The Ways of Words

Nick C. Ellis

School of Psychology,
University of Wales Bangor,
Bangor,
Gwynedd,
Wales LL57 2DG
UK

Submitted for inclusion in Albrechtsen, D., Henriksen, B. & Mees, I. (Eds.),
Festschrift for Kirsten Haastrup.
Odense: Odense University Press

For section: "The Contributors"

Nick C. Ellis is a Reader in Psychology at the University of Wales, Bangor. His research interests include first, second and foreign language acquisition, implicit and explicit learning, reading, spelling, language disorders, connectionism, and memory. He can be contacted by e-mail as: n.ellis@bangor.ac.uk.

Note:

Parts of this chapter appear in Ellis, N. C. (1998). Emergentism, Connectionism and Language Learning. *Language Learning, 49*.

Kirsten Haastrup is keenly interested in the ways of words. This fascination is implicit in her speech, a flawless and precise second language English, and explicit in her research: her 1991 book, *Lexical Inferencing Procedures or Talking about Words*, is an insightful, thorough and painstaking investigation into the cues which foreign-language learners use to infer the meanings of new words. Her dedication, her friendship and her scholarship have been widely influential. But I must say no more on this matter, since any deviation from modesty would not be the Danish way.

Instead, I will engage with her research interest. Haastrup (1991) investigated average- and high-achieving foreign language learners' conscious problem-solving strategies for determining the meanings of words they have never heard or seen before. She analysed participants' talk-aloud protocols to determine the sources of information which learners use: CONTEXTUAL CUES (collocational cues from the immediate co-text, information from other sentences in the text), INTRALINGUAL CUES I: the test word itself (its phonology, orthography, morphology [analysis of prefix, suffix and stem], word class, collocations, semantic remindings), and INTRALINGUAL CUES II: the syntax of the sentence (e.g., use of article, co-ordinate adjective, preposition or number cues). Haastrup's findings are rich in descriptive detail and implications for instruction: Less proficient learners were heavily influenced by formal similarity between words in the L2 and those in their L1; their lack of L2 experience often led them to miss many of the intralingual cues detailed above; they found it very difficult to integrate these various intralingual and contextual cues. Both levels of learner, on failing to identify any intralingual cues, relied on context, the higher-level participants using information from the texts, the lower-level participants relying more of world knowledge.

With increasing competence, much of incidental vocabulary learning does indeed come from inference during reading. Skilled language users possess the necessary sources of linguistic information, are proficient at reading and automatically accessing meaning, having long-ago left behind any conscious concerns with simple word-form, and properly infer the meanings of novel words from texts. Reading is a rich source of vocabulary acquisition (Stanovich & Cunningham, 1992; Sternberg, 1987). It is an ideal medium for it. Moderate-to-low-frequency words - precisely those that differentiate between individuals of high and low vocabulary size - appear much more often in common reading matter than they do in common speech. And there is opportunity for the reader to study the context, to form hypotheses at leisure and to check and cross validate them, thus to infer meanings. The word holds still on the page, whereas in speech it passes ephemerally. Reading gives whole language, not just isolated words. So much for the skilled reader.

Pity then, the beginning language learner, lacking in declarative knowledge of word form and syntactic cues, lacking the automatic access of information provided by fluent reading skills, lacking the metalinguistic subskills of hypothesis formation and the flexibility to test and reformulate their hypotheses. So many *potential* sources of evidence to help them bootstrap their way up into language, but, from the beginner's perspective, from the floor, so many daunting and difficult codes to be cracked. So little for the beginning language learner.

How is progress possible? Just how do foreign language learners acquire these different types of information? How do they crack the codes of inflectional and derivational morphology? How do they learn orthographic and phonological structure, grapheme-phoneme mappings, collocations, syntactic word class, lexical semantics? How do they automatise processing? How do they get to the level of proficiency where they can readily teach themselves the meanings of novel words?

Some aspects of vocabulary acquisition are conscious and explicit. Because the beginnings of bootstrapping are laboriously slow and unmotivating, there is good advantage to be had from explicitly learning the most frequent two to three thousand words of the language which will form up to 90% of the content of the average texts which the learner will meet (Nation & Waring, 1997), and there are useful mnemonic ways of speeding this initial learning and of spacing repetitions to minimise forgetting (Nation, 1990; Ellis, 1995, 1997). Could these words then serve as a useful source of data for inferencing new words from context? The primarily translation-equivalent information that list-learned vocabulary gives rather slim pickings in this regard. Reading sentences composed of these list-learned words might evoke relevant world knowledge, but not much else. Many of the other categories of useable information listed by Haastrup (1991) are missing. What of collocations, syntax, word class, and morphology? What of the richness of word-meaning variations? What of homonymy, multiple syntactic class membership, idioms, pragmatics and registers? Where does all of this information come from? It seems doubtful that skilled language users have consciously learned all of this. So whence does it derive? It seems more likely that we incidentally gather these types of information while comprehending language: learning processes abstract this information from the masses of language input data that we are exposed to during the tens of thousands of hours on task which leads to nativelike proficiency. From this vast database of processed language there emerges patterns of syntagmatic and paradigmatic associations which form the rule-like regularities of syntax and semantics.

