Usage-based and form-focused SLA: The implicit and explicit learning of constructions

Nick C. Ellis

Psycholinguistics substantiates that language acquisition is usage-based. The first half of this paper reviews psycholinguistic research showing how language processing is intimately tuned to input frequency at all levels of grain: input frequency affects the processing of phonology and phonotactics, reading, spelling, lexis, morphosyntax, formulaic language, language comprehension, grammaticality, sentence production, and syntax. That language users are sensitive to the input frequencies of these patterns entails that they must have registered their occurrence in processing. I consider the implications of these effects for a usage-based model, the nature of language representations, and the *implicit* learning of constructions.

The second half of the paper concerns *explicit* language learning. There are 'fragile' aspects of second languages which learners fail to acquire despite high frequency in the ambient language: where input fails to become intake. Such situations arise because learners fail to notice cues which are lacking in salience and redundant in cuing meaning, or because of interference where the features need to be processed in a different way from that usual in their L1. I consider the role of noticing and attention in the initial acquisition of constructions, the effectiveness of form-focused instruction, and the nature of the interface between explicit and implicit learning. I review research concerning the cognitive neuroscience of complementary memory systems, and demonstrate that while they are separate representational systems, nevertheless, explicit knowledge can affect implicit learning in a variety of ways.

In these ways I illustrate how a usage-based account bridges linguistics, applied linguistics, SLA, psycholinguistics and brain science. The usage-based insight opens the study of language acquisition into the broad enterprise of cognitive science.

1. Implicit probabilistic processing of linguistic constructions

Counting from 1 to 10 is early content in most second and foreign language courses and learners of English as a foreign or a second language are soon secure in the knowledge of what ‘wʌn’ means. But should they be so sure? Consider the following wʌns: ‘That's wʌn for the money, two for the show, three to get ready’; ‘To love wʌnself is the beginning of a lifelong romance’; ‘wʌnce upon a time...’; ‘Alice in wʌnderland’; ‘wʌn the battle, lost the war’; ‘How to win life's little games without appearing to try – wʌnUpmanship’; ‘the human brain is a wʌnderful thing, it starts working the minute you’re born and never stops until you’re faced with the blank word-processor screen when starting a new article.’ These are different wʌns. Form-meaning associations are multiple and probabilistic, and fluent language processing exploits prior knowledge of utterances and of the world in order to determine the most likely interpretation in any given context. This usually works very well and the practiced comprehender is conscious of just one interpretation – Alice in wʌn sense and not the other. But to achieve this resolution, the language processing mechanism is unconsciously weighing the likelihoods of all candidate interpretations and choosing between them. Thus there is a lot more to the perception of language than meets the eye or ear. A percept is a complex state of consciousness in which antecedent sensation is supplemented by consequent ideas which are closely combined to it by association. The cerebral conditions of the perception of things are thus the paths of association irradiating from them. If a certain sensation is strongly associated with the attributes of a certain thing, that thing is almost sure to be perceived when we get that sensation. But where the sensation is associated with more than one reality, unconscious processes weigh the odds, and we perceive the most probable thing: “*all brain-processes are such as give rise to what we may call figured consciousness*” (James 1890). Accurate and fluent language perception, then, rests on the comprehender having acquired the appropriately weighted range of associations for each element of the language input.

Language learning is the associative learning of representations that reflect the probabilities of occurrence of form-function mappings. Frequency is thus a key determinant of acquisition because ‘rules’ of language, at all levels of analysis from phonology, through syntax, to discourse, are structural regularities which emerge from learners’ lifetime analysis of the distributional characteristics of the language input. Learners have to *figure* language out. It is these ideas which underpin the last thirty years of investigations of cognition using connectionist and statistical models (Elman et al. 1996; Rumelhart and McClelland 1986), the competetion model of language learning and processing (Bates and MacWhinney 1987; MacWhinney 1987, 1997), the recent emphasis on frequency in language acquisition and processing (Bybee and Hopper 2001; Ellis 2002; Jurafsky 2002), and proper empirical investigations of the structure of language by means of corpus analysis (Sinclair 1991; Biber, Conrad, and Reppen 1998; Biber et al. 1999).

Fluent language processing is intimately tuned to input frequency and probabilities of mappings at all levels of grain: phonology and phonotactics, reading, spelling, lexis, morphosyntax, formulaic language, language comprehension, grammaticality, sentence production, and syntax. It relies on this prior statistical knowledge. Let me give an example or two from each domain just to illustrate the enormity of the learner's database of relevant knowledge. What follows is a very small sample from literally thousands upon thousands of published psycholinguistic demonstrations of learners' implicit statistical knowledge of language. You can track down more detail in Ellis (2002a, 2002b) if interested.

1.1. Orthographics

One of the earliest proofs, a defining study of psycholinguistics half a century ago, was the demonstration by Miller, Bruner, and Postman (1954) that we are sensitive to varying degrees of approximation to our native language. When young adults were shown strings of 8 letters for just a tenth of a second, they could, on average, report 53% of strings made up of letters randomly sampled with equal probabilities (zero-order approximations to English such as 'CVGJCDHM'). They could report 69% of strings where the letters were sampled according to their individual frequencies in written English (first-order approximations like 'RPITCQET'), 78% of second-order approximation strings which preserve common bigram sequences of English (e.g., 'UMATSORE'), and 87% of fourth-order approximating strings made up of common tetragrams in English (like 'VERNALIT'). Clearly, the participants' span of apprehendsion of more regular orthographic sequences was greater than for less regular ones. The advantage of first-order over zero-order demonstrates that our perceptual systems are sensitive to the fact that some letters occur in our written language more often than others and that our pattern-recognition units for letters have their thresholds tuned accordingly. The advantage of second-order over first-order shows that our pattern recognition system is tuned to the expected frequency of bigrams. The advantage of fourth-order over second-order demonstrates that we are tuned to orthographic chunks four letters long. These chunking effects extend upwards through the levels of the representational hierarchy, and we can rest assured that in 1954 the undergraduate participants in the Miller et al. study would have been able to report rather more than the first eight letters of the string 'One, two, three o'clock, four o'clock, rock…'

1.2. Phonotactics

We are very good at judging whether nonwords are nativelike or not, and young children are sensitive to these regularities when trying to repeat nonwords (Treiman and Danis 1988). Phonotactic competence simply emerges from using language, from the primary linguistic data of the lexical patterns that a speaker knows (Bailey and Hahn 2001). Frisch et al. (2001) asked native speakers to judge nonword stimuli for whether they were more or less like English words. The nonwords were created with relatively high or low probability legal phonotactic patterns as determined by the logarithm of the product of probabilities of the onset and rime constituents of the nonword. The mean wordlikeness judgments for these nonword stimuli had an extremely strong relationship with expected probability ($r = .87$). An emergentist account of phonotactic competence is thus that any new nonword is compared to the exemplars that are in memory: the closer it matches their characteristics, the more wordlike it is judged. The gathering of such relevant distributional data starts in infancy. Saffran, Aslin, and Newport (1996) demonstrated that 8 month-old infants exposed for only 2 minutes to unbroken strings of nonsense syllables (for example, *bidakupado*) are able to detect the difference between three-syllable sequences that appeared as a unit and sequences that also appeared in their learning set but in random order. These infants managed this learning on the basis of statistical analysis of phonotactic sequence data, right at the age when their caregivers start to notice systematic evidence of their recognizing words.

