Incidentals that build fluency:
Optimal word processing and its implications for vocabulary acquisition

Nick C. Ellis
University of Wales, Bangor

n.ellis@bangor.ac.uk

Submitted to *New Zealand Studies in Applied Linguistics*

3 June, 2003

Abstract

This article reviews recent research demonstrating that fluent language users are optimal language processors – their unconscious language processing systems are probability tuned to predict the linguistic constructions which are most likely to be relevant in the ongoing discourse context. The first section describes the rational analysis of language processing. The second section provides a summary of psycholinguistic evidence illustrating usage-based optimalization of language representations. The final section considers the implications of these findings for second and foreign language acquisition and instruction. In sum, sophisticated probabilistic information of form-function mappings is a necessary component of native fluency which can only be acquired from communicative usage. Yet, however necessary, these incidentals are not sufficient.

Optimal word processing

Consider the word processing programs you have known. You probably have strong views. Remember the one that crashed at 2am losing your only copy, the one that took ten minutes to search and replace, the one that required perverse three letter combination commands, and the one you have now that replaces spellings, styles and punctuations for you, whether you like it or not, and you still can't figure how to get it to stop? Besides the obvious requirements for speed and reliability, an optimal word processing program would really know what you wanted to do next, and would, like Jeeves, present you with your next needed command, file, or figure, ready on a silver platter. You could, like Bertie Wooster, occasionally exert your independence and decline. But Jeeves would take this with good grace, and would remain constant, knowing just what was needed in the circumstances, providing optimal recommendations while ever allowing you the choice to think better.

There are some technical developments in this direction. One relatively new feature of several programs is that of 'Open Recent'. When a user goes to open a document, the word-processor presents a list of recently opened files to select from. If the sought after file is included in the list, the user is spared the time and effort of searching through the file hierarchy. The program is no mind-reader, it does not know the goals of the user. It simply proffers alternatives using the heuristic that the more recently a file has been opened, the more likely it is to need to be opened now. This simple rule does a surprisingly good job of putting the appropriate files on the list. It does so, as we will see, because the world works that way, and so do people.

Another recent feature is that of predictive text entry. When entering options from a limited set, for example a journal name into a citation management system, we appreciate it if the system suggests the most likely completion of our first few letters, and plumps for that choice when the uniqueness point is reached. Predictive text-entry systems try to anticipate our need and to complete the word stem we have begun typing, using the predictions of a statistical language model: probable pieces of text are made quick and easy to select, improbable pieces of text (low frequency words or text with spelling mistakes) are made harder to choose. The language model learns all the time: if you use a novel word once, it is easier to write next time. These systems are the ergonomic operationalizations of the cohort theory of perception which describes how we recognize speech - we will discuss this shortly. The common feature of such mechanisms is that they tune the availability of selections on the basis of past history of use. They make more likely used options more readily available. Their facility is measurable: we like them, as long as they choose correctly; there is nothing more frustrating than correcting a dumb program's dumb guesses.

Generally, these programs are reasonably successful. They work, as said before, because the world works like that. Things that were likely yesterday tend too to be likely in the same context today, as Jeeves knew full well. (1) Something which has been frequently required in the past is likely to be required now, as reliable a need as a hot bath at the end of the day. (2) Something which has been recently required is likely to be required now: that rather garish cravat that Bertie has regrettably taken a liking to. (3) Something which has been often required in this particular context is likely to be required now: the hangover pick-me-up morning after rather too jolly a night at the club.

These principles have been formally analysed for information need: the 'information retrieval problem', as it has been investigated for borrowing books from libraries (Burrell, 1980), or accessing files from computers, can be addressed because the statistical structure of the environment enables an optimal estimation of the odds that a particular book or file will be needed. The probability that a particular piece of information will be relevant, its 'need probability', can be estimated using Bayesian evaluation procedures whereby the odds ratio for that particular piece of information is a product of its odds ratio given its particular history, i.e. its combined frequency and recency of occurrence, and the current context (Anderson, 1989). Taking such factors into account, an optimal estimation of an item's need probability is possible. Using such a procedure for all items in the set, the most likely ones can then be made more available in readiness, with a cost/accuracy trade off whereby the more items suggested, the greater the chance of a hit, but also the more false positives and the longer the list of items that have to be checked and discounted.

