constructions productively requires that utterances be syntactically analyzed or parsed (Gass, 1988; Gregg, and Harrington, this volume) and that the learner eventually comes to 'know' (implicitly) that individual words are exemplars of lexical categories. The way in which learners acquire knowledge of lexical categories, constructions, and rules is a central issue in SLA, but it is being viewed here as a question related to the contrast between implicit and explicit learning, rather than to the contrast between attended vs unattended input.[5]

[5] There are at least five ways in which lexical categories and constructions could be established in a second language. They may be innate or transferred from the L1 (not learned from L2 input in either case). Alternatively, they may be learned from input based on an implicit, associative, inductive learning mechanism (N. Ellis, this volume). Or they may be learned explicitly, either through instruction or through active, conscious hypothesis testing. Bley-Vroman (1997) proposes that only the L1 and categories 'evidently present' in the input can be the source of such construction. What is evidently present or obvious from input clearly needs to be independently defined. For derivational and inflectional morphology, Bybee (1985) has argued that morphemes whose meanings are centrally related to the meanings of the stems to which they are attached are more obvious and will be acquired earlier than morphemes whose meanings are only peripherally related to that of the stem. For verbs – which deal mainly with events (actions, processes, and states) – the most important semantic distinctions are (in order): aspect, tense, mood, number, and person, and they are predicted to be acquired in that order. Cognitive linguistics constitutes a more general attempt to relate linguistic and cognitive categories and discuss the relationships of these to attention (see N. Ellis, this volume, for discussion).

2 *Memory for language*

Nick C. Ellis

Working memory

Consider the flow of your conscious experience. At any point in time you may choose to focus your attention on (i) the speech of your conversation-partner, (ii) a section of conversation which you've recreated in your mind's speech from information previously stored in long-term memory (LTM), (iii) a part of the visual scene before your eyes, (iv) a part of a visual scene which you have recreated in your mind's eye from information previously stored in LTM. Of course, you have potential access to a wider range of percepts and memories than these (Baddeley, 1997; Baddeley & Hitch, 1974; Gathercole & Baddeley, 1993).

Baddeley and Hitch (1974) identified three components of working memory (WM):

1. The *central executive*, or *supervisory attentional system (SAS)*, regulates information flow within WM, allocates attention to particular input modalities or LTM systems, activates or inhibits whole sequences of activities guided by schemata or scripts, and resolves potential conflicts between ongoing schema-controlled activities. The processing resources used by the central executive are limited in capacity, and the efficiency with which the central executive can fulfil a role depends upon the other demands that are simultaneously placed upon it. We have limited attentional resources (see Schmidt, and Skehan & Foster, this volume). The central executive is supplemented by two slave systems, each specialised for the short term memory (STM) and manipulation of material within a particular domain: the *phonological loop* holds verbally coded information; the *visuo-spatial sketchpad* deals with visual and spatial material.
2. The phonological loop itself comprises two components: a *phonological store* which represents material in a phonological

code and which decays over time, and a process of *articulatory rehearsal* where inner speech can be used to refresh the decaying representations in the phonological store in order to maintain memory items. If you imagine someone telling you a new phone number they wish you to ring for them: as you hear the information, it is first registered in the phonological store whence it will rapidly decay; to combat this your natural inclination may be to repeat the string to yourself in order to maintain it while dialling; once you stop this rehearsal and start talking on the phone, it is likely that the number will rapidly decay from your memory.

3. The visuo-spatial sketchpad is involved in generating images, temporarily maintaining them, and manipulating information with visual or spatial dimensions. If you try to describe the route from your kitchen to your car, you will place heavy demands on this system – accessing mental maps, perhaps rotating perspective, zooming in on particular areas, and orienting through this mental space.

These slave systems are of limited capacity. Articulatory rehearsal is a serial process and can only say one thing at a time – try saying the phone number and 'the, the, the . . . ' simultaneously. Decay from the phonological store in conjunction with a serial refresh process results in the capacity of the phonological loop for novel verbal material being approximately the amount of this material which can be articulated in about two seconds. The visuo-spatial sketchpad can only focus on one image at a time – you may well have found that you closed your eyes in the previous route-description exercise so that there was less competition from concurrent visual perception.

Perception, as input to WM, is automatically filtered and patterned by our existing LTM schema. Consider three examples: (1) As children learn about analogue clocks they closely attend to the features and relative positions of hands and numerals; when experienced adults consult their watch they are aware of *the time*, and have no immediate access to such lower-level perceptual information (Morton, 1967). (2) You do not consciously see the letter features on this page (unless you choose to) as you read this paragraph – you do not see the ascenders, the dots of the i's or the crosses of the t's, you don't even see the letters, you see the words, or groups of words (more likely still, you may not even be aware of the words, instead being conscious of their meaning). It was far from so as you learnt to read. (3) When looking up a new phone-number, if it contains 'chunks', you cannot fail to perceive them. Thus your STM for a patterned phone number (0800–123999) is much better than that for a more random sequence (4957–632512) even though both strings contain the same number of digits. Therefore, our model of WM must acknowledge the intimate connections and mutual influences of long-term phonological memory and the phonological loop, and of long-term visual memory and the visuo-spatial sketchpad. These interactions underpin the development of automaticity and fluency (see DeKeyser, this volume). Thousands of experiments have investigated the properties of these components of WM (Baddeley, 1986) and the theory is considerably more than as it is described here. Readers should note that in the present characterisation of WM, in emphasising the filtering of input to WM through LTM, I lean more towards interactive views where STM reflects the activated and attended subsets of LTM (Cowan, 1995), as does Doughty (this volume) than would be the natural inclination of Alan Baddeley and his associates. Also there are alternative models of memory which differ in focus, emphasis or content (for review see Baddeley, 1997; Schacter, 1991). But most of them, like the WM model, acknowledge different modalities of storage, separations between activated short-term and consolidated long term representations, and the role of attentional processes in learning and recollection. Thus the WM model will do as amodal view, and the simple summary architecture described in Figure 1 serves as a foundation for describing the role of memory in language acquisition.

The essence of the Working Memory Model is that we have specialist systems for perceiving and representing, both temporarily and in the long term, visual and auditory information,[1] along with a limited resource attentional system. Given the rich sophistication of language and its linguistic descriptions, this may seem a rather banal starting point for a description of the memory systems that underpin language acquisition. But it will do. It has to do. Because that's just about all there is.

Constructivist approaches to language acquisition

Constructivist views of language acquisition hold that it is primarily these systems that the child uses in bootstrapping their way into language. None the less, simple learning mechanisms operating in these systems as they are exposed to language data as part of a rich human

[1] The working memory model concentrates on the major perceptual modalities of vision and audition. But it acknowledges that the SAS also has access to other slave systems such as those for representing motor schema, kinaesthetics, tactile information, emotion, etc. As we will discuss in the forthcoming section concerning Cognitive Linguistics, these other modalities also have influence on language.

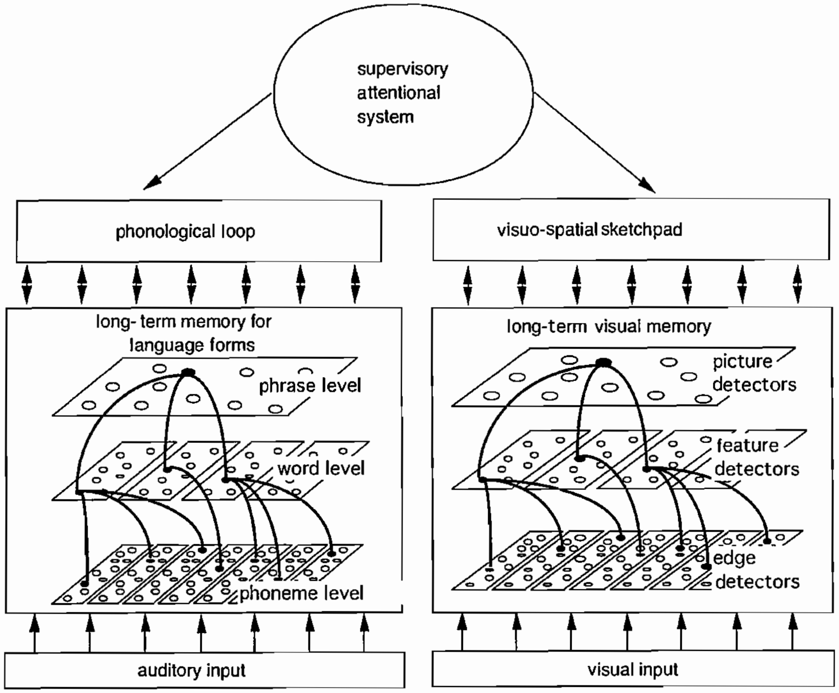

Figure 1. The model of working memory for language acquisition

environment by an organism eager to exploit the functionality of language is enough to drive the slow acquisition of complex language representations. Constructivists deny any innate linguistic universals.[2]

Fluent language users have had tens of thousands of hours on task. They have processed many millions of utterances involving tens of thousands of types presented as innumerable tokens. The evidence of language has ground on their perceptuo-motor and cognitive apparatus to result in complex competencies which can be described by formal theories of linguistics such as UG. It is more than a 'simplifying assumption' that language learning 'can be conceptualised as an instantaneous process' (Chomsky, 1976, pp. 14–15). It is an error which compounds into the fallacy of predeterminism. Language is *learned*. Various theories of language acquisition, including connectionist approaches (Levy, Bairaktaris, Bullinaria & Cairns, 1995; MacWhinney, this volume; McClelland, Rumelhart & Hinton, 1986), functional linguistics (Bates & MacWhinney, 1981 MacWhinney & Bates, 1989; see also Harrington and MacWhinney this volume), emergentist approaches (Elman, Bates, Johnson Karmiloff-Smith, Parisi, & Plunkett, 1996), and cognitive linguistic: (Lakoff, 1987; Langacker, 1987), believe that as the study of languag turns to consider ontogenetic acquisition processes, it favours a conclusion whereby the complexity of the final result stems from simpl learning processes applied, over extended periods of practice in th learner's lifespan, to the rich and complex problem space of languag evidence.

