SSLA, **24**, 297–339
DOI: 10.1017.S0272263102002140

REFLECTIONS ON FREQUENCY EFFECTS IN LANGUAGE PROCESSING

Nick C. Ellis
University of Wales, Bangor

This response addresses the following points raised in the commentaries: (a) complementary learning mechanisms, the distinction between explicit and implicit memory, and the neuroscience of "noticing"; (b) what must and what need not be noticed for learning; (c) when frequency fails to drive learning, which addresses factors such as failing to notice cues, perseveration, transfer from L1, developmental readiness, thinking too hard, pedagogical input, and practicing; (d) attention and form-focused instruction; (e) conscious and unconscious knowledge of frequency; (f) sequences of acquisition—from formula, through low-scope pattern, to construction; (g) the Fundamental Difference hypothesis; (h) the blind faith of categorical grammar; (i) Labovian variationist perspectives; (j) parsimony and theory testing; (k) universals and predispositions; and (l) *wanna*-contractions. It concludes by emphasizing that language acquisition is a process of dynamic emergence and that learners' language is a product of their history of usage in communicative interaction.

What you seize is what you get. There is more to the interpretation of a journal paper than meets the eye, too. The diversity in these commentaries reminds me of Doris Lessing's (1973) reactions to the range of letters from readers of her *Golden Notebook*:

Thanks to Neil Cheshire, Bill Croft, Albert Dudley, Kim Ellis, Ginny Gathercole, Dick Hudson, Carl James, Kathryn Kohnert, Julian Pine, Dennis Preston, Karen Roehr, Joyce Tang Boyland, Marilyn Vihman, and John Williams for their various helpful contributions. Thanks also to Albert Valdman, Bruce Anderson, and *SSLA* for creating the opportunity, and to the commentators for engaging so productively in this exchange.

Address correspondence to: Nick C. Ellis, School of Psychology, University of Wales, Bangor, Gwynedd LL57 2DG, UK; e-mail: n.ellis@bangor.ac.uk.

© 2002 Cambridge University Press 0272-2631/02 $9.50

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

> These incidents bring up again questions of what people see when they read a book, and why one person sees one pattern and nothing at all of another pattern, and how odd it is to have, as author, such a clear picture of a book, that is seen so very differently by its readers. (p. xxi)

Distinct experiences, distinct associations, distinct schemata, personal construct systems, distinct reminders: I am grateful to the commentators for their reflections, refractions, and foci that take forward the issue of frequency in language acquisition and representation, and I thank *SSLA* for the opportunity for this exchange. The range of the commentaries suggests that it is profitable to have frequency back on the agenda. I will address those arguments that I can before I run out of space.

POINT 1: COMPLEMENTARY LEARNING MECHANISMS—EXPLICIT AND IMPLICIT MEMORY AND THE NEUROSCIENCE OF "NOTICING"

My article concentrated on frequency effects in implicit learning, although for balance it reminded readers of the involvement of other conscious systems of cognition—explicit learning and memory for the initial registration of new information, and declarative memory systems, which allow the learning of pedagogical rules and explicit instruction:

> To the extent that language processing is based on frequency and probabilistic knowledge, language learning is implicit learning. This does NOT deny the importance of *noticing* (Schmidt, 1993) in the initial registration of a pattern-recognition unit. NOR does it deny a role for explicit instruction. (p. 145)

I did not know how to make NOT and NOR any bigger, but nevertheless the spotlight on implicit learning has obviously overshadowed the rest. Bley-Vroman reminds us that "Many things that are encountered only once or very rarely may strike the learner as salient, be noticed and processed deeply, and be incorporated into linguistic knowledge" (p. 213). This role of salience resonates through the commentaries of Larsen-Freeman and R. Ellis; Gass and Mackey discuss areas of SLA in which input is infrequent but acquisition is unimpeded; Eubank and Gregg remind us of instantaneous (one trial) learning and fast mapping; and Hulstijn makes the same point about a single encounter sometimes being sufficient for a new word (form-meaning connection) to be remembered forever. Understanding the ways in which explicit and implicit learning interact in language acquisition is a major and longstanding goal of our discipline, hence my editing of a book on this topic (Ellis, 1994a). It was kind of Hulstijn to begin his conclusions with an evaluation of the reasons for my current comparative neglect of explicit learning. He has my motivations to a T. It is clearly fitting to redress the balance now and to discuss the synergies between explicit and implicit learning and memory. The remainder of this first

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

section concerns the neuroscience of "noticing." This might seem overly psychological for some readers, using concepts perhaps remote from conventional SLA discourse. If this proves so and you feel like you are losing the plot, then skip this section and move on to more traditional issues. Please return here later, though: We are now at a stage at which there are important connections between SLA theory and the neuroscience of learning and memory.

Humans have two separable but complementary memory systems (Squire & Kandel, 1999). Understanding these has been a major psychological research effort over the last two decades: PsychINFO shows six articles on "implicit and explicit memory" between the years 1887 and 1966, 32 articles between 1967 and 1983, and, as I write, 1,077 articles since 1984. Explicit memory refers to situations in which recall involves a conscious process of remembering a prior episodic experience; it is tapped by tasks like recall and recognition during which the individual is consciously aware of the knowledge held. Examples include your answers to *What did you have for breakfast?*, *What's your dog called?*, *Tell me the story of* It's a Wonderful Life, and *Who are you? Tell me* your *story*. Implicit memory refers to situations in which there is facilitation of the processing of a stimulus as a function of a prior encounter with an identical or related stimulus but in which the subject at no point has to consciously recall the prior event; it is tapped by tasks like perceptual priming or in procedural skills—you don't have to remember when you last juggled to have improved as a result of the practice. Implicit and explicit memory are clearly dissociable: Bilateral damage to the hippocampus and related limbic structures results in profound anterograde amnesia, a failure to consolidate new explicit memories, along with a temporally graded retrograde amnesia. Both verbal and nonverbal memory are affected: Amnesic patients cannot recognize people met (perhaps many times) since their brain lesions, they cannot learn new names or concepts or arbitrary paired-associates, and they cannot remember any episode more than a few minutes after it has happened. Nevertheless, amnesic patients show normal implicit memory abilities: They learn new perceptual and motor skills, they show normal priming effects, and they evidence normal classical conditioning.

Thus, the hippocampus and related structures serve explicit memory, declarative learning (e.g., of verbal rules like "*i* before *e* except after *c*"), one-trial learning of randomly paired associates like *locomotive-dishtowel*, that the Welsh for *hawk* is *hebog*, that *hebog* is the name of our new house, and so forth. The hippocampus supports our autobiographical record of specific episodes. In contrast, there are the memory systems of the neocortex, including relatively peripheral primary sensory-input and motor-output systems, secondary association areas, as well as more central, highly interconnected frontal areas. The neocortical system underpins implicit learning and is the locus of the frequency effects discussed in my article. Whenever a stimulus is presented to our senses—say, a visually presented word—it produces a pattern of activity in the appropriate sensory system. This in turn gives rise to activity in the more central parts of the neocortical system, including those perhaps repre-

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

senting the visual appearance, meaning, and sound of the word. This in turn may give rise to an overt response, such as reading the word aloud. Any such event, any experience, produces a distributed pattern of activity in many parts of the cognitive system, and the information processing that we do occurs through the propagation of this activation through networks of neurons whose connection strengths have been tuned by prior experience. The neocortex underpins both the perception and the implicit memory of past experiences: We perceive the world through our memories of the world. Implicit memory is the result of small changes that occur in the synapses among the neurons that participate in this processing of the event. These small changes tend to facilitate the processing of the item if it is presented again at a later time, though the changes that are made on any given processing episode or event in the neocortex (as in the connectionist simulations of this implicit learning) are very subtle and, as such, are insufficient to serve as the basis for forming adequate associative links between arbitrarily paired items that have never occurred together before, new concepts, or new episodic records.

The hippocampal system subserves rapid explicit memory, one-off learning, the establishment of new conjunctions of arbitrarily different elements (Squire, 1992), the learning of separate discrete episodes—what *you* had for breakfast *this morning*—or the rote learning of the particular wording of Darwin's first usage of *profound ignorance* in his introduction to the *The Origin of Species* (1928):

> No one ought to feel surprise at much remaining as yet unexplained in regard to the origin of species and varieties, if he makes due allowance for our profound ignorance in regard to the mutual relations of all the beings which live around us. (p. 20)

For something as important as this, we do not want an average, an abstraction, or a gist: There is benefit in being able to keep some records straight, complete, and distinct.

The neocortical systems, in contrast, subserve implicit memory (Schacter, 1987), the gradual gathering of information about the mutual relations of all the beings involved in neuronal patterns of activation, the tuning of associative systems to reflect repeated patterns of activity and to generalize from them, semantic memory rather than episodic memory, the general satisfaction of breakfast, the prototypical meaning of *profound*, and the profundity of the notions of natural selection and emergence. To operate efficiently in the world we need to be able to identify general patterns by abstracting from instances; we need to classify and categorize.

Does the anterograde amnesiac's inability to consolidate new explicit memories have any consequences for their implicit learning? Well, no, and yes: Amnesiacs show normal levels of priming for stimuli whose pairing was already established before lesion—that is, for unitized preexisting memory representations. After being shown lists of common idioms (e.g., *sour-grapes*) or highly

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

related paired associates (e.g., *table-chair*), a short while later they will have no conscious memory of having seen them, and they will evidence their defining impairment of explicit memory if you ask them "What word went with *table*?" Yet if you tap their knowledge with an implicit task, say by using a free-association question such as "What is the first word that comes to mind when I say *table*?", they show normal levels of priming in that they are more likely to say *chair* than if they had not had that recent, yet unremembered, experience. They similarly show normal priming effects for other unitized preexisting memory representations when tested with lexical decision or perceptual identification tasks (see Schacter, 1987, for review)—thus, deficient explicit memory but preserved implicit memory. However, severe amnesiacs do not show priming effects for novel arbitrary associations. After being shown arbitrary paired-associates like *bell-cradle* as an example of information that does not have any preexisting representation as a unit in memory, again they forget the exposure and cannot later bring the associated response explicitly to mind, but, additionally, severe amnesiacs fail to show implicit memory for these new associations. (The balance of results is clearly in this direction, but see Dunn, 1998, pp. 104–105 for review of somewhat mixed effects; Schacter, 1987, p. 509; Squire, 1992, pp. 212–214.) To exhibit priming of new associations between two semantically unrelated words, subjects may need to access a link between the two words that was formed explicitly at the time of study. Deficient explicit memory for novel associations produces deficient implicit memory in consequence.

Such results are consistent with the idea that implicit learning is specialized for incremental cumulative change—the tuning of strengths of preexisting representations. New associations are best learned explicitly. If they are acquired implicitly, it is only after many, many repetitions:

> Amnesic patients with hippocampal damage should eventually be able to acquire new associations through repetition, as in the development of a habit. However their rate of learning would be grossly abnormal in comparison with normal subjects, and the acquired knowledge would be abnormal in other respects as well . . . for example, the knowledge should be relatively inflexible, that is accessible only when exactly the same cues are presented that were used during training. (Squire, 1992, p. 214)

Amnesia is not a failure to notice; it is a failure to consolidate an explicit memory as a result of noticing. Devoid of the memorial advantages of noticing, amnesiacs show normal implicit memory and tuning of their perceptual and motor systems for unitized preexisting memory representations but grossly slow acquisition of new associations. The parallels with the rate of SLA in the absence of noticing should not themselves go unnoticed. Schmidt's paradigm such case, Wes, was very fluent, with high levels of strategic competence but low levels of grammatical accuracy. He was described as being interested in the message, not the form, and as being impatient with correction. In discuss-

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

ing Wes's unconscious naturalistic acquisition of ESL in the five years since coming to America, Schmidt (1984) wrote:

> If language is seen as a medium of communication, as a tool for initiating, maintaining and regulating relationships, and carrying on the business of life, then W has been a successful language learner. . . . If language acquisition is taken to mean (as it usually is) the acquisition of grammatical structures, then the acquisition approach may be working, but very slowly. . . . Using 90% correct in obligatory contexts as the criterion for acquisition, none of the grammatical morphemes counted has changed from unacquired to acquired status over a five year period. (p. 5)

I will return to these issues of noticing and what can and cannot be learned without it in the following two sections.

The neurophysiological processes of consolidation of memories take place over an extended period of days. If an animal or human has an experience on a given day, and then the hippocampus is removed bilaterally immediately thereafter, there is a nearly total loss of memory for that experience. However, if the hippocampus is left intact for a period of time after the initial experience and then is removed, the subjects will show gradual increases in the degree of retention—this is Ribot's law (1882); hippocampal damage leads to a temporally graded retrograde amnesia. It is by bringing together the role of the hippocampus in—and the gradual time course of—consolidation, the differences between one-off episodic learning and gradual implicit learning as discussed by the present commentators, and the observations of unnatural catastrophic interference in connectionist networks, that McClelland (1998, 2000, 2001; McClelland, McNaughton, & O'Reilly, 1995) developed a cognitive neuroscience theory of the complementary interactions of hippocampal and neocortical learning systems. This suggests that memories are first registered via synaptic changes in the hippocampus involving a sparse pattern of activity in which the individual neurons represent specific combinations or conjunctions of elements of the event that gave rise to the pattern of activation. These changes support reinstatement of recent memories in the rich and highly distributed networks of activation in the neocortex, the neocortical synapses change a little on each reinstatement, and remote memory is based on accumulated changes. Models that learn via changes to connections help explain this organization. These models discover the structure in ensembles of items if learning of each item is gradual and interleaved with learning about other items. This suggests that the neocortex learns slowly to discover the structure in ensembles of experiences. The hippocampal system permits rapid learning of new items without disrupting this structure, and reinstatement of new memories interleaves them with others to integrate them into structured neocortical memory systems.

