LANGUAGE ACQUISITION, EMERGENTISM, AND THE BRAIN-CHANGING NORMS OF UNILATERAL INTERVENTIONISM

Author:
Thomas I. Nygren

Faculty Sponsor:
Yiqiang Wu,
Department of Special Education, Language, and Literacy

ABSTRACT AND INTRODUCTION

Since the late 1950s, Noam Chomsky has argued that language is so complex that human beings must be born with an innate language faculty. More recently, emergentist theories, which propose that language is learned from the environment and that the complexity of language emerges from regularities abstracted from billions of associations made during language use, have challenged that view. Moreover, the last decade has seen an explosion of research on the brain. The "memory-prediction framework" strongly supports the emergentist conception of first and second language acquisition and offers implications for second language instruction.

INNATISM VERSUS EMERGENTISM

Chomsky and his supporters, commonly known as innatists or nativists (Gass & Selinker, 2008, 159-190), believe that human beings have an innate faculty that guides them in learning language. Innatists begin by assuming that all children, unless physically impaired, learn their first language quickly and effortlessly, especially when compared to other domains of abstract knowledge such as physics or mathematics. By contrast, language learning for adults is typically slow, difficult, and rarely leads to native-like fluency. Furthermore, children learn language without any explicit instruction, and with only limited and partial language input from adults—a process often called the "poverty of the stimulus." More specifically, Chomsky notes two reasons for the inadequacy of the input children receive. First, there is incomplete *positive* evidence, since children never hear all the correct sentences that are possible, and second, a lack of *negative* evidence, because they rarely learn which sentences are ungrammatical.

Chomsky argues that all languages are fundamentally similar, thanks to inherent mental properties specific to language, which he terms "language acquisition devices." The universal underlying structure of all languages is Universal Grammar (UG), which consists of underlying principles that characterize all languages, combined with parameters that vary across languages. Parameters have a limited set of values, and the manner in which they are set determines the specific features of that language, such as the position of subject, verb, and object in a sentence. According to innatists, the combination of universal principles and language-specific parameters explains how children acquire language so effortlessly. Since UG comes "pre-wired" in the brain, children require only (1) sufficient language input to set the various parameters correctly and (2) the specific vocabulary of that language. Thus, innatists prefer the term " language acquisition" rather than "language learning," since a first language is not learned in the ordinary sense of the word. The apparent variety of languages of the world is thus merely a "surface structure" masking common underlying universal principles. The term "generative grammar" is another frequently used term, since language is understood to be generated from a set of innately determined rules.

Innatists share no clear consensus on how a second language (L2) is acquired (Gass & Selinker, 2008, 163-166). Some believe that L2 grammars are controlled by UG just as native grammars are. In this view, L2 learners have full access to the innate language faculty. Others argue that L2 grammars are not controlled by UG and therefore, fundamentally different—learners do not have access to UG (which is believed to be available only during childhood) and instead use general problem-solving abilities to learn another language. This helps explain why L2 learners often fail to attain native fluency. Innatists take various positions on whether first-language (L1) parameters can be reset to reflect the L2. However, despite this lack of consensus regarding the role of UG, most accept the "critical period hypothesis" that after a certain developmental age, generally around puberty, the ability of the brain to learn language changes so that L2 acquisition becomes fundamentally different in one way or another.

The innatist perspective, dominant and influential since the 1950s, has been challenged in the past 15 to 20 years by an opposing view.[1] Emergentism takes a broadly cognitive approach and, contrary to innatism, starts by proposing that language is learned from the environment without any need for an innate, language-specific, acquisition device. People learn language as they learn everything else. Emergentism takes a usage-based view of language development: the human brain, sensitive to the frequencies of events it experiences, creates form-meaning associations by abstracting regularities from probabilistic patterns of input. Thus, far from a "poverty of the stimulus," for emergentists input is everything.