1.3. Lexical recognition and production

The recognition and production of words is a function of their frequency of occurrence in the language. For written language, high frequency words are named more rapidly than low frequency ones (Forster and Chambers 1973), they are more rapidly judged to be words in lexical decision tasks (Forster 1976), and they are spelled more accurately (Barry and Seymour 1988). Auditory word recognition is better for high frequency than low frequency words (Luce 1986). Kirsner (1994) has shown that there are strong effects of word frequency on the speed and accuracy of lexical recognition processes (in speech perception, reading, object naming, and sign perception) and lexical production processes (speaking, typing, writing, and signing), in children and adults, in L1 and in L2.

Abstraction is an automatic consequence of aggregate activation of high-frequency exemplars, with regression towards central tendencies as numbers of highly similar exemplars increase. Thus there is a single voice advantage – words repeated in the same voice are better recognized than those in a different voice – and this advantage is greater for low frequency words: 'old' words which have been frequently experienced in various places by a variety of speakers inspire 'abstract' echoes, obscuring context and voice elements of the study trace (Goldinger 1998).

1.4. Phonological awareness

Children's awareness of the sounds of their language, particularly at the segmental levels of onset-rime and phoneme, is important in their acquisition of literacy (Ellis and Large 1987). It is an awareness that develops gradually. Thomson, Goswami, and Hazan (2003) demonstrated that 4-7 year old children are better able to identify the word with the odd sound in the Bradley and Bryant (1983) odd-one-out task when the spoken stimuli were from dense phonological neighborhoods where there are lots of words which share these rhymes (e.g., 'bag, rag, jack'), rather than when the stimuli came from sparse ones (e.g., 'pig, dig, lid'). The children were also better in short-term memory span tasks at remembering nonword triples from dense phonological neighborhoods (like 'cham, shen, deek') than triples like 'deeve, chang, shem' derived from sparse ones. These phonological neighborhood density effects are driven by vocabulary age, not by chronological age. Metsala and Walley (1998) proposed a 'lexical restructuring hypothesis' of these effects whereby, as vocabulary increases, more and more similar words are acquired; this drives an increasingly well-specified representation of these words in terms of subunits like onset and rime, and is an effect which occurs first in dense phonological neighborhoods. It is the learner's knowledge of individual lexical items which drives the abstraction process.

1.5. Spoken word recognition

The most general probabilistic tuning is that auditory word recognition is better for high frequency than low frequency words (Luce 1986). Thus the recognition units for high frequency words have been primed and are set at higher resting levels than those for infrequent words. But the speech signal unfolds over time and the processes of word recognition begin with the very onset of speech. The 'Cohort Model' of speech perception (Marslen-Wilson 1990) proposes that the initial phoneme of a word activates the set of all words in the lexicon which begin that way. Consider the recognition of the word *elephant* according to the cohort model. Phonemes are recognized categorically and on-line in a left-to-right fashion as they are spoken. Hearing /ɛ /, a large cohort of words might be activated in the unconscious mind of an educated English listener [aesthetic, any, .., ebony, ebullition, echelon, ... , economic, ecstasy, .., element, elephant, elevate, ..., entropy, entry,…, extraneous,…], if every English word beginning in this fashion, the cohort would comprise 324 recruits or more. As further information comes in, words inconsistent with the phoneme string are eliminated from the cohort. Thus at /ɛ l / the number of possible words in the cohort set might drop to a maximum of 28: [elbow, elder, eldest, elegance, elegiac, elegy, element, elemental, elementary, elephant, elephantine, elevate, elevation,…,]. At the next point in processing the spoken word, /ɛ l ɔ /, there are perhaps 12: [elegiac, elegy, element, elemental, elementary, elephant, elephantine, elevate, elevation, elevator, elocution, eloquent; N=12]. At /ɛ l ɔ f /, just 2: [elephant, elephantine]. And one more phoneme reduces any uncertainty, unambiguously signaling the single candidate [elephant]. This is the "uniqueness point," the point in a word at which it can be uniquely identified.

This model explains basic neighborhood effects in speech recognition whereby word recognition is harder when there are lots of words that begin in the same way. But the frequency tuning of individual word detectors affects cohort selection too. Marslen-Wilson (1990) proposed that activation in the cohort varies so that items are not simply "in or out." Rather, higher frequency words get more activation from the same evidence than do low frequency words. This assumption provides a means for accounting for lexical similarity effects, whereby a whole neighborhood of words is activated but the higher frequency words get more activation. Listeners are slower at recognizing low frequency words with high frequency neighbors because the competitors are harder to eliminate. In sum, the Cohort Model proposes that the initial phoneme activates a cohort of words starting with that phoneme, words in the cohort are activated according to their frequency, initial activation is bottom-up, and context effects play a top-down constraining role after initial cohort activation. Such effects demonstrate that our language processing system is sensitive both to the frequency of individual words and to the number of words which share the same beginnings (at any length of computation).

Language learners are sensitive to the frequencies and consistencies of mappings that relating written symbols and their sounds. To the extent that readers are able to construct the correct pronunciations of novel words or nonwords, they have been said to be able to apply sub-lexical "rules" which relate graphemes to phonemes (Coltheart et al. 1993) or larger orthographic units to their corresponding rimes or syllables (Ehri 1998; Goswami 1999; Glushko 1979; Treiman et al. 1995). For the case of adults reading English, words with regular spelling-sound correspondences (like *mint*) are read with shorter naming latencies and lower error rates than words with exceptional correspondences (cf. *pint*) (Coltheart 1978). Similarly, words which are consistent in their pronunciation in terms of whether this agrees with those of their neighbors with similar orthographic body and phonological rime (*best* is regular and consistent in that all *-est* bodies are pronounced in the same way) are named faster than inconsistent items (*mint* is regular in terms of its grapheme-phoneme conversion (GPC) rule, but inconsistent in that it has *pint* as a neighbor) (Glushko 1979). The magnitude of the consistency effect for any word depends on the summed frequency of its friends (similar spelling pattern and similar pronunciation) in relation to that of its enemies (similar spelling pattern but dissimilar pronunciation) (Jared, McRae, and Seidenberg 1990). Adult naming latency decreases monotonically with increasing consistency on this measure (Taraban and McClelland 1987). Because of the power law of learning, these effects of regularity and consistency are more evident with low frequency words than with high frequency ones where performance is closer to asymptote (Seidenberg et al. 1994).

