Applied Linguistics 27/2: 164–194
doi:10.1093/applin/aml015

Selective Attention and Transfer Phenomena in L2 Acquisition: Contingency, Cue Competition, Salience, Interference, Overshadowing, Blocking, and Perceptual Learning

NICK C. ELLIS

University of Michigan

If first language is rational in the sense that acquisition produces an end-state model of language that is a proper reflection of input and that optimally prepares speakers for comprehension and production, second language is usually not. This paper considers the apparent irrationalities of L2 acquisition, that is the shortcomings where input fails to become intake. It describes how 'learned attention', a key concept in contemporary associative and connectionist theories of animal and human learning, explains these effects. The fragile features of L2 acquisition are those which, however available as a result of frequency, recency, or context, fall short of intake because of one of the factors of contingency, cue competition, salience, interference, overshadowing, blocking, or perceptual learning, which are all shaped by the L1. Each phenomenon is explained within associative learning theory and exemplified in language learning. Paradoxically, the successes of L1 acquisition and the limitations of L2 acquisition both derive from the same basic learning principles.

N. C. Ellis (2006) reviewed how fluent language users are rational in their language processing, their unconscious language representation systems optimally prepared for comprehension and production, how language learners are intuitive statisticians, and how acquisition can be understood as contingency learning according to ΔP (the one-way dependency statistic) and probability contrast measures. But what of the shortcomings of naturalistic second language acquisition, the Basic Variety devoid of functional morphology, the fragile features of language that L2 learners fail to acquire despite thousands of occurrences in their input, the cases where input fails to become intake? Does this mean that L2 acquisition cannot be understood according to the general principles of associative learning that pertain in other aspects of human cognition? Is L2 acquisition irrational? In this article I explain why this is not the case.

First of all, associative learning has warts, too: there are situations where human and animal cognition appears to be irrational, although it is not when you look at it the right way. Analyses of human learning in complex situations where there are multiple co-occurring predictive cues demonstrate that biases can emerge from a learning mechanism which has evolved to yield normatively accurate beliefs concerning associative relationships yet which is sensitive to different effects of learned selective attention. There are bounds to human rationality dictated by limited cognitive resources (Simon 1957) and thus we behave in a manner that is as nearly optimal with respect to our goals as our cognitive resources will allow. Perception and learning are 'probabilistic achievements', but our perception of the world is shaped through the lenses of our prior analyses, beliefs, and preconceptions. The environment provides the setting, with all of the stimuli present, but we view those stimuli using previous experience as a lens; out of all the available stimuli, the cues that affect our learning are the stimuli that the individual pays most attention to. This is the essence of the 'probabilistic functionalism' and the associated 'lens theory' of Brunswik (1955). Our implicit models of the world induce a variety of perceptual biases, a range of phenomena of learned selective attention.

This paper argues that the linguistic forms that L2 learners fail to adopt and to use routinely thereafter in their second language processing are those which, however available as a result of frequency, recency, or context, fall short of intake because of one of associative learning factors of contingency, cue competition, or salience, or because of associative attentional tuning involving interference, overshadowing and blocking, or perceptual learning, all shaped by the L1.

The first part of this paper concerns the factors that affect both L1 and L2 acquisition: contingency, competition between multiple cues, and salience. Each of these is taken in turn, its processes are explained from within associative learning theory, and its effects are illustrated with examples from language learning. This section concludes by considering the combined operation of these factors as determined in a meta-analysis of the morpheme acquisition order studies (Goldschneider and DeKeyser 2001). The second half of the paper concerns the phenomena of L1 interference, overshadowing and blocking, and perceptual tuning, which all have a differential, more marked role on L2 acquisition, thus helping to provide non age-invoked biological explanations for why L2 acquisition stops short while L1 acquisition does not. Again, each factor is described from learning theory and exemplified in language learning, particularly L2 acquisition. The second section concludes with evidence of L1/L2 differences in morpheme acquisition order, illustrating these processes as they contribute to transfer and 'learned attention'. On reflection then, the apparent irrationalities of fragile L2 acquisition and fossilization provide further evidence that L2 acquisition is painted in the same colors as the rest of cognition, warts and all.

FACTORS AFFECTING BOTH L1 AND L2

Contingency and ΔP

Learning theory

N. C. Ellis (2006) reviewed how the associations between cues and outcomes can be understood as contingency learning according to ΔP (the one-way dependency statistic) which measures the directional association between a cue and an outcome, as illustrated in Table 1.

Language learning

Let us apply these concepts to those grammatical functors that are also absent from the Basic Variety, particularly those investigated in the 'morpheme order studies' which demonstrated there to be similar L2 and L1 orders of acquisition of these grammatical functors (Bailey *et al*. 1974; Brown 1973; Dulay and Burt 1973; Pica 1983). Although later studies showed some variability across L1/L2, and we will consider this in more detail towards the end of this paper, for the moment, concentrate on the commonalities. The morpheme order studies showed the order of acquisition to be broadly:

plural '-s'	'Book s'
progressive '-ing'	'John go ing'
copula 'be'	'John is here'/'John's here'
auxiliary 'be'	'John is going'/'John's going'
articles 'the/a'	'The books'
irregular past tense	'John went'
third person '-s'	'John like s books'
possessive ''s'	'John's book'

It is immediately abundantly clear that there are no one-to-one mappings between these cues and their outcome interpretations. Consider an ESL learner trying to learn from the naturalistic input illustrated in Figure 1. Plural '-s', third person singular present '-s', and possessive '-s', are all

Table 1: A contingency table showing the four possible combinations of events showing the presence or absence of a target Cue and an Outcome

	Outcome	No outcome
Cue	a	b
No cue	c	d

Notes: a, b, c, d represent frequencies, so, for example, a is the frequency of conjunctions of the cue and the outcome, and c is the number of times the outcome occurred without the cue. $\Delta P = P(O|C) - P(O|\neg C) = a/(a+b) - c/(c+d)$.