Apparent complexity may come more from the problem than fron the system which learns to solve it. Simon (1969) illustrated this b describing the path of an ant making its homeward journey on a pebbled beach. The path seems complicated. The ant probes, double back, circumnavigates and zigzags. But these actions are not deep an mysterious manifestations of intellectual power. Closer scrutiny reveals that the control decisions are both simple and few in numbe: An environment-driven problem solver often produces behaviour tha is complex only because a complex environment drives it. Languag learners have to solve the problem of language. Thus in this case, lik that of Simon's ant, it is all too easy to overestimate the degree of control, sophistication and innate neurological predisposition required i: its solution.

The complexity is in the language, not the learning process: 'Man universal or at least high-probability outcomes are so inevitable give: a certain "problem-space" that extensive genetic underwriting i unnecessary . . . Just as the conceptual components of language ma derive from cognitive *content*, so might the computational facts abou language stem from nonlinguistic *processing*, that is, from the multitude of competing and converging constraints imposed by perceptior production, and memory for linear forms in real time.' (Bates, 1984 188–190).

Constructivists are unhappy with nativist explanations simply because they may not be necessary – why posit predeterminism, lik magic, when simpler explanations might suffice (Sampson, 198(Tomasello, 1995)? They are additionally unhappy because the innateness hypothesis has no process explanation; our current theories c brain function, process and development do not readily allow for th inheritance of structures which might serve as principles or parameters of UG (Elman et al., 1996; Quartz & Sejnowski, 1997; cf. Gregg this volume). Without such a process explanation, innatist theorie are left with a 'and here a miracle occurs' step in their argumentatior

[2] Note that this claim concerns innate *linguistic* universals. It is a specific denial of representational innatism whilst acknowledging the influence of our particular endowment of perceptual transducers, inherited general neuronal architectural constraints, chronotopic constraints, computational capacity constraints and very general attentional biases (Elman et al., 1996; O'Grady, 1997).

Incompleteness of explanation is not a fatal flaw – Mendel was correct long before Crick and Watson provided a process explanation – but we do expect the gaps to be filled eventually, and current neuroscience makes implausible any assumptions about inherited parameter 'switches'. Finally, they are unhappy with arguments about learnability and the poverty of the stimulus. If the definition of language is revised to include statements of likelihood, Gold's (1967) [see editor's footnote] theorem cannot be proven, and the consequential limits to positive evidence-only learning are not established (Ellison, 1997; Quartz & Sejnowski, 1997).

Thus the constructivist view is that language learning results from general processes of human inductive reasoning being applied to the specific problem of language. There is no language acquisition device specifiable in terms of linguistic universals, principles and parameters, or language-specific learning mechanisms. Rather, language is cut of the same cloth as other cognitive *processes*, but it is special in terms of its cognitive *content*. Learners' language comes not directly from their genes, but rather from the structure of adult language, from the structure of their cognitive and social cognitive skills, and from the constraints on communication inherent in expressing non-linear cognition into the *linear* channel provided by the human vocal-auditory apparatus (Bates, Thal, & Marchman, 1991). Language is like the majority of complex systems that exist in nature and which empirically exhibit hierarchical structure (Simon, 1962). And like these other systems, its complexity emerges from simple developmental processes being exposed to a massive and complex environment. We are enlightened when we substitute a process description for a state description, when we describe development rather than the final state, when we focus on the language acquisition process (LAP) rather than language acquisition device (LAD).

Chunking as the LAP

The term *chunking* was coined by George Miller in his classical review of STM (Miller, 1956). It is the development of permanent sets of associative connections in long-term storage and is the process that underlies the attainment of automaticity and fluency in language. Newell (1990) argues that it is the overarching principle of human cognition: 'A chunk is a unit of memory organisation, formed by bringing together a set of already formed chunks in memory and welding them together into a larger unit. Chunking implies the ability to build up such structures recursively, thus leading to a hierarchical organisation of memory. Chunking appears to be a ubiquitous feature of human memory. Conceivably, it could form the basis for an equally ubiquitous law of practice.' (Newell, 1990: 7). *The power law of practice* describes the rate of acquisition of most skills (see DeKeyser, this volume): for example, Anderson (1982) showed that this function applies to cigar rolling, syllogistic reasoning, book writing, industrial production, reading inverted text, and lexical decision. The critical feature in this relationship is not just that performance, typically time, improves with practice, but that the relationship involves the power law in which the amount of improvement decreases as a function of increasing practice or frequency. For the case of language acquisition, Kirsner (1994) has shown that lexical recognition processes (both for speech perception and reading) and lexical production processes (articulation and writing) are governed by the relationship $T = PN^{-\alpha}$ where T is some measure of latency of response and N the number of trials of practice. Newell (1990; Newell & Rosenbloom, 1981) formally demonstrated that the following three assumptions of chunking as a learning mechanism could lead to the power law of practice. (1) People chunk at a constant rate: every time they get more experience, they build additional chunks. (2) Performance on the task is faster the more chunks that have been built that are relevant to the task. (3) The structure of the environment implies that higher-level chunks recur more rarely. Chunks describe environmental situations. The higher the chunk in the hierarchy, the more subpatterns it has; and the more subpatterns, the less chance there is of it being true of the current situation. For example, if one chunk is the trigram 'the' and another the bigram 'ir' then one will see each of these situations more frequently than the higher level chunk 'their'. These three assumptions interact as follows: the constant chunking rate and the assumption about speedup with chunking yields exponential learning. But as higher level chunks build up, they become less and less useful, because the situations in which they would help do not recur. Thus the learning slows down, being drawn out from an exponential towards a power law.

Although many lexical phenomena, because they involve relatively idiosyncratic memories, follow simply the power law of practice, other aspects of language acquisition, particularly those involving systemic generalisations, may seem at first sight to violate it. Thus there are classic exceptions like U-shaped learning curves and apparent backslidings in the acquisition of inflectional morphology and syntax more

Editor's footnote: Gold's theorem is a Mathematical Model of first language learnability describing the difficulties of inducing any target language from a 'hypothesis space', which is gradually narrowed on the basis of input to the learner, both in the form of 'text presentation' (positive evidence) and 'informant presentation' (negative evidence) (see Pinker, 1995: 147, and Saxton, 1997: 141, for discussion).

generally. Connectionist accounts of these phenomena hold that they result from the interactions in learning of many individual exemplars in a system, along with their components. The learning of each component follows a power law of practice, but systemic regularities may follow different non-linear growth curves which arise from interactions (both competitive and facilitatory) between the multiple components in the system and the combinations of their form-function mappings (Elman et al., 1996, Chapter 4). Thus, for example, a learner might learn *went* as the past tense of *go* initially as a lexical item. Its acquisition follows the power law. As time progresses more and more 'regular' past tense mappings are learned and the collaboration of these exemplars of form-past tense mapping pull the learner back to the *go-ed* form. There is not scope here to do more than introduce this matter – a more complete explanation for the particular case of regular and irregular morphosyntax can be found in Ellis and Schmidt (1998: 330–333) and we will return to the general issues in subsequent sections of this chapter which concern cognitive linguistics and the connectionist modelling.

So chunking, the bringing together a set of already formed chunks in memory and welding them together into a larger unit, is a basic associative learning process which can occur in all representational systems. Its operation in both phonological and visual LTM systems is acknowledged in Figure 1 by the arcs which gather recurring patterns of several units at a lower level together into one unit at the next-higher plane. The next section fills in some details of chunking in phonological memory as a key aspect of language acquisition.

Chunking in phonological memory

Language is sequential. Speech is a sequence of sounds. Writing is a sequence of symbols. Learning to understand a language involves parsing the speech stream into chunks which reliably mark meaning. The naturalistic learner doesn't care about linguists' analyses of language. They don't care about theories of grammar or whether words or morphemes are the atomic units of language. From a functional perspective, the role of language is to communicate meanings, and the learner wants to acquire the label-meaning relations.

This task is made more tractable by the patterns of language. Learners' attention to the evidence to which they are exposed soon demonstrates that there are recurring chunks of language. There are limited sets of sounds and of written alphabet. These units occur in more or less predictable sequences (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.). A key task for the learner is to discover these patterns within the sequence of language. At some level of analysis, the patterns refer to meaning. It doesn't happen at the lower levels: 't' doesn't 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.