But is this really possible? How can we test these notions? Human implicit learning is not something that we can assess using introspective protocols of the type used by Haarstrup -- that which is unconscious is not accessible to consciousness. What other sources of evidence could help us turn these possibilities, these 'seems', into something rather more definite? In the rest of this short chapter I will briefly review recent simulation work, done within the connectionist tradition, which demonstrates the types of lexical representation that can emerge from simple statistical analyses of sequences of language. There is insufficient space to do more than slide over the surface (see Elman, Bates et al. (1997), Redington & Chater (1997), Seidenberg (1997), Plunkett (1998), Ellis (in press), and MacWhinney (in press) for more detailed coverage). Nonetheless, I hope to persuade readers how rich a source of morphological, syntactic and semantic representations is the simple distribution of words in language.

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 (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. Computational practicalities dictate that, for the moment, this can only be done in a piecemeal fashion: there are now many separate models addressing the acquisition of morphology, phonological rules, novel word repetition, prosody, semantic structure, syntactic structure, etc. (see, e.g., Levy, Bairaktaris, Bullinaria & Cairns, 1995; MacWhinney & Leinbach, 1991; Rumelhart & McClelland, 1986; Plunkett, 1998). 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 last ten years of connectionist research has produced a range of substantive demonstrations of emergent language representations.

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 us 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, b; 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 since 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).

I will give illustration connectionist simulations from 5 broad areas of learning.

(1) Phonological and orthographic structure

Elman (1990) used a simple recurrent network to investigate the temporal properties of sequential inputs of language. Elman's network was fed one letter at a time and had to predict the next letter in the sequence (akin to asking you to predict the next letter at the following example choice points marked !: Once upo!n a !time). It was trained on 200 sentences where there was no word or sentence boundary information -- for example, part of the stream was:

Manyyearsagoaboyandgirllivedbytheseatheyplayedhappily...

The network abstracted a lot of information about the structure of English. It learned about orthographic sequential probabilities; it learned that there were common recurring units (which, it so happens, linguists identify as morphemes and words); it extracted word sequence information too. At times, when the network could not predict the actual next phoneme, it nonetheless predicted the correct category of phoneme: vowel/consonant, etc. (see also Elman & Zipser, 1988 where networks were trained 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.

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 (Ellis, 1996a). It is from these potential associations with other representations that other interesting properties of language emerge.

(2) 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, etc. (for example, 'dragon' occurred as a pattern in activation space which is 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 is hierarchical, soft and implicit. The network moves 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 give 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.

(3) Lexical semantics

Landauer and Dumais (1997) present a Latent Semantic Analysis (LSA) model which simulates first and second language learners' acquisition of vocabulary from text. The model simply treats words as being alike if they tend to co-occur with the same neighbouring 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 school-children. After the model had been trained by exposing it to text samples from over 30,000 articles from Groliers 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 is surprisingly good for a model which had no prior linguistic or grammatical knowledge and which could not see or hear, and thus could make no use of phonology, morphology or real-world perceptual knowledge. In this account, lexical semantic acquisition *emerges* from the analysis of word co-occurrence.

Syntactic and semantic information result from different grains of analysis of the same language input - local word co-occurrence data in sequence is important for the derivation of latent syntactic information, slightly more global word co-occurrence statistics, ignoring order, results in the derivation of lexical semantics. For lexical syntax and semantics alike, 'tell me thy company, and I will tell thee what thou art'.

(4) Morphosyntax

The processes described in (1) 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 we have 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 were Rumelhart and McClelland (1986) who 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 (Pinker & Prince, 1988; Lachter & Bever, 1988), some of which are 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. The successes of these recent models in capturing the regularities that are present (i) in associating phonological form of lemma with phonological form of inflected form (Daugherty & Seidenberg, 1994; MacWhinney & Leinbach, 1991; Marchman, 1993; Plunkett & Marchman, 1991), and (ii) between referents (+past tense or +plural) and associated inflected perfect or plural forms (Cottrell & Plunkett, 1994, Ellis & Schmidt, 1997), closely simulating 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, 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 that's needed for syntax falls quite naturally out of simple sequence analysis and the patterns of association between patterns of sequences and patterns of referents.