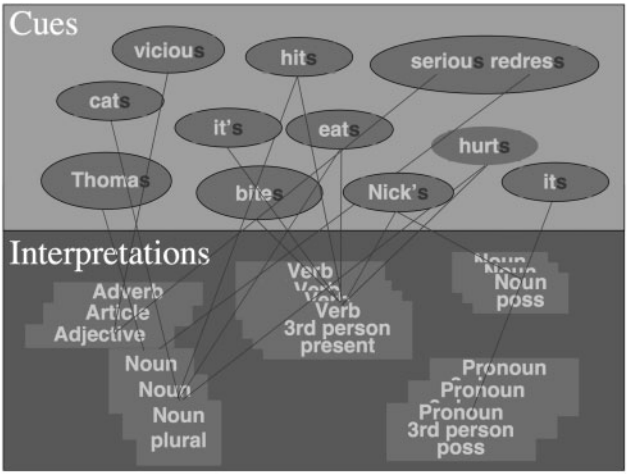

Figure 1: Some signs, signifiers, and signifieds

homophonous with each other as well as with the contracted allomorphs of copula and auxiliary 'be'. Therefore, if we evaluate '-s' as a cue for one particular of these functional interpretations, it is clear that there are many instances of the cue being there but that outcome not pertaining, (b) in Table 1 is of high frequency, and ΔP accordingly low. Evaluate the mappings from the other direction as well: plural '-s', third person singular present '-s', and possessive '-s' all have variant expression as the allomorphs [/s/,/z/,/ez/]. Therefore, if we evaluate just one of these, say /ez/, as a cue for one particular outcome, say plurality, then it is clear that there are many instances of that outcome in the absence of the cue, (c) in Table 1 is inflated, and ΔP concomitantly reduced. Thus a contingency analysis of these cue-interpretation associations suggests that they will not be readily learnable.

So much for the unreliable mappings between *–s* and its interpretations as plural, or third person plural, or copula. Most high-frequency grammatical functors are similarly highly ambiguous in their interpretations. Consider the range of meanings of the English preposition *in*, as illustrated in the story of the man who was wounded six times during World War I, the first time in Belgium, the second in the morning, the third in the leg, the fourth in his sleep, the fifth in August, and the sixth inadvertently. The semantic analysis of definite and indefinite reference shows its meaning to be highly variable and complicated (Diesing 1992; Faurud 1990; Hawkins 1978, 1991; Lyons 1999). Look at the number of pages of explanation given to 'the' in a grammar of English (Biber *et al*. 1999; Celce-Murcia and Larsen-Freeman 1999). The fuzziness and complexity of these mappings surely goes a long way to making ESL article acquisition so difficult. Finally, consider how the low ΔP for possessive '-s' compounded by interference from contracted 'it is' ensures, as experience of undergraduate essays attests, that the apostrophe is opaque in it's [sic] function and that native language learners can fail to sort out this system even after 18 years of experience. These are no one-to-one form to meaning mappings.

Multiple cues, the PCM, and cue-competition

Learning theory

N. C. Ellis (2006) next described how, in situations of multiple cues, learners operate according to the Probabilistic Contrast Model (PCM) (Cheng and Holyoak 1995; Cheng and Novick 1990) which assumes that potential causes are evaluated by contrasts computed over a 'focal set', that is the contextually-determined set of events that the reasoner *selects* to use as input to the computation of that contrast.

Language learning

Structuralist approaches, from their beginnings (Saussure 1916) to their modern counterparts (Croft 2001), analyze linguistic signs as a set of mappings between phonological forms and conceptual meanings or communicative intentions. The same meaning or intention can be expressed in a variety of ways, and there is considerable redundancy in language (Shannon 1948). So language is a prime example of a stimulus environment rich in multiple cues. MacWhinney (1997, 1987) explores the range of constructions that relate subject-marking forms to subject-related functions in the English sentence, 'The learner weighs the probabilities', are preverbal positioning ('learner' before 'weighs'), verb agreement morphology ('weighs' agrees in number with 'learner' rather than 'probabilities'), sentence initial positioning, and use of the article 'the'. Case-marking languages, unlike English, additionally include nominative and accusative cues in such sentences. The corresponding functional interpretations include actor, topicality, perspective, givenness, and definiteness. Different languages use these cues to lesser or greater degrees. The learning of sentence processing cues in a second language is a gradual process. It begins with L2 cue weight settings that are close to L1, and only gradually over time do these settings change in the direction of the native speaker's settings for L2. MacWhinney (2001) relates how his early investigations of L2 processing examined the comprehension of English sentences in academic colleagues: One subject was a native speaker of German who had lived in the United States for thirty years, was married to an American, and had published several important textbooks in experimental psychology written in English. Yet remarkably, this subject processed simple English sentences using the cue strength hierarchy of German: he used agreement and animacy cues whenever possible, largely ignoring word order when it competed with agreement and animacy. This first evidence for the preservation of a syntactic 'accent' in comprehension was subsequently supported in over a dozen studies across a wide variety of second language learning. For example, McDonald (1987) compared the cue use of English/Dutch and Dutch/English bilinguals with varying amounts of L2 exposure to that of native speaker controls. For all constructions tested (dative constructions, simple noun/verb/noun sentences, and relative clauses), it was found that with increasing exposure, cue usage gradually shifted from that appropriate to L1 to that appropriate to the L2. MacWhinney (2001) reports the connectionist model of Johnson and MacWhinney that successfully simulates these results using an Elman recurrent network.