The learner is searching for sequential patterns with reliable reference, and throughout this process, they are acquiring knowledge of the sequential aspects of language. From this perspective, language acquisition is essentially a sequence learning problem: the acquisition of word form, collocations, and grammatical class information all result from predominantly unconscious (or implicit) processes of analysis of sequence information (see forthcoming section on 'Sequences in learner talk'). Phonology, lexis, and syntax develop hierarchically by repeated cycles of differentiation and integration of chunks of sequences. With the benefit of hindsight, it comes as no surprise that language is acquired in this way. The formation of chunks, as stable intermediate structures, is the mechanism underlying the evolution and organisation of many complex hierarchical systems in biology, society, and physics (Dawkins, 1976; Simon, 1962).

Chunking and lexical acquisition

Learning lexical structure involves identifying the categorical units of speech perception, their particular sequences in particular words, and their general sequential probabilities in the language. Melton (1963) demonstrated for digit sequences like phone numbers that the more they are repeated in the phonological STM, the greater the LTM for these items, and in turn, the easier they are to repeat as sequences in STM. The same process of chunking allows us to bootstrap our way into lexis (Ellis, 1996a). Repetition of sequences in the phonological loop allows their consolidation in phonological LTM. Perception of frequent sequences, and the more frequent subsequences within them, allows their chunking in phonological LTM. The same cognitive system which does the LTM for phonological sequences does the perception of phonological sequences. Thus the tuning of phonological LTM to regular sequences allows more ready perception of input which contains regular sequences. Regular sequences are thus perceived as chunks and, as a result, individuals' phonological STM for regular sequences is greater than for irregular ones.

Experience of our environment leads to modification of our schemata, our schemata direct our exploration of the environment, our exploration samples the available information in the environment, and thus the cycle continues. The same systems which perceive language represent language. Thus the 'cycle of perception' (Neisser, 1976) is also the 'cycle of learning'; bottom–up and top–down processes are in constant interaction.

These processes result in sequences of language which are *potential* labels, but what about reference? In addition to implicit learning within input modalities, attentional focus in WM can result in the formation of cross-modal associations. The most basic principle of association is the Law of Contiguity: '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: 561). Nodes which are simultaneously or contiguously attended in WM tend to become associated in the long term. The implicit pattern-detection processes that occur *within* these modalities of representation entail that any such *cross-modal* associations typically occur between the highest chunked level of activated node. Thus, to extend Morton's (1967) example, the adult looking at their watch when post falls through their letter-box each morning learns an association that *mail-time* is 08.30, not one between envelopes and the big hand of their watch.

Similar processes occur within the language system. Consider for illustration two learners of differing levels of proficiency hearing the complaint 'I have a headache' while they observe salient visual input (Figure 2). The more proficient leaner, who knows the words *hEd* and *eik*, attends to the sequence of these *two* chunks along with the visual pattern. The less proficient learner, who has neither heard such words nor syllables before, has to attend to a much longer sequence of chunks: // h // E // d // ei // k //, and there is concomitantly greater chance of errors in sequencing (for example, Crystal (1987) describes a child who pronounced *blanket* as [bwati], [bati], [baki], and [batit] within a few hours of each other). Three occurrences for the more proficient learner might well result in three pairings of the image with // hEd // eik // and a concomitantly strengthened association between the visual and phonological representations. For the less proficient learner there might be much less commonality in the language sequences between trials, with // h // E // d // ei // k // on one trial, but // h // ei // k // on another and // E // d // h // ei // k // on a third (Treiman & Danis, 1988). No strong cross-modal association between the attended unit in the visual module and a common representation in the language module

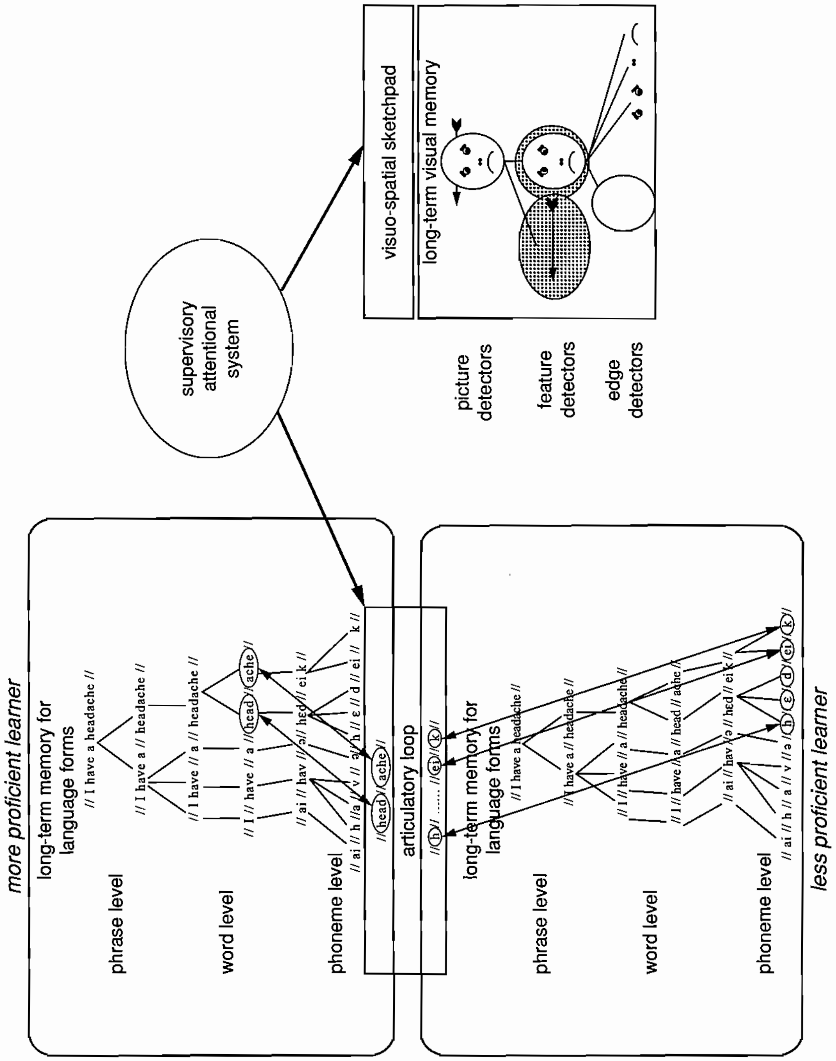

Figure 2. Cross-modal associations for less and more proficient learners hearing the term

can result. The more the units of language come as packaged wholes, the greater the possibility of attentional focus and resultant association.

The more any novel word, be it L1 or L2, is repeated in phonological WM, the more its regularities and chunks are abstracted, and the more accurately and readily these can be called to WM, either for accurate pronunciation as articulatory output or as labels for association with other representations. It is from these potential associations with other representations that interesting properties of language occur. Links with conceptual representations underlie reference and grounded semantics. Links with frequent local collocations underlie syntax and idiomatic meaning. Links with local and more distant lexical neighbours underlie lexical semantics. Links between L2 and simultaneously active L1 representations underlie translation and language transfer effects. These simple associations amass over the learner's language-input history into a web of multimodal connections which represent the complexities of language.

Chunking and idiom acquisition

The cycle of learning that underpins vocabulary acquisition operates at all levels of language. Language reception and production are mediated by learner's representations of chunks of language: 'Suppose that, instead of shaping discourse according to rules, one really pulls old language from memory (particularly old language, with all its words in and everything), and then reshapes it to the current context: "context shaping", as Bateson puts it, "is just another term for grammar"' (Becker, 1983: 218).

As we analyse language performance, so the underlying chunks become readily apparent. Sinclair (1991), as a result of his experience directing the Cobuild project, the largest lexicographic analysis of the English language to date, proposed *the principle of idiom* – 'a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments. To some extent this may reflect the recurrence of similar situations in human affairs; it may illustrate a natural tendency to economy of effort; or it may be motivated in part by the exigencies of real-time conversation. However it arises, it has been relegated to an inferior position in most current linguistics, because it does not fit the open-choice model' (Sinclair, 1991: 110). Rather than it being a rather minor feature, compared with grammar, Sinclair suggests that for normal texts, the first mode of analysis to be applied is the idiom principle, since most of text is interpretable by this

principle. Comparisons of written and spoken corpora demonstrate that collocations are even more frequent in spoken language (Butler, 1995). Collocations and stock phrases are viewed with just the same importance in SLA and FL research where they are known as holophrases (Corder, 1973), prefabricated routines and patterns (Hakuta, 1974), formulaic speech (Wong-Fillmore, 1976), memorised sentences and lexicalized stems (Pawley & Syder, 1983), lexical phrases (Nattinger & DeCarrico, 1992), or formulas (R. Ellis, 1994a).

Pawley and Syder (1983) give good reason to believe that much of language is in fact closed-class. They provide two sources of evidence: native-like selection and native-like fluency. Native speakers do *not* exercise the creative potential of syntactic rules of a generative grammar (Chomsky, 1965) to anything like their full extent. Indeed if they did, they would not be accepted as exhibiting native-like control of the language. While such expressions as (a) 'I wish to be wedded to you', (b) 'Your marrying me is desired by me', and (c) 'My becoming your spouse is what I want', demonstrate impeccable grammatical skill, they are unidiomatic, odd, foreignisms when compared with the more ordinary and familiar (d) 'I want to marry you'. Thus native-like selection is not a matter of syntactic rule alone. Speaking natively is speaking idiomatically, using frequent and familiar collocations, and learners thus have to acquire these familiar word sequences. That native speakers have done so is demonstrated not only by the frequency of these collocations in the language, but also by the fact that conversational speech is broken into fluent units of complete grammatical clauses of four to ten words, uttered at or faster than normal rates of articulation. A high proportion of these clauses, particularly of the longer ones, are entirely familiar memorized clauses and clause sequences which are the normal building-blocks of fluent spoken discourse (and at the same time provide models for the creation of (partly) new sequences which are memorable and in their turn enter the stock of familiar usages – for example 'I'm sorry to keep you waiting', 'Mr. Brown is so sorry to have kept you waiting', etc. can allow the creation of a lexicalised sentence stem 'NP be-*tense* sorry to keep-*tense* you waiting'). '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 LTM; by doing this he minimises 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' (Pawley & Syder, 1983: 192).

