strongly support the notion that acquisition of morphology is also a result of simple associative learning principles operating in a massively distributed system abstracting the regularities of association using optimal inference. Much of the information needed for syntax falls quite naturally out of simple sequence analysis and the patterns of association between patterns of sequences and patterns of referents.

Syntactic Constructions

Links between phonological chunks and conceptual representations underlie reference and grounded semantics; patterns in these cross-modal associations underlie the emergence of syntactic constructions (Goldberg, 1995).

> In cognitive linguistics the use of syntactic structures is largely seen as a reflection of how a situation is conceptualized by the speaker, and this conceptualization is governed by the attention principle. Salient participants, especially agents, are rendered as subjects and less salient participants as objects; verbs are selected which are compatible with the choice of subject and object, and evoke the perspective on the situation that is intended; locative, temporal and many other types of relations are highlighted, or "windowed for attention" by expressing them explicitly as adverbials. Although languages may supply different linguistic strategies for the realization of the attention principle, the underlying cognitive structures and principles are probably universal. (Ungerer & Schmid, 1996, p. 280)

The Competition Model (MacWhinney, 1987, 1997) emphasizes lexical functionalism where syntactic patterns are controlled by lexical items. Recent competition model studies have simulated the language performance data by using simple connectionist models that relate lexical cues and functional interpretations for sentence comprehension or production. Consider the particular cues that relate subject-marking forms to subject-related functions. The network's input are various combinations of cues. For example, in the input sentence "The boy loves the parrots," the cues are: preverbal positioning (*boy* before *loves*), verb agreement morphology (*loves* agrees in number with *boy* rather than *parrots*), sentence initial positioning, and use of the article *the*. Another potential cue for agency, absent here, is nominative case marking, but English has no case marking for nouns, only for pronouns (e.g., *I* vs. *me*). The outputs of the network are nodes representing functional interpretations, including actor, topicality, perspective, givenness, and definiteness. In the case of this particular sentence, all five output nodes are turned on. Hidden units which intervene between input and output layers allow the learning of non-linear associations between inputs and outputs. After such models are trained with sufficient exposures of sentences relating such cue combinations with their various functional interpretations, they abstract the regularities of the ways in which a particular language (in this case English, but the impact of these cues have been studied in more than a dozen languages) expresses agency. The linguistic strategy for expression of agency is abstracted from the input.

There are many attractive features of the competition model. It developmentally models the cues, their frequency, reliability, and validity, as they are acquired from representative language input. The competition part of the model shows how Bayesian cue use can resolve in activation of a single interpretative hypothesis from a rich network of interacting associations and connections (some competing, others, as a result of the many redundancies of language and representation, mutually reinforcing). It has been extensively tested to assess the cues, cue validity and numerical cue strength order in different languages. Finally, it goes a long way in predicting language transfer effects (MacWhinney, 1992).

Connecting Emergentism and Connectionism

The Competition model has been a good start for investigating the emergence of strategies for the linguistic realization of reference. However, if the communicative use of syntactic structures and the attention principle derive from the frequency and regularity of cross-modal associations between chunks of phonological surface form and (particularly visuo-spatial) imagery representations, then ultimately any model must properly represent human vision and spatial processing. These are not fixed and static, rather they are explored, manipulated, cropped and zoomed, and run in time like movies under attentional and scripted control (Kosslyn, 1983; Talmy, 1996a). Cognitive linguistics reminds one that the prominence of particular aspects of the scene and the perspective of the internal observer (i.e., the attentional focus of the speaker and the intended attentional focus of the listener) are key elements in determining regularities of association between elements of visuo-spatial experience and elements of phonological form. One cannot understand language acquisition by understanding phonological memory alone. All of the systems of working memory, all perceptual representational systems, attentional resources and supervisory systems are involved in collating the regularities of cross-modal associations underpinning language use.