It comes as no surprise then that Brian MacWhinney faced the problem of cue competition for the case of language before, indeed, Cheng and Holyoak (1995) were addressing the general question of inference in the face of multiple cues. The Competition Model (MacWhinney and Bates 1989; MacWhinney *et al*. 1984) was explicitly formulated to deal with competition between multiple linguistic cues to interpretation. MacWhinney *et al*. (1985) had found that the order in which these cues are initially acquired by children in their first language was predicted by a statistic called overall validity, related to ΔP, and that later use of cues can be predicted by another statistic called conflict validity (McDonald 1986). The overall validity of a cue is its availability (probability of presence in a sentence) times its reliability (probability of correctly indicating role assignment). Overall validity is computed using all the sentences in the language; conflict validity is computed for a cue using sentences in which the role assignment of that cue conflicts with the assignment of another cue. The Competition Model explains the transition of cue use from overall validity to conflict validity with a learning-on-error mechanism. A strength counter is maintained for each cue, and in deciding a role, the noun with the largest total cue strength is assigned to that role. When a role is assigned incorrectly, cues that could have predicted the correct answer have their strength increased. There is no increase in strengths in the case of a correct assignment. Initially, all cue strengths are small random values, so errors are made over a representative sampling of all sentences. Cue strengths are thus incremented proportionally to the ability of the cue to predict correct assignment over all sentences (overall validity). Errors continue to decrease, and at some point, sentences that do not have cues conflicting in the prediction of assignment do not produce errors. Then, cue strengths are incremented for sentences with conflicting cues (conflict validity). The competition model is language's own PCM. Its algorithm for probability contrast is somewhat different in detail to that of PCM, but its result is similar in that it first selects the most valid cue using statistical contingency analysis and then introduces cues thereafter on the basis of their potential to decrease error.

N. C. Ellis (2006) cited the experiments of Matessa and Anderson (2000) where, in the initial stages of language acquisition, learners tend to focus on only one cue at a time. Later on, after having tracked the use of this first cue, they add a second cue to the mix and begin to use the two in combination, and, as development proceeds, so additional cues are added if they significantly helped reduce errors of understanding. This is the behavior predicted by both the PCM and the Competition model. But a language learner might never get around to noticing low salience cues, particularly when the interpretation accuracy afforded by the other more obvious cues does well enough for everyday communicative survival. This is what is predicted by the Rescorla and Wagner (1972) model described in my companion piece: fossilization results from low *salience* cues whose *redundancy* denies them any more than low *outcome importance*; dV on any learning trial is negligible, and thus they never become integrated into a consolidated construction.

Salience

Learning theory

N. C. Ellis (2006) summarized the associative learning research demonstrating that selective attention, salience, expectation, and surprise are key elements in the analysis of all learning, animal and human alike. As the Rescorla–Wagner (1972) model encapsulates, the amount of learning induced from an experience of a cue–outcome association depends crucially upon the salience of the cue and the importance of the outcome.

Language learning

Many grammatical meaning–form relationships, particularly those that are notoriously difficult for second language learners like grammatical particles and inflections such as the third person singular '-s' of English, are of low salience in the language stream. The reason for this is the well documented effect of frequency and entrenchment in the evolution of language: grammaticalized morphemes tend to become more phonologically fused with surrounding material because their frequent production leads to lenition processes resulting in the loss and erosion of gestures (Bybee 2006; Jurafsky *et al*. 2001; Zuraw 2003). As Slobin (1992: 191) put it: 'Somehow it's hard to keep languages from getting blurry: speakers seem to 'smudge' phonology wherever possible, to delete and contract surface forms, and so forth.'

In informal and rapid speech, this tendency to give short shrift to function words and bound morphemes, exploiting their frequency and predictability, deforms their phonetic structure and blurs the boundaries between these morphemes and the words that surround them. Clitics, accent-less words or particles that depend accentually on an adjacent accented word and form a prosodic unit together with it, are the extreme examples of this: the /s/ of 'he's'. /l/ of 'I'll' and /v/ of 'I have' can never be pronounced in isolation. Thus grammatical function words and bound inflections tend to be short and low in stress, even in speech that is produced slowly and deliberately (Bates and Goodman 1997) and in speech directed to children (Goodman *et al*. 1990), with the result that these cues are difficult to perceive. When grammatical function words are clipped out of connected speech and presented in isolation, adult native speakers can recognize them only no more than 40 percent to 50 percent of the time (Herron and Bates 1997). This simple statistic has a profound consequence: If fluent native speakers can only hear these grammatical functors from the bottom-up evidence of input 40 percent-50 percent of the time, what chance have second language learners to hear them and thence learn their function?