Learners don't care about the units of language as long as they map onto accessible meanings (Peters, 1983) and language learning involves learning sequences of words (frequent collocations, phrases, and idioms) as much as it does sequences within words. For present purposes, such collocations can simply be viewed as big words – the role of WM in learning such structures is the same as for words. It is a somewhat more difficult task to the degree that these utterances are longer than words and so involve more phonological units to be sequenced. It is a somewhat less difficult task to the degree that the component parts cluster into larger chunks of frequently-encountered (at least for learners with more language experience) sequences comprising morphemes, words, or shorter collocations themselves (e.g., 'I've a _'). Despite these qualifications the principle remains the same – just as repetition aids the consolidation of vocabulary, so it does the long-term acquisition of phrases (Ellis & Sinclair, 1996).[3]

Chunking and creativity

There are two major aspects of language creativity. One is the ability to express ideas in a novel grammatical surface form (grammatical creativity). The other is the ability to express novel ideas (conceptual/semantic creativity). In both cases, the creative act involves the original and interesting combination of two or more pre-existing ideas or representations. It is the combination of pre-existing chunks. For interesting creative behaviour to occur, there must be a substantial knowledge base so that there is something to and from which transfer of information can occur. Thus a high level of prior knowledge acquisition is necessary. Quantifying this is difficult, but the 10-year rule has been suggested as an estimate of time-on-task needed for expertise, with, for example, 50,000 'kitchen facts' needed for an expert cook, and 50,000 board positions for a grand master at chess (Norman, 1980; Simon, 1980). There is some exploitable mileage in the observations (a) that the ball-park estimate of the size of vocabulary of an average college student is also around 50,000 words, and (b) that vocabulary development is a strong predictor of subsequent achievements in morphology and syntax (Marchman & Bates, 1994).

Memory chunks (schema, scripts, frames, stereotypes, etc.) lie at the core of creativity in all domains of cognition.[4] In drafting this chapter I am building structures from my pre-existing knowledge: sometimes they are low level clichés ('at the core of' in the previous sentence); sometimes they are high-level chunks (the earlier large quotation from Sinclair); sometimes they are surface form representations; sometimes they are conceptual ones (whole theories, like that of Neisser). Sometimes, 'thinks', like things, just click into place. Like pieces of a jigsaw, they have their own structure. My representations of my science, like those of my language, have evolved hierarchically by repeated cycles of differentiation and integration – identifying the smaller chunks and building up the larger ones. My available representations are tuned by the content and frequency of my lifespan intake. The intelligibility of this account for you, the reader, is determined by your system of constructs, which reflects your experience. The present account will thus sound more foreign to some linguist readers than to some psychologists. The point about the idiom principle is that maximally rapid intelligibility is afforded by the use of frequent, pre-existing chunks in the parole.

[3] A natural question might then be, 'Why then, in L1 and L2 acquisition, doesn't frequency in the input correlate in every case with acquisition orders for, say, morphemes?' The initial response to this is that frequency may not be the only explanation, but it takes us a large part of the way. Larsen-Freeman (1976) was the first to propose that the common acquisition order of English morphemes to which ESL learners, despite their different ages and language backgrounds, adhere is a function of the frequency of occurrence of these morphemes in adult native-speaker speech. But we must then qualify this answer as follows: for morpheme acquisition orders we are talking about the acquisition of regularities in *systems of form-function mapping* and thus there is something that overrides the mere effect of frequency of exposure to a particular form, and that is the effects of frequency of 'friends' (those exemplars which conspire in this form-function mapping) and 'enemies' (those which use different mappings). It is the summing of these competitions which result in the cue reliabilities, consistencies and validities which underpin relative learnabilities. Ellis and Schmidt (1998) and DeKeyser (1997) show that the power law applies to the acquisition of morphosyntax and that it is this acquisition function which underlies interactions of regularity and frequency in this domain. As described by MacWhinney, this volume, the Competition Model sees all of language acquisition as the process of acquiring from language input the particular cues which relate phonological forms and conceptual meanings or communicative intentions, and the determination of the frequencies, reliabilities and validities of these cues. This information then serves as the knowledge base for sentence production and comprehension in lexicalist constraint-satisfaction theories which hold that sentence processing is the simultaneous satisfaction of the multiple probabilistic constraints afforded by the cues present in each particular sentence (MacDonald, Pearlmutter & Seidenberg, 1994; MacWhinney & Bates, 1989).

[4] Besides schema, scripts, frames and stereotypes, there is one other general class of abstractions from input data, namely syntax. Does syntax also lie at the core of linguistic creativity? Well, it depends on what is meant by 'syntax'. The more syntax is conceived as systems of form-function mapping, as in systemic functional or lexical functional grammars, the more I would agree that syntax contributes, along with any of the other types of abstraction that are relevant to the communicative task in hand. But the more a conception of syntax isolates abstractions which relate signifiers, but divorces them from the signifieds, from semantics, the functions of language, and the other social, biological, experiential and cognitive aspects of the humankind who invented and used language ever since, the less I believe that such syntactic descriptions have any causal creative role (Ellis, 1998). Even so, there would still be some influence. Because humans, in their everyday making sense of the world, also abstract patterns which relate lexical forms. And thus, as shown by Epstein (1967), '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'!

Individual differences in phonological STM predict language aptitude

One advantage of this account is that individual differences in phonological STM ability explain individual differences in language learning aptitude. Individuals differ in their ability to repeat phonological sequences (this is known as phonological STM span). In part this can result from constitutional factors – some individuals are born with better phonological abilities than others. Language impaired and dyslexic individuals have poorer phonological STM spans (Ellis, 1990, 1996a; Ellis & Large, 1987; Gathercole & Baddeley, 1993). Other reasons for having a poor phonological STM span for a particular type of material are due to the learner not having any prior experience of the content of the to-be-recalled material, and thus having no LTM representations which aid the chunking of the incoming material. Examples of these interactions in the domain of language include the effects of: long-term lexical knowledge on STM for words (Brown & Hulme, 1992), long-term phonological knowledge on STM for non- and foreign-language words (Ellis & Beaton, 1993b; Gathercole & Baddeley, 1993; Treiman & Danis, 1988;), long-term grammatical knowledge on STM for phrases (Epstein, 1967), and long-term semantic knowledge on STM for word strings (Cook, 1979). This is also the reason why elicited imitation tests serve so well as measures of second-language competence (Bley-Vroman & Chaudron, 1994; Lado, 1965).

★ See Ellis 1996a

This ability to repeat verbal sequences (for example, new phone numbers or non-words like 'sloppendash') immediately after hearing them, is a good predictor of a learner's facility to acquire vocabulary and syntax in first, second, and foreign language learning. Ellis (1996a) reviews a wide range of evidence for this: (i) phonological STM span predicts vocabulary acquisition in L1 and L2, (ii) interfering with phonological STM by means of articulatory suppression disrupts vocabulary learning, (iii) repetition and productive rehearsal of novel words promotes their long term consolidation and retention, (iv) phonological STM predicts syntax acquisition in L1 and L2, (v) phonological rehearsal of L2 utterances results in superior performance in receptive skills in terms of learning to comprehend and translate L2 words and phrases, explicit metalinguistic knowledge of the detailed content of grammatical regularities, acquisition of the L2 forms of words and phrases, accuracy in L2 pronunciation, and grammatical fluency and accuracy (Ellis & Sinclair, 1996). Thus phonological sensitivity, chunking and segmentation are key components of language learning aptitude (see Sawyer & Ranta, this volume).

Let us consider L2 vocabulary learning as a detailed example. The novice L2 learner comes to the task with a capacity for repeating native words. The degree to which the relevant skills and knowledge are transferable to immediate L2 word repetition depends on the degree to which the phonotactic patterns in the L2 approximate to those o the native language (Ellis & Beaton, 1993b). Thus long-term knowledge affects phonological STM. The reverse is also true: repetitio of L2 forms promotes long-term retention (Ellis & Beaton, 1993a Ellis & Sinclair, 1996). As learners practise hearing and producin L2 words, so they automatically and implicitly acquire knowledge o the statistical frequencies and sequential probabilities of the phonotactics of the L2. In turn, as they begin to abstract knowledge of L regularities, they become more proficient at short-term repetition o novel L2 words. And so L2 vocabulary learning lifts itself up by it bootstraps. Although learners need not be aware of the *processes* o such pattern extraction, they will later be aware of the *product* o these processes since on the next time they experience that pattern i is the patterned chunk that they will be aware of, not the individua components.