Contrary to innatists, who think language inherently rule-driven, emergentists argue that rule-*like* behavior does not imply rule-*governed* behavior (Ellis, 1998). What innatists interpret as rules are regularities abstracted from billions of associations made while using language. Moreover, the apparent complexity of language does not imply the need for a complex system of rules; rather, the complexity emerges because the environment is complex. Similarly, the path of an ant on a pebbled beach appears complex despite a small set of simple control decisions (Ellis, 2001, 37). The "rules" of language merely describe dynamic, emergent patterns, without playing any causal role: "The brain does not process information by some imaginary computational rule-following any more than the gut does. . . . The notion of fixed cognitive categories, linguistic and non-linguistic both, is a myth: conceptual categories are dynamically construed" (Ellis, 2006a, 108). The fact that languages share certain fundamental properties is explained not by a universal innate language "instinct" but by the fact that all human brains work the same way and all people share common life experiences in the same physical world. In that limited sense, emergentists would agree that language acquisition is innate, because the human brain has evolved to extract regularities from complex input in a remarkably efficient and powerful way, but they would deny that language differs from any other input: "Instead of having a language instinct, humans are better described as having a communication instinct and an instinct for learning in general" (Ke & Holland, 2006, 694).

Emergentists do not believe that L2 learning differs from L1; the same usage-based mechanisms operate in all forms of learning. Thus, while acknowledging that L2 learning is more difficult than L1, they do not attribute the difference to a critical period. Instead, the difficulty of L2 acquisition arises from interference with patterns already entrenched from the L1. In the emergentist view, at birth a child's brain is, to a large extent, a *tabula rasa*, genetically designed to process the flood of input that streams in through the senses. By the time second language acquisition begins for most people, however, the brain is more like a "*tabula repleta*; it is no longer a plastic system, it is one that is already tuned and committed to the L1" (Ellis, 2006a, 109).

UNDERSTANDING HOW THE BRAIN WORKS: THE MEMORY- PREDICTION FRAMEWORK

In *On Intelligence* (2004), Jeff Hawkins presents a general theory of how the brain works, which he terms the "memory-prediction framework." Hawkins is primarily interested in artificial intelligence, but his explanation of how the brain processes and stores information is highly relevant for linguistics and language acquisition: the brain does not "compute" the answers to problems, but retrieves them from memory. Accordingly, Hawkins describes how a computer would solve the problem of a robot catching a ball (68-69). First, it would detect the motion of the onrushing ball and calculate the trajectory and speed of the ball using mathematical formulas and principles of physics. Then it would position the robot's arm and hand, readjusting as the ball got closer and the computer received better information about its trajectory. Even a fast computer would have difficulty computing the millions of lines of code required in the time available, and even then it might not catch the ball. The human brain, which processes signals much more slowly, relies on a completely different strategy. It retrieves a stored memory of the muscle commands required to catch the ball, learned over years of practice.

The brain still faces a problem: it has to be able to recognize that every throw and catch differ slightly. Hawkins explains this by describing four different attributes of neocortical memory (the neocortex, or cortex, is the outer layer of the brain responsible for most higher brain functions such as language): (1) the cortex stores sequences of patterns; (2) it stores patterns auto-associatively; (3) it stores them in invariantly; and (4) it stores them hierarchically.

First, while the capacity of memory is incomprehensibly vast—estimated at 30 billion neurons and at least 30 trillion synapses—it can only remember a few things at any one time, and it can only do so sequentially. How people remember illustrate this—one memory is linked to another, which prompts another memory and so on. Even a minimal cue—a whiff of perfume, the first bars of a song, a childhood photo—can conjure detailed memories from decades earlier. Furthermore, repeating a well known sequence, such as the alphabet, in a different order is difficult for most people. Hawkins powerfully observes that all inputs to the cortex are basically alike. Although each of the five senses gathers sensory input using different physical mechanisms, by the time neural signals reach the cortex they are all converted into sequential patterns of electrical-chemical neural "spikes."

Second, memory is auto-associative. This simply means that patterns are associated with themselves. Thus, the brain can retrieve a complete pattern even when given partial or distorted inputs. Just glimpsing a face in dim light is sufficient to recall a memory of the complete face. The same process occurs with temporal patterns: hearing a few notes of a song is enough to recall the whole song. In a noisy room, people don't actually hear all the words they perceive. The brain constantly completes patterns based on partial input.