Fluent language processors can perceive these elements in continuous speech because their language knowledge provides top-down support. But this is exactly the knowledge that learners lack. It is not surprising, therefore, that in L1 acquisition young children are unable to acquire grammatical forms until they have a critical mass of content words, providing enough top-down structure to permit perception and learning of those closed-class items that occur to the right or left of 'real words' (Bates and Goodman 1997: 51–2). Nor is it surprising that it is these elements that are difficult for second language learners, with the order of acquisition of these morphemes being pretty much the same in L2 as in L1 learners (Bailey *et al*. 1974; Brown 1973; Dulay and Burt 1973; Larsen-Freeman 1976). Indeed, lenition eventually influences the form of language as a whole, causing some grammatical markers to 'wear away' and creating a pressure for the development of others to replace them. McWhorter (2002) tags this a process of 'Defining Deviance Downwards': a generation that grows up hearing a sound produced less distinctly gradually comes to take this lesser rendition as the default. In following the general tendency to pronounce unaccented sounds less distinctly, they in turn pronounce their default version of the sound, already less distinct than the last generation's, even less distinctly. Eventually the default is no sound in that position at all. This erosion has a particularly dramatic effect in sounds such as suffixes or prefixes that perform important grammatical functions. In this way, while Latin had different forms for all six combinations of person and number in the present tense, French has just three different forms for the present tense of '-er' verbs (four for '-ir', '-re', and '-oir' type verbs), and modern English has just two. Thus do psycholinguistic and associative learning processes in usage affect both language learning and language change.

Frequency, salience and contingency in morpheme acquisition order

A frequency analysis would predict that grammatical functors, as closed class items of the language, are so frequent in the input that their frequency, recency, and context would guarantee their being learned (N. C. Ellis 2002a). Yet these same items have other properties which moderate their acquisition (N. C. Ellis 2002b). Each of the above explanations, low ΔP, low salience, redundancy, and low outcome importance, seems to have the potential to make them difficult to acquire. Can we weigh their respective contributions, or indeed know how factors like these might interact? There are many variables and such potential richness of language usage over time that it makes their dynamic interactions complex and difficult to predict. Nevertheless, there are good data which help to inform an answer.

Goldschneider and DeKeyser (2001) performed a detailed meta-analysis of the 12 'morpheme order studies' that, in the 25 years following Brown's (1973) descriptions of L1 acquisition, investigated the order of L2 acquisition of the grammatical functors, progressive '-ing', plural '-s', possessive '-s', articles 'a', 'an', 'the', third person singular present '-s', and regular past '-ed'. These studies show remarkable commonality in the orders of acquisition of these functors across a wide range of learners of English as a second language, yet, although each of the factors of input frequency, semantic complexity, grammatical complexity, phonological form, and perceptual salience has been historically considered for their sufficiency of cause, with input frequency being the favored account (Larsen-Freeman 1976), nevertheless, as Larsen-Freeman concluded, '[a] single explanation seems insufficient to account for the findings' (1975: 419).

Goldschneider and DeKeyser investigated whether instead a combination of five determinants (perceptual salience, semantic complexity, morphophonological regularity, syntactic category, and frequency) could account for the acquisition order. Their factors of frequency and perceptual salience were much as have been described here, with scores for perceptual salience being composed of three subfactors: the number of phones in the functor (phonetic substance), the presence/absence of a vowel in the surface form (syllabicity), and the total relative sonority of the functor. Their factor of morphophonological regularity relates to contingency since the two subfactors of conditioned phonological variation (for example, the [/s/,/z/,/ez/] allomorphs of plural '-s', possessive '-s', and third person singular '-s') and contractibility both result in multiple forms of the cue, and thus a less clear mapping between the outcome and one particular cue, whilst the third subfactor of homophony with other grammatical functors results in a less clear mapping between the cue and one particular outcome. Allomorphy and contractibility reduce ΔP by inflating *c*, homophony by inflating *b*. Oral production data from 12 studies, together involving 924 subjects, were pooled. On their own, each of these factors significantly correlated with acquisition order: perceptual salience $r = 0.63$, frequency $r = 0.44$, morphophonological regularity $r = 0.41$. When these three factors were combined with semantic complexity and syntactic category in a multiple regression analysis, this combination of five predictors jointly explained 71 percent of the variance in acquisition order, with salience having the highest predictive power on its own. Each of these factors of frequency, salience, and contingency is a significant predictor independently; together they explain a substantial amount of acquisition difficulty.

We must conclude that, to the extent that the order of acquisition of these morphemes is the same in L1 and L2, these factors play a similarly substantial role in first and second language acquisition. But, the studies meta-analysed in Goldschneider and DeKeyser pooled L2 learners from a variety of L1 backgrounds and did not concern the ways in which the nature of the first language might have a particular effect on the detailed path or rate of L2 acquisition. On top of the effects described in this first section concerning the learner and the language to be learned, there are discernable effects on second language learning resulting from transfer from the first language that the learner has already learned. The next section describes various associative learning processes that are involved in transfer and learned attention before gathering some experimental demonstrations of these particular effects of L1-specific transfer in L2 morpheme acquisition.

FACTORS SPECIAL TO L2

Interference

Learning theory

A hundred years and more ago, Müller and Pilzecker (1900) produced one of the earliest empirical demonstrations of forgetting due to *interference*: people were less likely to recall a memory item if in the interim the retrieval cue that was used to test that item had become associated to another memory. Memory for association A-B is worse after subsequent learning of A-C in comparison with control condition involving subsequent learning of unrelated material D-E. They called this effect *retroactive inhibition*, highlighting the manner in which the storage of new experiences interferes with memories encoded earlier in time. It is harder to remember the phone number, car registration, or whatever else you had ten years ago if you have acquired a new phone number, car registration, etc. in the interim. According to classical interference theory, such effects show that it is not the mere passage of time that causes forgetting (as trace decay explanations would hold), but rather it is what happens in that time, the storage of new experiences into memory. The next 50 years of research into interference theory, particularly that in the 'verbal learning tradition' (less kindly dubbed 'dust bowl empiricism') in the USA, demonstrated that it is the interactions of memories, particularly those of highly similar experiences, that are at the root of memory failures. 'Response competition theory' (McGeoch 1942) held that forgetting was a consequence of adding new associative structure, and it attributed interference effects to heightened competition arising from the association of additional traces to a retrieval cue (or to the strengthening of an existing competitor). These ideas continue today in models that emphasize how retrieval of a given item is impeded by competing associations (Anderson and Bjork 1994). Quite simply, when multiple traces are associated to the same cue, they tend to compete for access to conscious awareness (Anderson and Neely 1996; Baddeley, 1976: ch. 5, 1997; Postman 1971), and it is not just new memories that interfere with old, the competition runs both ways. So it is harder to learn a new phone number, car registration, or what-have-you, because the old ones tend to compete and come to mind instead—this effect of prior learning inhibiting new learning is called proactive inhibition (PI). Much of this work was succinctly summarized in Osgood's 'transfer surface' that draws together the effects of time of learning, similarity of material, and retention interval on negative (and positive) transfer of training (Osgood 1949).