Cognitive linguistics aims to understand how the regularities of syntax emerge from the cross-modal evidence that is collated during the learner's lifetime of using and comprehending language. The difficulties of this enterprise are obvious. Acknowledging the importance of embodiment, perspective and attention means that, to understand the emergence of language, researchers must also understand the workings of attention, vision and other representational systems. Then they must understand the regularities of the mappings of these systems onto particular languages. And the mappings in question are piecemeal—the detailed content of the mappings is important, not simply the modalities concerned—which is why cognitive linguistics focuses on particular constructions and representational aspects at a time, for example, motion event frames (Langacker, 1991; Talmy, 1996b) or spatial language (Bowerman, 1996).

Cognitive linguistics still has a long way to go in analyzing the argument-structure constructions and strategies that underpin grammar (e.g., Tomasello, in press, and commentaries in same issue). Connectionist simulations of the acquisition of constructions clearly have a lot further to go. Although connectionist work to date has had considerable success within lexis and morphosyntax, as Rispoli complains:

> Yes, we are interested in those areas, but we are also interested in a great deal more, such as the development of case and agreement systems, the development of negation, WH-questions, subject drop, subject-auxiliary inversion, relative clauses, cleft sentences, constraints on coreference, the semantics of verbs, the argument structure of predicates, phonological processes, metrical structure, and distinctive phonetic features, just to name a few. (Rispoli, in press)

Fodor is more damning of the enterprise to date: “no examples are given of how, even in sketch, an attested linguistic universal might be explained in this way” (Fodor, 1997, p. 4).

There is no denying that much that remains to be done. Nevertheless, researchers are never going to understand language by studying it in isolation, in the same way that one could never properly understand the game of soccer by investigating only the patterns of movement of the ball, or chess by analysing the interactions of just the white pieces. A proper understanding of these linguistic phenomena will only come when researchers reunite speakers, syntax and semantics, the signifiers and the signifieds. Language scientists have to be linguists, psychologists, physiologists, and computational neuroscientists at the same time. Recent computational work is beginning to relate linguistic constructions to more plausible models of visual perception of movement in space (Regier, 1996). Similarly, Narayanan (1997) has shown how the semantics of verbal aspect might be grounded in sensori-motor primitives abstracted from processes that recur in sensori-motor control (such as goal, periodicity, iteration, final state, duration, force and effort). The general enterprise of the L_0 project (Bailey, Feldman, Narayanan & Lakoff, 1997; Feldman et al., 1996) is well motivated. If one wants to understand emergence of language and believes in the constraints of embodiment, then one’s models have to realistically capture the physical and psychological processes of perception, attention and memory.

Space limitations prevent discussion of connectionist investigations of emergence in other language domains, and it is more important to leave this section with a nota bene. I have illustrated several different domains of emergence. But what must be remembered is that all of these emergent entities interact as well, so leading to interesting new emergent relations themselves. In language acquisition, as in evolution, there is "more in the conclusions than is contained in the premises." Thus might simple associations amass over the learner's language-input history into a web of multimodal connections that represent the complexities of language.

Conclusions

It's claimed that one can't have a theory about the development of something without having a theory of what that "something" is.[4] True, but so is the emergentist counter that one cannot properly understand something without knowing how it came about. Emergentists believe that simple learning mechanisms, operating in and across the human systems for perception, motor-action and cognition as they are exposed to language data as part of a communicatively-rich human social environment by an organism eager to exploit the functionality of language, suffice to drive the emergence of complex language representations. However, just about every content word in that previous sentence is a research discipline in itself.

I look to the next 50 years of language learning research for the details of these processes. This research must involve the interdisciplinary collaboration of the aforementioned -*ists* and the ever new -*istics* that will emerge from their mutual influences. It needs to understand the constraints on human cognition that arise from the affordances and the proscriptions in each of these domains, and how these constraints interact. And it needs dynamic models of the acquisition of representations, their interactions and the emergence of structure, all based on representative histories of experience. A crucial aspect of the exploration of these interactions in complex, dynamic, non-linear systems, where the timing of events can have dramatic influence on the developmental course and outcomes, will be the formal modelling of these processes. Then, researchers might have a chance of solving the riddle, more complex even than that of Samson (*Judges, 14*), of how, out of the strings came forth structure.

