4 Constructions, Chunking, and Connectionism: The Emergence of Second Language Structure

NICK C. ELLIS

1 Introduction and Overview

Constructivist views of language acquisition hold that simple learning mechanisms operating in and across human systems for perception, motor action, and cognition while exposed to language data in a communicatively rich human social environment navigated by an organism eager to exploit the functionality of language are sufficient to drive the emergence of complex language representations. The various tribes of constructivism – that is, connectionists (Christiansen and Chater, 2001; Christiansen, Chater, and Seidenberg, 1999; Levy, Bairaktaris, Bullinaria, and Cairns, 1995; McClelland, Rumelhart, and the PDP Research Group, 1986; Plunkett, 1998), functional linguists (Bates and MacWhinney, 1981; MacWhinney and Bates, 1989), emergentists (Elman, Bates, Johnson, Karmiloff-Smith, Parisi, and Plunkett, 1996; MacWhinney, 1999a), cognitive linguists (Croft and Cruse, 1999; Lakoff, 1987; Langacker, 1987, 1991; Ungerer and Schmid, 1996), constructivist child language researchers (Slobin, 1997; Tomasello, 1992, 1995, 1998a, 2000), applied linguists influenced by chaos/complexity theory (Larsen-Freeman, 1997), and computational linguists who explore statistical approaches to grammar (Bod, 1998; Jurafsky, 1996) – all share a functional-developmental, usage-based perspective on language. They emphasize the linguistic sign as a set of mappings between phonological forms and conceptual meanings or communicative intentions; thus, their theories of language function, acquisition, and neurobiology attempt to unite speakers, syntax, and semantics, the signifiers and the signifieds. They hold that structural regularities of language emerge from learners' lifetime analysis of the distributional characteristics of the language input and, thus, that the knowledge of a speaker/hearer cannot be understood as an innate grammar, but rather as a statistical ensemble of language experiences that changes slightly every time a new utterance is processed. Consequently, they analyze language acquisition processes rather than the final state or the language acquisition device (see Sorace, this volume; White, this volume). They work within the broad remit of cognitive science, seeking functional and neurobiological descriptions of the learning processes which, through exposure to representative experience, result in change, development, and the emergence of linguistic representations.

Section 2 of this review describes cognitive linguistic theories of construction grammar. These focus on constructions as recurrent patterns of linguistic elements that serve some well-defined linguistic function. These may be at sentence level (such as the imperative, the ditransitive, the yes-no question) or below (the noun phrase, the prepositional phrase, etc.). Whereas Government-Binding Theory denied constructions, viewing them as epiphenomena resulting from the interaction of higher-level principles-and-parameters and lower-level lexicon, cognitive linguistics – construction grammar in particular (Croft, 2001; Goldberg, 1995) – has brought them back to the fore, suspecting instead that it is the higher-level systematicities that emerge from the interactions of constructions large and small. Section 3 concerns the development of constructions as complex chunks, as high-level schemata for abstract relations such as transitives, locatives, datives, or passives. An acquisition sequence – from formula, through low-scope pattern, to construction – is proposed as a useful starting point to investigate the emergence of constructions and the ways in which type and token frequency affect the productivity of patterns. Section 4 presents the psychological learning mechanisms which underpin this acquisition sequence. It describes generic associative learning mechanisms such as chunking which, when applied to the stream of language, provide a rich source of knowledge of sequential dependencies ranging from low-level binary chunks like bigrams, through phonotactics, lexis, and collocations, up to formulae and idioms. Although a very basic learning mechanism, chunking results in hierarchical representations and structure dependency.

Emergentists believe that many of the rule-like regularities that we see in language emerge from the mutual interactions of the billions of associations that are acquired during language usage. But such hypotheses require testing and formal analysis. Section 5 describes how connectionism provides a means of evaluating the effectiveness of the implementations of these ideas as simulations of language acquisition which are run using computer models consisting of many artificial neurons connected in parallel. Two models of the emergence of linguistic regularity are presented for detailed illustration. Other simulations show how analysis of sequential dependencies results in grammatically useful abstract linguistic representations. The broad scope of connectionist and other distributional approaches to language acquisition is briefly outlined. The review concludes by discussing some limitations of work to date and provides some suggestions for future progress.

2 Construction Grammar

This section outlines cognitive linguistic analyses of the interactions between human language, perception, and cognition, and then focuses on construction grammar (Croft, 2001; Fillmore and Kay, 1993; Goldberg, 1995; Langacker, 1987; Tomasello, 1998a, 1998b) as an approach for analyzing the ways in which particular language patterns cue particular processes of interpretation. If words are the atoms of language function, then construction grammar provides the molecular level of analysis.

2.1 Cognitive linguistics

Cognitive linguistics (Barlow and Kemmer, 2000; Croft and Cruse, 1999; Goldberg, 1995; Lakoff, 1987; Lakoff and Johnson, 1980; Langacker, 1987, 1991; Talmy, 1988; Ungerer and Schmid, 1996) provides detailed qualitative analyses of the ways in which language is grounded in human experience and in human 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, depends on the perception and categorization of the real world around us. Since we constantly observe and play an active role in this world, we know a great deal about the entities of which it consists, and this experience and familiarity is reflected in the nature of language. Ultimately, everything we know is organized and related in some meaningful way or other, and everything we perceive is affected by our perceptual apparatus and our perceptual history. Language reflects this embodiment and this experience.

The different degrees of salience or prominence of elements involved in situations that we wish to describe affect the selection of subject, object, adverbials, and other clause arrangement. Figure/ground segregation and perspective taking, processes of vision and attention, are mirrored in language and have systematic relations with syntactic structure. Thus, paradoxically, a theory of language must properly reflect the ways in which human vision and spatial representations are explored, manipulated, cropped and zoomed, and run in time like movies under attentional and scripted control (Kosslyn, 1983; Talmy, 1996a). In language production, what we express reflects which parts of an event attract our attention; depending on how we direct our attention, we can select and highlight different aspects of the frame, thus arriving at different linguistic expressions. 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. In language comprehension, abstract linguistic constructions (like simple locatives, datives, and passives) serve as a "zoom lens" for the listener, guiding their attention to a particular perspective on a scene while backgrounding other aspects (Goldberg, 1995).

Thus, cognitive linguistics describes the regularities of syntax as emergent from the cross-modal evidence that is collated during the learner's lifetime of using and comprehending language.

Cognitive linguistics was founded on the principle that language cognition cannot be separated from semantics and the rest of cognition. The next section shows how it similarly denies clear boundaries between the traditional linguistic separations of syntax, lexicon, phonology, and pragmatics.

2.2 *Constructions*

Traditional descriptive grammars focus on constructions, that is, recurrent patterns of linguistic elements that serve some well-defined linguistic function. As noted earlier, these may be at sentence level (such as the imperative, the ditransitive, the yes-no question) or below (the noun phrase, the prepositional phrase, etc.). The following summary of construction grammar, heavily influenced by Langacker (1987) and Croft and Cruse (1999), illustrates the key tenets.

A construction is a conventional linguistic unit, that is, part of the linguistic system, accepted as a convention in the speech community, and entrenched as grammatical knowledge in the speaker's mind. Constructions may (i) be complex, as in [Det Noun], or be simple, as in [Noun] (traditionally viewed as "syntax"); (ii) represent complex structure above the word level, as in [Adj Noun], or below the word level, as in [NounStem-PL] (traditionally viewed as "morphology"); or (c) be schematic, as in [Det Noun], or specific, as in [the United Kingdom], traditionally viewed as "lexicon." Hence, "morphology," "syntax," and "lexicon" are uniformly represented in a construction grammar, unlike both traditional grammar and generative grammar. Constructions are symbolic. In addition to specifying the properties of an utterance's defining morphological, syntactic, and lexical form, a construction also specifies the semantic, pragmatic, and/or discourse functions that are associated with it. Constructions form a structured inventory of speakers' knowledge of the conventions of their language (Langacker, 1987, pp. 63–6), usually described by construction grammarians in terms of a semantic network, where schematic constructions can be abstracted over the less schematic ones which are inferred inductively by the speaker in acquisition. This non-modular semantic network representation of grammar is shared by other theories such as Word Grammar (Hudson, 1984, 1990). A construction may provide a partial specification of the structure of an utterance. Hence, an utterance's structure is specified by a number of distinct constructions. Constructions are independently represented units in a speaker's mind. Any construction with unique, idiosyncratic formal or functional properties must be represented independently in order to capture speakers' knowledge of their language. However, absence of any unique property of a construction does not entail that it is not represented independently and simply derived from other, more general or schematic constructions. Frequency of occurrence may lead to independent representation of even "regular" constructional patterns. This usage-based perspective implies that the acquisition of grammar is the piecemeal learning of many thousands of constructions and the frequency-biased abstraction of regularities within them.

Many constructions are based on particular lexical items, ranging from simple (*Howzat!* in cricket) to complex (*Beauty is in the eye of the beholder*). The importance of such lexical units or idiomatic phrases is widely acknowledged in SLA research when discussing holophrases (Corder, 1973), prefabricated routines and patterns (Hakuta, 1974), formulaic speech (Wong Fillmore, 1976), memorized sentences and lexicalized stems (Pawley and Syder, 1983), formulae (R. Ellis, 1994), sequences in SLA (N. Ellis, 1996, 2002), discourse management (Dörnyei and Kormos, 1998; Tannen, 1987), register (Biber and Finegan, 1994), style (Brewster, 1999), and lexical patterns and collocational knowledge (Carter, 1998; Hoey, 1991; Lewis, 1993; Schmitt, 2000). According to Nattinger (1980, p. 341), "for a great deal of the time anyway, language production consists of piecing together the ready-made units appropriate for a particular situation and . . . comprehension relies on knowing which of these patterns to predict in these situations." As Pawley and Syder (1983, p. 192) put it:

> In the store of familiar collocations there are expressions for a wide range of familiar concepts and speech acts, and the speaker is able to retrieve these as wholes or as automatic chains from the long-term memory; by doing this he minimizes the amount of clause-internal encoding work to be done and frees himself to attend to other tasks in talk-exchange, including the planning of larger units of discourse.

But other constructions are more abstract. Goldberg (1995) focuses on complex argument structure constructions such as the ditransitive (*Pat faxed Bill the letter*), the caused motion (*Pat pushed the napkin off the table*), and the conative (*Sam kicked at Bill*). She holds that these abstract and complex constructions themselves carry meaning, independently of the particular words in the sentence. For example, even though the verb *kick* does not typically imply transfer of possession, it works in the ditransitive *Pat kicked Bill the football*, and even though one is hard pressed to interpret anything but an intransitive *sneeze*, the caused motion *Pat sneezed the napkin off the table* is equally good. These abstract argument structure constructions thus create an important top-down component to the process of linguistic communication. Such influences are powerful mechanisms for the creativity of language, possibly even as manifest in derivational phenomena such as denominal verbs (*They tabled the motion*) and deverbal nouns (*Drinking killed him*) (Tomasello, 1998b).

Constructions show prototype effects. For example, for ditransitive constructions there is the central sense of agent-successfully-causes-recipient-to-receive-patient (*Bill gave/handed/passed/threw/took her a book*), and various more peripheral meanings such as future-transfer (*Bill bequeathed/allocated/granted/reserved her a book*) and enabling-transfer (*Bill allowed/permitted her one book*). Prototype effects are fundamental characteristics of category formation, again blurring the boundaries between syntax and lexicon and other cognitive domains (N. Ellis, 2002).

3 Learning Constructions

If linguistic systems comprise a conspiracy of constructions, then language acquisition, L1 or L2, is the acquisition of constructions. There is nothing revolutionary in these ideas. Descriptive grammars (e.g., Biber, Johansson, Leech, Conrad, and Finegan, 1999; Quirk, Greenbaum, Leech, and Svartvik, 1985) are traditionally organized around form–function patterns; so are grammars which are designed to inform pedagogy (e.g., Celce-Murcia and Larsen-Freeman, 1983). But what about the processes of acquisition? To date, construction grammar has primarily concerned descriptions of adult competence, although language acquisition researchers, particularly those involved in child language, are now beginning to sketch out theories of the acquisition of constructions which involve a developmental sequence from formula, through low-scope pattern, to construction.

3.1 Formulae and idioms

Formulae are lexical chunks which result from memorizing the sequence of frequent collocations. Large stretches of language are adequately described by finite-state grammars, as collocational streams where patterns flow into each other. Sinclair (1991, p. 110), then director of the Cobuild project, the largest lexicographic analysis of the English language to date, summarized this in 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.