Language learning

Prior proposals for understanding aspects of L2 acquisition in terms of transfer from L1 are well known. In the early 1950s Weinreich emphasized the importance of interference: 'those instances of deviation from the norms of either language which occur in the speech of bilinguals as a result of their familiarity with more than one language' (Weinreich 1953: 1). PI underpins a variety of fundamental phenomena of language leaning and language transfer.

Consider first the question of 'Why Ginny's jelly is not jello'. At a recent bilingualism seminar, Ginny Gathercole illustrated how hard it was to relearn a new meaning for an old word. Ginny was brought up in America where jelly is a fruit conserve that you spread on bread. In adulthood, Ginny married an Englishman for whom jelly is a gelatin pudding that wobbles on a plate, calls for plenty of cream, and is referred to in her native USA as jello. After living for extended periods in the UK, Ginny and her daughter were staying with friends in Edinburgh. Their host asked that they get some jelly from the cupboard, and Ginny sent her daughter to get a jar of conserve. Why, Ginny asked, after so many learning trails, had the referent of 'jelly' not been reset to jello? Ginny's jingoistic jelly jeopardy is a simple effect of PI. The prior A-B association interferes with the learning of A-C. It illustrates the usual effects whereby PI depends on the relative amounts of A-B and A-C learning, it depends upon the similarity of the original and subsequent association, and it increases with delay between A-C learning and recall. As Figure 2 illustrates, it can be understood in terms of response competition and trace-decay following the standard rational forgetting curve.

Proactive interference has, of course, much more pervasive effects than these in second language learning, as Robert Lado proposed in his Contrastive Analysis Hypothesis (CAH) (Lado 1957): 'We assume that the student who comes in contact with a foreign language will find some features of it quite easy and others extremely difficult. Those elements that are similar to his native language will be simple for him, and those elements that

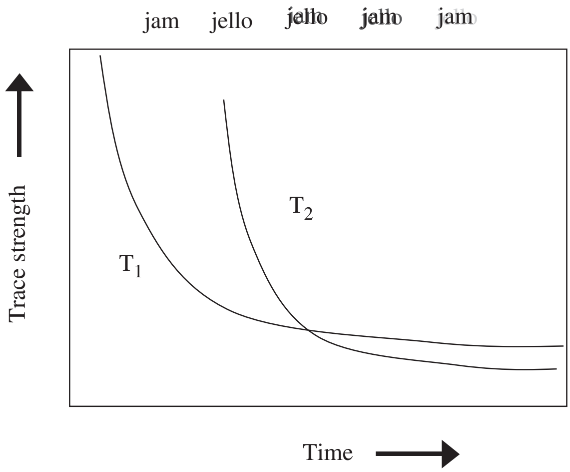

Figure 2: Why Ginny's Jelly is not Jello. The graph illustrates the trace strength for the first meaning (T1, jam) as it is first learned and subsequently decays over time, and for the second meaning (T2, jello) acquired later. T2 also decays in strength according to the forgetting curve. The top line is a visual analogue of the competition between these two responses as they come to mind with their respective trace strengths at different points in learning. Effects of Proactive Interference from T1 increase over time as a function of response competition and trace decay, with the T1 response coming to the fore again as the retention interval increases

are different will be difficult' (Lado 1957: 2). The CAH held that one could predict learner difficulty by reference to an utterance-by-utterance comparison of a learner's L1 and L2.

PI underlies the general negative transfer that makes the learning of second and subsequent language lexis generally difficult. It affords a positive edge to cognates and an extra negativity to *faux ami*. PI, along with its companions blocking and perceptual learning that I discuss next in this article, is a major process by which similarities and differences between languages can influence the acquisition of grammar, vocabulary, and pronunciation, and transfer has a justifiably rich history in the theoretical analysis of second and foreign language learning (C. James 1980; Odlin 1989). A survey of the influence of the CAH, made by a simple search of *Linguistics and Language Behavior Abstracts* to see how many articles abstracted in the last 30 years had the keyword descriptor 'contrastive analysis', produced a non-trivial 1,268.