(5) Syntactic constructions

Links between phonological chunks and conceptual representations underlie reference and grounded semantics; patterns in these cross-modal associations underlie the emergence of constructions (Goldberg, 1992).

"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 using simple connectionist models relating lexical cues (like word order, verb agreement morphology, case marking, etc.) and functional interpretations (like agency, topicality, perspective, givenness, etc.) for sentence comprehension or production. 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).

The competition model has been a good start for investigating the emergence of strategies for the linguistic realization of reference. But if we believe that 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 our models 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 us 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. We cannot understand language acquisition by understanding phonological memory alone. All of the systems of working memory, all perceptual representational systems, our 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 entails that to understand the emergence of language we must also understand the workings of attention, vision and other representational systems. And then we must understand the regularities of the mappings of these systems onto particular languages. And the mappings in question are piecemeal -- it's the detailed content of the mappings that is important, not simply the modalities concerned -- which is why cognitive linguistics focuses on particular constructions and representational aspects at a time, like, for example, motion event frames (Langacker, 1991; Talmy, 1996b) or spatial language (Bowerman, 1996).

We have to be linguist, psychologist, physiologist, and computational neuroscientist at the same time. There is recent computational work which is beginning to relate linguistic constructions to more plausible models of visual perception of movement in space (Regier, 1996). Similarly, Narayanan (1997) shows 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, Lakoff, Bailey, Narayanan, Regier & Stolcke, 1996) is well motivated -- if we want to understand emergence of language and we believe in the constraints of embodiment, then our 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*. We 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. Thus might simple associations amass over the learner's language-input history into a web of multimodal connections which represent the complexities of lexical representations.

Implicit and Explicit Vocabulary learning and teaching

To the almost complete degree that our fluent language use is unencumbered by metalinguistic descriptions of sufficient complexity to allow its generation, so the representation and processing that generates language must be unconscious. But what is the role of attention and consciousness in vocabulary *learning*?

I have argued that language understanding and language production utilise the many millions of associations that the learner has acquired in their history of language use. Thus language is learned in the course of using language, and the best predictor of language facility will simply be time-on-task. Research on implicit learning and implicit memory suggests that at least some of the relevant associations can be acquired from the input without the learner being consciously aware of the contingency, although the relevant aspects of the input must be attended for processing (see Ellis, 1994; Hsaio & Reber, 1997; Schmidt, 1994). In communicative situations, efforts are made to encourage the meanings are clearly attended and understood. But during reading, the meanings of the occasional new words may not readily be apparent. This is when the very conscious, problem-solving processes described by Haarstrup (1991) are most relevant.

Even associations which may eventually be acquired implicitly can be facilitated by conscious attention. In acquiring associations, some parts of the input environment can be made more salient, and learners are more likely to learn about the PARTS OF THE ENVIRONMENT which they selectively ATTEND. Thus there are ways of speeding learners' first or second language acquisition from a given amount of language exposure, to increase the quality of the learning (see Ellis & Laporte, 1997 for review). These ways, which include grammatical consciousness raising or input processing, as well as corrective feedback and recasts, promote the acquisition of sophisticated grammatical proficiency. There is some benefit in a focus on form in second language instruction (see R. Ellis, 1994; Long, 1988, 1991; Terrell, 1991, for reviews of instructional programs which incorporate these ideas into general language instruction and Nation, 1991; Carter, 1992; and Schmitt & McCarthy, 1997 for explicit vocabulary instructional programmes).

Communicative approaches give comprehensible input, time-on-task, and opportunity for relating form and function. Of course, all of this is necessary for developing the associations necessary for language learning. Naturalistic environments provide motivation and plenty of opportunity for output practice as well. These are situations which guarantee sufficient quantity of language. Focus on forms alone can teach some declarative pedagogical rules, but at its worst can be accompanied by too little time on the task of language use itself. At this worst it is insufficient to support language development. But focus on form instruction, which is rich in communicative opportunities and which also makes salient the associations between structures (which the learner is already at a stage to be able to represent) and functions, can facilitate language acquisition.