Rather than its being a somewhat minor feature compared with grammar, Sinclair suggests that, for normal texts, the first mode of analysis to be applied is the idiom principle, as most text is interpretable by this principle. Whereas most of the material that Sinclair was analyzing in the Bank of English was written text, comparisons of written and spoken corpora demonstrate that collocations are even more frequent in spoken language (Biber et al., 1999; Brazil, 1995; Leech, 2000). Parole is flat and Markovian because it is constructed "off the top of one's head," and there is no time to work it over. Utterances are constructed as intonation units which have the grammatical form of single clauses, although many others are parts of clauses, and they are often highly predictable in terms of their lexical concordance (Hopper, 1998). Language reception and production are mediated by learners' representations of chunks of language: "Suppose that, instead of shaping discourse according to rules, one really pulls old language from memory (particularly old language, with all its words in and everything), and then reshapes it to the current context: "'Context shaping', as Bateson puts it, 'is just another term for grammar'" (Becker, 1983, p. 218).

Even for simple concrete lexis or formulae, acquisition is no unitary phenomenon. It involves the (typically) implicit learning of the sequence of sounds or letters in the word along with separable processes of explicit learning of perceptual reference (N. Ellis, 1994c, 2001). Yet however multifaceted and fascinating is the learning of words (Aitchison, 1987; Bloom, 2000; N. Ellis and Beaton, 1993a, 1993b; Miller, 1991; Ungerer and Schmid, 1996), lexical learning has generally been viewed as a phenomenon that can readily be understood in terms of basic processes of human cognition. Learning the form of formulae is simply the associative learning of sequences. It can readily be understood in terms of the process of chunking which will be described in section 4.

The mechanism of learning might be simple, but the product is a rich and diverse population of hundreds of thousands of lexical items and phrases. The store of familiar collocations of the native language speaker is very large indeed. The sheer number of words and their patterns variously explains why language learning takes so long, why it requires exposure to authentic sources, and why there is so much current interest in corpus linguistics in SLA (Biber, Conrad, and Reppen, 1998; Collins Cobuild, 1996; Hunston and Francis, 1996; McEnery and Wilson, 1996). Native-like competence and fluency demand such idiomaticity.

3.2 Limited scope patterns

The learning of abstract constructions is more intriguing. It begins with chunking and committing formulae to memory. But there is more. Synthesis precedes analysis. Once a collection of like examples is available in long-term memory, there is scope for implicit processes of analysis of their shared features and for the development of a more abstract summary schema, in the same way as prototypes emerge as the central tendency of other cognitive categories.

Consider first the development of slot-and-frame patterns. Braine (1976) proposed that the beginnings of L1 grammar acquisition involve the learning of the position of words in utterances (e.g., *More car*, *More truck*, etc. allow induction of the pattern "more + recurring element"). Maratsos (1982) extended this argument to show that adult-like knowledge of syntactic constructions (including both syntactic relations and part-of-speech categories like verb and noun) can also result from positional analysis without the influence of semantic categories like agent and action. He proposed that this learning takes place through the amassing of detailed information about the syntactic handling of particular lexical items, followed by discovery of how distributional privileges transfer among them. The productivity of distributional analyses resultant from connectionist learning of text corpora will be described in section 5.

It is important to acknowledge the emphases of such accounts on piecemeal learning of concrete exemplars. Longitudinal child-language acquisition data suggest that, to begin with, each word is treated as a semantic isolate in the sense that the ability to combine it with other words is not accompanied by a parallel ability with semantically related words. An early example was that of Bowerman (1976), who demonstrated that her daughter Eva acquired the *more + X* construction long before other semantically similar relational words like *again* and *all-gone* came to be used in the similar pivot position in two-word utterances. Pine and Lieven (Lieven, Pine, and Dresner Barnes, 1992; Pine and Lieven, 1993, 1997; Pine, Lieven, and Rowland, 1998) have since demonstrated widespread lexical specificity in L1 grammar development. Children's language between the ages of 2 and 3 years is much more "low-scope" than theories of generative grammar have argued. A high proportion of children's early multi-word speech is produced from a developing set of slot-and-frame patterns. These patterns are often based on chunks of one or two words or phrases and they have "slots" into which the child can place a variety of words, for instance subgroups of nouns or verbs (e.g., *I can't + Verb*; *where's + Noun + gone?*). Children are very productive with these patterns and both the number of patterns and their structure develop over time. But they are lexically specific. Pine and Lieven's analyses of recordings of 2–3-year-old children and their mothers measure the overlap between the words used in different slots in different utterances. For example, if a child has two patterns, *I can't + X* and *I don't + X*, Pine and Lieven measure whether the verbs used in the X slots come from the same group and whether they can use any other CAN- or DO-auxiliaries. There is typically very little or no overlap, an observation which supports the conclusion that (i) the patterns are not related through an underlying grammar (i.e., the child does not "know" that *can't* and *don't* are both auxiliaries or that the words that appear in the patterns all belong to a category of Verb); (ii) there is no evidence for abstract grammatical patterns in the 2–3-year-old child's speech; and (iii) that, in contrast, the children are picking up frequent patterns from what they hear around them, and only slowly making more abstract generalizations as the database of related utterances grows.

Tomasello (1992) proposed the Verb Island hypothesis, in which it is the early verbs and relational terms that are the individual islands of organization in young children's otherwise unorganized grammatical system – in the early stages the child learns about arguments and syntactic markings on a verb-by-verb basis, and ordering patterns and morphological markers learned for one verb do not immediately generalize to other verbs. Positional analysis of each verb island requires long-term representations of that verb's collocations, and, thus, this account of grammar acquisition implies vast amounts of long-term knowledge of word sequences. Only later are syntagmatic categories formed from abstracting regularities from this large dataset in conjunction with morphological marker cues (at least in case-marking languages). Goldberg (1995) argues that certain patterns are more likely to be made more salient in the input because they relate to certain fundamental perceptual primitives, and, thus, that the child's construction of grammar involves both the distributional analysis of the language stream and the analysis of contingent perceptual activity:

> Constructions which correspond to basic sentence types encode as their central senses event types that are basic to human experience . . . that of someone causing something, something moving, something being in a state, someone possessing something, something causing a change of state or location, something undergoing a change of state or location, and something having an effect on someone. (Goldberg, 1995, p. 39)

Goldberg and Sethuraman (1999) show how individual "pathbreaking" semantically prototypic verbs form the seed of verb-centered argument structure patterns. Generalizations of the verb-centered instances emerge gradually as the verb-centered categories themselves are analyzed into more abstract argument structure constructions. The verb is a better predictor of sentence meaning than any other word in the sentence. Nevertheless, children ultimately generalize to the level of constructions, because constructions are much better predictors of overall meaning. Although verbs thus predominate in seeding low-scope patterns and eventually more abstract generalizations, Pine et al. (1998) have shown that such islands are not exclusive to verbs, and that the theory should be extended to include limited patterns based on other lexical types such as bound morphemes, auxiliary verbs, and case-marking pronouns.

3.3 Exemplar frequency and construction productivity

The research reviewed thus far has focused on piecemeal learning, the emergence of syntactic generalizations, and the elements of language which seed such generalizations. There is another important strand in L1 construction-learning research that concerns how the frequency of patterns in the input affects acquisition. Usage-based linguistics holds that language use shapes grammar through frequent repetitions of usage, but there are separable effects of token frequency and type frequency. Token frequency is how often in the input particular words or specific phrases appear; type frequency, on the other hand, counts how many different lexical items a certain pattern or construction is applicable to. Type frequency refers to the number of distinct lexical items that can be substituted in a given slot in a construction, whether it is a word-level construction for inflection or a syntactic construction specifying the relation among words. The "regular" English past tense *-ed* has a very high type frequency because it applies to thousands of different types of verbs, whereas the vowel change exemplified in *swam* and *rang* has a much lower type frequency. Bybee (Bybee, 1995; Bybee and Thompson, 2000) shows how the productivity of a pattern (phonological, morphological, or syntactic) is a function of its type rather than its token frequency. In contrast, high token frequency promotes the entrenchment or conservation of irregular forms and idioms – the irregular forms only survive because they are very frequent.

Type frequency determines productivity because: (i) the more lexical items that are heard in a certain position in a construction, the less likely it is that the construction is associated with a particular lexical item, and the more likely it is that a general category is formed over the items that occur in that position; (ii) the more items the category must cover, the more general are its criterial features, and the more likely it is to extend to new items; and (iii) high type frequency ensures that a construction is used frequently, thus strengthening its representational schema and making it more accessible for further use with new items (Bybee and Thompson, 2000).

3.4 The same sequence for SLA?

To what degree might this proposed developmental sequence of syntactic acquisition apply in SLA? SLA is different from L1A in numerous respects, particularly with regard to:

- i *mature conceptual development:*
  - a in child language acquisition knowledge of the world and knowledge of language are developing simultaneously whereas adult SLA builds upon pre-existing conceptual knowledge;
  - b adult learners have sophisticated formal operational means of thinking and can treat language as an object of explicit learning, that is, of conscious problem-solving and deduction, to a much greater degree than can children (N. Ellis, 1994a);
- ii *language input:* the typical L1 pattern of acquisition results from naturalistic exposure in situations where caregivers naturally scaffold development (Tomasello and Brooks, 1999), whereas classroom environments for second or foreign language teaching can distort the patterns of exposure, of function, of medium, and of social interaction (N. Ellis and Laporte, 1997);
- iii *transfer from L1:* adult SLA builds on pre-existing L1 knowledge (MacWhinney, 1992; Odlin, this volume), and, thus, for example, whereas a young child has lexically specific patterns and only later develops knowledge of abstract syntactic categories which guide more creative combinations and insertions into the slots of frames, adults have already acquired knowledge of these categories and their lexical membership for L1, and this knowledge may guide creative combination in their L2 interlanguage to variously good and bad effects. Nevertheless, unless there is evidence to the contrary, it is a reasonable default expectation that naturalistic SLA develops in broadly the same fashion as does L1 – from formulae, through low-scope patterns, to constructions – and that this development similarly reflects the influences of type and token frequencies in the input. (But see Doughty, this volume, for a discussion of how L1 and L2 processing procedures differ.)

There are lamentably few longitudinal acquisition data for SLA that are of sufficient detail to allow the charting of construction growth. Filling this lacuna and performing analyses of SLA which parallel those for L1A described in section 3.2 is an important research priority. But the available evidence does provide support for the assumption that constructions grow from formulae through low-scope patterns to more abstract schema. For a general summary, there are normative descriptions of stages of L2 proficiency that were drawn up in as atheoretical a way as possible by the American Council on the Teaching of Foreign Languages (ACTFL) (Higgs, 1984). These Oral Proficiency Guidelines include the following descriptions of novice and intermediate levels that emphasize the contributions of patterns and formulae to the development of later creativity:

> *Novice Low*: Oral production consists of isolated words and perhaps a few high-frequency phrases . . . *Novice High*: Able to satisfy partially the requirements of basic communicative exchanges by relying heavily on learned utterances but occasionally expanding these through simple recombinations of their elements . . . *Intermediate*: The intermediate level is characterized by an ability to create with the language by combining and recombining learned elements, though primarily in a reactive mode. (ACTFL, 1986, p. 18)

Thus, the ACTFL repeatedly stresses the constructive potential of collocations and chunks of language. This is impressive because the ACTFL guidelines were simply trying to describe SLA as objectively as possible – there was no initial theoretical focus on formulae – yet nonetheless the role of formulae became readily apparent in the acquisition process.

There are several relevant case studies of child SLA. Wong Fillmore (1976) presented the first extensive longitudinal study that focused on formulaic language in L2 acquisition. Her subject, Nora, acquired and overused a few formulaic expressions of a new structural type during one period, and then amassed a variety of similar forms during the next. Previously unanalyzed chunks became the foundations for creative construction (see also Vihman's, 1982, analyses of her young son Virve's SLA). Such observations of the formulaic beginnings of child L2 acquisition closely parallel those of Pine and Lieven for L1.