Working out how words work (i): distributional analysis of memorized collocations to derive word-class information and morphology

But of course, language learners do not simply recombine surfac elements of previous input in their novel productions. There i creativity in the transition to L1 competence and in L2 interlanguag which demonstrates the abstraction of systematicity from prior input a systematicity which is certainly rule-like if not rule-governed. Th following examples from my son Gabe's productions are utterance which he has certainly never heard before: 'Give it to he' (2 years 6 months), 'That's good of I' (2:11), 'We're coming soonly', 'Rain i falling straightly', 'I like you because you don't shout at me oftenly (3:0–3:2), 'Dad had a lot of them, but I had a lotter' (3:3). Such errors are not limited to learners – even Gabe's dad has been prone t overcorrect to produce such cringe-worthy utterances as 'Thank yo kindlily'. Learners abstract structural regularities from previously experienced utterances which share structural and functional similarit (see section 'Sequences in learner talk').

As we analyse word sequence chunks, so we discover that they hav characteristic structural types. Linguists call these regularities grammar. And if we take a bottom–up approach, and simply describe th distributional properties of words and morphemes in chunks, so we discover that something very close to traditional grammatical word-class and inflectional morphological information emerges. I will describe four demonstrations for illustration: (1) Kiss (1973) which was the first analysis of this type. (2) Elman (1990) which used a neurologically plausible connectionist model to analyze the sequential dependencies in miniature language systems. This work is important because it demonstrates how one single process of sequential acquisition can give rise to useful representations at all levels from phonetics, through phonotactics and lexis, and up to syntax. (3) The Cobuild Project (Sinclair, 1991), a corpus linguistic analysis of English-as-it-is-used to determine the collocational streams that are frequent in the input data. (4) Recent connectionist work concerning acquisition of regular and irregular morphology.

Kiss

Kiss (1973) provided the first computational model of the acquisition of grammatical word class from accumulating evidence of word distributions. An associative learning program was exposed to an input corpus of 15,000 words gathered from tape recordings of seven Scottish middle class mothers talking to their children who were between one and three years of age. The program read the corpus and established associative links between the words and their contexts (here defined as their *immediate successor*). Thus, for example, the program counted that *the* was followed by *house* 4.1% of the time, by *horse* 3.4%, by *same* 1%, by *put* never, etc., that *a* was connected to *horse* 4.2%, to *house* 2.9%, to *put* never, etc. Next a classification learning program analyzed this information to produce connections between word representations which had strengths determined by the degree of similarity between the words in terms of the degree to which they tended to occur together after a common predecessor (i.e., the degree of similarity based on their 'left-contexts'). This information formed a level of representation which was a network of word similarities. Finally the classification program analyzed this similarity information to produce a third network which clustered them into groups of similar words. The clusters that arose were as follows: (*hen sheep pig farmer cow house horse*) (*can are do think see*) (*little big nice*) (*this he that it*) (*a the*) (*you I*). It seemed that these processes discovered word classes which were nounlike, verblike, adjectivelike, articlelike, pronounlike, etc. Thus the third level of representation, which arises from simple analysis of word distributional properties, can be said to be that of word class. Kiss argues that in this way language learners can bootstrap their way into discovering word classes. More recent, and much larger, demonstrations show that such bootstrapping results from a variety of analysis methods including statistical, recurrent neural network, or self-organizing map models (Charniak, 1993; Elman, 1990; Finch & Chater, 1994; Sampson, 1987).

Elman

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

Elman's network had 5 input units (binary coding one letter at a time – thus *m* was represented as 01101, *a* as 00001, *e* as 00101, etc.), 20 hidden units, 5 output units (again coding individual letters), and 20 context units. It was fed one letter (or phoneme) at a time and had to predict the next letter in the sequence (akin to asking you to predict the next letter at the following example choice points marked!: Once upo!n a! time). It was trained on 200 sentences of varying length from 4 to 9 words. There was no word or sentence boundary information; thus part of the stream was:

Manyyearsagoaboyandgirllivedbytheseatheyplayedhappily . . .

The error patterns for a network trained on this task illustrate that the network abstracted a lot of information about the structure of English. If the error is high, it means that the network had trouble predicting this letter. Error tends to be high at the beginning of a word and decrease until the word boundary is reached. Thus the model is learning the word units of the language (compare these abilities in eight-month-old infants – Saffran, Aslin & Newport, 1996). Before it is exposed to the first letter in the word, the network is unsure what is to follow. But the identity of the first two phonemes is usually sufficient to enable the network to predict with a high degree of confidence subsequent phonemes in the word. The time course of this process is as predicted by cohort models of word recognition (Marslen-Wilson, 1993). Once the input string reaches the end of the word, the network cannot be sure which word is to follow, so the error increases – hence the saw-tooth shape of the error function. But some word sequence information is evidently extracted too. For example, the segmentation error '*aboy*' which the network made is simply a function of the distributional characteristics of these words in the language sample – *boy* often followed *a*, just like the phonemes within a word occurred together. Such word sequence information is important in the extraction of syntactic information. It is also implicated in the acquisition of formulaic phrases (once upon a time, etc.), and in the segmentation errors of language learners – *a nelephant for an elephant, ife* for *knife, dult* for *adult*, etc. There are both overshooting errors (*aboy* at 13th position) and undershooting errors (*the* rather than *they* at 39th position). The model also showed implicit categorization of units. Although at first the network's predictions were random, with time the network learned to predict, not necessarily the actual next phoneme, but the correct category of phoneme, whether it was a vowel or consonant, etc. (see also Elman & Zipser, 1988, on this with networks trained on a large corpus of unsegmented continuous raw speech without labels). Thus the network moves from processing mere surface regularities to representing something more abstract, but without this being built in as a pre-specified phonemic or other linguistic constraint.

Elman (1990) also trained a larger network of similar architecture (31 input nodes, 31 output nodes, hidden and context vectors 150 units each) with sequences of words following a simple grammar. A 27534 word sequence formed the training set and the network had to learn to predict the next word in the sequence. At the end of training, Elman cluster analyzed the representations that the model had formed across its hidden unit activations for each word+context vector. It is clear that the network discovered several major categories of words: large categories of verbs and nouns, smaller categories of inanimate or animate nouns, smaller still categories of human and non human animals, etc. (for example, 'dragon' occurs as a pattern in activation space which is in the region corresponding to the category animals, and also in the larger region shared by animates, and finally in the

area reserved for nouns). These categories emerge from the language input without any semantics or real world grounding. The category structure is hierarchical, soft, and implicit.

The same network architecture which discovered sublexical regularities of language discovers important grammatical and semantic information from the same processes of sequential analysis. The network moves from processing mere surface regularities to representing more abstract aspects such as word class, but without this being built in as a pre-specified syntactic or other linguistic constraint. Relatively general architectural constraints give rise to language-specific representational constraints as a *product* of processing the input strings. These linguistically relevant representations are an *emergent* property of the network's functioning.

The Cobuild project

The Cobuild project represents the largest descriptive enterprise of the English language as it is used. Over 250 million words of representative English have been analyzed for the sequential patterns that are present. The three key conclusions of this research are (i) that it is impossible to describe syntax and lexis independently, (ii) that syntax and semantics are inextricable, and (iii) that language is best described as being collocational streams where patterns flow into each other (often going over the clause boundary). 'Through the reliability and objectivity o the computer evidence, verbs can be divided according to the pattern and pattern can be seen to correlate with meaning – that is to say, verbs with similar patterns have similar meanings . . . We can now see tha this relation between meaning and pattern is inevitable – that meaning and usage have a profound and systematic effect on one another. (Sinclair, foreword to Cobuild Grammar Patterns: Verbs, 1991: iv) Thus the Collins Cobuild (1996) analysis of English verbs shows tha there are perhaps 100 major patterns of English verbs (of the type for example, V *by* amount: the verb is followed by a prepositiona phrase which consists of the preposition *by* and a noun group indicating an amount as in 'Their incomes have dropped by 30 per cent' 'The Reds were leading by two runs', etc.). Verbs with the same Comp (Complementizer) pattern share meanings (the above-illustrated pattern is used by three meaning groups: (i) the 'increase' and 'decrease group (inc. 'climb', 'decline', 'decrease', 'depreciate', etc.), (ii) the 'win and 'lose' group (inc. 'lead', 'lose', and 'win'), (iii) the 'overrun' grou (inc. 'overrun', 'overspend'). Any Comp pattern is describable *only* i terms of its lexis.

Perhaps surprisingly, Chomsky's recent accounts of syntax in the Minimalist Program for Linguistic Theory (MPLT) (Chomsky, 1995) shares this emphasis on lexis and sequence analysis. Chomsky (1989, emphasis added) stated: 'There is only one human language apart from the lexicon, and *language acquisition is in essence a matter of determining lexical idiosyncrasies*'. Within the MPLT, '*differences between languages are attributed to differences between the features of lexical items in the languages* and specifically between the features of lexical items belonging to the functional categories AGR and Tense... Vs and Ns are taken from the lexicon fully inflected with inflectional affixes... specific bundles of these features of the category AGR and T are lexical items and *differences between the sets of bundles available in the lexicon account for cross-linguistic syntactic differences between languages.*' (Marantz, 1995: 366).

Over the last twenty years theories of grammar have increasingly put more syntax into the lexicon, and correspondingly less into rules. The corpus linguistic approach and the MPLT alike both represent a natural culmination of this trend where lexis is at the very centre of syntax. In both accounts, syntax acquisition reduces to vocabulary acquisition – the analysis of the sequence in which words work in chunks.