Any 1,500-word summary of the relations between implicit and explicit memory needs an accompanying health warning: These issues are cutting edge and contentious, and psychology is no more protected than SLA from

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

the law that the more closely anything is examined the more complex it is seen to be. Nevertheless, these issues lie at the heart of both cognitive science and language acquisition. We stand a better chance of understanding them through our concerted efforts, and what links that already exist seem promising. For the moment, my best gloss is as follows: Noticing is important in the initial registration of memory traces. In terms of the usual naturalistic sequence of language acquisition proposed in my article, this first registered form is the formula. We need the hippocampal explicit memories for that. It is neocortical implicit learning, however, that allows the figuring, the generalization from variants on the theme, and the ultimate convergence on rulelike behavior, fluency, and nativelike performance.

POINT 2: WHAT MUST AND WHAT NEED NOT BE NOTICED FOR LEARNING?

Larsen-Freeman, R. Ellis, Bley-Vroman, Hulstijn, Gass and Mackey, heck, just about everyone stresses the importance of attention in learning. Research into the dissociations between implicit and explicit systems now spans experimental investigations of implicit and explicit learning of artificial grammars, complex control tasks and serial patterns (serial reaction time), as well as dissociations of implicit and explicit learning and memory systems in neuropsychological loss and in brain imaging studies. The contributions in N. Ellis (1994a) demonstrated how SLA and psychological research can profitably collaborate in the identification of what must and what need not be noticed for learning to take place. I will try here to briefly summarize some key findings since that date and to outline remaining controversies.

My target article sampled the body of psycholinguistic work showing implicit memory for preexisting representations. The crux of the matter clearly rests on what is a "unitized preexisting memory representation" and what is not. It seems clear that the activation of existing mental structures (representing letters, letter clusters, sounds, sound sequences, words, word sequences, grammatical constructions, etc.), whatever the depth of processing or the learner's degree of awareness as long as the form is attended to for processing, will result in facilitated activation of that representation in subsequent perceptual or motor processing.

However, in normal individuals it looks like noticing is necessary for new semantic associations to be established after just one or a few learning opportunities. Bowers and Schacter (1990) examined the degree to which people could perform implicit memory tests without noticing that the test items were previously encountered in the study phase of the experiment. Experiment 1 assessed priming of known single words using a stem completion task: In the study phase they saw a list of words (e.g., *reason*); at test they were asked to complete the stem (e.g., *rea__*) to make an English word. Implicit memory was indicated if participants previously exposed to the prime were relatively more likely to complete the stem with *reason* (rather than any other, e.g., *reading*,

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

reader, *reality*, etc.) than other participants who were not previously exposed. In this experiment, participants who reported awareness and those who reported unawareness that the test items were previously encoded in the study task showed equivalent priming. Experiments 2 and 3 assessed associative priming with the stem completion task (e.g., study *window-reason*, *officer-printer*; test *window-rea__* and *officer-rea__*). Implicit memory was shown if participants were more likely to complete with the word *reason* in the original context than in the reordered one. In this case, only subjects who were aware that the test items had been previously encountered showed associative priming effects. This clearly parallels the findings discussed previously, in which amnesiacs showed normal implicit memory for single words but did not show implicit memory for novel, arbitrary associations.

In Ellis (1994b) I summarized the existing research concerning lexical and sublexical levels of processing under the title "The Implicit Ins and Outs of Explicit Cognitive Mediation": You cannot learn the meaning of words without noticing the reference, but the surface forms of vocabulary are generally acquired implicitly. I still believe this to be a valid generalization, although there are exceptions even in L1. Failures of implicit learning of spelling irregularities make good examples: When I was a doctoral student, despite my research topic of developmental dyslexia and my many previous encounters with the word, it still took an overt correction from my supervisor to force a shift away from my idiosyncratic spelling *developemental*. It is the exceptions to more general patterns that require the noticing engendered by the direct negative evidence of correction, spell-checkers, or pedagogical rules.

Equally, there can be implicit learning of associations that are not unitized and preexisting and that have not been noticed. The work on implicit grammar learning shows this to be the case (for reviews, see Berry & Dienes, 1993; Cleeremans, Destrebecqz, & Boyer, 1998; Reber, 1993; Seger, 1994; Stadler & Frensch, 1997). Two separate unitized and preexisting representations that occur repeatedly in the same sequence, again and again, can accrete into one chunk even if their conjunction is not noticed. Hebb (1961) demonstrated that, when people were asked to report back random nine-digit sequences in a short-term memory task, if, unbeknownst to the participants, every third list of digits was repeated, memory for the repeated list improved over trials faster than memory for nonrepeated lists. The Hebb effect is the central mechanism of exemplar-based, implicit chunking accounts of linguistic form (Ellis, in press; Gobet et al., 2001; Perruchet & Pacteau, 1990; Redington & Chater, 1996; Servan-Schreiber & Anderson, 1990). Amnesiacs show as accurate and fast acquisition as normal controls in implicit learning of artificial grammars (Knowlton & Squire, 1996; Reber & Squire, 1998) and serial reaction time tasks alike (Dunn, 1998; Nissen, Willingham, & Hartman, 1989). The key determinants of implicit learnability here seem to involve adjacency and many repetitions.

Bley-Vroman (p. 210) acknowledges the phenomenon of collocations—"language production does have a statistical structure"—but worries that

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

these statistical facts are secondary and derivative of the really important stuff of meaning and that they have "little direct explanatory force." As my target article demonstrated again and again, however, knowledge of such statistical structure does indeed influence language processing. If there is need for yet one more example, this time specifically for collocations whose frequencies cannot have been explicitly tallied but whose strengthening has resulted from repetition in processing, consider the demonstrations of Schooler (1993; Schooler & Anderson, 1997; Anderson & Schooler, 2000) performed under the auspices of Anderson's (1991a, 1991b) Rational Analysis of Cognition model that I will discuss under point 5. The probability of a word occurring in, say, speech to children (from the CHILDES database), the *New York Times*, or the e-mail a person receives is predicted by its past probability of occurrence: There is a power (log-log linear) function relating probability of a word occurring in the headline in the *New York Times* on day *n* to how long it has been since the word previously occurred. Human memory is sensitive to this; thus, the probability of recalling an item, like the speed of its processing or recognition, is predicted by time since past occurrence and past occurrence frequency. There is a power function relating probability of recall (or recall latency) and retention interval, and these factors are similarly a joint power function of retention interval and frequency. So far, so rational—memory performance is tuned to the world. This is the power law of practice I describe in my target article. On top of this, however, there are effects of immediate context. In his analysis of the *New York Times* and the CHILDES databases, Schooler (1993) showed that a particular word was more likely to occur when other words that had occurred with it in the past were present. For instance, a headline one day mentioned Qaddafi and Libya, and sure enough a headline the next day that mentioned Qaddafi also mentioned Libya, in the same way that Bley-Vroman observed that *profound* and *ignorance* go together in the writings of Darwin as an example of collocations and the idiom principle. Schooler collected likelihood ratio measures of association between various words to assess the effect of this local context factor on memory and processing. As already described, in both the child language and the *New York Times* databases, a word was more likely to occur if it had occurred previously, but additionally, a word was more likely to occur in a headline if a string associate of it occurred, and these effects are additive in the way predicted by Bayesian probability. These context effects affect human processing time, too: Schooler showed that word fragment completion was faster for the second word of a strong context collocation (as in *profound-ig__*) than when the word was shown alone (*ign__*). The bottom line is that we process collocates faster, and we are more inclined therefore to identify them as a unit. These processing effects are crucial in the interpretation of meaning: It is thus that an idiomatic meaning can overtake a literal interpretation, and that familiar constructions can be perceived as wholes.

Bley-Vroman and Harrington and Dennis will be interested to note that there is an important role of subject encoding strategy here, which is why

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

Tulving's retrieval theory was originally called "encoding specificity": A critical factor in getting context effects is whether the subject encodes the memory with the context at study (Tulving & Thomson, 1973; see also Eich & Metcalfe, 1989). There is so much context that we cannot process everything, and thus statistical co-occurrence is not enough to ensure an association. A conjunction may not need to be noticed, but it does at least need to be processed.

Associations that are more complex than adjacency or immediate succession in artificial grammar learning experiments do require more conscious explicit learning and hypothesis testing to acquire. The experiments of Ellis, Lee, and Reber (1999) provided evidence that this is the case for some long-distance discontinuous dependencies in language acquisition. Cohen, Ivry, and Keele (1990) and Curran and Keele (1993) showed that unique sequences can be acquired implicitly in artificial grammar learning experiments, whereas ambiguous sequences require more attentional forms of learning. Likewise, Gomez (1997) demonstrated that learning can occur without awareness in cases of lesser complexity, such as learning first-order dependencies in artificial languages. However, more complex learning is linked to explicit learning, such as that involved in second-order dependencies or in transfer to stimuli with the same underlying syntax but new surface features.

Once learned, the tuning is implicit. But for the initial registration, some things you do have to notice for learning, and some things you do not. Research on this basic question is ongoing and far from settled. What determines whether things get "unitized" at encoding remains a basic question of cognitive science as well as a crucial one for a theory of implicit learning of language. We should continue to investigate the outcomes of language acquisition from meaning-focused and form-focused instruction. We should continue laboratory and field experiments of what aspects of language can be learned implicitly and what cannot. We should continue to monitor the work identifying what amnesiacs can and cannot learn and what people can learn implicitly and what they cannot. We have some initial answers, but we need a more complete classification. We must interrelate these findings. A failure of noticing must clearly be one cause of cases in which, despite high frequency in the input, second language learners fail to acquire a particular pattern or feature. I consider these in the next section.

POINT 3: WHEN FREQUENCY FAILS TO DRIVE LEARNING

Several of the commentators describe instances of failed language acquisition in the face of frequent evidence. Gass and Mackey discuss developmental sequences in SLA in which, for example, third-person singular *-s* is not acquired despite instruction and feedback. They speculate that this might result from the low salience of the form or, following Pienemann (1998), that it must wait until learners have the processing capacity necessary for forms at the preceding stage of acquisition. Larsen-Freeman makes related points about develop-

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

mental readiness, where, once the learner has a framework into which the new form fits, it becomes salient; otherwise it remains noise. Gass and Mackey also discuss cases of transfer in which acquisition paths are determined by the presence or absence of L2-like structures in his or her L1. Hulstijn considers postpubescent L2 learners who speak their L2 with nonnative pronunciation even when the period of exposure to L2 has been longer than the period of exposure to L1. In what follows, seven major factors that modulate the effect of frequency are discussed.

Failing to Notice Cues

Whereas some grammatical meaning-form relationships are both salient and essential to understanding the meaning of an utterance (e.g., Spanish interrogatives *qué* "what" and *quién* "who"), others, such as grammatical particles and many inflections like the third-person singular *-s* in English, are not. Inflections marking grammatical meanings such as tense are often redundant because they are usually accompanied by temporal adverbs that indicate the temporal reference. The high salience of these temporal adverbs leads L2 learners to attend to them and to ignore the grammatical-tense verb morphemes.

This is a prime motivation for explicit instruction. Thus, for example, processing instruction (VanPatten, 1996) aims to alter learners' default processing strategies by changing the ways in which they attend to input data, thereby maximizing the amount of intake of data in L2 acquisition. Terrell (1991) is another illustrative case. He characterized explicit grammar instruction as "the use of instructional strategies to draw the students' attention to, or focus on, form and/or structure" (p. 53). His binding-access framework postulated that learners' primary motivation is to understand language and therefore that the acquisition of grammatical form comes as a result of establishing a connection between meaning and form. As in cognitive linguistic theory, Terrell considered learners to be acquiring individual meaning-form relationships rather than grammatical rules. He recommended instruction as a way of increasing the salience of inflections and other commonly ignored features by first pointing them out and explaining their structure, then providing meaningful input that contained many instances of the same grammatical meaning-form relationship.

Perseveration

Hulstijn asks us to consider postpubescent L2 learners who speak their L2 with nonnative pronunciation despite massive exposure. If an English infant is open to the /r/~/l/ phonemic contrast, why do native speakers of Japanese, in which there is only a single alveolar liquid phoneme, fail to acquire the /r/~/l/ contrast when learning ESL despite high frequencies in the input? More generally, why is it that adults can still learn and adapt many skills, yet the ability

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

to adapt the perception and production of speech appears to diminish in adulthood?

The reason may have something to do with the rate of neuronal learning slowing with age. It is much more difficult to induce long-term potentiation (LTP, widely believed to be the synaptic substrate of learning) at single synapses in older animals than in younger ones (Kirkwood, Lee, & Bear, 1995). There are rational (see point 5) reasons for this: Cognition has evolved to enable us to learn about the world enough to be able to predict reasonably well what is likely to happen next. When each experience represents but a single, stochastic sample from a population, it is necessary to aggregate over many samples to obtain a decent estimate of the population statistic. The optimal procedure for estimating population statistics online is to make relatively large adjustments in response to initial observations and then to gradually reduce the size of the adjustments as sample size increases (McClelland et al., 1995, p. 447; H. White, 1989).

However, as Hulstijn suggests, a strong alternative explanation is that of entrenchment or neural commitment: The L2 learner's neocortex has already been tuned to the L1, incremental learning has slowly committed it to a particular configuration, and it has reached a point at which the network can no longer revert to its original plasticity (Elman et al., 1996, p. 389). This is the essence of Perceptual Magnet theory (Kuhl & Iverson, 1995), in which the phonetic prototypes of one's native language act like magnets or, in neural network terms, attractors (Cooper, 1999; van Geert, 1993, 1994), distorting the perception of items in their vicinity to make them seem more similar to the prototype. This and frequency form the major two limbs of Flege's (in press) triarchic theory of L2 speech performance, in which the three factors are neurological maturation, the interactions between phonic elements in the L1 and L2 phonetic subsystems, and age-related variations in input.