There are a few studies which focus on these processes in classroom-based SLA. R. Ellis (1984) described how three classroom learners acquired formulae which allowed them to meet their basic communicative needs in an ESL classroom, and how the particular formulae they acquired reflected input frequency – they were those which more often occurred in the social and organizational contexts that arose in the classroom environment. Weinert (1994) showed how English learners' early production of complex target-like German foreign language negation patterns came through the memorization of complex forms in confined linguistic contexts, and that some of these forms were used as a basis for extension of patterns. Myles, Hooper, and Mitchell (1998; Myles, Mitchell, and Hooper, 1999) describe the first two years of development of interrogatives in a classroom of anglophone French L2 beginners, longitudinally tracking the breakdown of formulaic chunks such as *comment t'appelles-tu?* (what's your name?), *comment s'appelle-t-il?* (what's his name?), and *où habites-tu?* (where do you live?), in particular the creative construction of new interrogatives by recombination of their parts, and the ways in which formulae fed the constructive process. Bolander (1989) analyzed the role of chunks in the acquisition of inversion in Swedish by Polish, Finnish, and Spanish immigrants enrolled in a 4-month intensive course in Swedish. In Swedish, the inversion of subject–verb after a sentence-initial non-subject is an obligatory rule. Bolander identified the majority of the inversion cases in her data as being of a chunk-like nature with a stereotyped reading such as *det kan man säga* (that can one say) and *det tycker jag* (so think I). Inversion in these sort of clauses is also frequent when the object is omitted as in *kan man säga* (can one say) and *tycker jag* (think I), and this pattern was also well integrated in the interlanguage of most of these learners. Bolander showed that the high accuracy on these stereotyped initial-object clauses generalized to produce a higher rate of correctness on clauses with non-stereotyped initial objects than was usual for other types of inversion clause in her data, and took this as evidence that creative language was developing out of familiar formulae.

Although there are many reviews which discuss the important role of formula use in SLA (e.g., Hakuta, 1974; Nattinger and DeCarrico, 1992; Towell and Hawkins, 1994; Weinert, 1995; Wray, 1992), there is clearly further need for larger-sampled SLA corpora which will allow detailed analysis of acquisition sequences. De Cock (1998) presents analyses of corpora of language-learner productions using automatic recurrent sequence extractions. These show that second language learners use formulae at least as much as native speakers and at times at significantly higher rates. There is much promise of such computer-based learner corpus studies (Granger, 1998), providing that sufficient care is taken to gather the necessarily intensive longitudinal learner data. There is also need to test the predictions of usage-based theories regarding the influences of type frequency and token frequency as they apply in SLA.

4 Psychological Accounts of Associative Learning

This section concerns the psychological learning mechanisms which underpin the acquisition of constructions. Constructivists believe that language is cut of the same cloth as other forms of learning. Although it differs importantly from other knowledge in its specific content and problem space, it is acquired using generic learning mechanisms. The Law of Contiguity, the most basic principle of association, pervades all aspects of the mental representation of language: "Objects once experienced together tend to become associated in the imagination, so that when any one of them is thought of, the others are likely to be thought of also, in the same order of sequence or coexistence as before" (James, 1890, p. 561).

4.1 Chunking

What's the next letter in a sentence beginning *T . . .* ? Native English speakers know it is much more likely to be *h* or a vowel than it is *z* or other consonants, and that it could not be *q*. But they are never taught this. What is the first word in that sentence? We are likely to opt for *the*, or *that*, rather than *thinks* or *theosophy*. If *The . . .* begins the sentence, how does it continue? "With an adjective or noun," might be the reply. And, if the sentences starts with *The cat . . .* , then what? And then again, how should we complete *The cat sat on the . . .* ? Fluent native speakers know a tremendous amount about the sequences of language at all grains. We know how letters tend to co-occur (common bigrams, trigrams, and other orthographic regularities). Likewise, we know the phonotactics of our tongue and its phrase structure regularities. We know thousands of concrete collocations, and we know abstract generalizations that derive from them. We have learned to chunk letters, sounds, morphemes, words, phrases, clauses, bits of co-occurring language at all levels. Psycholinguistic experiments show that we are tuned to these regularities in that we process faster and most easily language which accords with the expectations that have come from our unconscious analysis of the serial probabilities in our lifelong history of input (N. Ellis, 2002).

Furthermore, we learn these chunks from the very beginnings of learning a second language. N. Ellis, Lee, and Reber (1999) observed people reading their first 64 sentences of a foreign language. While they read, they saw the referent of each sentence, a simple action sequence involving colored geometrical shapes. For example, the sentence *miu-ra ko-gi pye-ri lon-da* was accompanied by a cartoon showing a square moving onto red circles. A linguistic description of this language might include the following facts: (i) that it is an SOV language; (ii) it has adjective–noun word order; (iii) grammatical number (singular/plural) agreement is obligatory, and in the form of matching suffix endings of a verb and its subject and of a noun and the adjective that modifies it; (iv) that the 64 sentences are all of the type: $[N]_{Subject}$ $[A\ N]_{Object}$ V; and (v) that lexis was selected from a very small set of eight words. But such explicit metalinguistic knowledge is not the stuff of early language acquisition. What did the learners make of it? To assess their intake, immediately after seeing each sentence, learners had to repeat as much as they could of it. How did their intake change over time? It gradually improved in all respects. With increasing exposure, performance incremented on diverse measures: the proportion of lexis correctly recalled, correct expression of the adjective–noun agreement, correct subject–verb agreement, totally correct sentence production, correct bigrams and trigrams, and, overall, conformity to the sequential probabilities of the language at letter, word, and phrase level. With other measures it was similarly apparent that there was steady acquisition of form–meaning links and of generalizable grammatical knowledge that allowed success on grammaticality judgment tests which were administered later (Ellis et al., 1999). To greater or lesser degree, these patterns, large and small, were being acquired simultaneously and collaboratively.

Acquisition of these sequential patterns is amenable to explanation in terms of psychological theories of chunking. The notion of chunking has been at the core of short-term memory research since Miller (1956) first proposed the term. While the chunk capacity of short-term memory (STM) is fairly constant at 7 ± 2 units, its information capacity can be increased by chunking, a useful representational process in that low-level features that co-occur can be organized together and thence referred to as an individual entity. Chunking underlies superior short-term memory for patterned phone numbers (e.g., *0800-123777*) or letter strings (e.g., *AGREEMENTS*, *FAMONUBITY*) than for more random sequences (e.g., *4957-632518*, *CXZDKLWQPM*), even though all strings contain the same number of items. We chunk chunks too, so *Ellis is wittering on about chunking again* is better recalled than *again wittering on is about Ellis chunking*, and, as shown by Epstein (1967) in a more rigorous but dreary fashion than Lewis Carroll's, *A vapy koobs desaked the citar molently um glox nerfs* is more readily read and remembered than *koobs vapy the desaked um glox citar nerfs a molently*:

> A chunk is a unit of memory organization, formed by bringing together a set of already formed elements (which, themselves, may be chunks) in memory and welding them together into a larger unit. Chunking implies the ability to build up such structures recursively, thus leading to a hierarchical organization of memory. Chunking appears to be a ubiquitous feature of human memory. (Newell, 1990, p. 7)

It operates at concrete and abstract levels, as we shall now see.

Sequences that are repeated across learning experiences become better remembered. Hebb (1961) demonstrated that, when people were asked to report back random nine-digit sequences in short-term memory task, if, unbeknownst to the participants, every third list of digits was repeated, memory for the repeated list improved over trials faster than memory for non-repeated lists. This pattern whereby repetitions of particular items in short-term memory result in permanent structural traces has since become known as the Hebb effect. It pervades learning in adulthood and infancy alike. Saffran, Aslin, and Newport (1996) demonstrated that 8-month-old infants exposed for only 2 minutes to unbroken strings of nonsense syllables (for example, *bidakupado*) are able to detect the difference between three-syllable sequences that appeared as a unit and sequences that also appeared in their learning set but in random order.

Chunks that are repeated across learning experiences also become better remembered. In early Project Grammarama experiments, Miller (1958) showed that learners' free recall of redundant (grammatical) items was superior to that of random items, and hypothesized that this was because they were "recoding" individual symbols into larger chunks which decreased the absolute number of units. *Structural patterns* that are repeated across learning experiences as well become better remembered. Reber (1967) showed that memory for grammatical "sentences" generated by a finite-state grammar improved across learning sets. More recent work reviewed by Manza and Reber (1997), Mathews and Roussel (1997), and others in Berry (1997) shows that learners can transfer knowledge from one instantiation to another, that is, learn an artificial grammar instantiated with one letter set (*GFBQT*) and transfer to strings instantiated in another (*HMVRZ*), so that if there are many letter strings which illustrate patterned sequences (e.g., *GFTQ*, *GGFTQ*, *GFQ*) in the learning set, the participants show faster learning of a second transfer grammar which mirrors these patterns (*HMZR*, *HHMZR*, *HMR*) than one which does not (*HMZR*, *VMHZZ*, *VZH*). Learners can also demonstrate cross-modal transfer, where the training set might be letters, as above, but the testing set comprises sequences of colors which, unbeknownst to the participant, follow the same underlying grammar. These effects argue for more abstract representations of tacit knowledge.

Hebb effects, Miller effects, and Reber effects all reflect the reciprocal interactions between short-term memory and long-term memory (LTM) which allow us to bootstrap our way into language. The "cycle of perception" (Neisser, 1976) is also the "cycle of learning," such that bottom-up and top-down processes are in constant interaction. Repetition of sequences in phonological STM results in their consolidation in phonological LTM as chunks. The cognitive system that stores long-term memories of phonological sequences is the same system responsible for perception of phonological sequences. Thus, the tuning of phonological LTM to regular sequences allows more ready perception of input which contains regular sequences. Regular sequences are thus perceived as chunks, and, as a result, language- (L1 or L2) experienced individuals' phonological STM for regular sequences is greater than for irregular ones. This common learning mechanism underpins language acquisition in phonological, orthographic, lexical, and syntactic domains.

But this analysis is limited to language form. What about language function? Learning to understand a language involves parsing the speech stream into chunks which reliably mark meaning. The learner does not 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. Learners' attention to the evidence to which they are exposed soon demonstrates the recurring chunks of language (to use written examples, in English *e* follows *th* more often than *x* does, *the* is a common sequence, *the* [*space*] is frequent, *dog* follows *the* [*space*] more often than it does *book*, *how do you do?* occurs quite often, etc.). At some level of analysis, the patterns refer to meaning. It does not happen at the lower levels: *t* does not mean anything, nor does *th*, but *the* does, and *the dog* does better, and *how do you do?* does very well, thank you. In these cases the learner's goal is satisfied, and the fact that this chunk activates some meaning representations makes this sequence itself more salient in the input stream. When the learner comes upon these chunks again, they tend to stand out as units, and adjacent material is parsed accordingly (see Doughty, this volume, for a detailed discussion of this).

What is "meaning" in such an associative analysis? At its most concrete, it is the perceptual memories which underpin the conscious experience which a speaker wishes to describe and which, with luck, will be associated with sufficient strength in the hearer to activate a similar set of perceptual representations. These are the perceptual groundings from which abstract semantics emerge (Barsalou, 1999; Lakoff, 1987). Perceptual representations worth talking about are complex structural descriptions in their own right, with a qualifying hierarchical schematic structure (e.g., a room schema which nests within it a desk schema which in turn nests within it a drawer schema, and so on). These visuo-structural descriptions are also acquired by associative chunking mechanisms, operating in a neural system for representing the visual domain. When we describe the structural properties of objects and their interactions we do so from particular perspectives, attending to certain aspects and foregrounding them, sequencing events in particular orders, etc., and so we need procedures for spotlighting and sequencing perceptual memories with language. The most frequent and reliable cross-modal chunks, which structure regular associations between perception and language, are the constructions described in sections 2 and 3. Chunking, the bringing together of a set of already formed chunks in memory and welding them into a larger unit, is a basic associative learning process which can occur in and between all representational systems.

4.2 Generic learning mechanisms

Constructivists believe that generic, associative-learning mechanisms underpin all aspects of language acquisition. This is clearly a parsimonious assumption. But additionally, there are good reasons to be skeptical of theories of learning mechanisms specific to the domain of language, first because innate linguistic representations are neurologically implausible, and second because of the logical problem of how any such universals might come into play:

i Current theories of brain function, process and development, with their acknowledgement of plasticity and input-determined organization, do not readily allow for the inheritance of structures which might serve, for instance, as principles or parameters of UG (Elman et al., 1996; Quartz and Sejnowski, 1997).

ii Whether there are innate linguistic universals or not, there is still a logical problem of syntactic acquisition. Identifying the syntactic category of words must primarily be a matter of learning because the phonological strings associated with words of a language are clearly not universal. Once some identifications have been successfully made, it may be possible to use prior grammatical knowledge to facilitate further identifications. But the acquisition of relevant phrase structure grammar requires knowledge of syntactic word class in the first place. This is a classic bootstrapping problem (Redington and Chater, 1998). Thus, in early L1 acquisition there simply is no specialized working memory system involved in the assignment of syntactic structure. Instead there is a general-purpose phonological memory, a process which stores enough verbal information to permit the analysis of distributional regularities which eventually results in word-class information and phrase-structure constructions (see also Doughty, this volume).