Current word-processing programs typically just use recency of usage in determining their suggestions. Information retrieval analysis suggests that it would be advantageous to take frequency of prior usage into account too – recently used items which also have a history of high frequency overall usage are more likely to be needed than those of infrequent overall history of usage. Context of use could also prove a useful guide: knowing who the particular user is and factoring that in provides a better estimate than averaging over all library or information users, as clients of Amazon will attest; similarly, knowing the user is working on a particular report makes it more likely that they will need the files pertaining to figures for that report, etc. (Schooler & Anderson, 1997). These are all ways of tuning a word processing package for optimal usage. As we will see, they are ways in which we human word processors are tuned for optimal operation too.

Analysis of the ways the world works demonstrates that optimal estimation of an item's need probability is possible. This insight has allowed, over the last twenty or thirty years or so, a sea change in our understanding of human cognition. It has stemmed largely from workers at Carnegie Mellon University, and it concerns the ways that human cognition is rational in that it optimally responds to the statistical contingencies of the world. A major characteristic of the environments that are relevant to human cognition is that they are fundamentally probabilistic (Anderson, 1991) -- as William James said over a century ago, "Perception is of definite and probable things" (James, 1890, p. 82). The rational analysis of memory (Anderson, 1989, 1990; Anderson & Milson, 1989; Schooler & Anderson, 1997) builds on these foundations and on the more recent investigations of information retrieval. Its central hypothesis is that human cognition optimally solves the problems that it faces. This means that if we can find, describe, and properly characterize the problem a cognitive system is trying to solve and find the optimal solution to this problem, then the rational analysis makes the strong prediction that the behaviour of the system will correspond in form to this solution. And the research that has followed in testing this claim suggests that human cognition is indeed rational in this sense, as the last half century's work in psycholinguistics similarly demonstrates that fluent language users are optimal processors of words. The words that we are likely to hear next, their most likely senses, the linguistic constructions we are most likely to utter next, the syllables we are likely to articulate next, the graphemes we are likely to read next, and the rest of what's coming next across all levels of language representation, are made more readily available to us by our language processing systems. In this way our unconscious language mechanisms are like Jeeves: they present up to consciousness the construction that's most likely to be relevant next. Their offering is usually appropriate, but consciousness, like Bertie Wooster, can decline if it has reason to think better. Fluent language users are optimal word processors. How about that!

Human language learning provides the optimal solution to the various problem-spaces posed by language form and the way it relates to its functions. But accounts of language learning rarely consider these aspects of language knowledge at all. Look in the index of your favourite books on acquisition. How many of them mention ‘probability’, or ‘frequency’, or ‘statistics’? Yet language acquisition is not simply a matter of learning linguistic constructions, it requires the tuning of the availability of these for optimal operation as well. In the next section I will briefly summarise some of the psycholinguistic evidence for this claim. In the final section I will introduce some of the implications for second and foreign language learning.

Probabilistic language processing

Counting from 1 to 10 is early content in most second and foreign language courses and an ESL or EFL student is soon secure in the knowledge of what ‘wZn’ means. But should they be so sure? Consider the following wZns: ‘That's wZn for the money, two for the show, three to get ready’; ‘To love wZnself is the beginning of a lifelong romance’; ‘wZnce upon a time...’; ‘Alice in wZnderland’; ‘wZn the battle, lost the war’; ‘How to win life's little games without appearing to try - wZnUpmanship’; ‘the human brain is a wZnnderful thing’. These are different ones. 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 practised comprehender is conscious of just one interpretation – Alice in wZn 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, p. 82). 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 & McClelland, 1986), the competition model of language learning and processing (Bates & MacWhinney, 1987; MacWhinney, 1987, 1997), the recent emphasis on frequency in language acquisition and processing (Bybee & Hopper, 2001; Ellis, 2002b; Jurafsky & Martin, 2000), and proper empirical investigations of the structure of language by means of corpus analysis (Biber, Conrad, & Reppen, 1998; Biber, Johansson, Leech, Conrad, & Finegan, 1999; Sinclair, 1991).