The third major attribute of neocortical memory, closely linked to its auto-associative nature, is its invariant storage of patterns. The brain stores important relationships independent of the specific details. Anything we see, such as a human face, is actually a stream of patterns of neural signals to the brain. If the angle of the face shifts even minutely, or the lighting changes, or if the person smiles or walks further away, a completely different pattern of inputs goes to the brain. A computer has enormous difficulty recognizing that this is the same object, but a human brain effortlessly accomplishes the task. How the brain accomplishes this remarkable feat is not well understood, but we know that it stores the relationships among subcomponents of the object, not an exact mental copy of the object. Thus, human faces are recognizable by the relationship of eyes to nose, mouth, etc., and the relative positions and dimensions of each subcomponent. The same is true of a melody; the brain stores the intervals between notes, not the actual notes, or else the same song in a different key would be unrecognizable.

Finally, the neocortex stores patterns hierarchically. In the functional regions of the brain, lower areas feed information to higher areas, and higher areas send feedback to lower areas. The lowest of the functional regions, the primary sensory areas, are where "raw" sensory information first arrives in the cortex. These regions send information to higher functional areas that process information in more specialized or abstract ways, and also associate data from different sensory inputs. Hawkins emphasizes that the hierarchical structure of the brain is not accidental; it is so because the world that human beings experience is hierarchical: language, for example, with patterns of phonemes forming words; words forming phrases; sentences, discourse; and so on.

Hawkins argues that the cortex's primary function is to make predictions, even terming it an "organ of prediction" (89). All regions of the cortex simultaneously try to predict their next experience, which means that the neurons involved in sensing become active before receiving sensory input (based on the pattern of input immediately prior). When the sensory input arrives, it is compared with what was expected. Our minds work by making probabilistic predictions.

An unexpected pattern causes the brain to take notice and incorrect predictions result in confusion. Specifically, if a lower functional region fails to predict accurately the next input pattern, it passes that information to the next higher region in the hierarchy until a region of the cortex can make sense of it. The brain thus extracts meaning from the streams of input constantly inundating the senses. For example, it tunes out background noise until it stops, or a different unexpected sound occurs. Similarly, people see what they expect to see and hear what they expect to hear, even when input is partial and distorted. If the brain cannot recognize a sequence of patterns by the time it reaches the highest functional level, it truly is a new and unexpected experience that may be stored as a new memory.

The memory-prediction framework posits that people are not born with innate knowledge;

rather, predictions are learned by experience through time and repetition as the cortex discovers the world's hierarchical relationships. Each region of the cortex, from the lowest to the highest, receives a stream of patterns. If a cortical region recognizes a sequence of patterns that always occurs together, it "names" that pattern and passes the name to the region above it in the hierarchy. This happens at every level in the cortex, from the lowest level of sensory input at which, for example, a region learns to recognize a visual edge at a certain angle, to a complex representation of a human face. Consequently, each higher region experiences less variability in its input, which forms the basis of the concept of invariant representations that are central to the memory-prediction framework. This explains why a pattern stored at a higher level can have different representations at lower ones. Thus, the concept "dog" that is stored at one level can be linked to both phonological and orthographic patterns at a lower level. This storage mechanism allows sharing and reuse of lower level patterns.

This cortical storage of patterns hierarchically plays a key role in learning. When something new is encountered, the brain processes it at a higher functional level because none of the lower levels recognizes it. For newborns, this is true of all sensory input. Gradually, through repetition, the cortex learns that certain sequences are related to one another and representations of objects move down the cortical hierarchy. As simple representations so move, the regions at the top are able to learn more complex patterns. Learning to read, for example: initially, a learner has to pay attention to every individual letter and sound out every syllable. But after years of practice, with one glance a person can recognize entire words and phrases. The same is true of any skill, from walking up a flight of stairs to playing the piano. It takes time for neural signals to flow up and down the cortical hierarchy; those memories stored at lower levels can be retrieved much more quickly and efficiently.