4.3 Trees from string: hierarchy and structure dependence

I have emphasized how large stretches of spoken language are adequately described by finite-state grammars, as collocational streams where patterns flow into each other. As Bolinger (1976, p. 1) puts it, "[o]ur language does not expect us to build everything starting with lumber, nails and blueprint, but provides us with an incredibly large number of prefabs, which have the magical property of persisting even when we knock some of them apart and put them together in unpredictable ways." Nativelike competence is indexed as much by fluent idiomaticity as by grammatical creativity, and chunking is the mechanism of learning which underpins the acquisition and perception of these formulaic sequences.

But eventually language learners do become open-class, generative, and grammatically creative in their language productions. Their language operations become structure dependent. Any blueprint we might posit as a summary model of their abilities needs at least the power of phrase-structure grammars for successful analysis, and the resultant descriptions are hierarchical in structure. Rules of phrase-structure grammar such as (i) *Sentence* → *NP* + *VP*, (ii) *NP* → *D* + *N*, (iii) *VP* → *Verb* + *NP*, (iv) *N* → {*man*, *ball*}, etc., by "rewriting" yield labeled bracketed phrase-structures such as *Sentence* (*NP* + *VP* (*Verb* + *NP*)), which are more usually represented as tree diagrams that more clearly show the hierarchy. Can chunking help us in understanding the acquisition of these more abstract hierarchical constructions? Constructivists believe so. They view such rules for constituent analysis as top-down, a posteriori linguistic descriptions of a system that has emerged bottom-up from usage-based analysis of the strings themselves. Top-down or bottom-up, either way, bracketing is the link between hierarchical structure and string. Inductive accounts thus require a learning mechanism which provides bracketing, and that is exactly what chunking is.

We have seen how this works in the examples of slot-and-frame acquisition described in section 3.2. Once a child has chunks for *(Lulu)*, *(Teddy)*, *(The ball)*, *(Thomas the Tank)*, and the like, then the following utterances are parsed as bracketed, *(The ball's) (Gone)*, *(Teddy's) (Gone)*, *(Thomas the Tank's) (Gone)*, and subsequent analysis of these and other related exemplars results in the more abstract pattern *(X) (Gone)*, where, in subsequent utterances, the object is consistently put in preverbal position. But the slot-filler in this position is itself made up of chunks which also will be analyzed further, sometimes a bare noun, *(Salad) (Gone)*, *(Peter Pan) (Gone)*, sometimes a noun phrase, *((Funny) (Man)) (Gone)*; the branches of the hierarchy grow; and possible combinations are determined categorically rather than lexically. As Tomasello concludes in his account of epigenesis in his daughter Travis's early language acquisition:

> It is not until the child has produced or comprehended a number of sentences with a particular verb that she can construct a syntagmatic category of "*cutter*", for example. Not until she has done this with a number of verbs can she construct the more general syntagmatic category of agent or actor. Not until the child has constructed a number of sentences in which various words serve as various types of arguments for various predicates can she construct word classes such as noun or verb. Not until the child has constructed sentences with these more general categories can certain types of complex sentences be produced. (Tomasello, 1992, pp. 273–4; see also Tomasello, 2000, on "analogy-making" and "structure-combining").

Likewise, Bolander's (1989) analysis of the role of chunking in the acquisition of Swedish subject–verb inversion after a sentence-initial non-subject, described in section 3.4, provides a clear illustration of the role of chunking in the integration and differentiation of second language structure. In sum, although a very basic learning mechanism, chunking results in hierarchical representations and structure dependency. In constructivist usage-based accounts, phonology, lexis, and syntax develop hierarchically by repeated cycles of differentiation and integration of chunks of sequences (Studdert-Kennedy, 1991).

Language has no monopoly on hierarchical structure. Instead, because the formation of chunks, as stable intermediate structures, is the mechanism underlying the evolution and organization of many complex systems in biology, society, and physics, hierarchical structure and structure dependence are in fact a characteristic of the majority of complex systems which exist in nature (Simon, 1962). It is the norm that animal behavioral sequences, from the grooming of blowflies to the goal-directed behavior of cormorants, exhibit hierarchical structure, so much so that hierarchical organization has been proposed as a general principle for ethology (Dawkins, 1976). Human behavioral sequences are no different – slips of action exhibit structure dependence (Reason, 1979), just as do slips of the tongue (Fromkin, 1980).

4.4 Emergentism

The study of language demonstrates many complex and fascinating structural systematicities. Generative linguistics provides careful descriptions of these regularities that are necessary for a complete theory of language acquisition. But they are not sufficient because they do not explain how learners achieve the state of knowledge that can be described in this way. Indeed, many cognitive scientists believe that such linguistic descriptions are something very different from the mental representations that underpin performance, that there has, at times, been an unfortunate tendency to raise these "rules" from explanandum to explanans, and that, instead, the complexities of language are emergent phenomena (MacWhinney, 1999a, 1999b). Like many scientific descriptions, the regularities of generative grammar provide well-researched patterns in need of explanation. 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 rules play no causal role in shifting even a grain of sand or a molecule of water. It is the interaction of water and rocks which smooths the irregularities and grinds the pebbles and sand. 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. The interactions that constitute language are associations, billions of connections which co-exist within a neural system as organisms co-exist within an eco-system. And systematicities emerge as a result of their interactions and mutual constraints.

Bod (1998) describes experience-based, data-oriented parsing models of language which learn how to provide appropriate linguistic representations from an unlimited set of utterances by generalizing from examples of representations of previously occurring utterances. These probabilistic models operate by decomposing the given representations into fragments and recomposing those pieces to analyze new utterances. Bod (1998, ch. 5) shows that any systematic restriction of the fragments seems to jeopardize the statistical dependencies that are needed for predicting the appropriate structure of a sentence. This implies that the productive units of natural language cannot be defined in terms of a minimal set of rules, constraints, or principles, but rather need to be defined in terms of a large, redundant set of previously experienced structures with virtually no restriction on size or complexity – the behavior of the society of syntax is determined by the interactions and associations of all of its members. If communities are excised or if new individuals join, the ecology changes. This conclusion is supported in L1 acquisition by the findings of Bates and Goodman (1997) that syntactic proficiency is strongly correlated with vocabulary size. Total vocabulary at 20 months predicts grammatical status at 28 months, and grammar and vocabulary stay tightly coupled across the 16–30-month range.

The representational database for language is enormous. It is the history of our language input and the multifarious syntagmatic and paradigmatic associations that were forged in its processing. We not only have representations of chunks of language, but we also have knowledge of the likelihood of their occurrence, and the regularity with which they are associated with other corresponding mental events. N. Ellis (2002) reviews the evidence that, in the course of normal language comprehension and production, unconscious learning processes strengthen the activations of representations and associations that are used in language processing. These processes effectively count the relative frequencies of use of the language representations (at all levels), and they strengthen the weights of the associations between those that are contiguously activated. The result is that we are tuned to our language input. Thus, our language processing evidences regularity effects in the acquisition of orthographic, phonological, and morphological form. There are effects of bigram frequency in visual word identification, of phonotactic knowledge in speech segmentation, of spelling-to-sound correspondences in reading, and of cohort effects in spoken word recognition. There are effects of neighbors and the proportion of friends (items which share surface pattern cues and have the same interpretation) to enemies (items which share surface pattern but have different interpretations) in reading and spelling, morphology, and spoken word recognition (see Kroll and Sunderman, this volume). At higher levels, it can be shown that language comprehension is determined by the listeners' vast amount of statistical information about the behavior of lexical items in their language, and that, at least, for English, verbs provide some of the strongest constraints on the resolution of syntactic ambiguities. Comprehenders know the relative frequencies with which individual verbs appear in different tenses, in active vs. passive structures, and in intransitive vs. transitive structures, the typical kinds of subjects and objects that a verb takes, and many other such facts. Such information is acquired through experience with input that exhibits these distributional properties; it is not some idiosyncratic fact in the lexicon isolated from "core" grammatical information. Rather, it is relevant at all stages of lexical, syntactic, and discourse comprehension. Comprehenders tend to perceive the most probable syntactic and semantic analyses of a new utterance on the basis of frequencies of previously perceived utterance analyses. Language users tend to produce the most probable utterance for a given meaning on the basis of frequencies of utterance representations.

This research, the mainstay of psycholinguistics (Altman, 1997; Gernsbacher, 1994; Harley, 1995), shows that our language processing systems resonate to the frequencies of occurrence that are usual in language input. Most, if not all, of this tuning is the result of implicit rather than explicit learning (Doughty, this volume; N. Ellis, 1994a, 1994b; N. Ellis et al., 1999) – the on-line conscious experiences of language learning involve language understanding rather than counting. Fluent language users have had tens of thousands of hours on task. They have processed many millions of utterances involving tens of thousands of types presented as innumerable tokens. The evidence of language has ground on their perceptuo-motor and cognitive apparatus to result in complex competencies which can be described by formal theories of linguistics.

4.5 Probabilistic parsing: chunks and their frequencies in language processing

The use of this probabilistic knowledge, and the way it is combined for multiple cue sources, is fruitfully explored in the competition model (Bates and MacWhinney, 1987; MacWhinney, 1987, 1997a). This emphasizes lexical functionalism where syntactic patterns are controlled by lexical items. Lexical items provide cues to functional interpretations for sentence comprehension or production. Some cues are more reliable than others. The language learner's task is to work out which are the most valid predictors. The competition model is the paradigmatic example of constraint-satisfaction accounts of language processing.

Consider the particular cues that relate subject-marking forms to subject-related functions in the English sentence, *The learner chunks the words*. They are preverbal positioning (*learner* before *chunks*), verb agreement morphology (*chunks* agrees in number with *learner* rather than *words*), sentence initial positioning, and use of the article *the*. Case-marking languages, unlike English, might additionally include nominative and accusative cues in such sentences. The corresponding functional interpretations include actor, topicality, perspective, givenness, and definiteness. Competition model studies analyze a corpus of exemplar sentences which relate such cue combinations with their various functional interpretations, thus to determine the regularities of the ways in which a particular language expresses, for example, agency. They then demonstrate how well these probabilities determine (i) cue use when learners process that language, and (ii) cue acquisition – the ease of learning an inflection is determined by its cue validity, a function of how often an inflection occurs as a cue for a certain underlying function (cue availability) and how reliably it marks this function (cue reliability) (MacWhinney, 1997a).

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 interpretive hypothesis from an interaction of cues. It has been extensively tested to assess the cues, cue validity, and numerical cue strength order in many different languages. Finally, it goes a long way in predicting language transfer effects (MacWhinney, 1992). Recent competition model studies have simulated the natural language performance data using simple connectionist models relating lexical cues and functional interpretations for sentence comprehension or production. Section 5 illustrates one of these studies, Kempe and MacWhinney (1998), in detail.

The use of this probabilistic knowledge is also made clear in Natural Language Processing (NLP) analyses of sentence processing. Computational implementations of generative grammars which are large enough to cover a non-trivial subset of natural language assign to many sentences an extremely large number of alternative syntactic analyses, yet fluent humans perceive only one or two of these when faced with the same input. Such models may be judged successful if the defining criterion is that it describes the space of possible analyses that sentences may get, but the combinatorial explosion of syntactic analyses and corresponding semantic interpretations is very problematic if the criterion is rather to predict which analyses human comprehenders actually assign to natural language utterances (Bod, 1998; Church and Patil, 1982; Martin, Church, and Patil, 1981). The NLP community has moved to the use of stochastic grammars to overcome these problems (Bunt and Nijholt, 2000; Charniak, 1993). Examples include stochastic context-free grammar (Sampson, 1986), stochastic unification-based grammar (Briscoe, 1994), stochastic head-driven phrase-structure grammar (Brew, 1995), stochastic lexical-functional grammar (Kaplan, 1999), and data-oriented parsing (Bod, 1998).