Interference theory primarily concerned the transfer of the content of associations. But more recent analyses demonstrate how from content, given enough of it, emerges principle, how form–meaning mappings conspire in biasing attention and process. As a ubiquitous process of learning, transfer pervades all language learning. As we pursue our researches, so we come to believe that the effects that we observe are ours and are special to our domain. In child language acquisition, the problem of referential indeterminacy (Quine 1960) led researchers to posit word learning constraints that might help limit learners' search space. It has been proposed that they are guided by general heuristics such as a tendency to believe that new words often apply to whole objects (the whole object constraint), that they more likely will refer to things for which a name is not already known (the mutual exclusivity constraint), that they more often relate to things distinguished by shape or function rather than by color or texture, and the like (Bloom 2000; Golinkoff 1992; Golinkoff *et al*. 1994; Gopnik and Meltzoff 1997; Markman 1989; Tomasello 2003). But mutual exclusivity is PI by another name. If a referent already has an associated name, it is harder to attach a new name to it. And so it is the things of the world that are not already labeled that attract new names more readily, in the same way that it is the empty pegs in the coat rack that are more likely get hats hung on them next. Kaminsky *et al*. (2004) tested the use of mutual exclusivity in a border collie named Rico. New objects were placed along with several familiar ones and the owner asked Rico to fetch it using a new name Rico had never heard before. He usually retrieved the new object, apparently appreciating that new words tend to refer to objects that don't already have names. A month later, Rico showed some retention of the words he had learned, with abilities comparable to three-year-old toddlers who were tested using similar designs.

Recent computational models provide concrete accounts of how such word learning principles emerge in development from more general aspects of cognition involving associative learning processes such as PI, learned attention and rational inference, that is from prior knowledge of the world and the ways language usually refers to it, and from the learner's existing repertoire of linguistic constructions (MacWhinney 1989; Merriman 1999; Regier 2003). Mutual exclusivity emerges as rational inference in Bayesian models (Regier 2003; Tenenbaum and Xu 2000) and in Competition models (MacWhinney 1989; Merriman 1999) of word learning and these simulations account for a variety of empirically observed mutual exclusivity effects.

Overshadowing and blocking

Learning theory

The emergence of a learning bias (Mutual Exclusivity) from prior learned content and associations (PI) illustrates how selective attention can be learned, how salience is a psychological property as well as a physical one. Associative learning research describes two general mechanisms that play a particular role in shaping our attention to language: *overshadowing* and *blocking*. In my companion paper, I explained the phenomenon of overshadowing whereby, when two cues are presented together and they jointly predict an outcome, the strength of conditioning to each cue depends upon their salience, with the most salient cue becoming associated with the outcome and the less salient one being overshadowed so that on its own it evinces little or no reaction (Kamin 1969). But cues also interact over time. As the Rescorla–Wagner (1972) model encapsulates, the more a cue is already associated with an outcome, the less additional association that outcome can induce. Equally, there is the phenomenon of *latent inhibition* whereby stimuli that are originally varied independently of reward are harder to later associate with reward than those that are not initially presented at all (Lubow 1973; Lubow and Kaplan 1997). Forms that have not previously cued particular interpretations are harder to learn as cues when they do become pertinent later. It makes good rational sense that evidence that an event is of no predictive value should encourage the learning system to pay less attention to that event in the future. As long as the world stays the same, that is.

Overshadowing as it plays out over time produces a type of learned selective attention known as blocking. Chapman and Robbins (1990) showed how a cue that is experienced in a compound along with a known strong predictor is blocked from being seen as predictive of the outcome. Their experiment, diagrammed in Table 2, had undergraduates make predictions about changes in a fictitious stock market. In the first period of the learning phase, whenever stock A rose in price (cue A), the market rose as well (outcome X). The movement of stock B during this period had no effect on the market. Thus A was a good predictor of outcome and B was not. In the second period of the learning phase, on some trials, stocks A and C rose together and the market increased, and on other trials, stocks B and D rose together and the market again rose. The number of cases of AC cue combination and BD cue combination in period 2 were the same, so they were equally good predictors of market growth. In a final testing phase, the learners were asked to rate on a scale from –100 (perfect predictor of market not rising) to +100 (perfect predictor of market rising) how well each stock predicted a change in the market.

Even though stocks C and D were associated with a rise in the market on exactly the same number of occasions with an actual ΔP for C and D of 0.57 calculated unconditionally over all trials, nevertheless, the learners judged

Table 2: The design and outcome of Chapman and Robbins (1990) cue interaction experiment illustrating 'Blocking'

Learning period 1	Learning period 2	Test phase	Mean judgment
A → X	AC → X	C	31
B → no X	BD → X	D	77

that cue D was a much better predictor (rating = 77) of market rise than was cue C (rating = 31). The prior learning of cue A 'blocked' the acquisition of cue C. Cue A was highly predictive of outcome in learning phase 1, with the result that in learning phase 2, cue C was to some extent overshadowed and ignored. In contrast, when cues were compounded with others which were not particularly informative (cue B), the target cue (D) received a normal association with the outcome.

The Probability Contrast Model (PCM) (Cheng and Holyoak 1995), introduced in my prior article, explains the deviations up and down from 57 as follows. The focal set for cue C is just the AC and A trials; and that for cue D is just the BD and B trials. ΔP turns out to be 0.0 for cue C because the outcome has the same probability on AC and A trails:

$$\Delta P_C = P(O/C.A) - P(O/-C.A) = 1.0 - 1.0 = 0.0$$

where P(O/C.A) is the probability of the outcome in the presence of both C and A and P(O/–C.A) is the probability of the outcome in the absence of A and the absence of C. But ΔP turns out to be 1.0 for cue D because the outcome probability differs on AC and A trails:

$$\Delta P_D = P(O/C.B) - P(O/-C.B) = 1.0 - 0.0 = 1.0$$

These are more extreme results than the 0.31 and 0.77 shown in Table 2, suggesting perhaps that the behavioral results had not yet reached asymptote. But the take-home message is clear: human statistical reasoning is bound by selective attention effects whereby informative cues are ignored as a result of overshadowing or blocking. Research shows this to routinely occur even in very simple learning situations like these (Kruschke 1993, 1996, 2001; Kruschke and Blair 2000; Kruschke and Johansen 1999; Shanks 1995)—they are not restricted to complex learning environments with dozens of cues and outcomes and intricate interactions.