Notes

[1]In enough caricature to clarify the schismogenesis.
[2]One consequence is the view that the mind does not know languages, but grammars: "The grammar in a person's mind/brain is real; it is one of the real things in the world. The language (whatever it may be) is not" (Chomsky, 1982, p. 5). Grammar is reified, while language is seen as an epiphenomenon.
[3]The current "Aims and Scope" of *Language Learning* expresses this as a list of contributing disciplines: "linguistics, psycholinguistics, cognitive science, ethnography, ethnomethodology, sociolinguistics, sociology, semiotics, educational inquiry, and cultural or historical studies" (See also Ellis, 1994a.)
[4]"No discipline can concern itself in a productive way with the acquisition and utilisation of a form of knowledge without being concerned with the *nature* of that system of knowledge"(Chomsky, 1977, p. 43).

References

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*, 369–406.

Bailey, D., Feldman, J., Narayanan, S., & Lakoff, G. (1997). Modelling embodied lexical development. *Proceedings of the Cognitive Science Conference 1997*. Pittsburgh, PA: Cognitive Science Society.

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

Bates, E. (in press). Construction grammar and its implications for child language research. *Journal of Child Language*.

Bates, E., & MacWhinney, B. (1981). Second language acquisition from a functionalist perspective. In H. Winitz (Ed.), Native language and foreign language acquisition, [Monograph]. *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.

Baum, E. B. (1988). Complete representations for learning from examples. In Abu-Mostafa (Ed.), *Complexity in information theory* (pp. 193–215). New York: Springer-Verlag.

Baum, E.B. (1989). A proposal for more powerful learning algorithms. *Neural Computation, 1*, 201–207.

Bowerman, M. (1996). Learning how to structure space for language: A cross-linguistic perspective. In P. Bloom, M. Peterson, L. Nadel, & M. Garrett (Eds.), *Language and space* (pp. 385–436). Cambridge, MA: MIT Press.

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

Chomsky, N. (1977). *Knowledge and responsibility*. New York: Pantheon Books.

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

Chomsky, N. (1982). *Some concepts and consequences of the theory of government and binding*. Cambridge, MA: MIT Press.

Chomsky, N. (1986). *Knowledge of language*. New York: Praeger.

Churchland, P. S., & Sejnowski, T. J. (1992). *The computational brain*. Cambridge, MA: MIT Press.

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

Cook, G., & Seidlhofer, B. (Eds.). (1995). *Principles and practice in applied linguistics*. Oxford: Oxford University Press.

Cottrell, G., & Plunkett, K. (1994). Acquiring the mapping from meaning to sounds. *Connection Science, 6*, 379–412.

Damasio, A.R., & Damasio, H. (1992). Brain and language. *Scientific American, 267* (3), 63–71.

Daugherty, K. G., & Seidenberg, M. S. (1994). Beyond rules and exceptions: A connectionist approach to inflectional morphology. In S. D. Lima, R. L. Corrigan, & G. K. Iverson (Eds.), *The reality of linguistic rules* (pp. 353–388). Amsterdam: John Benjamins.

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.

DeKeyser, R. M. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition, 19,* 195–222.

Ellis, N. C. (1994a). 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. (1994b). Vocabulary acquisition: The implicit ins and outs of explicit cognitive mediation. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 211–282). London: Academic Press.

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

Ellis, N. C. (1996b). Analyzing language sequence in the sequence of language acquisition: Some comments on Major and Ioup. *Studies in Second Language Acquisition, 18*, 361–368.