Since the late 1960s, theories of grammar have increasingly put more syntax into the lexicon, and correspondingly less into rules. The result is that lexical specifications now include not only a listing of the particular constructions that the word can appear in, but also the relative likelihoods of their occurrence. In stochastic models of parsing using lexicalist grammars, these probabilities are used to determine the levels of activation of candidate lexical frames, with the network of candidate unification links being set up between those that are activated, the most probable being favored. This, combined with a unification-based parser based on competitive inhibition, where candidate links that are incompatible compete for inclusion in the final parse by sending each other inhibitory signals that reduce the competitor's attachment strength (Vosse and Kempen, 2000), promises a model of language processing that is both effective and psychologically plausible.

5 Connectionism

Constructivists believe that the complexity of language emerges from associative learning processes being exposed to a massive and complex environment. But belief in syntax or other language regularities as emergent phenomena, like belief in innate linguistic representations, is just a matter of trust unless there are clear process, algorithm, and hardware explanations. A detailed transition theory is needed. If language is not informationally encapsulated in its own module, if it is not privileged with its own special learning processes, then we must eventually show how generic learning mechanisms can result in complex and highly specific language representations. We need dynamic models of the acquisition of these representations and the emergence of structure. And we need processing models where the interpretation of particular utterances is the result of the mutual satisfaction of all of the available constraints. For these reasons, emergentists look to connectionism, since it provides a set of computational tools for exploring the conditions under which emergent properties arise.

Connectionism has various advantages for this purpose: 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 (for introductions see Elman et al., 1996; McClelland et al., 1986; McLeod, Plunkett, and Rolls, 1998; Plunkett, 1998; Plunkett and Elman, 1997; Redington and Chater, 1998; Seidenberg, 1997).

Connectionist approaches to language acquisition investigate the representations that can result when simple associative learning mechanisms are exposed to complex language evidence. 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. 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 and McClelland, 1987, p. 196)

Connectionist implementations are computer models consisting of many artificial neurons that are connected in parallel. Each neuron has an activation value associated with it, often being between 0 and 1. This is 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 set the corresponding word forms. The units in the artificial network are typically multiply interconnected by associations with variable strengths or weights. These connections permit the level of activity in any one unit to influence the level of activity in all of the units that it is connected to (e.g., spreading activation). 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 on another set of units. These learning algorithms are intended to reflect basic mechanisms of neuronal learning, they are generic in that they are used for a wide variety of learning problems, and they do not encapsulate any aspects of cognitive learning mechanisms. The cognitive learning emerges from these neuronal mechanisms being exposed to large amounts of experience in a particular problem space. Thus, over the course of many presentations of many different picture–name pairs in our example simulation of vocabulary acquisition, if the connection strengths have been set appropriately by the learning algorithm, then it may be possible for units representing the detection of particular pictures to cause the units that represent the appropriate lexical labels for that stimulus to become activated. The network could then be said to have learned the appropriate verbal output for that picture stimulus.

There are various standard architectures of the models, 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). An example is illustrated in figure 4.1 (see box 4.1 below). 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 and 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 any difference, or error, is propagated back to the hidden unit weights, and, in turn, to the input weights, in a way that reduces the error.

Some models use localist representations, where each separate unit might, for example, represent a word or picture detector. Other models use distributed representations where different words are represented by different patterns of activity over the same set of units (in the same way as different patterns of activation over the set of detectors in the retina encode the reflections of all of our different visual inputs). Localist representations are clearly more akin to the units of traditional symbolic computation and linguistic description. But not all of language processing is symbol manipulation. Many of the representations that conspire in the semantics from which language is inextricable, in vision, in motor action, in emotion, are analog representations. There are interesting interactions between all levels of representation (in reading, for example, from letter features through letters, syllables, morphemes, lexemes . . .). These different levels interact, and processing can be primed or facilitated by prior processing at subsymbolic or pre-categorical levels, thus demonstrating subsymbolic influences on language processing. These processes are readily modeled by distributed representations in connectionist models. But note well, non-exclusivity of symbolic representation is by no means a denial of symbolic processes in language. Frequency of chunk in the input, and regularity and consistency of associative mappings with other representational domains, result in the emergence of effectively localist, categorical units, especially, but by no means exclusively, at lexical grain. It may well be that symbolic representations are themselves an emergent phenomenon (Deacon, 1997; MacWhinney, 1997b).

Perhaps the most exciting aspect of connectionist models is that, in the course of processing particular exemplars, they often acquire knowledge of the underlying structural regularities in the whole problem space. They develop representations of categories and prototypes. They generalize from this knowledge. This is why they are so relevant to usage-based accounts of language acquisition. There are now many separate connectionist simulations of a wide range of linguistic phenomena including acquisition of morphology, phonological rules, novel word repetition, prosody, semantic structure, syntactic structure, etc. (see for reviews: Allen and Seidenberg, 1999; Christiansen and Chater, 2001; Christiansen et al., 1999; N. Ellis, 1998; Elman et al., 1996; Levy et al., 1995; MacWhinney and Leinbach, 1991; Plunkett, 1998; Redington and Chater, 1998). These simple, small-scale 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. To the considerable degree that the processes of learning L1 and L2 are the same, these L1 simulations are relevant to SLA. The problem, of course, is determining this degree and its limits. Because ground is still being broken for first language, there has been rather less connectionist work directly concerning SLA, although the following provide useful illustrations: Broeder and Plunkett (1994), N. Ellis (2001), N. Ellis and Schmidt (1998), Gasser (1990), Kempe and MacWhinney (1998), Sokolik and Smith (1992), Taraban and Kempe (1999). I will concentrate on just two of these for detailed illustration.

Box 4.1 describes a model of the acquisition of regular and irregular inflectional morphology. There have been a number of compelling connectionist models of the acquisition of morphology. Rumelhart and McClelland (1986) presented the first connectionist model of the acquisition of morphology, in this case in the quasi-regular domain of the English past tense. The model generated U-shaped learning for irregular forms, like children tending to overgeneralize to produce past tense forms like *runned* and *drinked*. 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 and 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 and Bever, 1988; Pinker and Prince, 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. These recent models have been successful in capturing the regularities that are present (i) in associating phonological form of lemma with phonological form of inflected form (Daugherty and Seidenberg, 1994; MacWhinney and Leinbach, 1991; Marchman, 1993; Plunkett and Marchman, 1991), and (ii) between referents (+past tense or +plural) and associated inflected perfect or plural forms (Cottrell and Plunkett, 1994; N. Ellis and Schmidt, 1998), 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, as well as acquiring a default case allowing generalization on "wug" tests, even in test cases of "minority default inflections," as are found in the German plural system (Hahn and Nakisa, 2000). Such findings 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 is needed for syntax falls quite naturally out of simple sequence

Box 4.1 Connectionist simulations of longitudinal learning logs (N. Ellis and Schmidt, 1998)

Ellis and Schmidt (E & S) investigated the acquisition of a quasi-regular morphosyntactic domain by experimentally recording learners' language productions throughout learning, and then simulating acquisition using connectionist models exposed to the same language input. In fluent speakers, variables like frequency have much more observable an effect on the production of irregular items than of regular ones. Such observations underpin theories which hold that there are dual mechanisms involved in morphological inflection: regular items are computed procedurally by a suffixation rule in a grammatical processing module, while irregular items are retrieved from an associative memory. E & S gathered longitudinal acquisition data under precisely known circumstances to show how this pattern emerges as a natural result of associative learning, and, therefore, that frequency by regularity interactions does not implicate hybrid theories of morphosyntax. E & S further demonstrated that a simple connectionist model, as an implementation of associative learning, provided with the same language evidence, accurately simulated human SLA in this domain.

Alternative theoretical accounts:
Can human morphological abilities be understood in terms of associative processes, or is it necessary to postulate rule-based symbol processing systems underlying these grammatical skills?

Prasada, Pinker, and Snyder (1990) showed that when fluent English speakers see verb stems on a screen and are required to produce the past tense form, they take significantly less time for irregular verbs with high past tense frequencies (like *went*) than for irregular verbs with low past tense frequencies (like *slung*), even when stem frequencies are equated. However, there is no effect on latency of past tense frequency with regular verbs whose past tense is generated by adding *-ed*. Since frequency generally affects latency of retrieval from associative memory systems, this lack of frequency effect on regular forms has been taken as evidence that there must be symbol-manipulating syntactic mechanisms for language. Pinker's (1991) conclusion is that the language system responsible for morphological inflection is a hybrid: regular verbs (*walk–walked*) are computed by a suffixation rule in a neural system for grammatical processing, while irregular verbs (*run–ran*) are retrieved from an associative memory.

Rumelhart and McClelland (1986) pioneered an alternative connectionist approach to language acquisition by showing 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. This original past tense model spawned a number of revised and improved connectionist models of different aspects of the acquisition of morphosyntax. According to such accounts, there are no "rules" of grammar. Instead, the systematicities of syntax emerge from the set of learned associations between language functions and base and past tense forms, with novel responses generated by "on-line" generalizations from stored exemplars.

Recording acquisition of a quasi-regular morphosyntactic system:
E & S argued that it is difficult to understand learning and development from observations like those of Prasada et al. (1990) of the final state, when we have no record of

the content of the learners' years of exposure to language or of the developmental course of their proficiencies. To understand learning, one must study learning.

E & S therefore recorded adult acquisition of second language morphology using an artificial language where frequency and regularity were factorially combined. Learners' accuracy and latency in producing artificial language names for single or multiple items was recorded after *each* exposure. Plurality was marked by a prefix: half of the items had a regular plural marker 'bu-' (e.g., *car* = 'garth,' *cars* = 'bugarth'), the remaining items had idiosyncratic affixes (e.g., *horse* = 'naig,' *horses* = 'zonaig'). Frequency was factorially crossed with regularity, with half of each set being presented five times more often.

The acquisition data for both accuracy and latency evidenced frequency effects for both regular and irregular forms early on in the acquisition process. However, as learning progresses, so the frequency effect for regular items diminishes, whilst it remains for irregular items. The results, illustrated in the left-hand lower panel of figure 4.1, thus converge on the end point described by Prasada et al. (1990), but they additionally show how this end point is reached – the convergence of the latencies for high- and low-frequency regular plural responses indexes the rate of acquisition of the schema for the regular form, and the attenuation of the frequency effect for regular items is a simple consequence of the power law of learning.

Connectionist modeling of acquisition:
E & S describe a simple connectionist model which is exposed to the same exemplars in the same order as the human subjects. The model, shown in the top panel of figure 4.1, had input nodes representing the different referents of the language and whether any particular stimulus was singular or plural. The output units represented the stem forms for the referents and the various affixes for marking plurality. The model learned to associate each input with its appropriate name, chunking appropriately each affix and stem. The model acquired some patterns more slowly than others. The simulations closely paralleled human learning (see the right-hand lower panel of figure 4.1), explaining 78 percent of the variance of the human correctness data. There are initially frequency effects on both the regular and irregular forms, but with increased exposure, so the frequency effect for regular forms is attenuated.

Further simulations demonstrated how varying the computational capacity of the model affects the rate of acquisition of default case, as indexed by successful performance on "wug" tests (*Q*.: Here is a wug, here is another, what have we got? *A*.: A "buwug."); the presence or absence of frequency effects for regular items; and ability to acquire irregular items. These findings illuminate the difficulties of children with specific language impairment and individual differences in L2 learner aptitude.

Conclusions:
The connectionist system duplicated the human "rule-like" behavior, yet there are no "rules" in a connectionist network. Rather, frequency–regularity interactions are a natural and necessary result of the associative ways in which connectionist models learn. These data serve to remind one that regular, rule-like behavior does not imply rule-generation. Instead regularity effects can stem from consistency: regular affixes are more habitual and frequent, since consistent items all involve pairings between plurality and the regular affix. Thus, regularity is frequency by another name. These data and simulations demonstrate that adult acquisition of these aspects of L2 morphology, at least, is tractable using simple associative learning principles.