McClelland (2001) presented a connectionist simulation of such effects. A Kohonen self-organizing map network was taught the mappings between phonological input patterns and phonetic representation space. When the model was trained with exemplars from two relatively distinct neighborhoods (representing /r/ and /l/), it learned separate representations and could correctly classify examples into these categories. If, however, the network had previously been trained with exemplars from one wide neighborhood representing the single Japanese alveolar liquid, thereafter it learned to treat the two /r/ and /l/ classes of input as the same and "diabolically maintain[ed] this tendency, even when faced with input that would at first have caused it to represent the classes separately" (p. 112). The way to fix this problem, to kick the network into a state in which it could relearn the separation of the inputs, was to use inputs that exaggerated the differences between them. The exaggerated inputs activated distinct representations, and once this happened the mechanisms of synaptic modification strengthened separate representations of the exaggerated items. The difference between the items could then be gradually reduced, whereas Hebbian synaptic modification maintained the L2 distinct-

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

ness rather than the old L1 similarity. It was as if the network had, by the use of exaggerated input, been made to notice a distinction that had previously lacked salience and thus had been ignored.

McClelland (2001; Thomas & McClelland, 2001) referred to remediation studies that make use of exaggerated stimuli and adaptive training in this way. There are the interventions of Merzenich et al. (1996) and Tallal et al. (1996) with language learning impaired (LLI) children who have difficulty discriminating between similar phonemes, like /b/ and /d/. Merzenich et al. and Tallal et al. showed that LLI children can successfully learn these discriminations if trained using computer-enhanced speech, wherein the relevant differences between the phonemes are exaggerated. For the case of second language phonology, there are the studies training Japanese L1 speakers on /r/~/l/ discrimination. With frequent practice alone, improvement is slow: Bradlow, Pisoni, Akahane-Yamada, and Tohkura (1997) showed that 45 hours of training, during which Japanese learners were presented with a naturally spoken word (such as *rake* or *eagle*), and then given a two-alternative forced choice between the item and a minimal contrast foil (*rake* vs. *lake*; *eagle* vs. *eager*) with feedback, generated about a 20% improvement. However, learning can be induced more rapidly than this with the use of exaggerated stimuli that the learners could discriminate from the outset. Contrasts such as *rock* versus *lock* were computer synthesized into continua, and the contrast was made more exaggerated by extending their outer limits. Participants started with these discernible poles and then, as eight stimuli in succession were correctly identified, the discrimination was made more difficult. The use of exaggerated stimuli and adaptive training led to rapid learning, whereas the use of difficult stimuli with no adaptive modification produced little or no benefit (McCandliss, Conway, Fiez, Protopapas, & McClelland, 1998; McClelland).

In terms of noticing, the provision of exaggerated input makes the learner notice and become aware of a contrast that previously went unheard. In terms of complementary memory systems, the hippocampus plays an important role in allowing the contents of recent episodic memories to override whatever preexisting associations are present in neocortical connections and to seed distinct representations that, over time, the Hebbian mechanisms of synaptic modification will strengthen to reinforce a tendency to treat the stimuli as distinct. Learners need to be made to notice in order to make processing avoid attractors that were optimized for L1 but which now serve as magnets to local minima.

Transfer from L1

Even without the literally blinding effects of L1 attractors, there is competitive interference from L1 patterns of form-function mapping. Gass and Mackey (pp. 256–257) provide several important examples of transfer. The Competition Model (MacWhinney, 1992, 2001) has been applied to transfer from L1 just as it has to direct learning of L2, analyzing both processes in terms of cue reliability, cue cost, and form-function mappings.

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

The learning of sentence processing cues in a second language is a gradual process. It begins with L2 cue weight settings that are close to L1, and only gradually over time do these settings change in the direction of the native speaker's settings for L2. MacWhinney (2001) relates how his early investigations of L2 processing examined the comprehension of English sentences by academic colleagues:

> One subject was a native speaker of German who had lived in the United States for thirty years, was married to an American, and had published several important textbooks in experimental psychology written in English. Remarkably, we found that this subject processed simple English sentences using the cue strength hierarchy of German. This is to say that he used agreement and animacy cues whenever possible, largely ignoring word order when it competed with agreement and animacy. This first evidence for the preservation of a syntactic "accent" in comprehension has now been supported in over a dozen studies across a wide variety of second language learning. (p. 80)

A clear example is the study by McDonald (1987), which compared the cue use of English-Dutch and Dutch-English bilinguals with varying amounts of L2 exposure to that of native speaker controls. For all constructions tested (dative constructions, simple noun-verb-noun sentences, and relative clauses), it was found that with increasing exposure, cue usage in the L2 gradually shifted from that appropriate to L1 to that appropriate to the L2. MacWhinney (2001) reported on the connectionist model of Johnson and MacWhinney, which successfully simulated these results using an Elman recurrent network.

Developmental Readiness and Sequences of Acquisition

The issue of fixed sequences of acquisition is fundamental to SLA research, as is the interpretation of these phenomena to SLA theory. As Larsen-Freeman describes, there appear to be common developmental sequences of certain syntactic structures despite different learner L1 backgrounds, different exposures to language, and different teaching regimes. As we have seen, for the morpheme acquisition studies at least, frequency and salience seem to play a large role in determining these sequences (Goldschneider & DeKeyser, 2001). I believe that now we have recognized the ubiquitous influence of frequency and salience, henceforth we should expect them to hold and should simply adopt them as default explanatory variables. We should routinely factor out their effects in ways analogous to forcing their entry in the first two steps of a stepwise regression: What variance remains thereafter is the interesting stuff.

The aspects of language that then fit into this category are suggestive of explanations in terms of complexity and developmental readiness. The basic idea here is that complex structures are built of prior structures: A new construction can only be acquired if learners have already acquired the relevant representational building blocks or if they have sufficient working-memory capacity, phonological short-term memory span, or other aspects of general lan-

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

guage processing resource to be able to use the structure. I discuss the role of chunking in these processes in Ellis (1996, 2001, in press). As Gass and Mackey and R. Ellis remind us, Pienemann's (1985) Teachability Hypothesis is a good example of a theory that denies any possibility that instruction can alter the natural route of development of developmental features: "Instruction can only promote language acquisition if the interlanguage is close to the point when the structure to be taught is acquired in the natural setting" (p. 37). His processability theory (Pienemann, 1998) makes formal predictions regarding the structures that can be processed at a given level of L2 learning based on psychological mechanisms that underlie interlanguage and proposes a hierarchy of SLA processing procedures in the framework of lexical-functional grammar. Other examples of well-developed theories of this type that are based on different assumptions of grammar or psychological processing mechanism include Carroll's (1999, 2001) Autonomous Induction theory, which combines a model of induction with an interpretation of Universal Grammar, and O'Grady's (1997, 1998, 1999, in press) General Nativist theory of syntactic representations, which addresses issues of learnability and development as a consequence of an innate endowment for language that does not include an inborn grammar per se but instead consists of more general processing mechanisms and principles such as the general computational features (a) a propensity to operate on pairs of elements and (b) a propensity to combine functors with their arguments at the first opportunity (a storage-reducing "efficiency" strategy). Green (1998) sketched a similarly motivated account of L1 acquisition in terms of the induction of head-driven phrase structure grammar.

These are important developments. For me, their promise lies in their minimalism and commitment to an emergentist explanation. They remain strong as long as (a) they render to the input that which properly can be explained by the input, (b) they maintain an ecological view of grammar as a collaboration of constructions and eschew notions of transformation, movement, and the like, and (c) they permit a continuing interplay between top-down and bottom-up processes and between memorized structures and more open constructions:

> Language, as a complex, hierarchical, behavioral structure with a lengthy course of development . . . is rich in sequential dependencies: syllables and formulaic phrases before phonemes and features . . . , holophrases before words, words before simple sentences, simple sentences before lexical categories, lexical categories before complex sentences, and so on. (Studdert-Kennedy, 1991, p. 10)

Each of the language subsystems develops hierarchically by repeated cycles of differentiation and integration.

Thinking Too Hard

I'll give examples of this for second, first, and artificial language acquisition. These typically concern learners' noticing the wrong things, generating false hypotheses, and chasing hares.

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

Schmidt and Frota (1986) gathered instances in which an L2 learner's developing conscious understanding of the forms and functions of Brazilian Portuguese (recorded in a diary) matched the learner's performance in recorded interview data. These data included cases in which incorrect use could be traced to specific misanalyses of what was heard in input.

Peters and Menn (1993) described the false starts in Menn's son Daniel's native English acquisition. Between the ages of 2;0 and 2;9, Daniel slowly discovered the morphological status of English word-final sibilants. He used these sibilants way before he properly understood their function, and the early result was that he used a highly idiosyncratic phonologically conditioned rule that added [s ~ z] to the end of nonreduplicated two-syllable words ending in /r/ or having medial consonants and ending in /i/, and variably to other two-syllable words and monosyllables of a particular shape. These sibilants were added regardless of whether they were plural, possessive, or in the context to take a following auxiliary or copula, and they seemed to have come from his noticing in the input language several words of rather similar shape whose suffixed forms were frequent (e.g., *berries*, *pears*, *Peter's*, and *Barbara's*]. Peters and Menn described how this rule slowly crumbled from age 2;3 onward, to be replaced by context-sensitive pluralization and possessive rules that went through several stages of reanalysis and refinement before full productivity.

Comparison of the effects of implicit and explicit learning of artificial languages (Ellis, 1994a) shows that, when the material to be learned is fairly randomly structured with a large number of variables, and where the important relationships are not obvious, explicit learning can be ineffective and an implicit mode of learning is better. Reber (1976) investigated the effect of instructional set on learning of an artificial language. One group of subjects was given implicit instructions (memorize the examples for a memory test); the other was encouraged to undertake an explicit search for underlying rules. Although both groups could discriminate grammatical strings from ungrammatical strings, subjects in the explicit group were poorer at memorizing exemplars from the language, they learned less about the underlying structure, and they had a tendency to invent rules that were not accurate representations of the structure. Thus, the complex structure of this artificial grammar was too rich to be explicitly analyzed by the learners, and the instruction to explicitly search for rules disrupted performance as they searched in vain, elaborated irrelevant rule systems, and chased hares, thus interfering with implicit learning processes.

Pedagogical Input

As summarized in my target article, classroom environments for second or foreign language teaching can distort the patterns of exposure, function, medium, and social interaction. They can also provide adult learners with explicit rules, rote-learned conjugations and declensions, and deductive styles

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

of thinking about language. These, combined with a lack of sufficient authentic language practice, lead to the unnatural abilities scorned by Krashen (1985) and to learners' experiences of frustration "when all their technical knowledge of grammar seemed to leave them in the lurch on arriving at the Gare Du Nord" (Sharwood Smith, 1994).

Practicing

Tarone highlights the importance of learners' communicative intention, Larsen-Freeman that of speaker choice: "There are probabilistic tendencies in language, not absolutes" (Larsen-Freeman, p. 281). As I will show in the discussion under point 5, Anderson believes that human cognition is rational in that it optimally predicts the likelihood of events on the basis of Bayesian probability. The last two decades of his work have yielded an impressive cannonade of demonstrations of our sensitivity to environmental frequency. Nevertheless, he concluded a recent review chapter as follows:

> The effect of study strategy on memory also points to an Achilles' heel of the adaptive analysis. The adaptive analysis works best if one can conceive of memory as responding passively to external statistics in the environment. However, by different rehearsal patterns the subject can actively create his own unobserved environments with its own statistics. In the current terms, subjects can "trick" their memories into treating something as highly probable by giving it the statistics associated with a highly probable memory. This indicates that there are layers of adaptive consideration that go beyond just looking at environmental statistics. (Anderson & Schooler, 2000, p. 569)

POINT 4: ATTENTION AND FORM-FOCUSED INSTRUCTION

A central and longstanding theme in SLA research has concerned the interface between explicit and implicit knowledge. Krashen's (1985) Input Hypothesis is a noninterface position that posits that, although adults can both subconsciously acquire languages and consciously learn about language, nevertheless (a) subconscious acquisition dominates in second language performance; (b) learning cannot be converted into acquisition; and (c) conscious learning can be used only as a Monitor, that is, an editor to correct output after it has been initiated by the acquired system. Schwartz (1993, 1999) expands on this theme, presenting the argument that the language faculty is modular and that only positive data can affect the construction of an interlanguage grammar comparable to the knowledge system characterizing L1 acquisition. To the extent that language acquisition, representation, and processing are all tuned by frequencies in the input, Krashen's urgings for language acquisition environments that entail naturalistic, comprehensible input are well heeded; yet the noninterface position states that negative and explicit data result in a knowledge type that is not to be equated with linguistic competence.

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

A (perhaps overly) strong characterization of noninterface views is thus that explicit instruction or focus on form should have no effect on acquisition. As I discussed in my target article, however, this is clearly not the case. Norris and Ortega's (2000) comprehensive meta-analysis of the last 20 years' empirical work demonstrated that focused L2 instruction results in large target-oriented gains, that explicit types of instruction are more effective than implicit types, and that the effectiveness of L2 instruction is durable. However, many of the studies included in Norris and Ortega's prize-winning review involved outcome measures such as metalinguistic judgments or constrained constructed responses that may have allowed time and opportunity for the influence of explicit deliberation in learners' performance. How, by observing performance, can we get as clear as possible a sighting of competence? R. Ellis approaches this by looking only at those studies involving fluent implicit production. Norris and Ortega had initiated this contrast by comparing the studies that involved more explicit constrained constructed response measures (of which the average effect size was 1.20) with eight studies (13 treatments) that involved freely constructed production (of which the average effect size was a more modest but still notable 0.55). Even here, then, there was some interface, with a medium-sized effect of form-focused instruction (FFI) on acquisition. R. Ellis adds another three or four more recent studies to this list and reviews the effects of instruction here, concluding:

> FFI can contribute to the acquisition of implicit knowledge. . . . FFI involving extensive instruction directed at "simple" structures was more likely to succeed. However, limited instruction directed at complex structures also proved effective, provided that the target structures are readily available in noninstructional input. (p. 223)

It is hard to come to any firm conclusions about mechanisms of influence from such a small number of heterogeneous field experiments during which so much happens in the learning environment. These studies give the ecological validity required for the bridge between theory and educational practice, but laboratory experimentation is needed for finer-grained process analysis: In this enterprise, applied linguistics and experimental psychology go hand in hand (Ellis & Schmidt, 1997; Hulstijn & DeKeyser, 1997). Nevertheless, what are the likely processes of influence of instruction here? The factors outlined under points 1 and 3 (viz., the provision of pedagogic rules or explanations, exaggerated input, or orienting instructions to increase salience) may well have affected noticing and initial explicit acquisition of the structures, as R. Ellis concludes. As discussed under point 2, attention and maintenance in working memory may be necessary for the acquisition of complex structures involving discontinuous dependencies. Once these structures have been acquired, though, the tuning of such knowledge comes from their repeated use. It is thus interesting to note R. Ellis's conclusions about even limited instruction being effective provided that the structures were frequent in the noninstructional input.