1.6. Morphosyntax

Morphological processing, like reading and listening, shows effects of neighbors and false friends where, even within the regular paradigm, regular inconsistent items (e.g., *bake*-*baked* is similar in rhyme to neighbors *make*-*made*, and *take*-*took* which have inconsistent past tenses) are produced more slowly than entirely regular ones (e.g., *hate*-*hated*, *bate*-*bated*, *date*-*dated*) (Daugherty and Seidenberg 1994). These neighborhood effects, like all of the frequency effects across all domains of language processing that are so well modeled by connectionist simulations, attest the veracity of the core assumption of usage-based accounts: the language processing system is affected by every instance of usage, echoes of each usage are retained in memory, and the collaboration of these exemplars tunes the operations of the processing system. Ellis and Schmidt (1998) measured production of regular and irregular forms as learners practiced an artificial second language where regularity and frequency were factorially combined. Accuracy and latency data demonstrated frequency effects for both regular and irregular forms early in the acquisition process. However, as learning progressed, the frequency effect on regular items diminished whilst it remained for irregular items – a classic frequency by regularity interaction which is a natural result in connectionist models of morphological ability of simple associative learning principles operating in a massively distributed system abstracting the statistical regularities of association using optimal inference (MacWhinney and Leinbach 1991; Plaut et al. 1996).

1.7. Formulaic Language

Just as we learn the common sequences of sublexical components of our language, the tens of thousands of phoneme and letter sequences large and small, so also we learn the common sequences of words. Formulae are lexical chunks which result from binding frequent collocations (Pawley and Syder 1983). Large stretches of language are adequately described by finite-state-grammars, as collocational streams where patterns flow into each other. Sinclair (1991) summarized this as the Principle of Idiom "a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segments. To some extent this may reflect the recurrence of similar situations in human affairs; it may illustrate a natural tendency to economy of effort; or it may be motivated in part by the exigencies of real-time conversation." Rather than its being a rather minor feature, compared with grammar, Sinclair suggested that for normal texts, the first mode of analysis to be applied is the idiom principle, as most of text is interpretable by this principle. We process collocations faster and we are more inclined therefore to identify them as a unit (Schooler 1993; Bod 2001). 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. Much of language production consists of piecing together the ready-made units appropriate for a particular situation, and much of comprehension relies on knowing which of these patterns to predict in these situations.

1.8. Language Comprehension

The Competition Model (Bates and MacWhinney 1987; MacWhinney 1987, 1997) emphasizes lexical functionalism where syntactic patterns are controlled by lexical items. Lexical items provide cues to functional interpretations for sentence comprehension or production. Some cues are more reliable than others. The language learner's task is to work out which are the most valid predictors. The Competition Model is the paradigmatic example of constraint-satisfaction accounts of language comprehension. Consider the particular cues that relate subject-marking forms to subject-related functions in the English sentence*, The learner counts the words*. They are preverbal positioning (*learner* before *counts*), verb agreement morphology (*counts* agrees in number with *learner* rather than *words*), sentence initial positioning, and use of the article *the*. Case-marking languages, unlike English, would additionally include nominative and accusative cues in such sentences. The corresponding functional interpretations include actor, topicality, perspective, givenness, and definiteness. Competition model studies analyze a corpus of exemplar sentences which relate such cue combinations with their various functional interpretations, thus to determine the regularities of the ways in which a particular language expresses, for example, agency. They then demonstrate how well these probabilities determine (i) cue use when learners process that language, and (ii) cue acquisition – the ease of learning an inflection is determined by its cue validity, a function of how often an inflection occurs as a cue for a certain underlying function (cue availability) and how reliably it marks this function (cue reliability) (MacWhinney 1997).

For illustration of some more particular cues in sentence comprehension, consider the utterance "*The plane left for the …*" Does *plane* refer to a geometric element, an airplane, or a tool? Does *left* imply a direction, or is it the past tense of the verb *leave* in active or in passive voice? Odds on that your interpretation is along the lines in *The plane left for the East Coast*, and that you would feel somewhat led up the garden path by a completion such as *The plane left for the reporter was missing*. But less so by *The note left for the reporter was missing* (Seidenberg 1997). Why? Psycholinguistic experiments show that fluent adults resolve such ambiguities by rapidly exploiting a variety of probabilistic constraints derived from previous experience. There is the first-order frequency information: *plane* is much more frequent in its vehicle than its other possible meanings, *left* is used more frequently in active rather than passive voice. Thus the ambiguity is strongly constrained by the frequency with which the ambiguous verb occurs in transitive and passive structures, of which reduced relative clauses are a special type. On top of this there are the combinatorial constraints: *plane* is an implausible modifier of noun *left*, so *plane left* is not a high probability noun phrase, and is thus less easy to comprehend as a reduced relative clause than *note left* because it is much more plausible for a note to be left than to leave.

Studies of sentence processing show that fluent adults have a vast statistical knowledge about the behavior of the lexical items of their language. They know the strong cues provided by verbs, in English at least, in the interpretation of syntactic ambiguities. Fluent comprehenders know the relative frequencies with which particular verbs appear in different tenses, in active vs. passive and in intransitive vs. transitive structures, the typical kinds of subjects and objects that a verb takes, and many other such facts. This knowledge has been acquired through experience with input that exhibits these distributional properties and through knowledge of its semantics. This information is not just an aspect of the lexicon, isolated from 'core' syntax; rather, it is relevant at all stages of lexical, syntactic and discourse comprehension (McKoon and Ratcliff 1998; Seidenberg and MacDonald 1999). Frequent analyses are preferred to less frequent ones.

1.9. Implications for Language Learning and Instruction

There is no scope here for further review of psycholinguistic effects. I refer you to Altman (1997), Ellis (2002), Gernsbacher (1994), Harley (1995). McKoon and Ratcliff (1998) and Jurafsky (2002) for more complete treatment of these phenomena at all levels of language processing, in comprehension and production, in first and second language, from semantics, through syntax and grammaticality, right down to the tuning of infants' iambic/trochaic bias in their language-specific production of prosody. But what is here is surely enough to illustrate that the constructicon is huge indeed, involving tens of thousands of pieces, large and small, and mappings across several input and output modalities and to semantic and conceptual systems. And *all* of these associations are probability tuned.