Kruschke and Blair's (2000) explanation of blocking as being caused by rapidly shifting, learned attention echoes those of Kamin (1969) and Macintosh (1975). When learners are presented with cases of AC → X, since from before A predicts X, C is merely a distraction from a perfectly predictive cue. To avoid this error-inducing distraction, they shift their attention away from cue C to cue A, and consequently learn only a weak association from C to X. In contrast, a new control cue D which co-occurs with a cue which has no prior known significance, becomes associated with its outcome much more strongly. Blocking is a result of an automatically learned inattention. But this learned inattention can be pervasive and longstanding: once a cue has been blocked, further learning about that cue is attenuated (Kruschke and Blair 2000). Kruschke simulates these processes by building mechanisms of attention into his computational models of associative learning (ALCOVE (Kruschke 1992), ADIT (Kruschke 1996) and RASHNL (Kruschke and Johansen 1999)). In these models, each cue is gated by an attentional strength, and total attention is limited in capacity. The attention allocated to a cue affects both the associability of the cue and the influence of the cue on response generation. An exemplar unit is recruited for each distinct cue combination, with each exemplar unit encoding not only the presence or absence of cues, but also the attention paid to each cue. Thus, an exemplar unit does not record the raw stimulus, but the *stimulus as processed*.

In sum, the associative learning research demonstrates that:

- When two cues jointly predict an outcome, the more salient one may be learned and the less salient may not, the more salient cue thus overshadowing the other.
- Pre-exposure to an event which originally has no consequences of outcome will retard rather than enhance subsequent learning of it as a cue.
- A cue that is experienced in a compound along with a known strong predictor is blocked from being seen as predictive of the outcome. Once a cue has been blocked, further subsequent learning about that cue is attenuated.

Language learning

Not only are many grammatical meaning–form relationships low in salience, but they can also be redundant in the understanding of the meaning of an utterance. It is often unnecessary to interpret inflections marking grammatical meanings such as tense because they are usually accompanied by adverbs that indicate the temporal reference. Terrell illustrated it thus: 'if the learner knows the French word for ''yesterday'' then in the utterance *Hier nous sommes allés au cinéma* (Yesterday we went to the movies) both the auxiliary and past participle are redundant past markers. Furthermore, since the adverb *hier* has now marked the discourse as past, the past markers on subsequent verbs are also redundant' (Terrell 1991: 59).

I believe that this redundancy is much more influential in second rather than L1 acquisition. Children learning their native language only acquire the meanings of temporal adverbs quite late in development (Dale and Fenson 1996). However, the second language expression of temporal reference begins with a phase where reference is established by adverbials alone (Bardovi-Harlig 1992; Meisel 1987), and the prior knowledge of these adverbials can block subsequent acquisition of other cues. Schumann (1987) describes how L2 temporal reference is initially made exclusively by use of devices such as temporal adverbials ('tomorrow', 'now'), prepositional phrases ('in the morning . . .'), serialization (presenting events in their order of occurrence), and calendric reference ('May 12', 'Monday'), with the grammatical expression of tense and aspect emerging only slowly thereafter (Bardovi-Harlig 2000). Second language learners already know about temporal adverbs and narrative strategies for serialization, these strategies are effective in the communication of temporality, and thus the high salience of these means of expression leads L2 learners to attend to them and to ignore the phonologically reduced tense-markings.

Inflexions for number are similarly often overshadowed by the more obvious singularity of the clear subject of the verb. Pica (1983) describes how naturalistic L2 learners, but not instructed learners, tended to omit plural '-s' endings on nouns that are premodified by quantifiers. Like Schumann (1978), she observes how difficulty with this redundant marking of plurality is characteristic of L2 learners and pidgin speakers alike. There are many such examples. For each of them, take the relevant pair of high and low salience co-occurring forms, substitute them for cues A and B in Table 2, and they readily fit the requirements for the phenomena involving overshadowing. Thus, another pervasive reason for the non-acquisition of low salience cues in L2 acquisition is that of blocking, where redundant cues are overshadowed for the historical reasons that the learners' first language experience leads them to look elsewhere for their cues to interpretation. Under normal L1 circumstances, usage optimally tunes the language system to the input; under these circumstances of low salience of L2 form and blocking, all the extra input in the world might sum to naught, and we describe the learner as having 'fossilized' with an IL reminiscent of the Basic Variety.

The usual pedagogical reactions to these situations of overshadowing or blocking involve some means of retuning *selective attention*, some type of form-focused instruction or consciousness raising (Sharwood-Smith 1981) to help the learner to 'notice' the cue and to raise its salience. Schmidt summarized it as: '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, first by pointing them out and explaining their structure, and secondly by providing meaningful input that contains many instances of the same grammatical meaning–form relationship. VanPatten is similarly influenced by the fact that L2 speakers allocate more cognitive activation to meanings they consider to be more important to communication in the design of 'processing instruction' (VanPatten 1996) that 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. Likewise Doughty and Williams: 'For forms that are frequent in the input and yet still seem to lack salience for learners, it may be that other means are required to induce learners to notice' (Doughty and Williams 1998: 220). I review the range of mechanisms for the interface of explicit knowledge on implicit language learning in N. C. Ellis (2005); see also Doughty (2001), R. Ellis (2001) and Robinson (2001).