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. I cannot reference all of these studies, it would clog up many pages of the journal here, but you can track down more detail in Ellis (2002a,b) if interested.

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 apprehension 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…’.

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. Phonotactic competence simply emerges from using language, from the primary linguistic data of the lexical patterns that a speaker knows. When adults are asked to judge the degree to which a novel item is like the words of their language, unconscious processes compare the item to the lexical exemplars stored in memory, and the closer it matches their characteristics, the more wordlike it is judged. Thus phonotactic competence emerges from the collaboration of the pronunciation patterns of all of the words a person knows. The gathering of such relevant distributional data starts in infancy. 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. They manage 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.

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, they are more rapidly judged to be words in lexical decision tasks, and they are spelled more accurately. Auditory word recognition is better for high frequency than low frequency words. 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.

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. It is an awareness that develops gradually. Children are better able to identify the word with the odd sound in an odd-one-out task when the spoken stimuli are from dense phonological neighborhoods where there are lots of words which share these rhymes (e.g., 'bag, rag, jack'), rather than when the stimuli come from sparse ones (e.g., 'pig, dig, lid'). They are 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 neighbourhood density effects are driven by the children's vocabulary knowledge, not their chronological age. As vocabulary increases, more and more similar words are acquired and this drives an increasingly well-specified representation of these words in terms of subunits like onset and rime, an effect which occurs first in dense phonological neighborhoods. It is the learner's knowledge of individual lexical items which drives the abstraction process.

Spoken word recognition The speech signal unfolds over time and processes of word recognition begin with the very onset of speech. The initial phoneme of a word activates the set of all words in the lexicon which begin that way. Hearing 'w', a large cohort of English words are activated – wad, ouija, way, …wow, …, Wyoming. Then, as the speech signal unfolds, and more information is received, 'wZ', we narrow the set down, throwing out no-longer viable candidates like waddled, waffle, and wage. But the candidate set is still substantial, including worry, worrying, worryingly, wondrous, and wonder, besides one. This explains neighbourhood effects in speech recognition whereby word recognition is harder when there are lots of words that begin in the same way. Out of context, a particular word can only be identified once we have reached the uniqueness point. Hearing 'waI@', we would already be at the uniqueness point, since the only possible completion is Wyoming. But hearing 'wZn ', we still aren't there, there is still scope for our being wonder-struck. In the cohort model of speech recognition (Marslen-Wilson, 1990), 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 neighbourhood 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. 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).

Reading and Spelling Language learners are sensitive to the frequencies and consistencies of mappings that relate 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 link graphemes to phonemes or larger orthographic units to their corresponding rimes or syllables. 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). 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). 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). Adult naming latency decreases monotonically with increasing consistency on this measure.

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). Ellis and Schmidt (1998) measured production of regular and irregular forms as learners practised 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 & Leinbach, 1991; Plaut, McClelland, Seidenberg, & Patterson, 1996).

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 & Syder, 1983). Large stretches of language are adequately described by finite-state-grammars, as collocational streams where patterns flow into each other. Sinclair (1991, p. 110) 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). 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.

Language Comprehension The Competition Model (Bates & 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 & Ratcliff, 1998; Seidenberg & MacDonald, 1999). Frequent analyses are preferred to less frequent ones.