Fluent native speakers have figured out language by an implicit tallying of frequencies of occurrence and mapping. Language learners have to do the same: they simply cannot achieve the optimality of nativelike fluency without having acquired this probabilistic knowledge. Luckily, of course, they don't have to consciously count the occurrences and their interpretations. As is clear from introspection, this frequency information is acquired implicitly, it is an incidental product of usage. It doesn't seem like we spend our time counting the units of language, instead, when we use language, we are conscious of communicating. Yet in the course of conversation we naturally acquire knowledge of the frequencies of the elements of language and their mappings. As Hasher and Chromiak (1977) put it: "the processing of frequency may fall into the domain of what Posner and Snyder (1975) have called 'automatic processes.' That is, of processes which the organism runs off both without any awareness of the operation, with no intention of doing so, and with little effort, in the sense that the tagging of frequency has little impact on one's ability to simultaneously attend to other aspects of a situation, such as the interpretation of an ongoing conversation" (Hasher and Chromiak 1977: 173). This knowledge, at the very core of communicative competence, is acquired on the job of language processing. 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. Each activation results in an increment of facilitated processing. It's a power function which relates improvement and practice, rather than a linear one, but it's a process of counting and tuning nonetheless (Ellis 2002). Whatever else traditional grammar books, teachers, or other explicit pedagogical instruction can give us towards effective language learning, it is not this frequency information. A dictionary can't give you the odds, nor a grammar. The only source is the number of appropriate usages. Which is why an essential component of language experience and language instruction is communicative input and output.

In summary of the first half of this account of language acquisition, the bulk of language acquisition is implicit learning from usage. Implicit learning supplies a distributional analysis of the problem space: frequency of usage determines availability of representation according to the power law of learning, and this process tallies the likelihoods of occurrence of constructions and the relative probabilities of their mappings between aspects of form and interpretations, with generalization arising from conspiracies of memorized utterances collaborating in productive schematic linguistic constructions. In these ways, unconscious learning processes, which occur automatically during language usage, are necessary in developing the *rationality* of fluency (Anderson 1989; Ellis 2005; Jurafsky 2002).

2. Explicit attentive registration of linguistic constructions

A central and longstanding theme in second language research has concerned the interface between explicit and implicit knowledge. Krashen's (1985) Input Hypothesis was a non-interface position which posited that although adults can both subconsciously acquire languages and consciously learn about language, nevertheless (i) subconscious acquisition dominates in second language performance; (ii) learning cannot be converted into acquisition; and (iii) conscious learning can be used only as a Monitor, i.e. an editor to correct output after it has been initiated by the acquired system. The phenomena gathered thus far lend support to the importance of implicit/subconscious acquisition of language. Nevertheless, these incidentals are not sufficient. Many aspects of language are unlearnable, or at best only very slowly acquirable, from implicit processes alone. Which is why an attentive focus on the form-meaning relation is also necessary in the initial registration of pattern recognizers for constructions.

If implicit naturalistic acquisition was all there was to it, then second language acquisition would be as effective as first language acquisition, and would routinely proceed to an endpoint of fluent and proficient success for all individuals who engage naturalistically in communication in their L2. But this is not the case. It is a defining concern of second language research that there are certain aspects of language to which second language learners commonly prove impervious, where input fails to become intake (Corder 1967).

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

> 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)

Schmidt concluded his report of Wes with a call for research on the proposition that: "in addition to communicative effort, cognitive effort is a necessary condition for successful adult SLA" (Schmidt 1984: 14). Clearly he was suggesting a cognitive effort above and beyond the implicit learning that I have been describing so far. Six years later, Schmidt (1990) proposed in his *noticing* hypothesis that a conscious involvement, explicit learning, was required for the conversion of input to intake: it is necessary that the learner notices the relevant linguistic cues.

This idea has rightly become a cornerstone of second language research. A strong form of the noticing hypothesis is that attention must be paid to some aspect of the stimulus environment and that aspect must be noticed before a mental representation of it can first be formed. I believe that this is broadly correct, although with two provisos. The first is the strong form of the implicit tallying hypothesis which I have explained in the first half of this paper- that once a stimulus representation is firmly in existence, that stimulus need never be noticed again; yet as long as it is attended for use in the processing of future input for meaning, its strength will be incremented and its associations will be tallied and implicitly catalogued. The second is that implicit learning is clearly sufficient for the successful formation of new chunks from the binding of adjacent or successive items which are experienced repeatedly. Implicit learning is specialized for incremental cumulative change: (i) the tuning of strengths of preexisting representations, and (ii) the chunking of contiguous or sequential existing representations. Otherwise, new associations are best learned explicitly.

Attention is required in order to bind features to form newly integrated objects. Attention carves out for conscious experience the correct subset of conjunctions amidst the mass of potential combinations of the features present in a scene. Attentional focus is the solution to Quine's (1960) ‘gavagai’ problem that single words cannot be paired with experiences since they confront experience in clusters. Imagine a second language community who say ‘gavagai’ when confronted by a rabbit. Other things being equal, it is natural to translate the word as ‘rabbit,’ but why not translate it as, say, ‘undetached rabbit-part’ since any experience which makes the use of ‘rabbit’ appropriate would also make that of ‘undetached rabbit-part’ appropriate. But guided attention, focused by sharing the gaze and actions of another, scaffolded by interaction that creates some focus on form or consciousness-raising, makes salient the appropriate features. Explicit, episodic memory systems then rapidly and automatically bind together disparate cortical representations into a unitary representation of these new conjunctions of arbitrarily paired elements (Squire 1992) – a unitary representation that can then be recalled by partial retrieval cues at a later time. Thus attention, noticing, and explicit memory are key to the formation of new pattern recognition units.

The noticing hypothesis subsumes various ways in which SLA can fail to reflect the input (Ellis 2002b point 3). In what follows here I will consider just two of these: failing to notice cues because they are not salient, and failing to notice that cues need to be processed in a different way from that relevant to L1.

2.1. Failing to notice cues because they are not salient

While 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 morphological inflections like that third person singular *s* in English, are not. Inflections marking grammatical meanings such as tense are often redundant since they are usually accompanied by temporal adverbs which indicate the temporal reference. The high salience of these temporal adverbs leads L2 learners to attend to them and to ignore the grammatical tense.