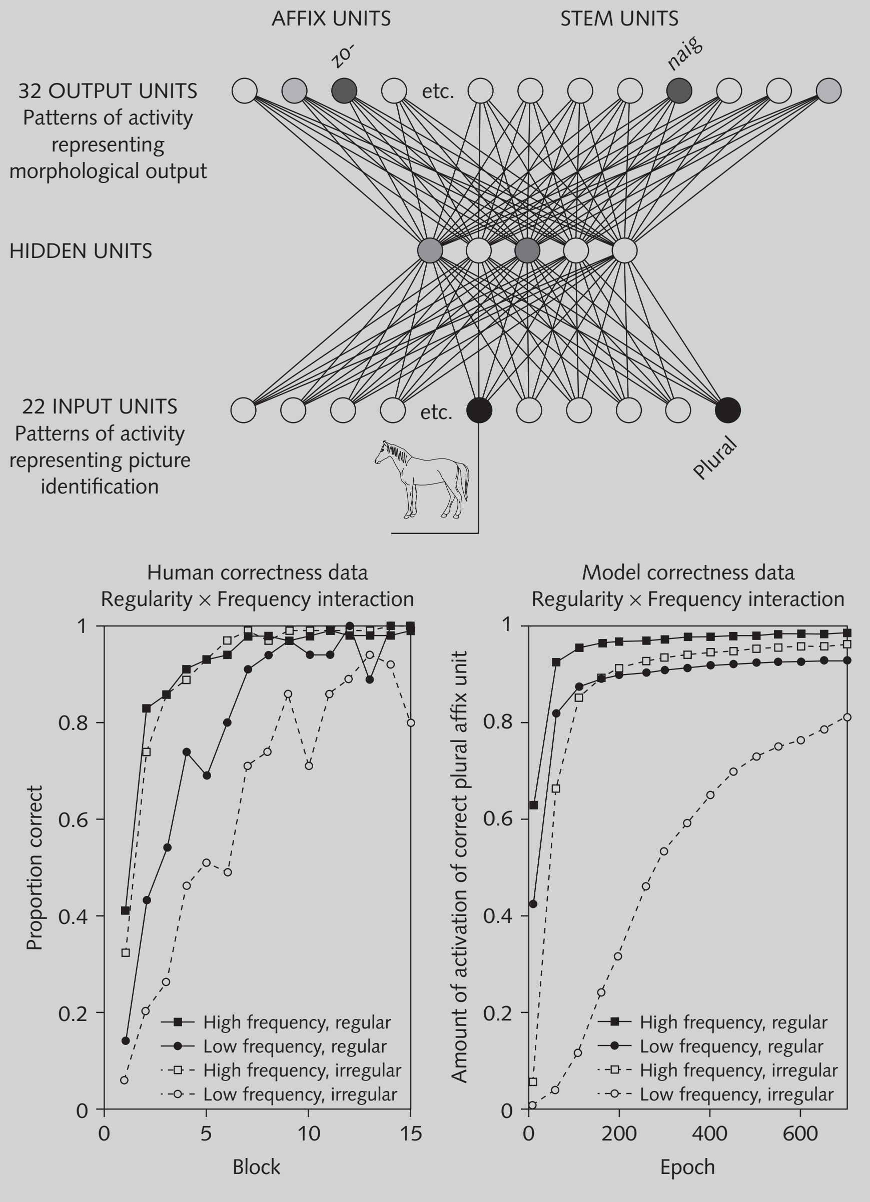

Figure 4.1 Human acquisition of high- and low-frequency, regular and irregular morphological inflections as a function of language exposure (lower left), a connectionist model for learning morphological inflection (top), and the acquisition functions of the model when exposed to the same pattern of language exemplars as the human learners (lower right).

Source: Adapted from Ellis and Schmidt (1998). Copyright 1998 by Psychology Press Ltd.

analysis and the patterns of association between patterns of sequences and patterns of referents.

The Ellis and Schmidt study in box 4.1 was selected for illustration because it clearly shows how this style of research strives to determine exactly what history of language exposure results in what learner competencies. Participants were taught an artificial second language in an experiment that measured their performance after each language experience so that their entire history of language input could be recorded. As shown in the detailed learning curves of figure 4.1, their resultant abilities in producing regular and irregular inflections of different frequencies of occurrence were assessed throughout learning. These results contradicted the findings of earlier studies which had restricted their observations to adult fluency. If we want to understand acquisition then we must study it directly. The study further demonstrated that a simple connectionist model, as an implementation of associative learning, when provided with the same relative frequencies of language evidence (something that was only possible because this history was determined in the experimental part of the study), accurately simulated human SLA in this domain.

The Kempe and MacWhinney study in box 4.2 again seeks to determine exactly what patterns are latent in learners' language input experience, but it assesses this in a different way. It illustrates the shared goals of connectionists and corpus linguists. Corpora of natural language are the only reliable sources of frequency-based data, and they provide the basis of a much more systematic approach to the analysis of language. For these reasons, we need large collections of representative language and the tools for analyzing these data. Corpus linguistics (Biber et al., 1998; McEnery and Wilson, 1996) bases its study of language on such examples of real-life performance data. Under normal circumstances, these natural language corpora provide the information that we need concerning the frequencies of different cues in language. However, Kempe and MacWhinney needed to estimate the language input to second language learners of German and Russian. In order to measure the validity of nominative and accusative cues in the two languages, they, therefore, analyzed a corpus of active transitive sentences from five textbooks widely used by learners of each language, and estimated the validity of these markers in the context of other surface cues such as word order, animacy of the nouns, and verb agreement. This showed that case marking in Russian is more complex than in German, but Russian case inflections are more reliable cues to sentence interpretation. Kempe and MacWhinney exploited the opposition of paradigm complexity and cue reliability in these two languages in order to contrast rule-based and associative theories of acquisition of morphology and to evaluate their predictions. Their connectionist model, as an implementation of associative learning and cue competition/constraint-satisfaction processing, was highly successful in predicting learners' relative acquisition rates.

Connectionist studies are important in that they directly show how language learning takes place through gradual strengthening of the associations between co-occurring elements of language, and how learning the distributional

Box 4.2 Connectionist learning from input corpus analysis (Kempe and MacWhinney, 1998)

Kempe and MacWhinney (K & M) investigated acquisition of the comprehension of morphological case marking by adult native speakers of English who were learning Russian or German as an L2. Their work compared acquisition of different languages using a fruitful combination of the methods of corpus analysis, psycholinguistic measurement of on-line performance, and connectionist simulations. Case marking in Russian is more complex than in German, but Russian case inflections are more reliable cues to sentence interpretation. K & M exploited the opposition of paradigm complexity and cue reliability in these two languages in order to contrast rule-based and associative theories of acquisition of morphology and to evaluate their predictions.

Alternative theoretical accounts:
Rule-based approaches to morphology view the learning of inflections as a process of discovering the grammatical dimensions underlying an inflectional paradigm (e.g., number, gender, person, case, or tense) through systematic hypothesis testing. According to such accounts, the more complex a paradigm, the longer it should take to learn.

Associative approaches to morphology view paradigms as epiphenomena that emerge from distributional characteristics of the language input. Learning takes place through gradual strengthening of the association between co-occurring elements of the language. According to these accounts, the ease of learning an inflection is determined by its cue validity, a function of how often an inflection occurs as a cue for a certain underlying function (cue availability) and how reliably it marks this function (cue reliability).

Quantifying paradigm complexity:
Complexity of paradigm in rule-based theories is determined by the number of dimensions, the number of cells, and the extent to which the cells in the paradigm are marked by unique inflections. Russian had more dimensions (animacy[2], number[2], gender[3], and case[6]) than German (number[2], gender[3], and case[4]). The crossings of these dimensions yields 72 cells in Russian, far more than the German system, which has only 24 cells. Average uniqueness of inflections is also lower in Russian. Russian is, thus, the more complex system by all three paradigm-based complexity measures. Rule-based accounts therefore predict that learners of German should do far better than learners of Russian in picking up case marking in the new language.

Quantifying cue validity using corpus analysis:
German and Russian differ in the extent to which they provide nominative and accusative markers as cues for agents and objects in sentences. In order to measure the validity of nominative and accusative cues in the two languages, K & M analyzed a corpus of active transitive sentences from five textbooks widely used by learners of each language, and estimated the validity of these markers in the context of other surface cues such as word order, animacy of the nouns, and verb agreement. Availability of a cue was computed as the total number of sentences in which a cue was present, divided by the total number of transitive sentences. Reliability of the cue

was the ratio of sentences in which the cue correctly signaled the agent, divided by the number of sentences in which the cue was present. Validity was the product of availability and reliability. These methods showed that the validity of case marking is much higher in Russian (.97) than in German (.56). Associative accounts therefore predict that learners of Russian, where case markers are readily available and reliable markers of thematic roles, should acquire case marking faster than learners of German.

Measuring acquisition as a function of exposure:
Learners of Russian and German were matched for language exposure on the basis of their knowledge of vocabulary, measured using a lexical decision task. Matching familiarity of learners of different languages is an accomplishment in itself (Kempe and MacWhinney, 1998).

As in other Competition Model studies, a computerized picture-choice task was used to probe the comprehension of L2 learners by varying the cues of case marking, noun configuration, and noun animacy, and determining the degree to which presence of a cue affected the accuracy and speed of learners' judgments of the agent of spoken sentences. As shown in figure 4.2a, the results demonstrated that learners of Russian used case marking at much earlier levels of language familiarity than learners of German.

Connectionist modeling of acquisition:
A small recurrent network (figure 4.2b) was used to model these cross-linguistic acquisition data. The four input units coded the following feature for each noun: animacy (±), nominative marking (±), accusative marking (±), and whether the input sentence is in English or in the L2. The input was restricted to the information for the first and second nouns of each sentence. In the output unit, an activation value of 1 was associated with the first noun as agent, 0 with second noun as agent. The network was first trained on a corpus of English transitive sentences where there was no case marking and the first noun was always the agent. Then it was trained on a representative sample of either Russian or German transitive sentences – essentially those same textbook sentences analyzed in the corpus analysis phase. The learning curves for this network's acquisition of Russian and German case marking are shown in figure 4.2c, where it is clear that, as in human learners, the network acquires the Russian system faster than the German one. The simulation data predicted 90 percent of the variance of the learner mean choice probabilities per pattern for Russian and 64 percent of the variance of the German choice data. It was also significantly successful in predicting on-line processing performance in terms of the human latency data.

Conclusions:
The match between simulation data and human performance supports the notion that adult SLA has a large associative component, and that the learning of inflectional morphology can be viewed as a gradual strengthening of the associations between co-occurring elements of language form and language function.

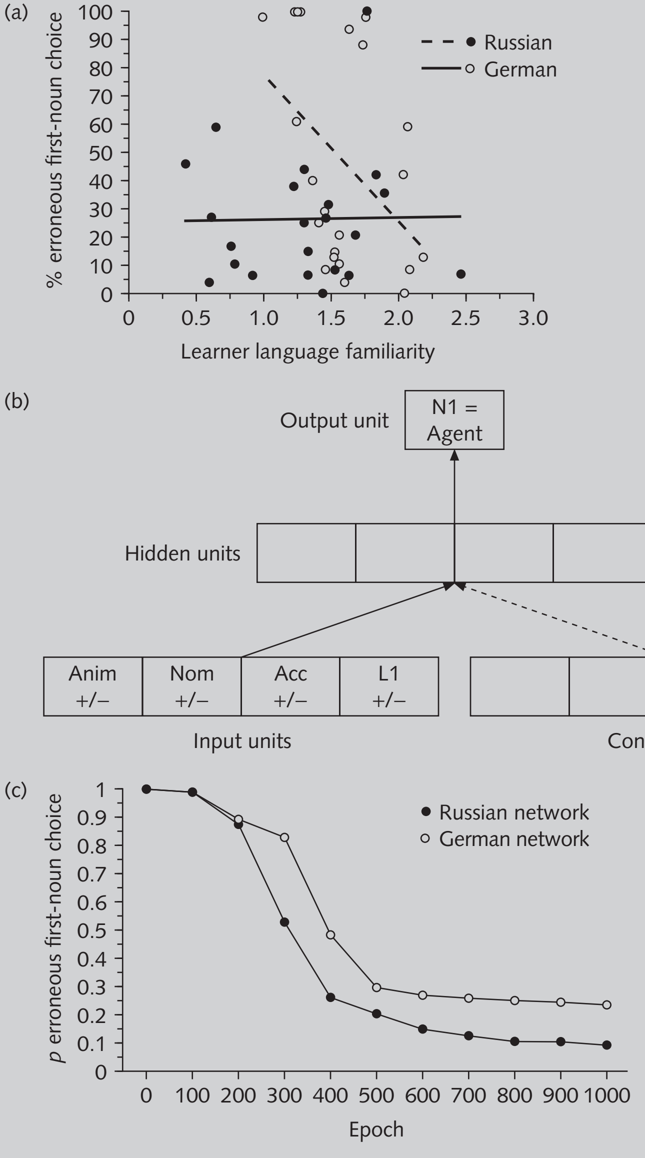

Figure 4.2 Acquisition data for Russian and German case marking, a connectionist model for learning case marking from representative language exposure, and the cross-linguistic acquisition functions for this model.