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

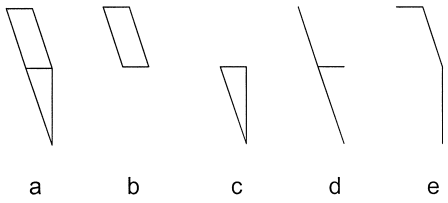


Figure 1. Figure segmentation (based on Pevtzow & Goldstone, 1994).

Boyland (1998) nicely illustrated the relevant learning processes with an analogy from the visual domain (see Figure 1). People who are shown a figure such as (1a) will usually be quicker in finding the embedded parallelogram (as in [1b]) and triangle (as in [1c]) than in finding the forked stick (as in [1d]) and turtle shell (as in [1e]). However, if they have been subjected to a prior categorization task involving a wide range of composite figures, but where a criterial feature is shaped as in (1d), then the figure is more easily segmented into (1d) plus (1e), rather than the usual parallelogram plus triangle. This demonstration, from Pevtzow and Goldstone (1994), showed how people learn to decompose complex objects based on their experience with component parts: Categorization training influences how an object is decomposed into parts. Once you are trained to see the object in that way, that is the way you see it (or at least first see it), and those are the features whose strengths are incremented on each subsequent processing episode. Goldstone (1998; Kersten, Goldstone, & Schaffert, 1998; Goldstone & Steyvers, 2001) presented a detailed analysis of the ways in which attentional persistence directs attention to attributes previously found to be predictive, elaborated a theory of conceptual and perceptual learning based on these mechanisms, and provided a connectionist model of the processes whereby category learning establishes detectors for stimulus parts that are diagnostic. These detectors, once established, bias the interpretation of subsequent objects to be segmented (Goldstone, 2000). The same effect occurs in all perceptual processing. As illustrated in my target article, it spans all levels of language, from the parsing of words in continuous speech to the parsing of *construct ion sin disco urse*, sorry, *constructions in discourse*.

It is also possible that FFI had other effects on the tuning processes involved in the studies reviewed by R. Ellis. Two additional mechanisms come to mind. The first is a quantitative effect resulting from learners' self-generated practice (see point 3, "Practicing"). The second is qualitative: Attention might increase the learning effectiveness of each exposure. Although there has not been a huge amount of work on its implementation in computer models (although see Kruschke, 2001), we know about the importance of attention on learning (Cowan, 1997) and emotion on learning (Schumann, 1998), and we are beginning to discover the neuroanatomical pathways involved in modula-

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

tory transmitters associated with arousal, attention, and emotion (Dudai, 1989, chap. 14; LeDoux, 2000; Mishkin & Appenzeller, 1990) to know that these factors must modulate the weight changes involved in implicit learning.

POINT 5: CONSCIOUS AND UNCONSCIOUS KNOWLEDGE OF FREQUENCY

My review of frequency effects began with evidence of human explicit knowledge of letter frequency (Hasher & Zacks, 1984) before concentrating on the wide range of psycholinguistic demonstrations of implicit knowledge as indexed by the ways in which the speed and accuracy of our language processing reflect the frequencies of patterns in our history of language usage. Harrington and Dennis question the idea of the language learner as an intuitive statistician by referring to the classic demonstrations of Tversky and Kahneman (1974) that, for example, if people are asked to estimate the proportion of English words that begin with the letter *k* versus words with a *k* in the third position, they answer that there are more words beginning with the letter *k* despite the actual fact that three times as many words have *k* in the third position. What is the reason for this? Well, we do not have direct, conscious access to the knowledge of letter positional frequencies that, as is clearly demonstrated in the psycholinguistic demonstrations of consistency, cohort, and neighborhood effects in reading, spelling, and word recognition, are indeed represented implicitly in our language processing systems—input systems that have developed through learning to the degree of automaticity that they are "informationally encapsulated" modules (Fodor, 1983). Instead, we must generate exemplars from memory and make the relative frequency judgment on the basis of the sample of words that consciously come to mind. We must rely on explicit memory as the source for our estimates here. Explicit recall, though, is determined by accessibility—how well the retrieval cues accord with memory organization and correlate with the contexts active during initial encoding (Tulving & Thomson, 1973). Words are more strongly associated to their first letter than to their third letter, and so, at retrieval with the focus of attention on *k*, activation will spread more from that letter to words beginning with it. This process will make words beginning with *k* more available than other words, and thus these words will be overrepresented in the sample that subjects take from memory to estimate the true proportion in the population. The same overestimation does not occur with words having *k* in the third position because words are not directly associated to the letters that occur in the third position (Anderson, 2000, chap. 10).

Kahneman and Tversky (1973) provided a range of such demonstrations that our explicit inductive reasoning can often be quite far off in terms of our judgments and use of probability, and they argued that these failings result from explicit retrieval factors such as salience, base rate neglect, framing effects, and heuristic ways of reasoning. Our cognitive processing, though, dis-

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

plays implicit knowledge of probabilities and the Bayesian principles of their combination, even in situations in which we might make errors when asked to make explicit judgments (e.g., the implicit reasoning demonstrations of Gluck & Bower, 1988; Chater & Oaksford, 1999; Oaksford & Chater, 1998, 2001; the range of psycholinguistic demonstrations marshaled in my review). Indeed, the generality of these effects has led Anderson (1991a, 1991b, 1993; Anderson & Schooler, 2000) to develop Rational Analysis and the ACT-R theory, which characterizes human categorization, memory, problem solving, causal inference, and perception as adaptive: These processes are rational in that they optimally estimate the likelihood of occurrence of different outcomes using Bayesian principles. The human associative learning system is rational in the sense that it accurately represents event relationships (Shanks, 1995). Perhaps not "life, the universe, and everything" as Eubank and Gregg accuse (p. 237), but the representation and exploitation of frequency information do underpin much of human cognition, and language has no exemption.

Biber and Reppen provide ample evidence of our failures to make accurate explicit estimates of the frequency of linguistic structures. I am reminded of Biber's demonstration, in his plenary address at the AAAL 2001 meeting in St. Louis, of the failures of our conscious intuitions about the relative frequencies of grammatical forms. Six hundred or more applied linguists were asked what the most frequent form of verb aspect in conversation is—progressive, simple, or perfect. A virtual votometer was discussed, but AAAL's resources ran instead to a simple show of hands. The vast majority of the audience opted for the progressive. Biber assures me that he has done this demonstration at least 20 times, with local and national conference audiences in the United States, as well as professional teacher audiences in the United Kingdom, France, Poland, and elsewhere. The results are consistent, with approximately 90% of teachers and language professionals believing that the progressive is by far the more frequent form in conversation. These beliefs fly in the face of the facts: Progressive and perfect aspect have about the same frequency in conversation, but simple aspect verbs are almost 20 times more common than either progressives or perfects (see Figure 3 in Biber and Reppen's commentary, p. 204).

Why do we get this wrong? An account similar to the *k* explanation discussed earlier seems plausible: We have no conscious access to the frequencies represented in our language processing systems, so we have to generate some exemplars in order to scrutinize them. We try to imagine a conversation. We think of what we are likely to say. What are we talking about? What are we doing while we are having this conversation? The particular perspective we set as we ask ourselves this question generates the aspect that comes to mind, and my question sets a progressive stage. Of course, in introspecting and elaborating this account, I am providing a post hoc rationalization; I am telling more than I can know (Nisbett & Wilson, 1977); I have no conscious access to the workings of the implicit processing systems, I only see the re-

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

sults of their workings, which is why introspection can be such a dubious data source for language science (Labov, 1996b) and why "one should not touch what one is likely to say with a ten-foot pole" (Preston, 1996, p. 24).

A final example, from Sinclair (1991), illustrates these memory-retrieval effects with lexical semantics:

> The commonest meaning of the commonest words are not the meanings supplied by introspection; for example, the meaning of *back* as "the posterior part of the human body, extending from the neck to the pelvis" (*Collins English Dictionary* (CED) 2nd edition 1986 sense 1) is not a very common meaning. Not until sense 47, the second adverbial sense, do we come to "in, to or towards the original starting point, place or condition", which is closer to the commonest use in our evidence [the Cobuild corpus]. (p. 112)

The core meaning of a word that comes to mind from introspection is the most frequent independent sense, a recall driven by ready imageability of reference; it is not the delexical sense that is typical of the most frequent words. The meanings of frequent words are difficult to identify and explain, and with very frequent words we are reduced to describe uses rather than meanings because the distinctive contribution made by that word to that meaning has become much reduced.

These failings of explicit knowledge of linguistic frequency data are important. Armchair intuitions are foundations of sand for the analysis of language representation and processing. Theories of language acquisition need to be based on empirical foundations (Labov, 1994; Sampson, 2001); just so, as Biber and Reppen persuasively demonstrate, should materials for language instruction and assessment.

POINT 6: SEQUENCES OF ACQUISITION—FROM FORMULA, THROUGH LOW-SCOPE PATTERN, TO CONSTRUCTION

I welcome the way that Bardovi-Harlig has examined the proposed developmental sequence—from formula, through low-scope pattern, to construction—with regard to new data concerning the acquisition of futurity in a longitudinal corpus of (predominantly) written samples from 16 ESL learners enrolled in a university Intensive English Program. Her study is exemplary in its design and, by investigating the degree to which this sequence, which was gleaned from naturalistic child L1 acquisition, generalizes to instructed L2 university-level learners, it is a properly challenging test of the limits.

Operationally defining formulaic use is criterial and complex. The standard definitions are those of Braine in his classic (1976) monograph and of Pine and Lieven (1993, pp. 556–557), who painstakingly examined rigidity of form and sequence of emergence as criteria. Bardovi-Harlig adopts the definitions of Weinert (1995) and Myles, Mitchell, and Hooper (1999) to likewise involve (a) frequency and invariance of form and (b) well-formedness that may be grammatically advanced compared to the rest of the learner's language; she

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

adds (c) a pragmatic component relating to situational dependence. Adoption of a defining requirement—that a formula must not have been separately evinced previously as its individual constituents—ideally requires a comprehensive longitudinal record of learner productions in order to track emergence. We can never achieve this, although for L1 Tomasello (2000b; Diessel & Tomasello, 1999), and Lieven and colleagues at the Max Planck Institute in Leipzig are getting admirably close by the collection, transcription, and analysis of "dense corpora"—massive amounts of data recording the input that some children receive and the output that they produce, with the goal of documenting in detail children's transition from item-specific to more abstract linguistic constructions. If only we could do the same for a few case studies of L2. As Rutherford and Thomas (2001) explained, it is vital that the enormous effort dedicated to the development of the tools and methods of the CHILDES project (MacWhinney, 2000), which has allowed such strides forward in child language acquisition research, be built on by researchers of SLA. The accuracy of anything less than a comprehensive record rests on sampling error. Nevertheless, Bardovi-Harlig provides a rich specimen of approximately 100 written texts over the year of observation for each of her 16 learners. The parallel logs completed by the instructors allow additional analysis of the effects of pedagogical input to the learners.

What does she find? With regard to the first productive uses of *going to*, for 5 of the 16 learners, the use of *I am going to write* stands out as a formulaic usage after this was introduced in a class-assigned model letter. Their productions over the months of observation suggest that this formula breaks into smaller parts, from the full *I am going to write about*, down to a likely still unanalyzed unit *going to* in which not only the verb but also person and number vary. This seems to be an example of learner production moving along the formulaic-creative dimension. Additionally, though, there are cases of formulaic use appearing at the same time or even after wider use. The analysis could only be done with the *going to* construction because *will* emerged early and spread so rapidly that its emergence could not be tracked properly in this corpus, which is why we could do with corpora as intensive as Tomasello's. Bardovi-Harlig reaches the measured conclusion that the proposed developmental sequence is a useful starting point but not fully descriptive of L2 development for her instructed adult language learners for whom formulaic language plays a limited but noteworthy role "because formulaic expressions do not dominate initial stages, nor are they excluded from subsequent stages" (p. 190).

The use of formulaic language at subsequent stages of development is not a surprise. I guess I should have been clearer in what my proposed developmental sequence was a sequence of! It was intended as the sequence that culminates with a construction that *can* be used creatively. Nevertheless, there is clear cognitive economy to be had from the use of formulas at all stages of development. For the reasons outlined with regard to the idiom principle and formulaic language production in my target article, much of language is drawn

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

off the shelves of memory. Frequency of occurrence leads to entrenchment and the independent representation of even "regular" constructional patterns. Erman and Warren's (2000) analysis suggested that about half of fluent native text is constructed according to the idiom principle. De Cock (1998) examined corpora of language-learner productions using automatic recurrent sequence extractions that showed second language learners using formulas at least as much as native speakers and at times at significantly higher rates. There is much promise in such computer-based learner corpus studies (Granger, 1998), provided that sufficient trouble is taken in the necessarily intensive gathering of longitudinal learner data.

What of the fact that formulaic language does not dominate initial sequences here? That is certainly more of a challenge to the generality of the proposed principle to the writing of classroom-instructed university ESL students. What other factors might underlie this? As Bardovi-Harlig quotes from my target article, "classroom environments . . . can distort the patterns of exposure, function, medium, and social interaction" (Ellis, p. 170) and these distortions will distort from the naturalistic sequence. Textbook-driven learning combined with a deductive emphasis will disrupt the sequence, at the extreme producing learners who compute by a mix of pedagogical rule, direct translation, and online declension. She continues, "Instruction is just one source of input in a mixed environment, and it, like input from other sources, is subject to influence of salience, complexity, timing with respect to learner developmental level, and communicative need" (Bardovi-Harlig, p. 197). I agree that all of these factors can cause deviations from the naturalistic child sequence, and under point 3, I gathered a number of other potential influences. All of these factors might well be expected to play a role with adult classroom-instructed university ESL students; the closer we move to naturalistic child acquisition, the more the default sequence is expected to hold.