The remedy is explicit learning. In these situations, some type of form-focused instruction or consciousness raising (Sharwood-Smith 1981) can help the learner to 'notice' the cue in the first place. Schmidt summarized it thus: "since many features of L2 input are likely to be infrequent, non-salient, and communicatively redundant, intentionally focused attention may be a practical (though not theoretical) necessity for successful language learning" (Schmidt 2001). Terrell characterized explicit grammar instruction as "the use of instructional strategies to draw the students' attention to, or focus on, form and/or structure" (Terrell 1991), with instruction targeted at increasing the salience of inflections and other commonly ignored features by firstly pointing them out and explaining their structure, and secondly by providing meaningful input that contains many instances of the same grammatical meaning-form relationship. An example is 'processing instruction' (VanPatten 1996) which aims to alter learners' default processing strategies, to change the ways in which they attend to input data, thus to maximize the amount of intake of data to occur in L2 acquisition. Once consolidated into the construction, it is this new cue to interpretation of the input whose strengths are incremented on each subsequent processing episode. The cue doesn't have to be repeatedly noticed thereafter; once consolidated, mere use in processing for meaning is enough for implicit tallying.

2.2. Preservation and transfer-The magnetism of L1

Other common situations where implicit learning does not take place in SLA involve L1 entrenchment. The initial state of the neural stuff involved in language processing is one of plasticity whereby structures can emerge from experience as the optimal representational systems for the particular L1 they are exposed to. Infants between 1 and 4 months of age can perceive the phoneme contrasts of every possible language, but by the end of their first year they can only distinguish the contrasts of their own (Werker and Tees 1984; Werker and Lalonde 1988). In contrast to the newborn infant, the starting disposition of the neural stuff for second language acquisition is already tuned to the L1 and is set in its ways. What might be examples of two separate phonemic categories, /r/ and /l/, for an L1 English language speaker are all from the same phonemic category for an L1 Japanese speaker. And in adulthood the Japanese native cannot but perceive /r/ and /l/ as one and the same. The same form category is activated on each hearing and incremented in strength as a result. And whatever the various functional interpretations or categorizations of these assorted hearings, their link to this category is strengthened every time, rightly or wrongly. The phonetic prototypes of one's native language act like perceptual magnets, or attractors, distorting the perception of items in their vicinity to make them seem more similar to the prototype (Kuhl and Iverson 1995). Under normal L1 circumstances, usage optimally tunes the language system to the input. A sad irony for an L2 speaker under such circumstances of transfer is that more input simply compounds their error; they dig themselves ever deeper into the hole begun and subsequently entrenched by their L1.

Proven remedies here make use of exaggerated stimuli and adaptive training (McClelland, Fiez, and McCandliss 2002). This, like errorless learning techniques more generally, ensures that subsequent responding correctly differentiates the new contrast rather than compounding the old confusion (Baddeley and Wilson 1994; Baddeley 1992; Evans et al. 2000). Contrastive pairs such as “rock” vs. “lock” are made more exaggerated by extending their outer limits beyond the normal range until L2 learners can perceive the difference. They start with these discernible poles and then, as repeated occurrences are correctly identified, the discrimination is made more difficult. The use of such exaggerated stimuli and adaptive training leads to rapid learning, while the use of difficult stimuli with no adaptive modification produced little or no benefit (McCandliss et al. 2002; McClelland 2001).

Other examples of learner’s first language experience leading them to look elsewhere for their cues to interpretation include English learners of Chinese who have difficulty with tones, and Japanese learners of English with the article system, both problems resulting from zero use in the L1. Similarly, with case marking, word order, agreement, and noun animacy, along with other cues, all helping to identify the subject of a sentence to lesser or greater degree in different languages, learners carry their L1 cue strength hierarchy across to their L2, only gradually resetting the ordering after considerable L2 experience (MacWhinney 1987), if at all (MacWhinney 2001). Under normal L1 circumstances, usage optimally tunes the language system to the input; under these circumstances of low salience of L2 form, all the extra input in the world might sum to naught, and we describe the learner as having ‘fossilized.’ Again, the instructional techniques that are commonly marshaled in such circumstances accord to the general principle of explicit learning in SLA: If you can change the cues that learners focus upon in their language processing, so you change what their implicit learning systems tune.

And the data show that these forms of attentional focus are effective and that language acquisition can be speeded by such provision. Reviews of the experimental and quasi-experimental investigations into the effectiveness of L2 instruction (Doughty and Williams 1998; Ellis and Laporte 1997; Hulstijn and DeKeyser 1997; Lightbown, Spada, and White 1993; Long 1983; Spada 1997), particularly the comprehensive meta-analysis of Norris and Ortega (2000), demonstrate 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. This is not to say that just providing learners with pedagogical rules will make them into fluent language users. Far from it (Krashen and Terrell 1983; Krashen 1985), because then the learner neither gets the exemplars nor the tuning. Pegagogical rules are only properly effective when demonstrated in operation with a number of illustrative exemplars of their application (Ellis 1993) and when they can subsequently thus affect input processing in usage.

We learn language while using language. When things go right, when routine communication comes easy and fluent, this time on task tunes our skills without us giving much thought to it. When thinks go wrong, when communication breaks down, we try hard to negotiate meaning, and we learn a lot about linguistic construction in the process. Implicit learning of language occurs during fluent comprehension and production. Explicit learning of language occurs in our conscious efforts to negotiate meaning and construct communication. There is a wide range of attentive processes of working memory which contribute to noticing and the consolidation of a pattern-recognition unit, a unitized representation of a linguistic construction. I review the range of these in Ellis 2005,

2.3. Brain processes, complementary memory systems, and interface: Towards a cognitive science of usage-based acquisition

These are some of the psycholinguistic processes involved in second language acquisition. One can view them from many perspectives, focusing variously on learner, language, input, sociolinguistic context, cognitive representations and processes, or brain. I want to close by briefly considering related research in cognitive neuroscience into the ways the brain processes and represents language. There are important insights to be had about these psycholinguistic processes of language acquisition from current work in cognitive science (including the use of connectionist models of learning and representation) and neuroscience (including cognitive neuropsychology and brain imaging).

Humans have two separable but complementary memory systems (Squire and Kandel 1999). Explicit memory refers to situations where recall involves a conscious process of remembering a prior episodic experience; it is tapped by tasks like recall and recognition where the individual is consciously aware of the knowledge held. Explicit memories include all situations where we remember the context of learning, declarative learning (for example, of verbal rules like '<u>i</u> before *e* except after *c*'), one-trial learning that the Quinean for *rabbit* is *gavagai*, and our autobiographical record of specific episodes. Implicit memory is where there is facilitation of the processing of a stimulus as a function of a prior encounter with an identical or related stimulus but where 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, or spelled 'receive,' 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. Amnesic patients cannot learn new names or concepts or arbitrary paired-associates, they cannot remember any episode more than a few minutes after it has happened. But amnesic patients show normal implicit memory abilities: they learn new perceptual and motor skills, they show normal priming effects, they evidence normal classical conditioning.