The Memory-Prediction Framework and Language Acquisition

As the discussion above indicates, the memory-prediction framework does not support innatist theories of language acquisition. Given the uniformity of the neocortex and the fact that it processes neural signals from all sensory input identically, there is no reason to treat the language learning process differently from any other. According to Hawkins, "language fits nicely into the memory-prediction framework without any special language sauce or dedicated language machinery" (Hawkins, 2004, 181). In fact, the cortex is ideally designed to process and store the highly hierarchical system of language. Nor is there evidence for the inheritance of structures or mechanisms in the brain that might correspond to UG parameter setting. The remarkable plasticity of the brain and the fact that different areas of the cortex can be trained to process any form of input further undermine the notion of a specialized language acquisition device. The way the brain processes a stream of sounds pouring into the auditory nerve, parsing it into phonemes, words, phrases, and sentences, and then matching the patterns against invariant patterns stored in memory to extract meaning is no different from its parsing of input from the optic nerve. As it solved the problem of catching a ball, the brain interprets and produces sentences not by using a "central processor" that computes the correct answer according to predefined rules (which would take far too long given the speed of neural signals) but by retrieving the relevant set of hierarchically stored invariant memory patterns.

The Poverty of the Stimulus Argument

From a memory-prediction perspective, "poverty of the stimulus," is ruled out by definition, since all learning results from environmental input (Lieven and Tomasello, 2008; and Goldberg and Casenhiser, 2008). Children are exposed to large numbers of words, between 10 and 30 million by age three (Sampson, 2005, 78). Why, innatists ask, should this quantity be regarded as inadequate? Nor is it clear that the time of acquisition is all that short compared to the level of attainment, as innatists drawing an analogy to learning physics or mathematics claim. Emergentists view this as an inappropriate analogy, and suggest that a better comparison would be with learning to perform physical activities such as walking, throwing a ball, or pouring liquid (Sampson, 38). Moreover, there is growing evidence that the language learning process is consistent with a usage-based memory-prediction framework. Furthermore, although innatists have long assumed that children receive poor quality language, empirical studies have undermined this claim by showing that adult input to children is surprisingly well formed (MacWhinney, 2004, 884). One study of "motherese" found that only one utterance out of 1,500 could be considered nonfluent (Newport et al., cited in Sampson, 2005, 43). According to the memory-prediction framework, frequency of input is a key variable in understanding how form-meaning associations are created by the cortex. This conclusion is well established by emergentists (Bybee, 2008; Ellis, 2002; Lieven and Tomasello, 2008; and Mellow, 2006).

Further support for the memory-prediction framework comes from connectionist approaches, which have successfully demonstrated how neural networks can exhibit many of the properties observed in language learning. In connectionist models, neural links, simulated using computer software, become strengthened or weakened by activation or non-activation.[2]

The Critical Period Hypothesis

All theories of language acquisition must explain why second language acquisition (L2A) is so much more difficult than first (L1A). Most innatists accept the critical period hypothesis, whereas emergentists reject it, arguing that the basic learning mechanisms in the brain are the same for all kinds of learning at any age. The memory-prediction framework strongly supports the emergentist perspective on the critical period. Birdsong recently (2006) reviewed evidence for L2A age effects. Although more than two dozen experiments show an overall negative correlation between the age at which a learner is immersed in L2 and the level of ultimate L2 proficiency, the results are inconsistent when early arrivals are compared to late arrivals. In addition, the timing of the effects does not match with the theory, with the decline in performance coming as much as ten years after the end of puberty. Birdsong concludes that "the behavioral data are generally inconsistent with either a period of peak sensitivity whose end coincides with the end of maturation or with a leveling off of sensitivity whose beginning coincides with the end of maturation" (19).

Moreover, "the facts relating to native-like attainment in L2A do not lend themselves to simple generalization" (Birdsong, 2006, 19). Although innatists assume that native fluency is virtually impossible for L2 learners, more recent studies suggest that, when length of residence and contact with native speakers are controlled for, native fluency, though atypical, is not rare (20). Even in the area of pronunciation, studies have shown that native attainment is possible. Studies comparing brain activity for L1 and L2 processing have found that different regions of the brain are used for low proficient L2 speakers, but that, consistent with memory-prediction models, these differences converge as L2 proficiency increases.