Conclusions

Haastrup (1991) first focused on the ways in which higher-level language learners teach themselves new vocabulary whilst reading. Then she determined the reasons why lower level language learners have difficulties doing this. She showed that lower-level language learners lack many of the interlingual cues to word meaning, lack the automaticity in reading which frees cognitive resources for meaning problem-solving and inference, and lack strategies for combining different sources of information. The research which I have described here focused on the earlier stages of acquisition of these interlingual cues of word form, morphology, syntactic class, collocations, idioms, function and meaning. These are complementary questions to those of Haastrup, using quite different research methods. The particulars of the methods one chooses to adopt aren't terribly important, nor even are the specific questions. What is ultimately good, as Kirsten Haastrup exemplifies, is a concern for learners, a belief in scientific method and academic community, scrupulous description and analysis, energy, and, most of all, a readiness for fascination.

References

Bailey, D., Feldman, J., Narayanan S., & Lakoff, G. (1997). Modelling Embodied Lexical Development. *Proceedings of the Cognitive Science Conference 1997*. Cognitive Science Society, Pittsburgh, PA.

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.

Carter, R. (1987). *Vocabulary: Applied Linguistic Perspectives.* London: Allen & Unwin.

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

Daugherty, K. G., & Seidenberg, M. S. (1994). Beyond rules and exceptions: A connectionist approach to inflectional morphology. In S. D. Lima, R. L. Corrigan, and 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.

Ellis, N. C. (1994). 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. (1995). The psychology of foreign language acquisition: Implications for CALL. *International Journal of Computer Assisted Language Learning (CALL), 8*, 103-128.

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). Vocabulary acquisition: Word structure, collocation, grammar, and meaning. In M. McCarthy & N. Schmidt (Eds.) (1997), *Vocabulary: description, acquisition and padagogy* (pp. 122-139). Cambridge: Cambridge University Press.

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

Ellis, N. C., & Laporte, N. (1997). Contexts of Acquisition: Effects of Formal Instruction and Naturalistic Exposure on Second Language Acquisition. In A. M. B. de Groot and J. F. Kroll (Eds.), *Tutorials in Bilingualism: Psycholinguistic Perspectives* (pp. 53-83). Hillsdale, N.J.: Lawrence Erlbaum.

Ellis, N. C., & Schmidt, R. (1998). Rules or Associations in the Acquisition of Morphology? The Frequency by Regularity Interaction in Human and PDP Learning of Morphosyntax. *Language and Cognitive Processes, 13*, 307-336.

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

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: A Bradford Book.

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). L0: 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.

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

Haastrup, K. (1991). *Lexical inferencing procedures or Talking about words.* Tübingen: Gunter Narr Verlag Tübingen.

Hsaio, A. T., & Reber, A. S. (1997). The role of attention in implicit sequence learning: Exploring the limits of the cognitive unconscious. In M. A. Stadler & P. A. Frensch (Eds.), *Implicit Learning Handbook* (pp. 471-494). London: Sage Publications.

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.

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.

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

Long, M. (1988). Instructed interlanguage development. In L. Beebe (Ed.), *Issues in second language acquisition: Multiple perspectives* (pp. 115-141). New York: Newbury House.

Long, M. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, D. Coste, R. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspectives* (pp. 39-52). Amsterdam: John Benjamins.

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

MacWhinney, B. (Ed.) (1998). *The emergence of language*. 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. (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, 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 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.

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

Nation, I.S.P. (1990). *Teaching and Learning Vocabulary.* New York: Newbury House.

Nation, I.S.P. & Waring, R. (1997). Vocabulary size, text coverage and word lists. In M. McCarthy & N. Schmidt (Eds.) (1997), *Vocabulary: description, acquisition and padagogy* (pp. 6-19). Cambridge: Cambridge University Press.

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. (Ed.) (1998) Connectionist Models of Language [Special issue].*Language and Cognitive Processes*, *13*.

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

Redington, M. & Chater, N. (1997). Probabilistic and distributional approaches to language acquisition. *Trends in Cognitive Sciences, 1,* 273-281.

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

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

Schmidt, R. (1994). Implicit learning and the cognitive unconscious: Of artificial grammars and SLA. In N. Ellis (Ed.), *Implicit and Explicit Learning of Languages* (pp. 165-210). London: Academic Press.

Schmitt, N. & McCarthy, M. (Eds.) (1997). *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press.

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, Mass.: MIT Press.

Stanovich, K.E., & Cunningham, A.E. (1992). Studying the consequences of literacy within a literate society: The cognitive correlates of print exposure. *Memory and Cognition,* **20**, 51-68.

Sternberg, R.J. (1987). Most vocabulary is learned from context. In M.G. McKeown & M.E. Curtis (Eds.), *The Nature of Vocabulary Acquisition* (pp. 89-105). Hillsdale, N.J.: Lawrence Erlbaum.

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.

Terrell, T. (1991). The role of grammar instruction in a communicative approach. *The Modern Language Journal, 75* , 52-63.

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