Neural systems in the hippocampus and related limbic structures allow the consolidation of explicit memories. The hippocampus rapidly and automatically binds together disparate cortical representations into a unitary representation which can then be recalled by partial retrieval cues at a later time. Thus the hippocampal system confers a sense of unity to a particular experience (i.e., an episodic memory) – otherwise, these experiences would remain just a jumble of loosely connected features and facts (Squire 1992; Squire and Kandel 1999). By forming unitized memory representations, the hippocampal region performs the information-processing function of forming pattern-recognition units for new stimulus configurations, of consolidating new bindings; these are then adopted by other brain regions in the neocortex where they subsequently partake in implicit tuning (Gluck, Meeter, and Myers 2003; O'Reilly and Norman 2002).

The neocortical system underpins implicit learning and is the locus of the frequency effects. 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 representing the visual appearance, the meaning, the sound of the word; and 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. But 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, or new concepts, or new episodic records.

Recent brain imaging studies support this view of complementary memory systems in the cortex and hippocampus. Hippocampal structures in the medial temporal lobes are very active early in training, when subjects are learning about stimulus – stimulus regularities and evolving new stimulus representations, but less active later in training when other brain regions (including the striatum and basal ganglia) are using these representations to perform on the task (Poldrack et al. 2001). Other imaging studies also demonstrate hippocampal system activations during the encoding of memories, with these encoding activations indexing stimulus novelty in that they are greater for stimuli seen initially rather than repeatedly (Tulving et al. 1994; Stern et al. 1996). Repeated memories result in activation elsewhere: lesion and imaging studies provide convergent evidence that implicit memory as indexed by different forms of repetition priming reflect process-specific plasticity in separate neocortical regions, with visual, auditory, and tactual priming being mediated by changes in visual, auditory, and somatosensory neocortices respectively (Gabrieli 1998). Thus, repetition priming in a given domain appears to reflect experience-induced changes in the same neural networks that subserved initial perceptual processing in that domain, with these changes facilitating the subsequent reprocessing of the stimuli.

The two complementary memory systems, the hippocampal system and the neocortical sensori-motor areas, allow the co-existence of instances and abstractions, thus solving the two basic knowledge functions of an organism which needs to be able to acquire both specifics (Where did you park your car today? What is the L2 phrase for 'Two beers, please'?) and generalizations (What's the script for purchasing petrol at the garage? How does the L2 form a plural?), and they prevent the problem of catastrophic interference suffered by purely implicit connectionist mechanisms (McClelland 1998, 1995; O'Reilly and Norman 2002). The neocortex has a slow learning rate to gradually integrate new information with existing knowledge, using overlapping distributed representations to extract the general statistical structure of the environment. In contrast, the hippocampus learns rapidly, assigning distinctive sparse representations to input patterns to encode the episodic details of specific events while minimizing interference.

Further such research into these complementary learning and memory systems, as well as into the unique contributions of the attentional systems of the prefrontal cortex in binding features to form newly integrated object representations, and how neuronal synchrony is related to perceptual integration, buildup of coherent representations, attentional selection, and awareness (Cleeremans 2003; Ellis2005, 2006) gives promise, I think, for understanding the cognitive neuroscience of the ways that linguistic constructions are first noticed and registered, and thence figured and tuned into the system (Ellis 2003, 2008. As the focus of GURT 2003 rightly affirmed, these issues lie at the heart of language acquisition and cognitive science both.

References

Anderson, John R.
1989 A rational analysis of human memory. In *Varieties of Memory and Consciousness: Essays in Honour of Endel Tulving,* Henry L.I. Roediger and Fergus I.M. Craik (eds.), 195–210. Hillsdale, NJ: Erlbaum.

Altman, Gerry T.
1997 *The Ascent of Babel*. Oxford: Oxford University Press.

Baddeley, Alan D.
1992 Implicit memory and errorless learning: A link between cognitive theory and neuropsychological rehabilitation? In *Neuropsychology of Memory*, Larry R. Squire and Nelson Butters, (eds.), 309–314. New

York: Guillford Press.

Baddeley, Alan D., and Barbara A. Wilson
1994 When implicit learning fails: Amnesia and the problem of error elimination. *Neuropsychologia* 32 (1): 53–68.

Bailey, Todd M., and Ulrike Hahn
2001 Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language* 44: 568–591.

Barry, Christopher, and Philip H. Seymour
1988 Lexical priming and sound-to-spelling contingency effects in nonword spelling. *Quarterly Journal of Experimental Psychology A* 40 (1, Sect A): 5–40.

Bates, Elizabeth, and Brian MacWhinney
1987 Competition, variation, and language learning. In *Mechanisms of Language Acquisition*, Brian MacWhinney (ed.), 157–193. Hillsdale, NJ: Erlbaum.

Biber, Douglas, Susan Conrad, and Randi Reppen
1998 *Corpus Linguistics: Investigating Language Structure and Use.* New York: Cambridge University Press.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan
1999 *Longman Grammar of Spoken and Written English.* Harlow, UK: Pearson Education.

Bod, Rens
2001 Sentence memory: Storage vs. computation of frequent sentences. Paper presented at CUNY-2001, at Philadelphia, PA.

Bradley, L., and P. E. Bryant
1983 Categorising sounds and learning to read: A causal connection. *Nature* 310:419–421.

Bybee, Joan, and Paul Hopper (eds).
2001 *Frequency and the Emergence of Linguistic Structure*. Amsterdam/Philadelphia: John Benjamins.

Cleeremans, Axel (ed.)
2003. *The Unity of Consciousness: Binding, Integration, and Dissociation*. Oxford: Oxford University Press.

Coltheart, Max
1978 *Lexical access in simple Reading Tasks*. Strategies of information processing, Geoffrey Underwood, (ed.), 151–216. San Diego, CA: Academic Press.

Coltheart, Max., Brent Curtis, Paul Atkins, and Michael Haller
1993 Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review* 100:589–608.

Corder, Stephen P.
1967 The significance of learners' errors. *International Review of Applied Linguistics* 5:161–169.