Inflectional morphology

As described at the beginning of this section, distributional analysis generates words, fuzzy word-class clusters with prototypical structure (like 'nounlike' rather than 'noun'), and letter sequences which are fairly reliable morphological markers (like *-s, -ing, -ed,* etc. in English). If particular combinations of these are reliably associated with particular temporal perspectives (for tense and aspect) or number of referents (for noun plural marking) for example, then we have the information necessary for the beginnings of a system which can generate inflectional morphology. But how could an associative system ever generalise to allow it to mark tense, aspect, case, etc. for words it has never previously experienced in a marked form? How could an associative system abstract regularities in order to operate grammatically with novel words? 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? This question has generated considerable debate in the literature over the past decade, much of it focusing on the behaviour of 'regular' and 'irregular' inflectional morphology. There are broadly two contrasting accounts. Dual-processing models (for example Marcus, Brinkmann, Clahsen, Wiese & Pinker, 1995; Pinker & Prince, 1988; Prasada, Pinker & Snyder, 1990) take the differences in behaviour of regular and irregular inflections to represent the separate underlying processes by which they are produced: regular inflections are produced by rules (for example, for the past tense 'add *-ed* to a Verb'), while irregular inflections are listed in memory. Associative accounts, whether connectionist (e.g., MacWhinney & Leinbach, 1991; Plunkett & Marchman, 1993; Rumelhart & McClelland, 1986) or schema-network (Bybee, 1995) models, assume that both regular and irregular inflections arise from the same mechanism, a single distributed associative network, with the differences in behaviour being due to statistical distributional factors.

There have been a number of compelling PDP models of the acquisition of morphology. The pioneers were Rumelhart and McClelland (1986) who showed that a simple learning model reproduced, to a remarkable degree, the characteristics of young children learning the morphology of the past tense in English: the model generated the so-called U-shaped learning curve for irregular forms; it exhibited a tendency to overgeneralize, and, in the model as in children, different past-tense forms for the same word could co-exist at the same time. Yet there was no 'rule' – 'it is possible to imagine that the system simply stores a set of rote-associations between base and past-tense forms with novel responses generated by "on-line" generalisations from the stored exemplars.' (Rumelhart & McClelland, 1986: 267). This original past-tense model was very influential. It laid the foundations for the connectionist approach to language research; it generated a large number of criticisms (Lachter & Bever, 1988; Pinker & Prince, 1988), some of which are undeniably valid; and, in turn, it thus spawned a number of revised and improved PDP models of different aspects of the acquisition of the English past tense. The successes of these recent models in capturing the regularities that are present in associating phonological form of lemma with phonological form of inflected form (Daugherty & Seidenberg, 1994; MacWhinney & Leinbach, 1991; Marchman, 1993; Plunkett & Marchman, 1991), and between referents (+past tense or +plural) and associated inflected perfect or plural forms (Cottrell & Plunkett, 1994, Ellis & Schmidt, 1997); in closely simulating the error patterns, profiles of acquisition, differential difficulties, false-friends effects, reaction times, and interactions of regularity and frequency that are found in human learners (both L1 and L2); as well as in acquiring a default case allowing generalisation on 'wug' tests, all 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. That morphology can be described as being rule-like behaviour does not imply that morphology is rule-governed (Ellis, 1996b; Harris, 1987).

Much of the information that's needed for syntax falls quite naturally out of simple sequence analysis.

Working out how words work (ii): distributional analysis of word co-occurrences to derive lexical semantic information

Recent work by Landauer and Dumais (1997) demonstrates that a large part of semantics can also come from sequence analysis. Although much of semantics comes from the grounding of lexical meaning in conceptual/imagery-perceptual representations, there is clearly another source of lexical meaning, particularly important for more abstract words, which arises from a word's associations with the other words with which it tends to co-occur. This is the aspect of meaning which drives the collocational analysis of meaning (see earlier section concerning 'the Cobuild project') stemming from Firth's (1957; Bazell, Catford, Halliday, & Robins, 1966) dictum: 'You shall know a word by the company it keeps.' The lexical context which surrounds a lexeme is crucial to the determination of its meaning and its grammatical role. The telling evidence that this is a potent source of lexis is the fact that people who read more know more vocabulary (Anderson, Wilson, & Fielding, 1988). This relationship between print exposure and vocabulary appears to be causal in that it holds even when intelligence and even reading comprehension ability – an excellent measure of general verbal ability – is controlled (Stanovich & Cunningham, 1992).

Landauer and Dumais (1997) present a theory and mechanism of acquired similarity and knowledge representation called Latent Semantic Analysis (LSA) which simulates both L1 and L2 acquisition of vocabulary from text. By inducing global knowledge indirectly from local co-occurrence data in a large body of representative text, LSA acquired knowledge about the full vocabulary of English at a comparable rate to school children. Yet LSA has no prior linguistic or perceptual similarity knowledge, and is based solely on a general mathematical learning method that achieves induction by extracting about 300 dimensions to represent words-as-letter-strings in the context of other words-as-letter-strings. Conceptually, LSA can be viewed as a large symmetrical three-layered connectionist network linking every word type encountered in layer 1, through several hundred hidden units in layer 2, to a layer 3 which comprises nodes for every text window context ever encountered. After the model had been trained by exposing it to text samples from over 30 thousand articles from Groliers Academic American Encyclopaedia, it was tested with 80 items from the synonym portion of the Test of English as a Foreign Language (TOEFL). Applicants to US colleges from non-English speaking countries who took tests containing these items averaged 64.5% correct on this test. LSA got 64.4% correct.

LSA closely mimics the behaviour of a group of moderately proficient English readers with respect to judgements of meaning similarity Yet it acquired this competence without any other information tha simple exposure to words (as sequences of letters) and other suc words as tend to co-occur as their neighbours. Interestingly, the inpu to LSA was, as Landauer and Dumais put it, 'a simple bag of words' all information from word-order was ignored, and there was therefor no explicit use of grammar or syntax. All it had was the frequency profile of co-occurring words in 30,000 text samples comprising roughl 150 words each. It appears that word co-occurrence data in sequenc is important for the derivation of syntactic information (see previou section), but that simple word co-occurrence statistics, ignoring orde is at least sufficient for the derivation of lexical semantics. It remain to be seen if the performance of LSA given ordered information i superior to that given unordered input.

The performance of LSA will be surprising to many readers. Th model could not see or hear, and thus could make no use of phonology, morphology, or real-world perceptual knowledge. It was provide with no prior linguistic or grammatical knowledge. But from a larg corpus of simple lexical-string co-occurrence data it acquired later semantic information to allow it to perform at levels expected of good ESL learner.

In this account, lexical semantic acquisition reduces to the analys of word co-occurrences, i.e. the words that tend to chunk togethe (ignoring order).

Collocations, slot-and-frame patterns, and sequences in learner talk

We have seen that as powerful computers are used for distribution analysis of large language corpora, so they demonstrate the underlyin chunks of language and the ways in which lexical items, with their pa ticular valences and subcategorization requirements, operate in thes patterns. Is there parallel evidence that *learners* acquire collocations on their path to fluency, and that their analyses of these chunks gives them the information about lexical idiosyncrasies that allows later more open-class productions?

Collocations and patterns in L1 acquisition

Tomasello (1992) begins his book, *First Verbs: A Case Study of Early Grammatical Development*, with the following observation from Wittgenstein: 'Language games are the forms of language with which a child begins to make use of words... When we look at the simple forms of language the mental mist which seems to enshroud our ordinary use of language disappears. We see activities, reactions, which are clear-cut and transparent. On the other hand we recognize in these simple processes forms of language not separated by a break from our more complicated ones. We see that we can build up the more complicated forms from the primitive ones by gradually adding new forms.' (Wittgenstein, The Blue Book).

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

> 'There is individuality and contextedness everywhere, signs of broad-based rules nowhere. T did bring order and systematicity to her language during her 2nd year of life, but it was a gradual, constructive process. It did not resemble in any way the instantaneous and irrevocable setting of parameters...
>
> T's earliest three-or-more-word sentences (18–21 months) were almost all structured by verbs. The vast majority of these involved straight-forward coordinations of already produced word combinations (93%), preserving in almost all cases the established ordering patterns of the constituents (99%).
>
> T began marking the syntagmatic relations in these three-or-more-word sentences through the use of contrastive word order and prepositions. She did this, however, on a verb-by-verb basis. By far the best predictor of the arguments and argument markings that T used with a particular verb at a particular time was previous usage of that verb, not same time usage of other verbs' (Tomasello, 1992: 264–266).

Tomasello concludes:

> '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' (Ibid.: 273–274).

Other analyses of child language corpora point to similar conclusions. For example, Lieven, Pine and Dresner Barnes (1992) show formulae to be both frequent (children's first 100 words typically contain about 20 formulae) and productive (in providing templates which, following analysis, are converted into lexically based patterns). Pine and Lieven (1997) and Lieven, Pine and Baldwin (1997) show that a lexically based positional analysis can account for the structure of a considerable proportion of children's early multiword corpora. The corpus-analyses of Pine and Lieven (1997) suggests that the development of, for example, an adult-like determiner category may be a gradual process involving the progressive broadening of the range of lexically-specific frames in which different determiners appear. These data are all broadly consistent with constructivist models of children's early grammar development.