I have found one other analysis of the role of formulas in the SLA of adults on intensive classroom courses. Bolander (1989) analyzed the role of chunks in the acquisition of inversion in Swedish by Polish, Finnish, and Spanish immigrants enrolled in a 4-month intensive course in Swedish. In Swedish, the inversion of subject and verb after a sentence-initial nonsubject is an obligatory rule. Bolander identified the majority of the inversion cases in her data as being of a chunklike nature with a stereotyped reading, such as *det kan man säga* "that can one say" and *det tycker jag* "so think I". Inversion in this sort of clause is also frequent when the object is omitted, as in *kan man säga* "can one say" and *tycker jag* "think I", and this pattern was also well integrated in the interlanguage of most of these learners. Bolander showed that the high accuracy on these stereotyped initial-object clauses generalized to produce a higher rate of correctness on clauses with nonstereotyped initial objects than was usual for other types of inversion clauses in her data, and she took this as evidence that creative language was developing out of familiar formulas.

My target article was passed on to Kathryn Kohnert of the University of

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

Minnesota, who kindly contacted me identifying an important "missing link" in that my target article failed to properly integrate research on child SLA. I admit this inadvertence. Kohnert emphasizes that, besides being interesting in their own right, studies of child SLA and bilingual language acquisition give us greater scope for investigating the parallels between L1 and L2 acquisition in naturalistic contexts untainted by instruction, and they are important windows on issues of transfer and the relations between conceptual and language development.

The good sense of Kohnert's urgings becomes clear if we follow her advice and look to studies of child SLA that bear on the role of formulas in the acquisition sequence. McLaughlin (1995, p. 7) summarized, "One of the hallmarks of child second language development is the use of formulas." Hakuta (1976), analyzing an impressive corpus of 2 hours every 2 weeks for 60 weeks of child Uguisu's SLA, identified a large number of so-called prefabricated patterns that were used "without knowledge of their underlying structure, but with the knowledge as to which particular situations call for what patterns" (p. 331). These chunks slowly become analyzed into their component structures, and, as the prefabricated pattern becomes gradually replaced by a more analyzed form, so there is greater variability in the learner's interlanguage (p. 329). Wong-Fillmore (1976) presented another extensive longitudinal study that focused on formulaic language in child SLA using data from five Mexican children, each recorded interacting with the observer and an English-speaking friend. She emphasized the vital social role that formulas serve in opening communication channels with speakers of the L2 and thus in securing L2 input. She also emphasized the crucial cognitive role of formulas as an entry to grammar, a databank to be used for later analysis and reorganization. The most effective language learner, Nora, also had the greatest number of formulas, acquiring and overusing a few formulaic expressions of a new structural type during one period and then amassing a variety of similar forms during the next, with previously unanalyzed chunks becoming the foundations for creative construction: "Perhaps having a few formulaic expressions of a particular type permitted her to notice, interpret, and pick up like expressions" (p. 508). Wong-Fillmore maintained that a sufficient set of exemplars is needed before a learner is in a position to perform the analysis, which is the prerequisite to acquisition. In the first 2 months of the 8 months of observation, the different children's utterances ranged from 53% to 100% formulaic, in the last 2 months from 37% to 81%. Vihman (1982) came to similar conclusions from the analysis of her young daughter Virve, native in Estonian, as she acquired ESL from 21 months of age.

I hope that there will be more studies of the role of formulas in SLA. The investigations initiated 25 years ago by Hakuta and Wong-Fillmore suggest that in child SLA the developmental sequence—from formula, through low-scope pattern, to construction—may be as clearly in evidence as it is in L1 acquisition. Bardovi-Harlig's investigation suggests a more limited but noteworthy

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

role of formulas in adult instructed ESL. We need more such studies to identify the degree to which this sequence is tempered by the learning environment, the learner's age, and their existing conceptual and linguistic abilities.

POINT 7: THE FUNDAMENTAL DIFFERENCE HYPOTHESIS

Even within the generative tradition, not all researchers who hold that child language acquisition is guided by UG believe the same to be true for adults. The no-access hypothesis, as typified by Clahsen (1988) and Bley-Vroman (1990), claims that "child first language and adult second language acquisition are guided by distinct cognitive principles" (Clahsen, p. 47). The basic claim is that L2 acquisition is governed by cognitive faculties that are separate and distinct from the domain-specific language faculty. Clahsen suggested that processes similar to Slobin's (1973) Operating Principles can account for L2 learning (see also Andersen, 1989, 1993, for further application of such operating principles, as well as Andersen, 1993, and Shirai, 2000, for analyses of acquisition of linguistic patterns of tense and aspect as prototypes reflecting input frequency). Bley-Vroman suggested that L2 learning strategies derive from Piaget's Formal Operating Principles, including the capacity for distributional analysis, analogy, and hypothesis formation and testing (p. 54). These linguists, at least, are open to more psychological explanations of SLA.

I thank Bley-Vroman for the input of his study with Yoshinaga (2000), which they claim demonstrated that high-proficiency nonnative speakers of English whose first language is Japanese show acceptability judgments of English multiple *wh*-questions that (a) were not like those of native speakers, (b) were not a result of transfer from L1, but (c) were a consequence of the learners' frequency of experience of exemplars of that type. Mind you, if this proposed cause of frequency does indeed hold for the L2 learners, one wonders why it might not for the natives as well, given that their data (Bley-Vroman & Yoshinaga, Table 3) showed performance of the L1 and L2 groups to correlate at $r = 0.92$, $rho = 1.0$ across the six sentence types. On these figures, at least, if it is good enough for one, it ought to be good enough for the other, and however provocative, I doubt that the cluster solutions for the two language groups are significantly different. They explain their findings as follows:

> Non-parameter settings accounts, such as the Fundamental Difference Hypothesis, hold that SLA depends on individual learners' noticing specific structural patterns in the input and adding them to their developing structure-store (Bley-Vroman, 1990: 42–43; 1996). Learning of structural patterns will in general be conservative: a learner only incorporates a pattern into the grammar if it is encountered (and noticed) in the input. A learner who has not encountered (and noticed) target language multiple *wh*-questions will not be inclined to accept them as grammatical. The system is also assumed to possess some sort of strengthening mechanism: the more a pattern is encountered and noticed, the more it will be accepted. High-proficiency learners may then come to accept multiple *wh*-questions to the extent they have been encountered. . . . With time, one would expect some

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

> learners to come to distinctions in acceptability (pattern strength, actually) which would reflect, though dimly, the distinctions made by native speakers. (p. 22)

They give an example of noticing and the importance of exemplars in one Japanese student's comment that he had heard *Let's see who was standing where* in the movie *Twelve Angry Men*, and that his judgments were based on whether he had heard a similar pattern before (Bley-Vroman & Yoshinaga, p. 23).

The similarities with the account I have been presenting here are quite clear. The process of second language learning is the acquisition of constructions, of specific structural patterns (see also Bley-Vroman, 1990, 1996, 1997). Noticing is important for initial registration (I agreed with this under points 1 and 3, with the exceptions made under point 2). The system is assumed to posses a strengthening mechanism that is sensitive to frequency of encounter (the frequency argument made in my target article). We do differ in the specifics at this point: For Bley-Vroman, strengthening only takes place with each noticed encounter; for me, strengthening is contingent on mere use—the language has to be processed for meaning, and strengthening occurs provided that construction is used, whether it is noticed or not. The role of exemplars in his model is clear. Grammaticality judgments are made on the basis of pattern strength.

Of course, we disagree as to the extent to which these processes are relevant to L1 acquisition—paradoxically perhaps, what is a fundamental difference for Bley-Vroman entails that, for me, L1 and L2 acquisition are fundamentally the same. Nevertheless, Bley-Vroman and Yoshinaga (2000) closed their paper by stating:

> The possibility that learners are basing their judgments in part on the patterns which they notice in the input requires both a more elaborated theory of patterns, and more extensive empirical studies of frequency and other factors which may affect noticing. Corpus investigation is obviously indicated. (p. 24)

For L2 and L1, my sentiments exactly!

POINT 8: THE BLIND FAITH OF CATEGORICAL GRAMMAR

Eubank and Gregg attest that "in fact grammatical categories are categorical (Newmeyer, 1998), and they enter into systematic, lawlike relations" (p. 244). "Lawlike" rather than lawful or rule governed may be an unintended concession, but in other respects this doctrinaire assertion denies both the empirical evidence and much of current linguistic opinion. Consider, for example, Huddleston (1984), a textbook with no particular theoretical axe to grind, which nevertheless emphasizes the internal structure of categories:

> One preliminary point that should be made is that not all nouns are identical with respect to their grammatical properties: we find rather that we

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

> have a core of what we may regard as central or prototypical nouns sharing a considerable number of distinctive properties, with other words classified as nouns by virtue of sharing a significant number of these properties and/or of being significantly more like a prototypical noun than like a prototypical adjective, verb or whatever. For example, *umbrage*. . . . (p. 54)

Huddleston (1984) continued: "We shall see that there are numerous places in the grammar where it is necessary to recognize categories with a clear prototypical core but a somewhat fuzzily delimited periphery" (p. 72). If we allow the data to influence our ideas and so perform a distributional analysis of "nouns," then we find that different "nouns" vary as to their morphosyntactic behavior and that some "nouns" display more behavior characteristic of "nouns" than others, which is why it would be hard to find any descriptive linguist (Quirk, Greenbaum, Crystal, and more recently Aarts, Biber, and others) who disagreed with this view.

"Knowledge of language is knowledge" (Goldberg, 1995, p. 5), hence "linguistic categories should be of the same type as other categories in our linguistic system" (Lakoff, 1987, p. 58). Taylor (1998) considered the example of the English category of adjective. The systematic lawlike relations that apply to all the items that we should like to include in such a category, and that exclude everything else, must presumably include that adjectives can be used predicatively (i.e., in association with the copula, as in *this house is large*) and attributively (i.e., before the noun, thereby restricting its referential possibilities, as in *large house*). Many words that we would intuitively want to call adjectives, though, can only be used attributively (e.g., *former colleagues* vs. **colleagues who are former*; *a corporate decision* vs. **a decision that is corporate*); others can only be used predicatively (e.g., *the child is asleep* vs. **the asleep child*). Similar analyses could be done in relation to almost any word class—for example, adverbs, auxiliary verbs (among which *dare*, *need*, and possessive *have* are marginal [R. Hudson, personal communication, July, 2001]), and verbs (with the continuum from so-called true intransitives, through middles and reflexives, to ditransitives at the other end [Kemmer, 1993]). As Taylor (p. 197) concluded: "Prototype effects arise not only with respect to the referential possibilities of lexical items, but also with respect to the very categories of linguistic structure itself." *Zoop*.

How does the language learner get access to these categories in the first place? The cognitive approach, of course, believes that, like other categories, they are induced from exemplars. The UG alternative is to assume an innate universal set of categories that are not learned and certain flashes of insight that trigger parameter setting from instances. Putting aside the total lack of specification about how such categories might be neurally represented, genetically transmitted, or effective in processing, there is still the logical problem of syntactic acquisition: Identifying the syntactic category of words must primarily be a matter of learning because the phonological strings associated with words of a language are clearly not universal. Once some identifications

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

have been successfully made, it may be possible to use prior grammatical knowledge to facilitate further identifications. The acquisition of relevant phrase structure grammar, however, requires knowledge of syntactic word class in the first place. In the same way, the setting of a head-direction parameter requires the child first to identify which units are the heads and which are the complements; but if their linguistic skills are sophisticated enough to do this, surely they will already know what order they are in. The so-called linking problem is a classic bootstrapping problem (Redington & Chater, 1998; Tomasello, 2000a).

And what about the explanatory power? Eubank and Gregg tell us that "[s]tandard (generative) accounts of language structure and language acquisition . . . have achieved explanatory coverage of vast amounts of data from countless different languages" (pp. 244–245). How should we test this claim? The general, language-universal method for identifying syntactic elements (categories and relations) that are the basic units of syntactic theory is distributional analysis, in which syntactic categories are defined by the occurrence or nonoccurrence of their members in different types of utterances. As distributional analysis is applied crosslinguistically, though, it is abundantly clear that the constructions used to define the categories in question in one language (say, English) are missing in the other language to be analyzed. For example, inflection for number, gender, and case are commonly used to identify nouns, and inflection for agreement, tense, aspect, and mood to identify verbs, yet Vietnamese, for example, lacks all morphological inflection; the distributional analysis of inflections cannot tell us whether or not Vietnamese has parts of speech, let alone the parts of speech that English does. Even when the relevant constructions exist in the language in question, they give widely different distributions and hence widely different categories than those found in English. The result is that different criteria are used to identify syntactic categories in different languages, leading to "crosslinguistic methodological opportunism" that is unprincipled, ad hoc, and "not a rigorous scientific method for discovering the properties of Universal Grammar" (Croft, 2001, chap. 1). The conflicting distributional patterns found within and across languages lead to the rejection of the notion that parts of speech are global patterns of particular language grammars, let alone categories of UG (Croft, 2000; 2001, chap. 2). Croft's (2001) *Radical Construction Grammar* proposed instead that language-particular categories are construction-specific; that is, they are defined by their position in the conceptual space of form-function mappings—the ways the language achieves the pragmatic functions of reference, predication, and modification in referring to objects, properties, and actions. Thus, grammatical categories are emergent and prototypical: "The language-particular categories which linguists prefer to call Noun, Verb and Adjective are what Lakoff (1987) calls *radial categories*. Radial categories are categories with internal structure, typically with a prototype with extensions that are conceptually motivated but linguistically conventional" (Croft, 2001, p. 104).