There is no scope here for further review of psycholinguistic effects. I refer you to Altman (1997), Ellis (2002a), Gernsbacher (1994) and Harley (1995) 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 language construction kit 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.

Implications for Language Learning and Instruction

Fluent native speakers have figured out language by counting 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 & Chromiak, 1977) put it: "the processing of frequency may fall into the domain of what (Posner & 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 & Chromiak, 1977). 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, 2002a). 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 conclusion, 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. 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 a focus on form is also necessary in the initial registration of pattern recognizers for constructions. But that's another story (Ellis, in press).

In sum, sophisticated probabilistic information of form-function mappings is a necessary component of native fluency which can only be acquired from usage. Yet, however necessary, these incidentals are not sufficient.

References

Altman, G. T. (1997). *The ascent of Babel*. Oxford: Oxford University Press.

Anderson, J. R. (1989). A rational analysis of human memory. In H. L. I. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 195-210).

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.

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

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review, 96*, 703-719.

Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of language aquisition* (pp. 157-193).

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. New York: Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow, UK: Pearson Education.

Burrell, Q. L. (1980). A simple stochastic model for library loans. *Journal of Documentation, 36*, 115-132.

Bybee, J., & Hopper, P. (Eds.). (2001). *Frequency and the emergence of linguistic structure*. Amsterdam: Benjamins.

Ellis, N. C. (2002a). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24*, 143-188.

Ellis, N. C. (2002b). Reflections on frequency effects in language processing. *Studies in Second Language Acquisition, 24*, 297-339.

Ellis, N. C. (in press). *How does explicit knowledge affect implicit language learning?* Unpublished manuscript.

Ellis, N. C., & Schmidt, R. (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*, 307-336.

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.

Gernsbacher, M. A. (1994). *A handbook of psycholinguistics*. San Diego, CA: Academic Press.

Harley, T. A. (1995). *The psychology of language: From data to theory*. Hove, UK: Taylor & Francis.

Hasher, L., & Chromiak, W. (1977). The processing of frequency information: An automatic mechanism? *Journal of Verbal Learning and Verbal Behavior, 16*, 173-184.

Jurafsky, D., & Martin, J. H. (2000). *Speech and language processing: An introduction to natural language processing, speech recognition, and computational linguistics*. Englewood Cliffs, NJ: Prentice-Hall.

MacWhinney, B. (1987). The competition model. In B. MacWhinney (Ed.), *Mechanisms of language aquisition* (pp. 249-308).

MacWhinney, B. (1997). Second language acquisition and the competition model. In A. M. B. De Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 113-142).

MacWhinney, B., & Leinbach, J. (1991). Implementations are not conceptualizations: Revising the verb learning model. *Cognition, 40*, 121-157.

Marslen-Wilson, W. (1990). Activation, competition, and frequency in lexical access. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing* (pp. 148-172). Cambridge, MA: ACL-MIT Press.

McKoon, G., & Ratcliff, R. (1998). Memory-based language processing: Psycholinguistic research in the 1990s. *Annual Review of Psychology, 49*, 25-42.

Miller, G. A., Bruner, J. S., & Postman, L. (1954). Familiarity of letter sequences and tachistoscopic identification. *Journal of General Psychology, 50*, 129-139.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191-225). London: Longman.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 94*, 523-568.

Posner, M. I., & Snyder, C. R. (1975). Attention and cognitive control. In G. H. Bower & J. T. Spence (Eds.), *Psychology of learning and motivation* (Vol. 3, pp. 55-85). San Diego, CA: Academic Press.

Rumelhart, D. E., & McClelland, J. L. (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2: Psychological and biological models). Cambridge, MA: MIT Press.

Schooler, L. J. (1993). *Memory and the statistical structure of the environment.* 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.

Seidenberg, M. S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science, 275*, 1599-1603.

Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science, 23*, 569-588.

Sinclair, J. (1991). *Corpus, concordance, collocation.* Oxford: Oxford University Press.