Ellis, N. C. (1997). The epigenesis of language: Acquisition as a sequence learning problem. In A. Wray & A. Ryan (Eds.), *Evolving models of language* (pp. 41–57). British Studies in Applied Linguistics, *12*. Clevedon, UK: Multilingual Matters.

Ellis, N. C. (in press). Memory for language. In P. Robinson (Ed.), *Cognition and second language instruction*. Cambridge: Cambridge University Press.

Ellis, N. C., & Schmidt, R. (1997). Morphology and longer-distance dependencies: Laboratory research illuminating the A in SLA. *Studies in Second Language Acquisition*, *19*, 145–171.

Ellison, T. M. (1997, July). *Acquisition, learnability and distributional information*. Paper presented at the 5th International Cognitive Linguistics Conference, Amsterdam.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*, 179–211.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Elman, J. L., & Zipser, D. (1988). Learning the hidden structure of speech. *Journal of the Acoustic Society of America, 83*, 1615–1626.

Feldman, J., Lakoff, G., Bailey, D., Narayanan, S., Regier, T., & Stolcke, A. (1996). L_0: The first five years of an automated language acquisition project. *Artificial Intelligence Review, 10*, 103–129.

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.

Fisher, C., Gleitmann, H., & Gleitmann, L. R. (1991). On the semantic content of subcategorization frames. *Cognitive Psychology, 23*, 331–392.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Fodor, J. A. (1997). [Review of the book *Rethinking innateness: A connectionist perspective on development*.] *Times Literary Supplement, May 16*.

Gold, E. (1967). Language identification in the limit. *Information and Control, 16*, 447– 474.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure.* Chicago: University of Chicago Press.

Greenberg, J. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. Greenberg (Ed.), *Universals of language* (pp. 73–113). Cambridge, MA: MIT Press.

Gross, M. (1975). On the relations between syntax and semantics. In E. L. Keenan (Ed.), *Formal semantics of natural language* (pp. 389–405). Cambridge: Cambridge University Press.

Kirsner, K. (1994). Implicit processes in second language learning. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 283–312). London: Academic Press.

Kosslyn, S. (1983). *Ghosts in the mind's machine: Creating and using images in the brain.* New York: Norton.

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.

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

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104,* 211–240.

Langacker, R. W. (1987). *Foundations of cognitive grammar: Vol. 1. Theoretical prerequisites.* Stanford, CA: Stanford University Press.

Langacker, R. W. (1991). *Foundations of cognitive grammar: Vol. 2. Descriptive application.* Stanford, CA: Stanford University Press.

Larsen-Freeman, D. (1976). An explanation for the morpheme acquisition order of second language learners. *Language Learning, 26,* 125–134.

Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied Linguistics, 18,* 141–165.

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

Lloyd Morgan, C. (1925). *Emergent evolution*. London: Williams and Norgate.

Long, M. (1983). Native speaker/nonnative speaker conversation and the negotiation of comprehensible input. *Applied Linguistics, 4,* 126–141.

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition, 19,* 37–66.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review, 101,* 676–703.

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

MacWhinney, B. (1992). Transfer and competition in second language learning. In R. J. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 371–390). Amsterdam: North Holland.

MacWhinney, B. (1995). *The CHILDES project: Tools for analyzing talk* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

MacWhinney, B. (1997). Second language acquisition and the competition model. In A.M.B. de Groot and J.F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 113–144). Hillsdale, NJ: Lawrence Erlbaum.

MacWhinney, B. (Ed.). (1998). *The emergence of language*. Hillsdale, NJ: Lawrence Erlbaum.

MacWhinney, B. (in press). Emergent language. In M. Darnell, E. Moravcsik, F. Newmeyer, M. Noonan, & K. Wheatley (Eds.), *Functionalism and formalism in linguistics*. New York: John Benjamins.

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

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

Marchman, V. A. (1993). Constraints on plasticity in a connectionist model of the English past tense. *Journal of Cognitive Neuroscience*, *5*, 215–234.