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

POINT 9: LABOVIAN VARIATIONIST PERSPECTIVES ON LANGUAGE ACQUISITION

I am grateful to Tarone (this volume) and R. Ellis (1999) for pointing me to the variationist perspectives on SLA, particularly Preston (1996, 2000), Tarone (1988), and the readings in Gass, Madden, Preston, and Selinker (1989) and Bayley and Preston (1996). I should have known more of this substantial body of work. I am very glad I do now. It is clear that sociolinguistic and psycholinguistic approaches to SLA share basic assumptions about language representation and acquisition, a commitment to linguistic data, and a belief in multiple interacting influences on linguistic variation. Psychologists share a fascination with Tarone's (1988, p. vii) question "what does it mean to say that learner language is systematic when there is massive evidence that the organization of learners' language performance varies from one situation to another?" and they will readily join with Preston (1996, p. 25): "The discovery and weighting of influencing factors is the most valuable area of interaction between variation linguistics and SLA and ought to be of interest to anyone who has any interest in a developing (or even fossilized) interlanguage."

Early sociolinguistic research (characterized as Level I in Preston, 2000) concerned how linguistic communities affected learner performance and how speech partners affect selection. From these observations came the notion of variable linguistic competence: "It is not a hypothesis that children do probability matching [during language learning]. It is simply a description of the observed facts" (Labov, 1994, p. 583). Probability matching is the behavior of choosing among alternatives in proportion to their success. Labov (chap. 20) provided as good an introductory review of this as you could find at the time. So, sociolinguists developed the notion of the variable rule and accordingly their own inferential statistical modeling packages such as VARBRUL (Preston, 1996; Young & Bayley, 1996) to determine the factors that influence learner selection of a particular construction. The commonalities with exemplar-based and connectionist approaches are even stronger in so-called Level II sociolinguistic studies (Preston, 2000), in which variationists tease out the influence of one linguistic (not social) factor on another. VARBRUL uses logistic multivariate regression to test for linguistic dependencies; connectionist models are essentially nonlinear statistical analyses of the same thing. When variation is considered from the processing perspective, "from the point of view of a probabilistic device, one applied *each time* a variant is selected" (Preston, 2000, p. 145, emphasis in original), there are clear parallels with the usage-based accounts of cognitive linguistics and the exemplar-counting psycholinguistics theories of priming and Hebbian learning.

The two fields also share ideas on learning failure. Labov (1996a) presented an analysis of the "role of misperception in language learning" that called for an improved method of contrastive analysis so as to analyze the processes of transfer—one that takes probabilities of occurrence into account (cf. the Competition Model). He considered why Puerto Rican L2 and L1 speakers

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

treat /r/ differently in regard to the native English rule of consonant-cluster simplification. For speakers with a Spanish base, /r/ functions as a consonant rather than the glide of native English speakers. Labov suggested that Spanish ESL speakers continue to perceive the English /r/ with their prior Spanish perceptual categories and that these misperceptions

> may give rise to an asymmetry between the first and second language learner's pool of probabilities upon which the learning process depends. . . . The adult non-native speaker is no doubt more likely than the native child to misperceive the phonetic production of the native adult speaker. It follows that the non-native speaker must be continually storing forms that do not correspond to what was actually produced. [cf. the discussion of *attractors* under point 3, "Perseveration"] . . . what practical methods might follow to facilitate language learning? . . . If we believe that the storage of misperceived productions has led to the stabilizing of non-native underlying forms and rules, it seems reasonable to search for some ways of destabilizing them. The basic strategy would be to reverse the negative effects of these misperceived cases by a strong emphasis on the contexts where phonological contrasts are easiest to perceive [cf. the use of *exaggerated input* under point 3, "Perseveration"]. (pp. 250–251)

It is clear that psycholinguists and sociolinguists feel the same about language, though our research is complementary rather than overlapping because we started from different places. However similar are our broad sketches, we provide detail in different areas. I hope that Preston agrees that the research I have reviewed on exemplar frequency effects in implicit L2 processing, along with that on complementary memory systems, illustrates how what is known of the "'real' psycholinguistic factors of memory, attention, access, processing and the like" (Preston, 2000, fn. 3) readily fills in the gaps that he identified in his variationist perspective on SLA.

It will be all too clear that I share Tarone's belief in the creative power of language play, particularly puns. There is danger, though, in trying too hard. Searching for a good pun, like scanning a Necker cube, can often result in a mental block. Those that are consciously fashioned can be just too labored—in writing this I must have crafted five different creations, then five again, trying each in turn to raise a laugh; nonetheless, no pun in ten did. Those that instead come fluently to mind are usually to be preferred. Arthur Koestler called puns two strings of thought tied with an acoustic knot—a lowly form of association for the lowest and most groveling kind of wit, albeit an important foundation for humor. Thus, the associations in punning, unlike those of the semantically knotted metaphors that pervade all language (Lakoff & Johnson, 1980), are more likely to leave the originator in deep, hot, or troubled water.

POINT 10: PARSIMONY AND THEORY TESTING

Eubank and Gregg remind us of potential pitfalls in the application of Occam's razor in theory development by quoting Marcus, who in turn cited Crick. I will

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

rephrase in Halperin's (1998) revision of a maxim of Mencken: "For every Complex Problem there exists a Unique Solution which is elegant, simple, and wrong." It is true that it is all too tempting to claim parsimonious broad coverage where we think we can get away with it. I wonder if advocates of emergence are more prone to it than those of the Unique Solution of UG? Was not the driving force of UG that it should be minimal, parsimonious, and highly deductive? My favorite example of an application of parsimony guaranteed to raise eyebrows is from a section entitled "Universal Symbols in Children's Rules" in Pinker (1984):

> No doubt many people will wonder what I could possibly mean, then, when I say that a child has (say) a SUBJ symbol in his or her grammar at a time when there is still only one privilege of occurrence for subjects, namely appearing preverbally, and where the other procedures that it triggers have not yet gone into action. In such cases the symbol itself is not doing any of the work that we normally expect of an entity before associating linguistic symbols like SUBJ to it. . . . In calling a part of a rule SUBJ, I am simply asserting that *in the future* it will interact with certain mechanisms in predetermined ways, regardless of its limited role at that time. . . . An analogy with embryology may help. While an embryo is still a mostly undifferentiated mass of cells, it is possible to identify regions that will develop into mature organs. A clump of cells absorbing a large amount of some stain, or forming a slight bulge, may turn into the wing of a bird. Would one say that this bulge is "really" a wing? In one sense, no; in another, yes. . . . One says that a collection of cells "is" a wing in order to express the claim that there is continuity of mechanism, and it is for precisely this reason that I say that young children's rules have symbols that "are" SUBJs. (pp. 164–165, emphasis in original)

Perhaps this quote helps clarify why Eubank and Gregg, though professing an interest in the acquisition of second language, concentrate so on the property theory. By such reasoning, the property theory becomes the theory of acquisition, and argument falls more under the influence of Procrustes than of William of Occam. Once continuity of mechanism is assumed, one just has to look backwards and identify the bulge. A continuity hypothesis raises explanandum to explanans.

POINT 11: UNIVERSALS AND PREDISPOSITIONS

Eubank and Gregg begin by stating:

> It is obvious that children know which form-function pairings are possible in human-language grammars and which are not, regardless of exposure. Examples of this kind abound; to name but one, children know that PERSON—very frequent in input—can be linked to function morphemes but that ROUGH SHAPE—also very frequent in input—cannot be so linked. (p. 238)

Although this may be obvious to Eubank and Gregg, perhaps it is less so to language learners themselves: Somebody should tell the hundreds of millions

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

of speakers of Japanese, Korean, Vietnamese, Chinese, and many of the indigenous languages of North and South America—all of which have various classifier systems in which nouns are marked with morphemes that accord with their shape, like "round" and "long cylindrical"—that they should immediately stop using these morphemes functionally to link quantifiers with base nouns because it is so obviously impossible.

The last 20 years of research on classifiers has shown them to be important clues to cognitive processes of categorization. It is notable that taxonomy-oriented approaches to classifier analysis do not work well and that prototype theory has proved much more effective. Lakoff's (1987) *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*, one of the foundation stones of cognitive linguistics, takes its title from the radially structured categories of Dyirbal, an aboriginal language of Australia in which nouns are marked by classifiers. Whenever a Dyirbal speaker uses a noun in a sentence, the noun must be preceded by a variant of one of four words: *bayi*, *balan*, *balam*, or *bala*. These words classify all objects in the Dyirbal universe. *Balam*, for example, is used with edible fruits and plants, honey, cigarettes, wine, and cake; *bayi* is used with males, the moon, most fish, and animals. If a subset of nouns has some particularly important property that the rest of the set does not have, then such members appear in a different class from the rest of the set, thus marking this property. The important property is most often "harmfulness"—harmful things are classified by *balan*. This is a beautiful example of how languages grammaticalize the notions that are relevant to language users. The classes that tend to be more universally grammaticalized are *functionally* defined—they are those that are across-the-board relevant to human experience and communication, online decodable, and for which there are high-frequency and general lexical items that can seed grammaticalization. However, as Slobin (1997, section 4.3) concludes in his last chapter of volume 5 of his *Universals of Language Acquisition* set:

> The reason why languages have no grammatical markers for quantified categories of "fixed distance, size, contour and angle" (Talmy, 1988, p. 171) is simply because human beings do not regularly code, store, and report their experience in these terms—not because these categories are a priori excluded from the grammatical module. I would suggest, then, that anything that is important and salient enough for people to want to refer to it routinely and automatically most of the time, and across a wide range of situations, CAN come to be grammatically marked. (p. 308)

The work on classifiers is well known. I do take the general point Eubank and Gregg are making here though. It was Lakoff who developed these ideas of embodiment into the central theme of cognitive linguistics, arguing that experience is made possible and structured by preconceptual structures—or "directly meaningful concepts"—that are roughly the same for all human beings, thus providing certain fixed points in the objective evaluation of situations. Basic-level structures arise as a result of our capacities for gestalt

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

perception, mental imagery, and motor movement and manifest themselves as basic-level categories such as hunger and pain, water, wood and stone, people and animals, houses, tools and artifacts, and such. Image schemas are spatial mappings such as source-path-goal, center-periphery, and container. I could only give pointers to these areas at the very end of my target article and barely a little more introduction in Ellis (1998), yet I hope my rather long sentence of summary in the (1998) work made it clear that I have no doubts that, as Hulstijn rightly emphasizes, both embodiment (our innate anatomy and physiology) and function (our environmental and sociocultural universals) influence our cognition and language acquisition:

> Emergentism holds that simple learning mechanisms, operating in and across the human systems for perception, motor-action and cognition, as they are exposed to language data as part of a communicatively-rich human social environment by an organism eager to exploit the functionality of language, suffice to drive the emergence of complex language representations. (Ellis, 1998, p. 657)

What I find more difficult to credit is the notion of innate linguistic universals, their putative influence in acquisition (see Bowerman's [1990] critique of "semantic bootstrapping"), and their supposed mechanisms of representation, neural encoding (Lieberman, 2000), and inheritance.

My skepticism here, however, does not deny the need for an acquisition theory. It was Gregg (1993) himself who made the important and oft-quoted demand that a theory of SLA must contain both property theory, explaining the structure of linguistic knowledge, and transition theory, explaining how one knowledge state changes into another, yet attempts to relate SLA to current acquisition theories in cognitive neuroscience seem perversely to reliably rouse his antipathy (e.g., Eubank & Gregg, 1995). We are told that the acquisition theory provided by the UG framework is "*the only one there is*" (Eubank & Gregg, 1995, p. 51, emphasis in the original), that "the theory to be reduced calls the shots" (p. 53), and that a transition theory to explain such linguistic principles as binding theory (anaphora), trace theory (*wanna*-contraction), or level-ordering rules (compounds) should either

> (a) try to show how principles and rules are acquired, or else—*because as we have seen there is little reason to think they can be acquired*—it would (b) assume that they are inherent elements of the mind/brain and try to show how specific input interacts with them to produce a specific grammar. (p. 52, emphasis added)

Despite the façade of open-mindedness, Gregg's call for a transition theory rings hollow when it becomes clear how all options but one are closed. These pages provide the steps by which Eubank and Gregg make their magic and conjure up an autistic, self-contained acquisition theory from the explanandum, the property theory itself.

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

Gregg is right, though: We do need a sensible theory of language acquisition—one that ventures outside UG, engages with other approaches to L1 and L2 acquisition, and is consonant with the rest of cognitive neuroscience.

POINT 12: *WANNA*-CONTRACTIONS—DUNNO ABOUT TRACES—GONNA CONSIDER THEM LEXICALLY

Gass and Mackey (pp. 255–256) refer to L. White's (1989) discussion of the distribution of *wanna*-constructions (e.g., *Who do you wanna see?* vs. **Who do you wanna see the doctor?*) as illustrating abstract syntactic properties of UG, and they argue that it is not clear how frequency-based accounts can allow learners to make abstractions of the sort necessary to account for such complexities of language. Eubank and Gregg likewise want a transition theory of the acquisition of *wanna*-contractions that accords with their favored account of *wh*-traces, where contraction is not possible "across a trace." But the *wh*-trace explanation of *wanna*-contractions is by no means the only possible analysis. A number of arguments against the Trace Mechanism Proposal (that each movement transformation leaves a trace) as an account of the distribution of *wanna*-contractions were first presented by Postal and Pullum (1978), who proposed an alternative account based on relational grammar. There have since been a number of variants on this theme by Bolinger (1981), Hudson (1981), Pullum (1997), and most influentially by Sag and Fodor (1995, 1996) in terms of head-driven phrase structure grammar. These accounts share the notion that transitive *want NP to* and intransitive *want to* are different lexicalized constructions, and *wanna* is only possible (or is very much more likely) as a contraction of the latter, a very natural shortening given that the sequence *want + to* happens much more often in the intransitive construction than in the transitive one. Bybee (this volume) and Boyland (1998) clearly demonstrate how the morphosyntactic reanalyses that result in late grammaticalization changes such as *going to* > *gonna*, *want to* > *wanna*, and *I don't know* > *I dunno* come about. These articulatory reductions occur in only the more frequent phrases (e.g., there is *t*/*d*-deletion in most of the phrases where *don't* often occurs; there is additional vowel reduction to schwa only when preceded by its most frequent collocate *I*). It is crucial to note here that analysis of these phonological fusions in terms of purely structural factors fails: "The class of verbs before which *don't* reduces is not definable in any way except by frequency of occurrence with *I don't*" (Bybee, p. 216). Thus, the phenomenon is entirely graded: There is no binary categorization of *want to* versus *wanna*; instead, the degree of reduction is a continuous function of the frequency of the target word and the conditional probability of the target given the previous word and that of the target given the next word (Jurafsky, Bell, Gregory, & Raymond, 2001). If you look at the real language data, there are no clear-cut categories to be had here either. As Bybee explains, this, like the evidence given under point 2, shows that experience with collocations is registered in memory, that chunks of words are processing units, and that these

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

different lexicalized constructions are then free to differentially attach to the particular meanings to which they most commonly refer. The new forms take on new analyses and have functions different from those of the original. *Wanna* has evolved to be a subject-control verb.