Collocations and patterns in SLA

No observation is entirely theory-free. Yet we are fortunate to have some descriptions of stages of L2 proficiency which were drawn up in as atheoretical way as possible by the American Council on the Teaching of Foreign Languages (ACTFL) (Higgs, 1984). The ACTFL (1986) *Oral Proficiency Guidelines* include the following descriptions of novice and intermediate levels which emphasise 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 Mid:** Oral production continues to consist of isolated words and learned phrases within very predictable areas of need, although quantity is increased. Vocabulary is sufficient only for handling simple, elementary needs and expressing basic courtesies. Utterances rarely consist of more than two or three words and show frequent long pauses and repetition of interlocutor's words.

> **Novice High:** Able to satisfy partially the requirements of basic communicative exchanges by relying heavily on learned utterances but occasionally expanding these through simple recombinations of their elements. Can ask questions or make statements involving learned material . . . Speech continues to consist of learned utterances rather than of personalized, situationally adapted ones . . . Pronunciation may still be strongly influenced by first language.
> **Intermediate:** Characterized by an ability to create with the language by combining and recombining learned elements, though primarily in a reactive mode; initiate, minimally sustain, and close in a simple way basic communicative tasks; and ask and answer questions.
> **Intermediate-Mid:** Able to handle successfully a variety of uncomplicated, basic communicative tasks and social situations. Can talk simply about self and family members. Can ask and answer questions and participate in simple conversations on topics beyond the most immediate needs; e.g., personal history and leisure-time activities. *Utterance length increases slightly, but speech may continue to be characterized by frequent long pauses, since the smooth incorporation of even basic conversational strategies is often hindered as the speaker struggles to create appropriate language forms. Pronunciation may continue to be strongly influenced by first language* and fluency may still be strained.’ (ACTFL, 1986, emphases added).

Thus the ACTFL repeatedly stresses the constructive potential of collocations and chunks of language which are slowly analysed on a word-by-word basis to allow the determination of L2 grammatical word class and grammatical dependencies. 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 none the less the role of formulae became readily apparent in the acquisition process. Wong-Fillmore (1976) presented the first extensive longitudinal study which 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 unanalysed chunks became the foundations for creative construction. These observations closely parallel those of Lieven et al. (1997) for L1 acquisition. But Nora was just one child, and there is clearly need for larger sampled, detailed collection and analysis of SLA corpora, although there is some recent progress: Myles, Mitchell and Hooper (1999) studied the first two years of development of interrogatives in anglophone French L2 beginners and tracked the breakdown of interrogative chunks, the creative construction of interrogatives, and the ways in which formulae fed the constructive process (see also Myles, Hooper & Mitchell (1998) for the constructions

that stem from three other formulae during this period). Other useful reviews of formulae in SLA include Hakuta (1974), Nattinger and DeCarrico (1992), Towell and Hawkins (1994), Weinert (1995), and Wray (1992).

Working out how words work (iii): cognitive linguistics and grounded lexical meaning

Much of traditional linguistics views language as a closed modula
system where syntax can be described as a body of logical rules fo
generating the sentences of a language that are grammatically correct
This enterprise had largely studied syntax in isolation from semantics
the functions of language, or the other social, biological, experiential
or cognitive aspects of the humankind who invented and used languag
ever since. But the meaning of the words of a given language, and hov
they can be used in combination, depends on the perception and cat
egorization of the real world around us. Since we constantly observ
and play an active role in this world, we know a great deal abou
the entities of which it consists, and this experience and familiarit
is reflected in the nature of language. Language reflects our *exper*
ence and our *embodiment*. The different degrees of *salience* or *prom*
nence of elements involved in situations which we wish to describ
affect the selection of subject, object, adverbials, and other clause a
rangement. Figure/ground segregation, which originated from Gesta
psychological analyses of visual perception, and *perspective takin*
again very much in the domains of vision and attention, are mirrore
in language and have systematic relations with syntactic structur
We have expectations of the world which are represented as con
plex packets of related information (schemata, scripts, or *frames* fo
for example, how the parts of a chair inter-relate, a trip to the de
tist, buying and selling, indeed, everything we know – Schank, 1982
What we express reflects which parts of an event attract our *attentio*
Depending on how we direct our attention, we can select and higl
light different aspects of the frame, thus arriving at different linguist
expressions.

All of these concerns – the experiential grounding of languag
our embodiment which represents the world in a very partic
lar way, the relations between our perceptual and imagery repr
sentations and the language which we use to describe them, o
perspective and attentional focus – lie at the heart of an altern
tive view of language: cognitive linguistics (Fillmore, 1977; Lako

1987; Lakoff & Johnson, 1980; Langacker, 1987, 1991; Talmy, 1988):

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

It seems possible that much of our knowledge of language structure emerges from the analysis of chunks of language form. The research described in earlier sections demonstrates how orthographic and phonological regularities, lexical and morphological form, collocations, word class, and even aspects of lexical semantics might so arise. However, other aspects of language, like the grounding of lexical semantics and the communicative use of syntactic structures, derive from the frequency and regularity of cross-modal associations between chunks of phonological surface form and, particularly visuo-spatial, imagery representations. But these visual representations are not fixed and static; rather they are explored, manipulated, cropped and zoomed, and run in time like movies under attentional and scripted control (Kosslyn, 1983; Talmy, 1996a). Cognitive linguistics reminds us that the prominence of particular aspects of the scene and the perspective of the internal observer (i.e. the attentional focus of the speaker and the intended attentional focus of the listener) are key elements in determining regularities of association between elements of visuo-spatial experience and elements of phonological form. We cannot understand language acquisition by understanding phonological memory alone. All of the systems of WM, the slave systems and the SAS, are involved in collating the regularities of cross-modal associations underpinning language use. Cognitive linguistics aims to understand how the regularities of syntax emerge from the cross-modal evidence that is collated during the learner's lifetime of using and comprehending language. The difficulties of this enterprise are obvious. Acknowledging the importance of embodiment, perspective, and attention entails that to understand the emergence of language we must also understand the workings of attention, vision, and other representational systems (Talmy, 1997). And then we must understand the regularities of the mappings of these systems onto particular languages. And the mappings in question are piecemeal; it's the content of the mappings that is important, not simply the modalities concerned. Which is why cognitive linguistics focuses on one particular construction and representational aspect at a time, for example motion event frames (Langacker, 1991; Talmy, 1996b), spatial language (Bowerman, 1996; Regier, 1996), verb aspect (Narayanan, 1997), or inflectional morphology.

Attention and language learning: implicit and explicit learning, memory and instruction

One area of WM which, because of its centrality in language learning and language use, deserves much more consideration than it has been given here, is the role of attention in language learning. Since our fluent language use is unencumbered by metalinguistic descriptions of sufficient complexity to allow its generation, so much of the representation and processing that generates language must be unconscious. But what is the role of attention and consciousness in language learning? This complex and long-standing question cannot be dealt with adequately here, but it is nevertheless important to highlight some relations between this chapter's emphasis on memory and theories of attention in language learning and processing.

Attention, as the SAS, is the most central element of the Working Memory Model, yet it is the least-well understood. One must look elsewhere to the research on consciousness and attention (Ellis, 1994a, and chapters in this volume by Doughty, Hulstijn, Robinson, and Schmidt) for more detailed specification of its role in language learning.

Language understanding and language production utilise the many millions of associations that the learner has acquired in their history of language use. Thus language is learned in the course of using language, and the best predictor of language facility will simply be time-on-task. Research on implicit learning and implicit memory suggests that at least some of the relevant associations can be acquired from the input without the learner being consciously aware of the contingency, although the relevant aspects of the input must be attended for processing (see Ellis, 1994b; Hsaio & Reber, 1998; Schmidt, this volume). In acquiring associations, some parts of the input environment can be made more salient, and learners are more likely to learn about the part of the environment which they selectively attend. Thus there are ways of speeding learners' L1 or L2 acquisition from a given amount of language exposure, to increase the quality of the learning (see Ellis & Laporte, 1997, for review). These ways, which include grammatical consciousness raising or input processing, as well as corrective feedback and recasts, promote the acquisition of sophisticated grammatical proficiency. There is some benefit in a focus on form in L2 instruction (see Doughty & Williams, 1998b; R. Ellis, 1994a; Long, 1988, 1991; Terrell, 1991, for reviews of instructional programs which incorporate these ideas).

Communicative approaches give input, time-on-task, and opportunity for relating form and function. Of course, all of this is necessary for developing the associations necessary for language learning. Naturalistic environments provide motivation and plenty of opportunity for output practice as well. These are situations which guarantee sufficient quantity of language. But without any focus on form, formal accuracy is an unlikely result. Focus on forms alone can teach some declarative rules of grammar, but at its worst can be accompanied by too little time on the task of language use itself (see Skehan & Foster, Doughty, and Robinson, this volume). At its worst it is insufficient to support language development. But focus on form instruction, which is rich in communicative opportunities and which also makes salient the associations between structures (which the learner is already at a stage to be able to represent) and functions, can facilitate language acquisition. Instruction must build on the levels of representation which the learner has already acquired. Just as there is the issue of Learnability in L1, so there is that of Teachability in L2 – any empirical findings about natural (in terms of epistemics, not biology – see Ellis, 1996a, b) developmental sequences should be respected in the design of instructional materials (Pienemann, 1985), and attempts to teach structures or strategies which build on still-to-be acquired representations are likely to fail.