Source: Adapted from Kempe and MacWhinney (1998). Copyright 1998 by Cambridge University Press.

characteristics of the language input results in the emergence of rule-like, but not rule-governed, regularities. They are ways of looking at the effects of type and token frequency in the input and at how cue validity, a function of how often a surface form occurs as a cue for a certain underlying function (cue availability) and how reliably it marks this function (cue reliability), affects the emergence of regularities. Given that connectionist models have been used to understand various aspects of child language acquisition, the successful application of connectionism to SLA suggests that similar mechanisms operate in children and adults, and that language acquisition, in its essence, is the distributional analysis of form–function mappings in a neural network that attempts to satisfy simultaneously the constraints of all other constructions that are represented therein.

6 Current Limitations, Future Directions

"No discipline can concern itself in a productive way with the acquisition and utilization of a form of knowledge without being concerned with the nature of that system of knowledge" (Chomsky, 1977, p. 43). While this may be true, so is the emergentist counter that one cannot properly understand something without knowing how it came about. This brings us back to our opening stance. Constructivist views of language acquisition hold that simple learning mechanisms operating in and across human systems for perception, motor action, and cognition, while exposed to language data in a communicatively rich human social environment navigated by an organism eager to exploit the functionality of language, are sufficient to drive the emergence of complex language representations. The problem, though, is that just about every content word in this sentence is a research discipline in itself and that in our attempt to reunite speakers, syntax, and semantics, we have to be linguist, psychologist, physiologist, computational neuroscientist, and much more besides. At present there is far too little interdisciplinarity of research effort.

My sincere hope is that the material reviewed here convinces readers of the promise of these constructivist approaches to language acquisition. Clearly, there is much further to go. We need more-detailed longitudinal SLA corpora which will allow a proper tracking of the developmental sequences of constructions. We need more connectionist investigations of the emergence of linguistic structures from exemplars. Current connectionist models often use "test-tube" fragments of language and, thus, have low input representativeness. However good their contact with the data, more research is needed to explore the degrees to which these initial promising results can be scaled up to deal with the complexities of real language. Most connectionist work to date concerns L1 acquisition, and there needs to be far more work using this approach in SLA. If we wish to understand the emergence of language and we believe in the constraints of embodiment, then our models have to capture realistically the physical and psychological processes of perception, attention, and memory; the visual, motor, and other modalities which underpin conceptual knowledge; the limits of working memory; and all the rest.

There needs to be much more cross-talk between SLA and cognitive linguistic, child language, NLP, psycholinguistic, and connectionist research. The study of SLA must go forward within the broader remit of cognitive science. It is from these mutually supportive and naturally symbiotic interdisciplinary associations that eventually a more complete understanding of SLA will emerge.

ACKNOWLEDGMENT

Thanks to the Chester Language Development Group for feedback on a draft of this chapter and for pointing me at the right stuff in the first place.

REFERENCES

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

Aitchison, J. 1987: *Words in Mind*. Oxford: Blackwell.

Allen, J. and Seidenberg, M. S. 1999: The emergence of grammaticality in connectionist networks. In B. MacWhinney (ed.), *The Emergence of Language*. Mahwah, NJ: Lawrence Erlbaum Associates, 115–52.

Altman, G. T. M. 1997: *The Ascent of Babel: An Exploration of Language, Mind, and Understanding*. Oxford: Oxford University Press.

Barlow, M. and Kemmer, S. (eds) 2000: *Usage Based Models of Language*. Stanford, CA: CSLI Publications.

Barsalou, L. W. 1999: Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577–609.

Bates, E. and Goodman, J. 1997: On the inseparability of grammar and the lexicon: evidence from acquisition, aphasia and real-time processing. *Language and Cognitive Processes*, 12, 507–86.

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

Bates, E. and MacWhinney, B. 1987: Competition, variation, and language learning. In B. MacWhinney (ed.), *Mechanisms of Language Acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates, 157–93.

Becker, A. L. 1983: Toward a post-structuralist view of language learning: a short essay. *Language Learning*, 33, 217–20.

Berry, D. C. (ed.) 1997: *How implicit is implicit learning?* Oxford: Oxford University Press.

Biber, D. and Finegan, E. (eds) 1994: *Sociolinguistic Perspectives on Register*. New York: Oxford University Press.

Biber, D., Conrad, S., and Reppen, R. 1998: *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

Biber, B., Johansson, S., Leech, G., Conrad, S., and Finegan, E. 1999: *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education.
Bloom, P. 2000: *How Children Learn the Meanings of Words*. Cambridge, MA: MIT Press.
Bock, J. 1986: Syntactic persistence in language production. *Cognitive Psychology*, 18, 355–87.
Bod, R. 1998: *Beyond Grammar: An Experience-Based Theory of Language*. Stanford, CA: CSLI Publications.
Bolander, M. 1989: Prefabs, patterns, and rules in interaction? Formulaic speech in adult learners' L2 Swedish. In K. Hyltenstam and L. K. Obler (eds), *Bilingualism across the Lifespan: Aspects of Acquisition, Maturity, and Loss*. Cambridge: Cambridge University Press, 73–86.
Bolinger, D. 1976: Meaning and memory. *Forum Linguisticum*, 1, 1–14.
Bowerman, M. 1976: Semantic factors in the acquisition of rules for word use and sentence construction. In D. Morehead and A. Morehead (eds), *Normal and Deficient Child Language*. Baltimore, MD: University Park Press, 99–180.
Braine, M. D. S. 1976: The ontogeny of English phrase structure: the first phase. In C. A. Ferguson and D. I. Slobin (eds), *Studies of Child Language Development*. New York: Holt, Rinehart and Winston, 407–20.
Brazil, D. 1995: *A Grammar of Speech*. Oxford: Oxford University Press.
Brent, M. R. 1999: Speech segmentation and word discovery: a computational perspective. *Trends in Cognitive Sciences*, 3, 294–301.
Bresnan, J. 1999: Linguistic theory at the turn of the century. Plenary presentation, 12th World Congress of Applied Linguistics. Tokyo, August 1–6.
Brew, C. 1995: Stochastic HPSG. *Proceedings of the European chapter of ACL '95*. Dublin, 83–99.
Brewster, J. 1999: Exploring gendered talk: some effects of interactional style. In A. Tosi and C. Leung (eds), *Rethinking Language Education*. London: CILT, 196–212.
Briscoe, T. 1994: Prospects for practical parsing of unrestricted text: robust statistical parsing techniques. In N. Oostdijk and P. de Haan (eds), *Corpus-Based Research into Language*. Amsterdam: Rodopi, 67–95.
Broeder, P. and Plunkett, K. 1994: Connectionism and second language acquisition. In N. Ellis (ed.), *Implicit and Explicit Learning of Languages*. London: Academic Press, 421–54.
Bunt, H. and Nijholt, A. 2000: *Advances in Probabilistic and Other Parsing Technologies*. Dordrecht: Kluwer Academic.
Bybee, J. 1995: Regular morphology and the lexicon. *Language and Cognitive Processes*, 10, 425–55.
Bybee, J. and Thompson, S. 2000: Three frequency effects in syntax. *Berkeley Linguistics Society*, 23, 65–85.
Carter, R. 1998: *Vocabulary: Applied Linguistic Perspectives*. Second edition. London: Routledge.
Celce-Murcia, M. and Larsen-Freeman, D. 1983: *The Grammar Book*. London: Newbury House.
Charniak, E. 1993: *Statistical Language Learning*. Cambridge, MA: MIT Press.
Chomsky, N. 1977: *Knowledge and Responsibility*. New York: Pantheon Books.
Christiansen, M. H. and Chater, N. 2001: Connectionist psycholinguistics: capturing the empirical data. *Trends in Cognitive Sciences*, 5, 82–8.
Christiansen, M. H., Allen, J., and Seidenberg, M. S. 1998: Learning to segment speech using multiple cues: a connectionist model. *Language and Cognitive Processes*, 13, 221–68.
Christiansen, M. H., Chater, N., and Seidenberg, M. S. (eds) 1999: Connectionist models of human

language processing: progress and prospects. Special issue of *Cognitive Science*, 23, 415–634.

Church, K., and Patil, R. 1982: Coping with syntactic ambiguity or how to put the block in the box on the table. *Computational Linguistics*, 8, 139–49.

De Cock, S. 1998: A recurrent word combination approach to the study of formulae in the speech of native and non-native speakers of English. *International Journal of Corpus Linguistics*, 3, 59–80.

Collins Cobuild 1996: *Grammar Pattern 1: Verbs*. London: HarperCollins.

Corder, S. P. 1973: *Introducing Applied Linguistics*. Harmondsworth: Penguin.

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

Croft, W. 2001: *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

Croft, W. and Cruse, A. 1999: Cognitive linguistics. Ms. University of Manchester.

Daugherty, K. G., and 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*. Amsterdam: John Benjamins, 353–88.

Dawkins, R. 1976: Hierarchical organisation: a candidate principle for ethology. In P. P. G. Bateson and R. A. Hinde (eds), *Growing Points in Ethology*. Cambridge: Cambridge University Press, 7–54.

Deacon, T. W. 1997: *The Symbolic Species*. New York: Norton.

Dörnyei, Z. and Kormos, J. 1998: Problem-solving mechanisms in L2 communication: a psycholinguistic perspective. *Studies in Second Language Acquisition*, 20, 349–85.

Ellis, N. C. (ed.) 1994a: *Implicit and Explicit Learning of Languages*. London: Academic Press.

Ellis, N. C. 1994b: Implicit and explicit processes in language acquisition: an introduction. In N. Ellis (ed.), *Implicit and Explicit Learning of Languages*. London: Academic Press, 1–32.

Ellis, N. C. 1994c: Vocabulary acquisition: the implicit ins and outs of explicit cognitive mediation. In N. Ellis (ed.), *Implicit and Explicit Learning of Languages*. London: Academic Press, 211–82.

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

Ellis, N. C. 1998: Emergentism, connectionism and language learning. *Language Learning*, 48, 631–64.

Ellis, N. C. 2001: Memory for language. In P. Robinson (ed.), *Cognition and Second Language Instruction*. Cambridge: Cambridge University Press.

Ellis, N. C. 2002: Frequency effects in language acquisition: a review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24, 143–88.

Ellis, N. C. and Beaton, A. 1993a: Factors affecting the learning of foreign language vocabulary: imagery keyword mediators and phonological short-term memory. *Quarterly Journal of Experimental Psychology*, 46A, 533–58.

Ellis, N. C. and Beaton, A. 1993b: Psycholinguistic determinants of foreign language vocabulary learning. *Language Learning*, 43, 559–617.

Ellis, N. C. and 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. Hillsdale, NJ: Lawrence Erlbaum Associates, 53–83.
Ellis, N. C. and 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, 307–36.
Ellis, N. C., Lee, M. W., and Reber, A. R. 1999: Phonological working memory in artificial language acquisition. Ms. University of Wales, Bangor.
Ellis, R. 1984: Formulaic speech in early classroom second language development. In J. Handscome, R. Orem, and B. Taylor (eds), *On TESOL '83: The Question of Control*. Washington, DC: TESOL, 57–71.
Ellis, R. 1994: *The Study of Second Language Acquisition*. Oxford: Oxford University Press.
Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., and Plunkett, K. 1996: *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press.
Epstein, W. 1967: The influence of syntactical structure on learning. In N. J. Slamecka (ed.), *Human Learning and Memory: Selected Readings*. New York: Oxford University Press, 391–5.
Fillmore, C. J. and Kay, P. 1993: *Construction Grammar Coursebook*. Berkeley, CA: University of California Press, chapters 1–11 (Reading Materials for Ling. X20).
Fromkin, V. A. 1980: *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen and Hand*. New York: Academic Press.
Gasser, M. 1990: Connectionism and universals of second language acquisition. *Studies in Second Language Acquisition*, 12, 179–99.
Gernsbacher, M. A. (ed.) 1994: *Handbook of Psycholinguistics*. San Diego: Academic Press.
Givón, T. 1998: The functional approach to language. In M. Tomasello (ed.), *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure*. Mahwah, NJ: Lawrence Erlbaum Associates, 41–66.
Goldberg, A. E. 1995: *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
Goldberg, A. and Sethuraman, N. 1999: Learning argument structure generalizations. Ms. University of Illinois, Urbana-Champaign.
Granger, S. (ed.) 1998: *Learner English on Computer*. London: Longman.
Hahn, U. and Nakisa, R. C. 2000: German inflection: single or dual route? *Cognitive Psychology*, 41, 313–60.
Hakuta, K. 1974: Prefabricated patterns and the emergence of structure in second language acquisition. *Language Learning*, 24, 287–98.
Harley, T. A. 1995: *The Psychology of Language: From Data to Theory*. Hove: Lawrence Erlbaum Associates.
Hebb, D. O. 1961: Distinctive features of learning in the higher animal. In J. F. Delafresnaye (ed.), *Brain Mechanisms and Learning*. Oxford: Blackwell, 37–46.
Higgs, T. V. (ed.) 1984: *Teaching for Proficiency: The Organizing Principle*. Lincolnwood, IL: National Textbook Company.
Hoey, M. P. 1991: *Patterns of Lexis in Text*. Oxford: Oxford University Press.
Hopper, P. J. 1998: Emergent grammar. In M. Tomasello (ed.), *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure*. Mahwah, NJ: Lawrence Erlbaum Associates, 155–76.
Hudson, R. 1984: *Word Grammar*. Oxford: Blackwell.
Hudson, R. 1990: *English Word Grammar*. Oxford: Blackwell.
Hunston, S. and Francis, G. 1996: *Pattern grammar: A corpus driven approach to the*

lexical grammar of English. Amsterdam: John Benjamins.