Vihman (1982, pp. 277–279) gave a clear account of bilingual 2-year-old Virve's early piecemeal acquisition of the different forms of *want*. *Wanna* was first used exclusively between 2;1 and 2:5, whether followed by an adverb, verb, or noun. It tended to be pronounced in an English fast-speech style. At 2;5, /uant/ appeared pronounced in an Estonian accent; thereafter, the recorded instances showed *wanna* followed by a verb, whereas /uant/ was correctly followed by a noun or pronoun complement. For Virve, *wanna* and *want* were different traces linked to different functions.

Thus, the alternation *want to* ~ *wanna* is no more interesting than *going to* ~ *gonna*, where *going to* has evolved to be pronounced *gonna* only when it is used as a future marker and not when it is used to mean traveling from one place to another. No more interesting, nor less interesting either: The fact that speakers are not consciously aware of these distinctions provides additional testament to the argument made in the discussion under point 7 that construction tuning and strengthening occurs automatically with use; it is not additionally necessary to notice the form-function pairing.

I am grateful to Bybee for her commentary. She has done more than most to explore the linguistic consequences of frequency, and her recent book (Bybee & Hopper, 2001) provides a firm footing for a truly empirical functional linguistics.

CONCLUSIONS

In the passage from his introduction to *The Origin of Species* quoted in the first section, Darwin highlighted our lack of knowledge concerning the mutual relations of the beings in an eco-culture—their surroundings, habits, and modes of life—as the reason for our profound ignorance of the origin of species and varieties. I am thankful to Bley-Vroman for being directed back to this. The frequency effects on language discussed in my target article are the result of Hebbian learning in the human language-processing system trying to discover the mutual relations of all the perceptible beings of language.

In *The Descent of Man*, Darwin (1890) claimed:

> Languages, like organic beings, can be classed into groups under groups: and they can be classed either naturally according to descent or artificially by other characters. . . . We see variability in every tongue, and new words are continually cropping up; but as there is a limit in the powers of the memory, single words like whole languages gradually become extinct. As Max Moiler has well remarked that struggle for life is constantly going on amongst the words and grammatical forms in each language. (p. 60)

As in language, so in the language learner; as Hakuta (1976) reminded us,

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

> [the process of SLA] is a dynamic, fluid process in which the system of the learner is constantly shifting: shifting in a slow and gradual manner either toward the maintenance of an internal consistency within the structures which the learner possesses, or in the direction of external consistency, where the learner attempts to fit the internal system into what is heard in the input. (p. 331)

Darwinian principles in interlanguage; Darwinian principles in the cognitive processing systems, in neurophysiology, and in connectionism; Darwinian principles in sociolinguistic competence; Darwinian principles in language change; Darwinian principles in the origin and ontogenic evolution of language competence: *Yes, and* . . . all the commentators are right—more questions remain than have been answered, but we are discovering some useful relations.

Things are coming together, probably.

REFERENCES

Andersen, R. W. (1989). The theoretical status of variation in interlanguage development. In S. Gass, C. Madden, D. Preston, & L. Selinker (Eds.), *Variation in second language acquisition: Psycholinguistic issues* (pp. 46–64). Clevedon, UK: Multilingual Matters.

Andersen, R. W. (1993). Four operating principles and input distribution as explanations for underdeveloped and mature morphological systems. In K. Hyltenstam & A. Viborg (Eds.), *Progression and regression in language* (pp. 309–339). New York: Cambridge University Press.

Anderson, J. R. (1991a). The adaptive nature of human categorization. *Psychological Review*, *98*, 409–429.

Anderson, J. R. (1991b). Is human cognition adaptive? *Behavioral and Brain Sciences*, *14*, 471–517.

Anderson, J. R. (1993). *Rules of the mind*. Mahwah, NJ: Erlbaum.

Anderson, J. R. (2000). *Cognitive psychology and its implications* (5th ed.). New York: Worth Publishers.

Anderson, J. R., & Schooler, L. J. (2000). The adaptive nature of memory. In E. Tulving & F. I. K. Craik (Eds.), *The Oxford handbook of memory* (pp. 557–570). Oxford: Oxford University Press.

Bayley, R., & Preston, D. R. (1996). *Second language acquisition and linguistic variation*. Amsterdam: Benjamins.

Berry, D. C., & Dienes, Z. (1993). *Implicit learning: Theoretical and empirical issues*. Mahwah, NJ: Erlbaum.

Bley-Vroman, R. (1990). The logical problem of foreign language learning. *Linguistic Analysis*, *20*, 3–49.

Bley-Vroman, R. (1996, May). *Conservative pattern accumulation in foreign language learning*. Paper presented at EUROSLA 6, Nijmegen, The Netherlands. [Available: http://www.lll.hawaii.edu/esl/bley-vroman/eurosla.txt].

Bley-Vroman, R. (1997, October). *Features and patterns in foreign language learning*. Plenary address, Second Language Research Forum, Michigan State University, East Lansing. [Available: http://www.lll.hawaii.edu/bley-vroman/slrf97.pdf].

Bley-Vroman, R. & Yoshinaga, N. (2000). The acquisition of multiple *wh*-questions by high proficiency non-native speakers of English. *Second Language Research*, *16*, 3–26.

Bolander, M. (1989). Prefabs, patterns, and rules in interaction? Formulaic speech in adult learners' L2 Swedish. In K. Hyltenstam & L. K. Obler (Eds.), *Bilingualism across the lifespan: Aspects of acquisition, maturity, and loss* (pp. 73–86). New York: Cambridge University Press.

Bolinger, D. (1981). Consonance, dissonance, and grammaticality: The case of *wanna*. *Language and Communication*, *1*, 189–206.

Bowerman, M. (1990). Mapping thematic roles onto syntactic functions: Are children helped by innate linking rules? *Linguistics*, *28*, 1253–1289.

Bowers, J. S., & Schacter, D. L. (1990). Implicit memory and test awareness. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 404–416.

Boyland, J. T. (1998). How developing perception and production contribute to a theory of language change: Morphologization < expertise + listening < development. In M. C. Gruber, D. Higgins, K. S. Olsen, & T. Wysocki (Eds.), *Proceedings of the 34th meeting of the Chicago Linguistic Society* (pp. 27–38). Chicago: Chicago Linguistic Society.

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, *101*, 2299–2310.
Braine, M. D. S. (1976). Children's first word combinations. *Monographs of the Society for Research in Child Development*, *41*, 1–96.
Bybee, J., & Hopper, P. (Eds.). (2001). *Frequency and the emergence of linguistic structure*. Amsterdam: Benjamins.
Carroll, S. E. (1999). Putting "input" in its proper place. *Second Language Research*, *15*, 337–388.
Carroll, S. E. (2001). *Input and evidence: The raw material of second language acquisition*. Amsterdam: Benjamins.
Chater, N., & Oaksford, M. (1999). Ten years of the rational analysis of cognition. *Trends in Cognitive Science*, *3*, 57–65.
Clahsen, H. (1988). Parameterized grammatical theory and language acquisition: A study of the acquisition of verb placement and inflection by children and adults. In S. Flynn & W. O'Neil (Eds.), *Linguistic theory in second language acquisition* (pp. 47–75). Dordrecht: Kluwer.
Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences*, *2*, 406–416.
Cohen, A., Ivry, R. I., & Keele, S. W. (1990). Attention and structure in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 17–30.
Cooper, D. (1999). *Linguistic attractors: The cognitive dynamics of language acquisition and change*. Amsterdam: Benjamins.
Cowan, N. (1997). *Attention and memory: An integrated framework*. Oxford: Oxford University Press.
Croft, W. (2000). Parts of speech as typological universals and as language particular categories. In P. M. Vogel & B. Comrie (Eds.), *Approaches to the typology of word classes* (pp. 65–102). Berlin: Mouton de Gruyter.
Croft, W. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
Curran, T., & Keele, S. W. (1993). Attention and non-attentional forms of sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 189–202.
Darwin, C. (1890). *The descent of man, and selection in relation to sex* (2nd ed.). London: John Murray.
Darwin, C. (1928). *The origin of species*. London: Everyman's Library.
De Cock, S. (1998). A recurrent word combination approach to the study of formulae in the speech of native and non-native speakers of English. *International Journal of Corpus Linguistics*, *3*, 59–80.
Diessel, H., & Tomasello, M. (1999). Why complement clauses do not have a *that*-complementizer in early child language. *Proceedings of the 25th meeting of the Berkeley Linguistic Society*, (pp. 106–120). Berkeley: University of California Press.
Dudai, Y. (1989). *The neurobiology of memory: Concepts, findings, trends*. Oxford: Oxford University Press.
Dunn, J. (1998). Implicit memory and amnesia. In K. Kirsner & C. Speelman (Eds.), *Implicit and explicit mental processes* (pp. 99–117). Mahwah, NJ: Erlbaum.
Eich, E., & Metcalfe, J. (1989). Mood dependent memory for internal versus external events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 443–455.
Ellis, N. C. (Ed.). (1994a). *Implicit and explicit learning of languages*. San Diego, CA: Academic Press.
Ellis, N. C. (1994b). Vocabulary acquisition: The implicit ins and outs of explicit cognitive mediation. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 211–282). San Diego, CA: Academic Press.
Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, *18*, 91–126.
Ellis, N. C. (1998). Emergentism, connectionism, and language learning. *Language Learning*, *48*, 631–664.
Ellis, N. C. (2001). Memory for language. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 33–68). New York: Cambridge University Press.
Ellis, N. C. (in press). Constructions, chunking, and connectionism: The emergence of second language structure. In C. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition*. Oxford: Blackwell.
Ellis, N. C., Lee, M. W., & Reber, A. R. (1999). *Phonological working memory in artificial language acquisition*. Unpublished manuscript, University of Wales, Bangor.
Ellis, N. C., & Schmidt, R. (1997). Morphology and longer-distance dependencies: Laboratory research illuminating the A in SLA. *Studies in Second Language Acquisition*, *19*, 145–171.
Ellis, R. (1999). Item versus system learning: Explaining free variation. *Applied Linguistics*, *20*, 460–480.

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.
Eubank, L., & Gregg, K. R. (1995). "Et in amygdala ego"? UG, (S)LA, and neurobiology. *Studies in Second Language Acquisition*, *17*, 35–57.
Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text*, *20*, 29–62.
Flege, J. (in press). Interactions between the native and second-language phonetic systems. In P. Burmeister, T. Piske, & A. Rohde (Eds.), *An integrated view of language development: Papers in honor of Henning Wode*. Trier, Germany: Wissenschaftlicher Verlag Trier.
Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
Gass, S., Madden, C., Preston, D., & Selinker, L. (Eds.). (1989). *Variation in second language acquisition: Psycholinguistic issues*. Clevedon, UK: Multilingual Matters.
Gluck, M. A., & Bower, G. H. (1988). From conditional reasoning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, *8*, 37–50.
Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C.-H., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Science*, *5*, 236–243.
Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
Goldschneider, J. M., & DeKeyser, R. M. (2001). Explaining the "natural order of L2 morpheme acquisition" in English: A meta-analysis of multiple determinants. *Language Learning*, *51*, 1–50.
Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, *49*, 585–612.
Goldstone, R. L. (2000). A neural network model of concept-influenced segmentation. In L. R. Gleitman & A. K. Joshi (Eds.), *Proceedings of the 22nd annual Conference of the Cognitive Science Society* (pp. 172–177). Mahwah, NJ: Erlbaum.
Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General*, *130*, 116–139.
Gomez, R. L. (1997). Transfer and complexity in artificial grammar learning. *Cognitive Psychology*, *33*, 154–207.
Granger, S. (Ed.). (1998). *Learner English on computer*. London: Longman.
Gregg, K. R. (1993). Taking explanation seriously; or, Let a couple of flowers bloom. *Applied Linguistics*, *14*, 276–294.
Green, G. (1998). Modeling grammar growth: Universal Grammar without innate principles or parameters. In A. Sorace, C. Heycock, & R. Shillcock (Eds.), *Proceedings of the GALA '97 Conference on Language Acquisition* (pp. 338–345). Edinburgh: Human Communication Research Centre, University of Edinburgh, Scotland.
Hakuta, K. (1976). A case study of a Japanese child learning ESL. *Language Learning*, *26*, 321–352.
Halperin, P. M. (1998). *Scribblings: Definitions for risk management*. [Available: http://www.btinternet.com/~phalperin/Scribblings/riskmdef.htm].
Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist*, *39*, 1372–1388.
Hebb, D. O. (1961). Distinctive features of learning in the higher animal. In J. F. Delafresnaye (Ed.), *Brain mechanisms and learning* (pp. 37–46). Oxford: Blackwell.
Huddleston, R. (1984). *Introduction to the grammar of English*. New York: Cambridge University Press.
Hudson, R. (1981). *Wanna* and the lexicon. *The Nottingham Linguistic Circular*, *10*, 132–154.
Hulstijn, J., & DeKeyser, R. (Eds.). (1997). Testing SLA theory in the research laboratory [Special issue]. *Studies in Second Language Acquisition*, *19*.
Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229–254). Amsterdam: Benjamins.
Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*, 237–251.
Kemmer, S. (1993). *The middle voice*. Amsterdam: Benjamins.
Kersten, A. W., Goldstone, R. L., & Schaffert, A. (1998). Two competing attentional mechanisms in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1437–1458.
Kirkwood, A., Lee, H.-K., & Bear, M. F. (1995). Long-term potentiation and experience-dependent plasticity in visual cortex are coregulated by age and experience. *Nature*, *375*, 328–331.
Knowlton, B. J., & Squire, L. R. (1996). Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 169–181.
Krashen, S. D. (1985) *The input hypothesis: Issues and implications*. London: Longman.
Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology*, *45*, 812–863.