Nonetheless, there is no question that aging affects all types of learning. Performance begins to decline in early adulthood, around age 20, well after the age of maturation, and remains generally linear and continuous for the rest of one's life (Birdsong, 2006, 29). Some of these factors influence second language use. However, neurological research does not generally support the critical period hypotheses of innatists, or the notion that native L2 attainment is impossible. The reality is much more complex, with different factors at work.

What explanations do emergentists propose to explain the difficulty of L2 acquisition, and how well do these fit within the memory-prediction framework? Their answers begin by assuming that L1 already has been entrenched in the brain and therefore creates interference and transfer problems. For emergentists, the initial state for L2A is *tabula repleta* not *tabula rasa* (Ellis, 2006a). Ellis (2006b) describes how interference occurs, invoking factors of cue competition, salience, interference, overshadowing and blocking, and perceptual learning. The Competition Model developed by MacWhinney (2001) posits competition among input cues of varying reliability and availability.

Ellis (2006b) emphasizes the well documented effect of proactive inhibition. Prior learning inhibits new learning, as for example, the knowledge of old phone numbers makes it harder to learn new ones. Ellis also explains two related concepts, overshadowing and blocking, which together play a role in shaping how people pay attention to language. When two cues jointly predict an outcome, the less salient cue is overshadowed, minimizing its effect. Eventually, overshadowing leads to blocking: the less salient cue is actively ignored. L2A learners often face redundant cues, such as the adverb "yesterday" coinciding with the past-tense marker "ed." Since the adverb is more salient, the past-tense marker may be blocked. Finally, Ellis shows how learners become more sensitive to cues or stimuli that are psychologically significant and less sensitive to redundant ones that do not play an important role in comprehension. Adult Japanese speakers have difficulty distinguishing the English phonemes /r/ and /l/. Because the two sounds represent the same phoneme in Japanese, the cortex learns to ignore the difference in the sounds. This outcome is precisely what emergentists would predict: if the cortex has already stored representations of a particular phonetic pattern that have reliably predicted the correct meaning in the past, it will take considerable time and repetition to establish new synaptic links.

Automatization and Attention

For emergentists, automaticity and attention are central to learning language. In fact, achieving fluency can be understood as the process of transitioning from conscious attention to language forms to using language effortlessly and "automatically," not unlike the process of learning to type or drive a stick-shift. Automatic tasks can be hard to suppress, as many people have experienced after moving to a new house and finding themselves driving to the old address. The automatic process of learning a language does not appear to be any different from other cognitive domains (DeKeyser, 2001, 150). "Chunking" closely related to automaticity, is defined as "the development of permanent sets of associative connections in long term storage" (Ellis, 2001, 38). Newell (1990) argues that chunking is the "overarching principle of human cognition: A chunk is a unit of memory organization, formed by bringing together a set of already formed chunks of memory and welding them together into a larger unit. Chunking implies the ability to build up such structures recursively, thus leading to a hierarchical organization of memory" (quoted in Ellis, 2001, 38-39). As language learners are exposed to input, both visual and phonological, they detect that certain sequences occur together with greater frequency than others. Regular sequences are perceived as chunks and associated with particular meanings. More advanced learners recognize chunks at higher levels. The same process applies at higher levels in recognizing idioms, stock phrases, and collocations.[3] In fact, some researchers argue that a large proportion of language production consists of the retrieval and recombination of such chunks. Speaking with native fluency requires knowing which collocations to use in which situations; if sentences were generated only by rules, as generative linguists assume, then one would expect much more diversity of usage than actually occurs.

Hawkins describes how the cortex stores "sequences of sequences," exactly the way Newell describes chunks in the reference above, and argues that the two basic components of learning are "forming the classifications of patterns and building sequences" (Hawkins, 2004, 165). A young brain is slower to recognize inputs because the relevant memories are stored higher up in the cortical hierarchy, and information has to flow all the way up and down many times to resolve conflicts. But with repetitive learning, representations of objects move down the cortical hierarchy and can be recognized with much less effort, allowing higher regions to learn more complex and subtle patterns. Thus, in learning music, one initially focuses on every note, but over time recognizes common sequences and then entire phrases. The same holds true with language.