Daugherty, Kim G., and Mark S. Seidenberg.
1994 Beyond rules and exceptions: A connectionist approach to inflectional morphology. In *The Reality of Linguistic Rules*, Susan D. Lima, Roberta L. Corrigan, and Gregory K. Iverson, (eds.), 353–388. Amsterdam/Philadelphia: John Benjamins.

Doughty, Catherine, and Jessica Williams (eds.)
1998 *Focus on Form in Classroom Second Language Acquisition*. New York: Cambridge University Press.

Ehri, Linnea C.
1998 Word reading by sight and by analogy in beginning readers. In *Reading and Spelling: Development and Disorders*, Charles. Hulme, and R. Malatesha Joshi, (eds.), 87–112. Mahwah, NJ: Erlbaum.

Ellis, Nick C.
1993 Rules and instances in foreign language learning: Interactions of explicit and implicit knowledge. *The European Journal of Cognitive Psychology* 5 (3): 289–318.
2002 Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition* 24 (2): 143–188.
2002 Reflections on frequency effects in language processing. *Studies in Second Language Acquisition* 24 (2): 297–339.

(2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition*. Oxford: Blackwell.

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

(2006). Selective attention and transfer phenomena in SLA: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics, 27*(2), 1-31.

(2008). Optimizing the input: Frequency and sampling in Usage-based and Form-focussed Learning. In M. H. Long & C. Doughty (Eds.), *Handbook of second and foreign language teaching*. Oxford: Blackwell.

Ellis, Nick C., and Nadine Laporte.
1997 Contexts of acquisition: Effects of formal instruction and naturalistic exposure on second language acquisition. In *Tutorials in Bilingualism: Psycholinguistic Perspectives*, Annette M. DeGroot

and Judith F. Kroll, (eds.), 53–83. Mahwah, NJ: Erlbaum.

Ellis, Nick C., and B. Large.
1987 The development of reading: As you seek so shall you find. *The British Journal of Psychology* 78 (1): 1–28.

Ellis, Nick C., and Richard Schmidt.
1998 Rules or associations in the acquisition of morphology? The frequency by regularity interaction in human and PDP learning of morphosyntax. *Language & Cognitive Processes* 13 (2–3): 307–336.

Elman, Jeffrey L., Elizabeth A. Bates, Mark H. Johnson, Annett Karmiloff-Smith, Domenico Parisi, and Kim Plunkett.
1996 *Rethinking Innateness: A Connectionist Perspective on Development.* Cambridge, MA: MIT Press.

Evans, J. J., B. A. Wilson, U. Schuri, J. Andrade, A. D. Baddeley, O. Bruna, T.

Forster, Kenneth
1976 Accessing the mental lexicon. In *New Approaches to Language Mechanisms*, Roger J. Wales and Edward Walker (eds.), 257–287. Amsterdam: North Holland.

Forster, K. I., and S. Chambers.
1973 Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior* 12: 627–635.

Frisch, Stefan A., Nathan R. Large, Bushra Zawaydeh, and David B. Pisoni.
2001 Emergent phonotactic generalizations in English and Arabic. In *Frequency and the Emergence of Linguistic Structure*, J. Bybee and P. Hopper, (eds.), 159–180. Amsterdam/Philadelphia: John Benjamins.

Gabrieli, John D. E.
1998 Cognitive neuroscience of human memory. *Annual Review of Psychology* 49:87–115.

Gernsbacher, Morton A.
1994 *A Handbook of Psycholinguistics*. San Diego, CA: Academic Press.

Gluck, Mark A., Martijn Meeter, and Catherine E. Myers.
2003 Computational models of the hippocampal region: linking incremental learning and episodic memory. *Trends in Cognitive Science* 7:269–276.

Glushko, Robert J.
1979 The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance* 5:674–691.

Goldinger, Stephen D.
1998 Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105:251–279.

Goswami, Usha
1999 Causal connections in beginning reading: The importance of rhyme. *Journal of Research in Reading* 22:217–240.

Harley, Trevor A.

1995 *The Psychology of Language: From Data to Theory.* Hove, UK: Taylor & Francis.

Hasher, Lynn, and Walter Chromiak.

1977 The processing of frequency information: An automatic mechanism? *Journal of Verbal Learning and Verbal Behavior* 16: 173–184.

Hulstijn, Jan, and Robert DeKeyser (eds.)

1997 Testing SLA theory in the Research Laboratory. Special issue, *Studies in Second Language Acquisition* 19.

James, William

1890 *The Principles of Psychology*. Vol. 2. New York: Holt.

Jared, D., K. McRae, and M. S. Seidenberg

1990 The basis of consistency effects in word naming. *Journal of Memory and Language* 29:687–715.

Jurafsky, Dan

2002. Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. In *Probabilistic Linguistics,* Rens Bod, Jennifer Hay and Stephanie Jannedy (eds.), 39–96. Harvard, MA: MIT Press.

Kirsner, Kim

1994 Implicit processes in second language learning. In *Implicit and Explicit Learning of Languages*, Nick C. Ellis (ed.), 283–312. San Diego, CA: Academic Press.

Krashen, Stephen D.

1985 *The Input Hypothesis: Issues and Implications*. London: Longman.

Krashen, Stepen D., and Tracy Terrell.

1983 *The Natural Approach: Language Acquisition in the Classroom*. Oxford: Pergamon Press.

Kuhl, Patricia K., and Paul Iverson.

1995 Linguistic experience and the "perceptual magnet effect." In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, Winifred Strange (ed.), 121–154. Timonium, MD: York Press.

Lightbown, Patsy M., Nina Spada, and Lydia White (eds.)

1993 The Role of Instruction in Second Language Acquisition. Special issue, *Studies in Second Language Acquisition* 15.

Long, Michael H.

1983 Does second language instruction make a difference? A review of research. *TESOL Quarterly* 17:359–382.

Luce, Paul A.

1986 A computational analysis of uniqueness points in auditory word recognition. *Perception and Psychophysics* 39:155–158.

MacWhinney, Brian (ed.)

1987 The Competition Model. In *Mechanisms of Language Acquisition.*

1987 Applying the Competition Model to bilingualism. *Applied Psycholinguistics* 8 (4): 315–327.

1997 Second language acquisition and the Competition Model. In *Tutorials in Bilingualism: Psycholinguistic Perspectives*, Annette M. B. De Groot and Judith F. Kroll (ed.), 113–144. Lawrence Erlbaum Associates: Mahwah, NJ.

2001 The competition model: The input, the context, and the brain. In *Cognition and Second Language Instruction*, Peter Robinson (ed.)., 69–90. New York: Cambridge University Press.