James, W. 1890: *The Principles of Psychology*. New York: Holt.

Jurafsky, D. 1996: A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, 20, 137–94.

Kaplan, R. M. 1999: Computational and theoretical linguistics. Plenary presentation, 12th World Congress of Applied Linguistics. Tokyo, August 1–6.

Kempe, V. and MacWhinney, B. 1998: The acquisition of case-marking by adult learners of Russian and German. *Studies in Second Language Acquisition*, 20, 543–87.

Kosslyn, S. 1983: *Ghosts in the Mind's Machine: Creating and Using Images in the Brain*. New York: Norton.

Lachter, J. and 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. and Johnson, M. 1980: *Metaphors We Live By*. Chicago: University of Chicago Press.

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. 1997: Chaos/complexity science and second language acquisition. *Applied Linguistics*, 18, 141–65.

Leech, L. 2000: Grammars of spoken English: new outcomes of corpus-oriented research. *Language Learning*, 50, 675–724.

Levy, J. P., Bairaktaris, D., Bullinaria, J. A., and Cairns, P. (eds) 1995: *Connectionist Models of Memory and Language*. London: UCL Press.

Lewis, M. 1993: *The Lexical Approach: The State of ELT and the Way Forward*. Hove: English Language Teaching Publications.

Lieven, E. V. M., Pine, J. M., and Dresner Barnes, H. 1992: Individual differences in early vocabulary development: redefining the referential–expressive dimension. *Journal of Child Language*, 19, 287–310.

MacWhinney, B. (ed.) 1987: *Mechanisms of Language Acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

MacWhinney, B. 1992: Transfer and competition in second language learning. In R. J. Harris (ed.), *Cognitive Processing in Bilinguals*. Amsterdam: North-Holland, 371–90.

MacWhinney, B. 1997a: Second language acquisition and the competition model. In A. M. B. de Groot and J. F. Kroll (eds), *Tutorials in Bilingualism: Psycholinguistic Perspectives*. Hillsdale, NJ: Lawrence Erlbaum Associates, 113–44.

MacWhinney, B. 1997b: Lexical connectionism. In P. Broeder and J. Murre (eds), *Cognitive Approaches to Language Learning*. Cambridge, MA: MIT Press.

MacWhinney, B. (ed.) 1999a: *The Emergence of Language*. Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney, B. 1999b: The emergence of language from embodiment. In B. MacWhinney (ed.), *The Emergence of Language*. Mahwah, NJ: Lawrence Erlbaum Associates, 213–56.

MacWhinney, B. and Bates, E. (eds) 1989: *The Crosslinguistic Study of Sentence Processing*. New York: Cambridge University Press.

MacWhinney, B. and Leinbach, J. 1991: Implementations are not conceptualizations: revising the verb learning model. *Cognition*, 40, 121–57.

Manza, L. and Reber, A. S. 1997: Representing artificial grammars: transfer across stimulus forms and modalities. In D. C. Berry (ed.), *How Implicit is Implicit Learning?* Oxford: Oxford University Press, 73–106.

Maratsos, M. 1982: The child's construction of grammatical categories. In E. Wanner and L. R. Gleitman (eds), *Language Acquisition: The State of the Art*. Cambridge: Cambridge University Press, 240–66.

Marchman, V. A. 1993: Constraints on plasticity in a connectionist model of the English past tense. *Journal of Cognitive Neuroscience*, 5, 215–34.

Martin, W. K., Church, K., and Patil, R. 1981: *Preliminary Analysis of a Breadth-First Parsing Algorithm: Theoretical and Experimental Results*. Tech. Rep. MIT/LCS/TR 261. Cambridge, MA: MIT, Laboratory for Computer Science.

Mathews, R. C. and Roussel, L. G. 1997: Abstractness of implicit knowledge: a cognitive evolutionary perspective. In D. C. Berry (ed.), *How Implicit is Implicit Learning?* Oxford: Oxford University Press, 13–47.

McClelland, J. L. Rumelhart, D. E., and 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. and Wilson, A. 1996: *Corpus Linguistics*. Edinburgh: Edinburgh University Press.

McLeod, P., Plunkett, K., and Rolls, E. T. 1998: *Introduction to Connectionist Modelling of Cognitive Processes*. Oxford: Blackwell.

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

Miller, G. A. 1958: Free recall of redundant strings of letters. *Journal of Experimental Psychology*, 56, 485–91.

Miller, G. A. 1991: *The Science of Words*. New York: Scientific American Library.

Myles, F., Hooper, J., and Mitchell, R. 1998: Rote or rule? Exploring the role of formulaic language in classroom foreign language learning. *Language Learning*, 48, 323–64.

Myles, F., Mitchell, R., and Hooper, J. 1999: Interrogative chunks in French L2: a basis for creative construction. *Studies in Second Language Acquisition*, 21, 49–80.

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

Nattinger, J. R. 1980: A lexical phrase grammar for ESL. *TESOL Quarterly*, 14, 337–44.

Nattinger, J. R. and DeCarrico, J. S. 1992: *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.

Neisser, U. 1976: *Cognition and Reality: Principles and Implications of Cognitive Psychology*. San Francisco: Freeman.

Newell, A. 1990: *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Pawley, A. and Syder, F. H. 1983: Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In J. C. Richards and R. W. Schmidt (eds), *Language and Communication*. London: Longman, 191–225.

Pine, J. M. and Lieven, E. V. M. 1993: Reanalyzing rote-learned phrases: individual differences in the transition to multi-word speech. *Journal of Child Language*, 20, 551–71.

Pine, J. M. and Lieven, E. V. M. 1997: Slot and frame patterns in the development of the determiner category. *Applied Psycholinguistics*, 18, 123–38.

Pine, J. M., Lieven, E. V. M., and Rowland, C. F. 1998: Comparing different models of the development of the English verb category. *Linguistics*, 36, 807–30.

Pinker, S. 1991: Rules of language. *Science*, 253, 530–5.

Pinker, S. and 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: *Language and Cognitive Processes*, 13. Special issue on connectionist models of language.

Plunkett, K. and Elman, J. L. 1997: *Exercises in Rethinking Innateness*. Cambridge, MA: MIT Press.

Plunkett, K. and Marchman, V. 1991: U-shaped learning and frequency effects in a multi-layered perceptron: implications for child language acquisition. *Cognition*, 38, 3–102.

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

Prasada, S., Pinker, S., and Snyder, W. 1990: Some evidence that irregular forms are retrieved from memory but regular forms are rule-governed. Paper presented at the Thirty First Meeting of the Psychonomic Society. New Orleans, November.

Quartz, S. R. and Sejnowski, T. J. 1997: The neural basis of cognitive development: a constructivist manifesto. *Behavioral and Brain Sciences*, 20, 537–56.

Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. 1985: *A Comprehensive Grammar of the English Language*. London: Longman.

Reason, J. T. 1979: Actions not as planned: the price of automatisation. In G. Underwood and R. Stevens (eds), *Aspects of Consciousness. Vol. 1: Psychological Issues*. London: Academic Press, 67–89.

Reber, A. S. 1967: Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 77, 317–27.

Redington, M. and Chater, N. 1998: Connectionist and statistical approaches to language acquisition: a distributional perspective. *Language and Cognitive Processes*, 13, 129–92.

Rumelhart, D. and 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*. Cambridge, MA: MIT Press, 272–326.

Rumelhart, D. E. and 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 Associates, 195–248.

Saffran, J. R., Aslin, R. N., and Newport, E. L. 1996: Statistical learning by 8-month-old infants. *Science*, 274, 1926–8.

Sampson, G. 1986: A stochastic approach to parsing. In *Proceedings of COLING '86*. Bonn, 151–5.

Schmitt, N. 2000: *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press.

Seidenberg, M. S. 1997: Language acquisition and use: learning and applying probabilistic constraints. *Science*, 275, 1599–603.

Simon, H. A. 1962: The architecture of complexity. *Proceedings of the American Philosophical Society*, 106, 467–82. Reprinted in H. A. Simon (1969), *The Sciences of the Artificial*. Cambridge, MA: MIT Press.

Simon, H. A. 1969: *The Sciences of the Artificial*. Cambridge, MA: MIT Press.

Sinclair, J. 1991: *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Slobin, D. I. 1997: The origins of grammaticizable notions: beyond the individual mind. In D. I. Slobin (ed.), *The Crosslinguistic Study of Language Acquisition. Vol. 5*. Mahwah, NJ: Lawrence Erlbaum Associates, 265–323.

Sokolik, M. E. and Smith, M. 1992: Assignment of gender to French nouns in primary and second language acquisition: a connectionist model. *Second Language Research*, 8, 39–58.

Studdert-Kennedy, M. 1991: Language development from an evolutionary perspective. In N. A. Krasnegor, D. M. Rumbaugh, R. L. Schiefelbusch, and M. Studdert-Kennedy (eds), *Biological and Behavioral Determinants of Language Development*. Hillsdale, NJ: Lawrence Erlbaum Associates, 5–28.

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 and S. Thompson (eds), *Grammatical Constructions: Their Form and Meaning*. Oxford: Oxford University Press, 235–87.

Talmy, L. 1996b: Fictive motion in language and "ception." In P. Bloom, M. Peterson, L. Nadel, and M. Garrett (eds), *Language and Space*. Cambridge, MA: MIT Press, 211–75.

Tanenhaus, M. and Trueswell, J. 1995: Sentence comprehension. In J. L. Miller and P. D. Eimas (eds), *Speech, Language and Communication*. New York: Academic Press, 217–62.

Tannen, D. 1987: Repetition and variation as spoken formulaicity in conversation. *Language*, 63, 574–605.

Taraban, R. and Kempe, V. 1999: Gender processing in native (L1) and non-native (L2) Russian speakers. *Applied Psycholinguistics*, 20, 119–48.

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, 131–56.

Tomasello, M. (ed.) 1998a: *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure*. Mahwah, NJ: Lawrence Erlbaum Associates.

Tomasello, M. 1998b: The return of constructions: review of the book *Constructions: A Construction Grammar Approach to Argument Structure. Journal of Child Language*, 75, 431–47.

Tomasello, M. 2000: Do young children have adult syntactic competence? *Cognition*, 74, 209–53.

Tomasello, M. and Brooks, P. 1999: Early syntactic development: a construction grammar approach. In M. Barrett (ed.), *The Development of Language*. London: UCL Press, 116–90.

Towell, R. and Hawkins, R. 1994: *Approaches to Second Language Acquisition*. Clevedon: Multilingual Matters.

Ungerer, F. and Schmid, H. J. 1996: *An Introduction to Cognitive Linguistics*. Harlow: Addison Wesley Longman.

Vihman, M. 1982: Formulas in first and second language acquisition. In L. Obler and L. Menn (eds), *Exceptional Language and Linguistics*. San Diego: Academic Press, 261–84.

Vosse, T. and Kempen, G. 2000: Syntactic structure assembly in human parsing: a computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, 75, 105–43.

Weinert, R. 1994: Some effects of a foreign language classroom on the development of German negation. *Applied Linguistics*, 15, 76–101.

Weinert, R. 1995: Formulaic language in SLA: a review. *Applied Linguistics*, 16, 180–205.

Wong-Fillmore, L. 1976: *The second time around*. Doctoral dissertation. Stanford University.

Wray, A. 1992: *The Focussing Hypothesis*. Amsterdam: John Benjamins.