Because the sequences are stored in hierarchical relationships, higher regions of the cortex can maintain more abstract information while lower regions are focusing on details. The memory-prediction framework posits that each region only passes up an invariant representation of whatever pattern it recognizes. This allows the brain to process ambiguous information and still retrieve the correct memory and pass it to the next region. Thus, the brain can "fill in" missing details such as words or morphemes only partially heard.

For emergentists, attention, or noticing, plays a critical role in language learning and processing. Schmidt (2001) insists that "the concept of attention is necessary in order to understand virtually every aspect of second language acquisition" (3). Without attention, input is a meaningless stream of speech, incomprehensible to the L2 learner.[4] Factors that affect attention include perceptual salience, frequency, awareness of gaps between learner and native-speaker speech, and any other factor that makes the speaker consciously aware of some aspect of the input. Indeed, the brain, emergentists assume, has limited capacity for attention and therefore is selective in its focus. Beginning learners thus tend to be cognitively overloaded, and only as basic skills become automatic can they focus on other details and cues in the language input.

Hawkins explains how this process works in the cortex (159-160). Each region of the brain is constantly receiving input and using known sequences of patterns to predict the next.[5] When an unexpected pattern arrives, it is automatically passed to the next higher cortical region. The higher region may be able to understand the new pattern, but if not, it moves up the cortical hierarchy until a higher region can interpret it as a predictable pattern. Then, a new prediction is sent back down. If that prediction also fails, the process repeats itself. The difficulty of predicting the next item in an unfamiliar pattern is what people experience as "noticing," or in the context of language, something "not sounding right" or "not making sense." At the very top of the cortical hierarchy sits the hippocampus, which is essential for forming new memories.[6]

CONCLUSION

Hawkins, of course, acknowledges that the memory-prediction framework is a theory that will inevitably be revised and enhanced as more is learned about the brain. However, unless emergentism is wildly off the mark, it is difficult to reconcile it with innatism. Memory-prediction theory, which emphasizes spatial and temporal patterns of input, hierarchical memory structures that store sequences of sequences, sequential processing, invariant representations, and predictive feedback loops that notice new input patterns, provides a strong neurological basis for understanding, and illuminating, emergentist, usage-based views of language.

Unlike innatist theories, which pay relatively little attention to the process of learning, emergentism is explicitly processional (Ellis, 2001; and MacWhinney, 2006). Consequently, it has clear pedagogical implications for second language instruction. Specifically, it provides a theoretical rationale for the value of explicit language instruction. A considerable body of research now shows that focused L2 instruction is effective and has long lasting results (Ellis, 2005, 307). Indeed, selective attention can raise learners' awareness of low salience features of the language and help them accurately generalize, hastening the language learning process. Form-focused instruction can also enrich and strengthen the links between words and concepts. Additionally, pedagogical approaches that lower factors that may block attention and increase learners' engagement and motivation are likely to be effective. Similarly, encouraging language production and interaction with native speakers enhances learning.

Emergentists acknowledge that much of language learning is implicit, and that sufficient quantities of appropriate input are essential for language learning to become automatic and fluent. However, instructors can maximize the value of input by emphasizing high frequency words and repeating them in different contexts. They can also recognize the importance of teaching language in chunks, presenting new words with their most frequent collocations and phrases. Indeed, emergentism provides a strong theoretical framework for a balanced approach to language teaching that incorporates both explicit attention to form and as much input as possible in realistic, communicative contexts. This approach, long intuitively embraced by second language teachers, is fully consistent with the latest advances in neuroscience.

NOTES

1. Emergentism is closely tied to a cluster of related views including associationism, connectionism, constructionism, and cognitivism. This paper uses the term emergentism unless a more specific term is called for.
2. Ellis (2001) summarizes the results of a number of these studies, beginning with Kiss (1973) who showed how the concept of word classes (noun, verb, adjective, etc.) could emerge from a simple model of associative learning. Elman (1990) is particularly interesting from the memory-prediction perspective. This computer model attempted to predict the next letter in a sequence of input letters, which is consistent with Hawkins's explanation of the role of prediction. The models showed evidence of identifying word boundaries in the input and discover word classes. Other influential connectionist models have demonstrated how associative networks relying on parallel distributed processing can reproduce the acquisition of inflectional morphology, such as regular and irregular past tense forms (Ellis, 2001, 54-56).