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the "perceptual magnet effect." In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 121–154). Timonium, MD: York Press.
Labov, W. (1994). *Principles of linguistic change: Internal factors*. Oxford: Blackwell.
Labov, W. (1996a). Some notes on the role of misperception in language learning. In R. Bayley & D. R. Preston (Eds.), *Second language acquisition and linguistic variation* (pp. 245–252). Amsterdam: Benjamins.
Labov, W. (1996b). When intuition fails. In L. McNair, K. Singer, L. Dolbrin, & M. Ancon (Eds.), *Proceedings of the 32nd meeting of the Chicago Linguistic Society* (pp. 77–106). Chicago: Chicago Linguistic Society.
Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.
LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience*, *23*, 155–184.
Lessing, D. (1973). Introduction to the 1973 edition of *The Golden Notebook*. London: Panther.
Lieberman, P. (2000). *Human language and our reptilian brain: The subcortical bases of speech, syntax, and thought*. Cambridge, MA: Harvard University Press.
MacWhinney, B. (1992). Transfer and competition in second language learning. In R. J. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 371–390). Amsterdam: North Holland.
MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk. Vols. 1–2* (3rd ed.). Mahwah, NJ: Erlbaum.
MacWhinney, B. (2001). The competition model: The input, the context, and the brain. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 69–90). New York: Cambridge University Press.
McCandliss, B. D., Conway, M., Fiez, J. A., Protopapas, A., & McClelland, J. L. (1998). Eliciting adult plasticity: Both adaptive and non-adaptive training improves Japanese adults' identification of English /r/ and /l/. *Society for Neuroscience Abstracts*, *24*, 1898.
McClelland, J. L. (1998). Complementary learning systems in the brain: A connectionist approach to explicit and implicit cognition and memory. *Annals of the New York Academy of Sciences*, *843*, 153–169.
McClelland, J. L. (2000). Connectionist models of memory. In E. Tulving & F. I. M. Craik. (Eds.), *The Oxford handbook of memory* (pp. 583–596). Oxford: Oxford University Press.
McClelland, J. L. (2001). Failures to learn and their remediation: A Hebbian account. In J. L. McClelland & R. S. Siegler (Eds.), *Mechanisms of cognitive development: Behavioral and neural perspectives. Carnegie Mellon symposia on cognition* (pp. 97–121). Mahwah, NJ: Erlbaum.
McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*, 419–437.
McDonald, J. L. (1987). Sentence interpretation in bilingual speakers of English and Dutch. *Applied Psycholinguistics*, *8*, 379–413.
McLaughlin, B. (1995). *Fostering second language development in young children: Principles and practices* [Educational Practice Report No. 14]. Washington, DC: National Center For Research On Cultural Diversity and Second Language Learning, George Washington University. [Available: http://www.ncbe.gwu.edu/miscpubs/ncrcdsll/epr14.htm].
Merzenich, M., Jenkins, W., Johnston, P. S., Schreiner, C., Miller, S. L., & Tallal, P. (1996). Temporal processing deficits of language-learning impaired children ameliorated by training. *Science*, *271*, 77–80.
Mishkin, M., & Appenzeller, T. (1990). The anatomy of memory. In R. Llinás (Ed.), *The workings of the brain: Development, memory, and perception* (pp. 88–104). New York: Scientific American Press.
Myles, F., Mitchell, R., & Hooper, J. (1999). Interrogative chunks in French L2: A basis for creative construction. *Studies in Second Language Acquisition*, *21*, 49–80.
Newmeyer, F. J. (1998). *Language form and language function*. Cambridge, MA: MIT Press.
Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231–259.
Nissen, M. J., Willingham, D., & Hartman, M. (1989). Explicit and implicit remembering: When is learning preserved in amnesia? *Neuropsychologia*, *27*, 341–352.
Norris, J., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*, 417–528.
Oaksford, M., & Chater, N. (1998). *Rationality in an uncertain world: Essays on the cognitive science of human reasoning*. Hove, UK: Psychology Press and Erlbaum.

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

Oaksford, M., & Chater, N. (2001). The probabilistic approach to human reasoning. *Trends in Cognitive Science*, *5*, 349–357.
O'Grady, W. (1997). *Syntactic development*. Chicago: University of Chicago Press.
O'Grady, W. (1998). The acquisition of syntactic representations: A general nativist approach. In W. Ritchie & T. Bhatia (Eds.), *Handbook of language acquisition* (pp. 157–93). San Diego, CA: Academic Press.
O'Grady, W. (1999). Toward a new nativism. *Studies in Second Language Acquisition*, *21*, 621–633.
O'Grady, W. (in press). The radical middle: Nativism without Universal Grammar. In C. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition*. Oxford: Blackwell.
Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General*, *119*, 264–275.
Peters, A. M., & Menn, L. (1993). False starts and filler syllables: Ways to learn grammatical morphemes. *Language*, *69*, 742–777.
Pevtzow, R., & Goldstone, R. L. (1994). Categorization and the parsing of objects. In A. Ram & K. Eiselt (Eds.), *Proceedings of the 16th annual meeting of the Cognitive Science Society* (pp. 717–722). Mahwah, NJ: Erlbaum.
Pienemann, M. (1985). Learnability and syllabus construction. In K. Hyltenstam & M. Pienemann (Eds.), *Modelling and assessing second language acquisition* (pp. 23–75). Clevedon, UK: Multilingual Matters.
Pienemann, M. (1998). *Language processing and second language development: Processability theory*. Amsterdam: Benjamins.
Pine, J. M., & Lieven, E. V. M. (1993). Reanalyzing rote-learned phrases: Individual differences in the transition to multi-word speech. *Journal of Child Language*, *20*, 551–571.
Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.
Postal, P. M., & Pullum, G. K. (1978). Traces and the description of English complementizer contraction. *Linguistic Inquiry*, *9*, 1–29.
Preston, D. (1996). Variationist perspectives on second language acquisition. In R. Bayley & D. R. Preston (Eds.), *Second language acquisition and linguistic variation* (pp. 1–46). Amsterdam: Benjamins.
Preston, D. (2000). A variationist perspective on SLA: Psycholinguistic concerns. In R. Kaplan (Ed.), *Oxford handbook of applied linguistics* (pp. 141–159). Oxford: Oxford University Press.
Pullum, G. K. (1997). The morpholexical nature of English *to*-contraction. *Language*, *73*, 79–102.
Reber, A. S. (1976). Implicit learning of synthetic languages: The role of instructional set. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 88–94.
Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. Oxford: Oxford University Press.
Reber, P. J., & Squire, L. R. (1998). Encapsulation of implicit and explicit memory in sequence learning. *Journal of Cognitive Neuroscience*, *10*, 248–263.
Redington, M. & Chater, N. (1996). Transfer in artificial grammar learning: A reevaluation. *Journal of Experimental Psychology: General*, *125*, 123–138.
Redington, M., & Chater, N. (1998). Connectionist and statistical approaches to language acquisition: A distributional perspective. *Language and Cognitive Processes*, *13*, 129–192.
Ribot, T. (1882). *Diseases of memory*. New York: Appleton-Century-Crofts.
Rutherford, W., & Thomas, M. (2001). The Child Language Data Exchange System in research on second language acquisition. *Second Language Research*, *17*, 195–212.
Sag, I. A., & Fodor, J. D. (1995). Extraction without traces. In R. Aranovich, W. Byrne, S. Preuss, & M. Senturia (Eds), *Proceedings of the 13th West Coast Conference on Formal Linguistics* (pp. 365–384). Stanford, CA: CSLI Publications.
Sag, I. A., & Fodor, J. D. (1996). A traceless account of extraction phenomena. *Languages*, *30*, 8–31.
Sampson, G. (2001). *Empirical linguistics*. London: Continuum.
Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 501–518.
Schmidt, R. W. (1984). The strengths and limitations of acquisition: A case study of an untutored language learner. *Language, Learning, and Communication*, *3*, 1–16.
Schmidt, R. (1993). Awareness and second language acquisition. *Annual Review of Applied Linguistics*, *13*, 206–226.
Schmidt, R., & Frota, S. (1986). Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese. In R. R. Day (Ed.), *Talking to learn: Conversation in second language acquisition* (pp. 237–322). Rowley, MA: Newbury House.

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

Schooler, L. J. (1993). *Memory and the statistical structure of the environment*. Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA.
Schooler, L. J., & Anderson, J. R. (1997). The role of process in the rational analysis of memory. *Cognitive Psychology*, *32*, 219–250.
Schumann, J. H. (1998). The neurobiology of affect in language. *Language Learning*, *48*(Suppl. 1), xiii–341.
Schwartz, B. D. (1993). On explicit and negative data effecting and affecting competence and linguistic behavior. *Studies in Second Language Acquisition*, *15*, 147–163.
Schwartz, B. D. (1999). Let's make up your mind: "Special nativist" perspectives on language, modality of mind, and nonnative language acquisition. *Studies in Second Language Acquisition*, *21*, 635–655.
Seger, C. A. (1994). Implicit learning. *Psychological Bulletin*, *115*, 163–196.
Servan-Schreiber, E., & Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 592–608.
Shanks, D. R. (1995). *The psychology of associative learning*. New York: Cambridge University Press.
Sharwood Smith, M. (1994). The unruly world of language. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 33–44). San Diego, CA: Academic Press.
Shirai, Y. (2000, July). *Input-based prototype formation as a model of grammatical and semantic development*. Paper presented at the 12th International Conference on Korean Linguistics, Prague, Czech Republic.
Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In C. A. Ferguson and D. I. Slobin (Eds.), *Studies of child language development* (pp. 175–208). New York: Holt, Reinhart, & Winston.
Slobin, D. I. (1997). The origins of grammaticizable notions: Beyond the individual mind. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition: Vol. 5* (pp. 265–323). Mahwah, NJ: Erlbaum.
Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, *99*, 195–231.
Squire, L. R., & Kandel, E. R. (1999). *Memory: From mind to molecules*. New York: Scientific American Library.
Stadler, M. A., & Frensch, P. A. (Eds.). (1997). *Handbook of implicit learning*. Thousand Oaks, CA: Sage Publications.
Studdert-Kennedy, M. (1991). Language development from an evolutionary perspective. In N. A. Krasnegor, D. M. Rumbaugh, R. L. Schiefelbusch, & M. Studdert-Kennedy (Eds.), *Biological and behavioral determinants of language development* (pp. 5–28). Mahwah, NJ: Erlbaum.
Tallal, P., Miller, S. L., Bedi, G., Byma, G., Wang, X., Nagarajan, S. S., Schreiner, C., Jenkins, W. M., & Merzenich, M. M. (1996). Language comprehension in language-learning impaired children improved with acoustically modified speech. *Science*, *271*, 81–84.
Talmy, L. (1988). The relation of grammar to cognition. In B. Rudzka-Ostyn (Ed.), *Topics in cognitive linguistics* (pp. 166–205). Amsterdam: Benjamins.
Tarone, E. (1988). *Variation in interlanguage*. London: Edward Arnold.
Taylor, J. (1998). Syntactic constructions as prototype categories. In M. Tomasello (Ed.), *The new psychology of language* (pp. 177–202). Mahwah, NJ: Erlbaum.
Terrell, T. (1991). The role of grammar instruction in a communicative approach. *The Modern Language Journal*, *75*, 52–63.
Thomas, A. G., & McClelland, J. L. (2001). *When learning fails: A Hebbian account of loss of plasticity*. [Available: http://www.cnbc.cmu.edu/~thomas/public/tm.pdf].
Tomasello, M. (2000a). Do young children have adult syntactic competence? *Cognition*, *74*, 209–253.
Tomasello, M. (2000b). The item based nature of children's early syntactic development. *Trends in Cognitive Sciences*, *4*, 156–163.
Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 359–380.
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.
van Geert, P. (1993). A dynamic systems model of cognitive growth: Competition and support under limited resource conditions. In L. B. Smith & E. Thelen (Eds.), *A dynamic systems approach to development: Applications* (pp. 265–331). Cambridge, MA: MIT Press.
van Geert, P. (1994). *Dynamic systems of development: Change between complexity and chaos*. New York: Harvester Wheatsheaf.

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press

VanPatten, B. (1996). *Input processing and grammar instruction in second language acquisition*. New York: Ablex.
Vihman, M. (1982). Formulas in first and second language acquisition. In L. Obler & L. Menn (Eds.), *Exceptional language and linguistics* (pp. 261–284). San Diego, CA: Academic Press.
Weinert, R. (1995). Formulaic language in SLA: A review. *Applied Linguistics*, *16*, 180–205.
White, H. (1989). Learning in artificial neural networks: A statistical perspective. *Neural Computation*, *1*, 425–464.
White, L. (1989). *Universal Grammar and second language acquisition*. Amsterdam: Benjamins.
Wong-Fillmore, L. (1976). *The second time around*. Unpublished doctoral dissertation, Stanford University, Stanford, CA.
Young, R., & Bayley, R. (1996). VARBRUL analysis for second language acquisition research. In R. Bayley & D. R. Preston (Eds.), *Second language acquisition and linguistic variation* (pp. 253–306). Amsterdam: Benjamins.

https://doi.org/10.1017/S0272263102002140 Published online by Cambridge University Press