Marslen-Wilson, W. C. (1993). Issues of process and representation. In G. Altmann & R. Shillcock (Eds.), *Cognitive models of speech processing*. Hillsdale, NJ: Lawrence Erlbaum.

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

McEnery, T., & Wilson, A. (1996). *Corpus linguistics*. Edinburgh: Edinburgh University Press.

Miikkulainen, R. (1993). *Subsymbolic natural language processing*. Cambridge, MA: MIT Press.

Narayanan, S. (1997). Talking the talk is like walking the walk: A computational model of verb aspect. *Proceedings of the Cognitive Science Conference 1997*. Pittsburgh, PA: Cognitive Science Society.

Peterson, S. E., Fox, P. T., Posner, M. I., Mintun, M. A., & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single word processing. *Nature*, *331*, 585–589.

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., & Elman, J. L. (1997). *Exercises in rethinking innateness*. Cambridge, MA: MIT Press.

Plunkett, K., & Marchman, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition, 48,* 21–69.

Posner, M.I., & Raichle, M.E. (1994). *Images of Mind*. New York: Scientific American Library.

Preston, D. (1989). *Sociolinguistics and second language acquisition*. Oxford: Blackwell.

Quartz, S. R., & Sejnowski, T. J. (in press). The neural basis of cognitive development: A constructivist manifesto. *Behavioral and Brain Sciences*.

Redington, M., & Chater, N. (1998). Connectionist and statistical approaches to language acquisition: A distributional perspective. *Language and Cognitive Processes, 13,* 129–192.

Regier, T. (1996). *The human semantic potential: Spatial language and constrained connectionism*. Cambridge, MA: MIT Press.

Rispoli, M. (in press). [Review of the book *Rethinking innateness: A connectionist perspective on development*]. *Journal of Child Language*.

Rumelhart, D. E., & McClelland, J. (1986). On learning the past tense of English verbs. In J. L. McClelland and D. E. Rumelhart (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models* (pp. 272–326). 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: Lawrence Erlbaum.

Schlaggar, B. L., & O'Leary, D. D. (1991). Potential of visual cortex to develop an array of functional units unique to somatosensory cortex. *Science, 252*, 1556–1560.

Seidenberg, M. S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science, 275*, 1599–1603.

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, MA: MIT Press)

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

Simon, P. (1983). Cars are Cars. On *Hearts and Bones* [cassette]. WEA Records Ltd.

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

Sur, M., Pallas, S. L., & Roe, A. W. (1990). Cross-modal plasticity in cortical development: Differentiation and specification of sensory neocortex. *Trends in Neuroscience, 13,* 227–233.

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science, 12*, 49–100.

Talmy, L. (1996a). The windowing of attention in language. In M. Shibatani & S. Thompson (Eds.), *Grammatical constructions: Their form and meaning* (pp. 235–287). Oxford: Oxford University Press.

Talmy, L. (1996b). Fictive motion in language and “ception.” In P. Bloom, M. Peterson, L. Nadel, & M. Garrett (Eds.), *Language and space* (pp. 211–275). Cambridge, MA: MIT Press.

Tarone, E. (1988). *Variations in interlanguage*. London: Edward Arnold.

Tarone, E. (1997). Analyzing IL in natural settings: A sociolinguistic perspective on second-language acquisition. *Culture & Cognition, 30*, 137–150.

Tarone, E., & Swain, M. (1995). A sociolinguistic perspective on second-language use in immersion classrooms. *Modern Language Journal, 79*, 166–178.

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.

Tomasello, M. (in press). [Review of the book *Constructions: A construction grammar approach to argument structure*.] *Journal of Child Language*.

Tomasello, M., & Brooks, P. J. (in press). Early syntactic development: A construction grammar approach. In M. Barrett (Ed.), *The development of language*. Hove, UK: Psychology Press.

Ungerer, F., & Schmid, H. J. (1996). *An introduction to cognitive linguistics*. Harlow, UK: Addison Wesley Longman.

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