MacWhinney, Brian, and Jared Leinbach

1991 Implementations are not conceptualizations: Revising the verb learning model. *Cognition* 40 (1–2): 121–157.

Marslen-Wilson, William

1990 Activation, competition, and frequency in lexical access. In *Cognitive Models of Speech Processing*, Gerry T. M. Altmann (ed.), 148–172. Cambridge, MA: ACL-MIT Press.

McCandliss, Bruce D., Julie A. Fiez, Athanassios Protopapas, Mary Conway, and James L. McClelland.

2002 Success and failure in teaching the [r]-[l] contrast to Japanese adults. *Cognitive, Affective & Behavioral Neuroscience* 2 (2): 89–108.

McClelland, James L.

1995 Constructive memory and memory distortions: A parallel distributed processing approach. In *Memory distortions: How Minds, Brains, and Societies Reconstruct the Past*, Daniel L. Schacter (ed.), 69–90.

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.

2001 Failures to learn and their remediation: A Hebbian account. In *Mechanisms of Cognitive Development: Behavioral and Neural Perspectives*, James L. McClelland and Robert S. Siegler (eds.), 97–121. Mahwah, NJ: Erlbaum.

McClelland, James L., Julie A. Fiez, and Bruce D. McCandliss

2002 Teaching the /r/-/l/ discrimination to Japanese adults: Behavioral and neural aspects. *Physiology & Behavior* 77 (4–5): 657–662.

McKoon, Gail, and Roger Ratcliff

1998 Memory-based language processing: Psycholinguistic research in the 1990s. *Annual Review of Psychology* 49:25–42.

Metsala, Jamie L., and Amanda C. Walley

1998 Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In *Word Recognition in Beginning Literacy,* Metsala, Jamie L , Linnea C Ehri (eds.), 89–120.

Miller, G. A., J. S. Bruner, and L. Postman

1954 Familiarity of letter sequences and tachistoscopic identification. *Journal of General Psychology* 50: 129–139.

Norris, John, and Louis Ortega

2000 Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning* 50:417–528.

O'Reilly, Randall C., and Kenneth A. Norman

2002 Hippocampal and neocortical contributions to memory: advances in the complementary learning systems framework. *Trends in Cognitive Science* 6:505–510.

Pawley, Andrew, and Frances H. Syder.

1983 Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In *Language and Communication*, Jack C. Richards and Richard W. Schmidt, (eds.), 191–226. London: Longman.

Plaut, David C., James L. McClelland, Mark S. Seidenberg, and Karalyn Patterson

1996 Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review* 94:523–568.

Poldrack, R.A., J. Clark, E. J. Pare-Blagoev, D. Shohamy, J. Creso Moyano, C. Myers, and M. A. Gluck

2001 Interactive memory systems in the human brain. *Nature* 414:546–550.

Posner, Michael I., and Charles R. Snyder

1975 Attention and cognitive control. In. Gordon H. Bower and Janet T. Spence, (eds.). San Diego, CA: Academic Press.

Quine, Willard V. O

1960 *Word and object*. Cambridge, MA: MIT Press.

Rumelhart, David E., and James L. McClelland (eds.)

1986 *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Vol. 2, *Psychological and Biological Models*. Cambridge, MA: MIT Press.

Saffran, Jenry R., Richard N. Aslin, and Elissa L. Newport

1996 Statistical learning by 8-month-old infants. *Science* 274:1926–1928.

Schmidt, Richard

1984 The strengths and limitations of acquisition: A case study of an untutored language learner. *Language, Learning, and Communication* 3:1–16.

1990 The role of consciousness in second language learning. *Applied Linguistics* 11:129–158.

2001 Attention. In *Cognition and second language instruction*, Peter Robinson, (ed.), 3–32. Cambridge: Cambridge University Press.

Schooler, Lael J.

1993 Memory and the statistical structure of the environment. Ph.D. diss., Department of Psychology, Carnegie Mellon University.

Seidenberg, Mark S.

1997 Language acquisition and use: Learning and applying probabilistic constraints. *Science* 275: 1599–1603.

Seidenberg, Mark S., and Maryellen C. MacDonald.

1999 A probabilistic constraints approach to language acquisition and processing. *Cognitive Science* 23: 569–588.

Seidenberg, Mark S., David C. Plaut, Alan S. Petersen, James L. McClelland, and Ken McRae

1994 Nonword pronunciation and models of word recognition. *Journal of Experimental Psychology: Human Perception & Performance* 20 (6): 1177–1196.

Sharwood-Smith, Michael

1981 Consciousness raising and the second-language learner. *Applied Linguistics* 2:159–168.

Sinclair, John

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

Spada, Nina

1997 Form-focused instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching Research* 30: 73–87.

Squire, Larry R.

1992 Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review* 99: 195–231.

Squire, Larry R., and Eric R. Kandel.

1999 *Memory: From Mind to Molecules*. New York: Scientific American Library.

Stern, Chantal E., Suzanne Corkin, R. Gilberto Gonzalez, Alexander R. Guimaraes, John R. Baker, Peggy J. Jennings, Cindy A. Carr, Robert M. Sugiura, Vasanth Vedantham, and Bruce R. Rosen.

1996 The hippocampus participates in novel picture encoding: evidence from functional magnetic resonance imaging. *Proceedings of the National Academy of Sciences of USA*. 93 (16): 8660–8665.

Taraban, Roman, and James L. McClelland

1987 Conspiracy effects in word pronunciation. *Journal of Memory & Language* 26 (6): 608–631.

Terrell, Tracy

1991 The role of grammar instruction in a communicative approach. *The Modern Language Journal* 75:52–63.

Thomson, Jennifer, Usha Goswami, and Valerie Hazan.

2003 The Effect of Neighbourhood Density on Phonological Awareness in Dyslexic and Normally-Developing Readers. Poster presented at the First Workshop of the Centre for Human Communication, at University College London.

Treiman, Rebecca, and Catalina Danis.

1988 Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14:145–152.

Treiman, Rebecca, John Mullennix, Ranka Bijeljac-Babic, and E. Daylene Richmond-Welty.

1995 The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General* 124 (2): 107–136.

Tulving, Endel, Shitij Kapur, Fergus I. M. Craik, Morris Moscovitch, and Sylvain Houle.

1994 Hemispheric encoding/retrieval asymmetry in episodic memory: Positron emission findings. *Proceedings of the National Academy of Sciences of the United States of America*. 91 (6): 2016–2020

VanPatten, Bill

1996 *Input Processing and Grammar Instruction in Second Language Acquisition*. New York: Ablex.

Werker, Janet E., and Chris E. Lalonde.

1988 Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology* 24:672–683.

Werker, Janet E., and Richard C. Tees.

1984 Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7:49–63.