Conclusions and overview: the different connections and the need for connectionist simulations

The proper study of language acquisition is to chart the course by which perceptual, motoric, and cognitive functions induce structure, from undifferentiated novice performance to that remarkably differentiated nativelike competence. The history of linguistic science demonstrates language to be an extremely complicated set of evidence. To remind us of the enormity of the representational database, consider just the four lines of nursery-book language shown in Figure 3. Processing this simple 18 word sentence taps into a rich abundance of

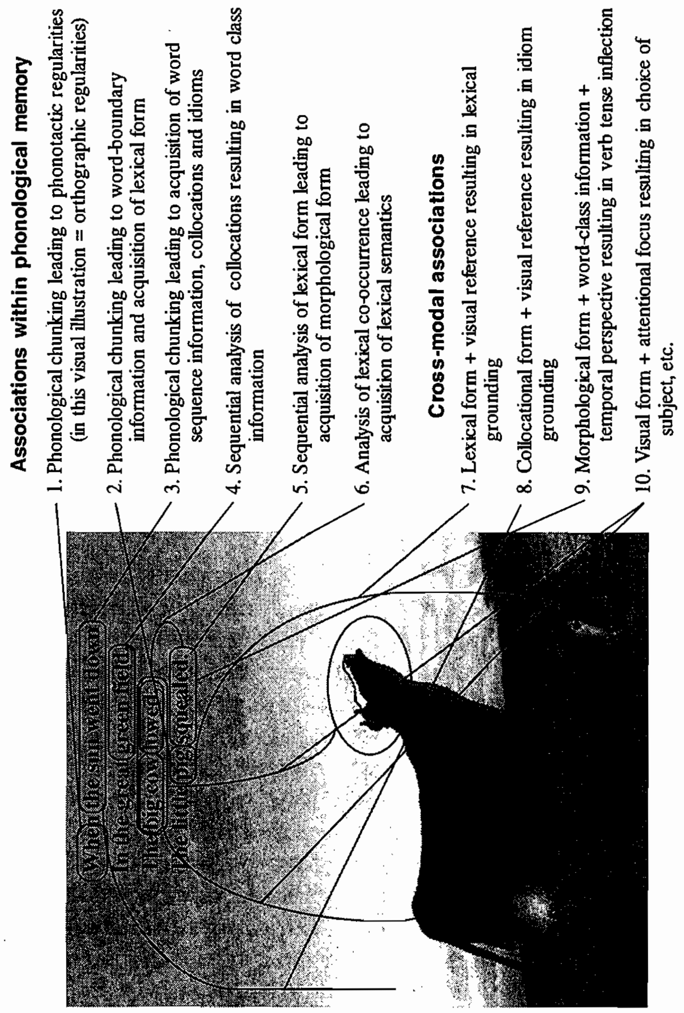
When the sun went down
In the green field
The big cow lowed.
The little pig squealed.
Associations within phonological memory
1. Phonological chunking leading to phonotactic regularities (in this visual illustration = orthographic regularities)
2. Phonological chunking leading to word-boundary information and acquisition of lexical form
3. Phonological chunking leading to acquisition of word sequence information, collocations and idioms
4. Sequential analysis of collocations resulting in word class information
5. Sequential analysis of lexical form leading to acquisition of morphological form
6. Analysis of lexical co-occurrence leading to acquisition of lexical semantics
Cross-modal associations
7. Lexical form + visual reference resulting in lexical grounding
8. Collocational form + visual reference resulting in idiom grounding
9. Morphological form + word-class information + temporal perspective resulting in verb tense inflection
10. Visual form + attentional focus resulting in choice of subject, etc.

Figure 3. An overview of the range of language-relevant associations. (Original picture taken from Big Red Barn *(p. 21) by Margaret Wise Brown (pictures by Felicia Bond), HarperFestival, A Division of Harper Collins Publishers. Text copyright 1956, 1989 by Robert Brown Rauch. Illustrations copyright 1989 by Felicia Bond.)*

associations, some associative chunks in phonological memory, some cross-modal associations. The types of association are also illustrated in Figure 3. Since the 'cycle of perception' is also the 'cycle of learning', the processing of the sentence itself also results in some language acquisition; associations which are used in the processing of this sentence are forged or strengthened in LTM, making them more accessible in the future (short term implicit memory or priming effects (Ellis, 1994b) and long term life-span practice effects (Kirsner, 1994). If we just consider the orthographic chunks that are potentially activated from this 72 character sequence, there are 71 bigrams to be processed, each much more likely than chance although individually varying in their likelihood; 70 trigrams; 18 lexemes, which are common but varying in their individual likelihood of occurrence; three common verb inflections; 17 biword sequences, and collocations ranging in size from *the sun*, through *when the sun went down*, up to, for me, *the father who has read these words so many nights before*, the whole sentence, etc. If we then consider the various ways in which these varying sizes of chunk map onto phonology, syntax and semantics it is clear that there is a combinatorial explosion of associations potentially activated in processing this simple sentence. Although there are not many modalities of representation involved here, the many representations within each modality, richly interconnected by associations of varying weight, affords a massively complex database.

This type of analysis will surely leave many linguists aghast, observing that there are just too many associations, that anything is possible from such complexity, and that it is yet more psycholinguistic research lacking any but the most naive linguistic content (see Gregg, this volume). This would be fair comment indeed. None the less, the research outlined here, along with four decades or so of psycholinguistics, demonstrates that language learners do indeed possess this richness of association. The computational work outlined in this chapter also shows that analysis of these associations results in generalization and emergent linguistically relevant representations. Thus it seems worthwhile investigating how connectionist or constraint probability models might allow single interpretations to result from the competition between all of these available cues, in the same way that somehow, as we read the nursery text in Figure 3, we settle on-line on one interpretation.

We need to understand the active competition between the cues available in input and their representational associations. And we need dynamic models of the acquisition of these representations and the emergence of structure. Connectionism provides a set of computational tools for exploring the conditions under which such emergent properties arise. Advantages of connectionism include: neural inspiration; distributed representation and control; data-driven processing with prototypical representations emerging rather than being innately pre-specified in nativist accounts; 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. We have already discussed the millions of potential connections that map between orthographic structure and phonological structure. When connectionist models of reading acquisition learn these mappings, so highly plausible models of reading acquisition arise (Seidenberg & McClelland, 1989). We have described the associations between stem form and past tense form involved in inflectional morphology. When connectionist models learn these mappings, so close simulations of acquisition and performance result. The problem with modelling the totality of language knowledge is that we need to understand, thus to be able to model, representation in other modalities as well – the visual, motor and other systems which underpin conceptual knowledge. This is very hard, although some useful progress is being made: see, e.g., Regier (1996) on spatial language; Narayanan (1997) on how the semantics of verbal aspect are grounded in sensori-motor primitives abstracted from processes that recur in sensori-motor control (such as goal, periodicity, iteration, final state, duration, force, and effort); more general work on embodied language development by the L$_0$ project (Bailey, Feldman, Narayanan & Lakoff, 1997; Feldman, Lakoff, Bailey, Narayanan, Regier & Stolcke, 1996); and MacWhinney's competition model (this volume).

Gregg (this volume) is correct in reminding us of the essential need for linguistic analysis, and the UG framework has provided the most complete description of language competence to date. But, like Studdert-Kennedy (1991), I believe that UG is neither a prescription nor a program for development, but rather it is a partial and *a posteriori* description of the phenotypic product of the developmental system. In this view, UG is a consequence, not a condition of development. The types of association described in this chapter are pretty much all we have as far as language representation is concerned. But out of these associations comes systematicity. There is a more relevant Universal which concerns process and learning rather than content: it is to be found in efforts to rationalize intelligence in terms of models of optimal (Bayesian) inference in the presence of uncertainty. My money's on connectionism to help us solve the riddle, more complex even than that of Samson (*Judges, 14*), of how, out of the strings, came forth syntax.

3 *The competition model: the input, the context, and the brain*

Brian MacWhinney

Introduction

Language learning is a three-way interaction between the input, the learner, and the interactional context (Bloom, 1974). This three-way interaction provides a general framework for understanding first and second language acquisition (SLA), in both naturalistic and formal contexts. In order to elaborate this general framework, we need to model its three components:

1. The input. We need to know how the linguistic input can be structured to maximize effective learning. What aspects of the phonology, syntax, semantics, and morphology of the input does the learner use to 'crack the code' of the new language?
2. The learner. We need to understand exactly how the cognitive abilities of the learner shape the process and outcome of L2 instruction.
3. The context. Traditionally, the classroom environment maintains a rather uniform structure in which interaction is controlled by the instructor. How does this framework affect learning and how can it be varied to improve the learning process?

This paper will examine these three components within the framework of the Competition Model (MacWhinney, 1987; MacWhinney & Bates, 1989). To quantify the role of the input in L2 learning, the model relies on the concepts of cue reliability and availability. To characterize the cognitive abilities of the learner, the model relies on findings from cognitive neuroscience. To understand the role of the context, the model elaborates the concepts of environmental and social support. The Competition Model views both L1 and L2 learning as constructive, data-driven processes that rely not on universals of linguistic structure, but on universals of cognitive structure. It attributes development to learning and transfer, rather than to the principles and parameters of Universal Grammar (Chomsky, 1965).