3. Other terms used by researchers for the same concept include holophrases, prefabricated routines, formulaic speech, lexical phrases, and formulas.
4. Corder (1967) uses the term "intake" for input that is internalized by the learner. Gass and Selinker call this "apperceived input" (Gass & Selinker, 2008, 482). Ellis (2005) treats in depth the interactions of explicit and implicit language knowledge.
5. This phenomenon, by which neurons become active in advance of receiving the next input, is called "priming" by linguists and psychologists and is well established experimentally.
6. Research in L2 acquisition repeatedly has demonstrated the importance of social and cultural factors, as well as attitude and motivation. Unlike innatists, who focus on linguistic structure without paying much heed to external factors, emergentists incorporate these factors into their models through the concept of selective attention. Learners can choose to what they pay attention, which may appear irrational to observers, but is rational from the perspective of the learner, especially when context and environment are considered. These factors can account for individual differences in L2A.

REFERENCES

Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning, 56* (Supplement 1), 9-49.

Bybee, J. (2008). Usage-based grammar and second language acquisition. In P. Robinson & N. C. Ellis (eds.), *Handbook of cognitive linguistics and second language acquisition.* (216-236). New York: Routledge.

Corder, S. P. (1967).The significance of learners' errors. *International Review of Applied Linguistics, 5,* 161-170.

DeKeyser, R. M. (2001). Automaticity and automatization. In P. Robinson (ed.), *Cognition and second language instruction* (125-151). Cambridge: Cambridge University Press.

Ellis, N. C. (1998). Emergentism, connectionism and language learning. *Language Learning, 48* (4), 631-664.

Ellis, N. C. (2001). Memory for language. In P. Robinson (ed.), *Cognition and second language instruction* (33-68). Cambridge: Cambridge University Press.

Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition, 24,* 143-188.

Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition, 27,* 305-352.

Ellis, N. C. (2006a). Cognitive perspectives on SLA: The associative-cognitive CREED. *AILA Review, 19*, 100-121.

Ellis, N. C. (2006b). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics, 27* (2), 164-194.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14,* 179-212.

Gass, S. M. & Selinker, L. (2008). *Second language acquisition: An introductory course*. (3rd. ed.). New York: Routledge.

Goldberg, A. E. & Casenhiser, D. (2008). Construction learning and second language acquisition. In P. Robinson & N. C. Ellis (eds.), *Handbook of cognitive linguistics and second language acquisition.* (197-215). New York: Routledge.

Hawkins, J., with Blakeslee, S. (2004). *On intelligence.* New York: Times Books, Henry Holt and Company.

Ke, J. & Holland, J. H. (2006). Language origin from an emergentist perspective. *Applied Linguistics, 27* (4), 691-716.

Kiss, G. R. (1973). Grammatical word classes: A learning process and its simulation. In G. G. Bower (ed.), *The psychology of learning and motivation: Advances in research and theory:* Vol. 7. New York: Academic Press.

Lieven, K. & Tomasello, M. (2008). Children's first language acquisition from a usage-based perspective. In P. Robinson & N. C. Ellis (eds.), *Handbook of cognitive linguistics and second language acquisition.* (168-196). New York: Routledge.

MacWhinney, B. (2001). The competition model: the input, the context, and the brain. In P. Robinson (ed.), *Cognition and second language instruction* (69-90). Cambridge: Cambridge University Press.

MacWhinney, B. (2004). A multiple process solution to the logical problem of language acquisition. *Journal of Child Language, 31*, 883-914.

MacWhinney, B. (2006). Emergentism—Use often and with care. *Applied Linguistics, 27* (4), 729-740.

Mellow, J. D. (2006). The emergence of second language syntax: A case study of the acquisition of relative clauses. *Applied Linguistics, 27* (4), 645-670.

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

Sampson, G. (2005). *The 'language instinct' debate.* (Revised ed.). London: Continuum.

Schmidt, R. (2001). Attention. In P. Robinson (ed.), *Cognition and second language instruction* (3-32). Cambridge: Cambridge University Press.