Perceptual learning

Learning theory

Our perceptual systems change their structure during their history of processing the stimuli to which they are exposed even in the absence of any overt consequences. William James (1890) discusses the case of the novice wine-taster who starts out being unable to distinguish claret and burgundy, but who, after repeated exposure to these wines, comes to find them highly distinct. As a simple consequence of usage, without there being any contingency between the perceptual stimuli they process and any other outcomes or events, perceptual systems alter their sensitivity to stimulus features, becoming more sensitive to those which are psychologically significant dimensions of variation amongst the stimuli, and becoming less sensitive to those redundant characteristics which do not play any role in accurate classification. This tuning which automatically emerges as a result of experience of exemplars is called perceptual learning (Fahle and Poggio 2002; Goldstone 1998; Seitz and Watanabe 2003; Watanabe *et al*. 2001). Whereas the associative learning effects detailed above relate to specific cues or constructions and their interpretations, perceptual learning is more to do with the organization of the whole system and the dimensions of the underlying psychological space. As more and more instances are processed, so the representations of these exemplars become sorted and positioned in psychological space so that similar items are close together and dissimilar ones are far apart. The dimensions that define this space are to a large degree emergent—as in the statistical techniques of principle components or factor analysis, they come forward in the analysis as the major defining characteristics of the data under scrutiny (Elman *et al*. 1996; Nosofsky 1986, 1987). In fact, we investigate these psychological representation spaces using statistical techniques of multidimensional scaling (MDS) (Nosofsky 1992) rather than factor analysis, and we simulate the emergence of these structures using connectionist techniques, particularly those involving self-organizing maps (SOM) (Kohonen 1998).

Nosofsky (1986) describes animal learning and human categorization research evidencing attention shifts toward the use of dimensions that are useful for the tasks in hand: the dimensions that are relevant for categorizations are psychologically ‘stretched’ with the result that learners become more sensitive to these dimensions and are better able to make discriminations involving them. But in addition to important dimensions acquiring distinctiveness, irrelevant dimensions are psychologically ‘shrunk’ acquiring equivalence and becoming less distinguishable. During category learning, people show a trend toward emphasizing features that reliably predict experimental categories. For example, Livingston and Andrews (1995) showed that in undergraduates' learning to categorize complex schematic drawings: (1) the sequence of encounter of exemplars caused variation in feature salience, with bottom-up perceptual factors being critical to development of hypotheses about a category; (2) variation in feature salience was related to performance on categorization tasks; and (3) nonoptimal feature salience assignments were revised given sufficient experience in the domain: in particular, learners tended to revise faulty hypotheses by adjusting feature salience so as to maximize outcomes, but this revision was much more difficult when it required a complete reassignment of feature salience values.

Goldstone (1994; Goldstone and Steyvers 2001) presented a range of experiments involving perceptual learning of shapes showing that physical differences between categories become emphasized with training. After learning a categorization in which one dimension was relevant and a second dimension was irrelevant, learners made same/different judgments about whether two shapes were physically identical. Ability to discriminate between stimuli in this judgment task was greater when they varied along dimensions that were relevant during categorization training, and was particularly elevated at the boundary between the categories. Further research showed that category learning systematically distorts the perception of category members by shifting their perceived dimension values away from members of opposing categories (Goldstone 1995). Goldstone's research thus provides evidence for three influences of categories on perception: (1) category-relevant dimensions are sensitized; (2) irrelevant variation is deemphasized; and (3) relevant dimensions are selectively sensitized at the category boundary.

A related perceptual phenomenon is that of *feature imprinting*. If a stimulus part is important, varies independently of other parts, or occurs frequently, people may develop a specialized detector for that part. Efficient representations are promoted because the parts have been extracted due to their prevalence in an environment, and thus are tailored to that environment. Hock *et al*. (1987) showed that configurations of dots are more likely to be circled as coherent components of patterns if they were previously important for a categorization. Schyns and Rodet (1997) demonstrated that unfamiliar parts (arbitrary curved shapes within an object) that were important in one perceptual task were more likely to be used to represent subsequent categories. Their learners were more likely to represent a conjunction of two parts, X and Y, in terms of these two components (rather than as a whole unit, or a unit broken down into different parts) when they received previous experience with X as a defining part for a different category. Pevtzow and Goldstone (1994) similarly showed how people learn to decompose complex objects based on their experience with component parts: categorization training influences how an object is decomposed into parts. Once you are trained to see the object in that way, that is the way you see it (or that is the way you *first* see it), and those are the features whose strengths are incremented on each subsequent processing episode.

Goldstone (1998; Goldstone and Steyvers 2001; Kersten *et al.* 1998) presented a detailed analysis of the ways in which attentional persistence directs attention to attributes previously found to be predictive, elaborated a theory of conceptual and perceptual learning based on these mechanisms, and provided a connectionist model of the processes whereby category learning establishes detectors for stimulus parts that are diagnostic, and these detectors, once established, bias the interpretation of subsequent objects to be segmented (Goldstone 2000). These cognitive, computational, and neurophysiological results indicate that the building blocks used to describe stimuli are adapted to input history. Feature and part detectors emerge that capture the regularities implicit in the set of input stimuli. However, the detectors that develop are also influenced by task requirements and strategies. In general, whether a functional detector is developed will depend on both the objective frequency and subjective importance of the physical feature (Sagi and Tanne 1994; Shiu and Pashler 1992).

In sum, perceptual learning research demonstrates:

- The dimensions that are relevant for categorizations are psychologically 'stretched' with the result that learners become more sensitive to these dimensions and are better able to make discriminations involving them.
- Relevant dimensions are selectively sensitized at the category boundary.
- Irrelevant dimensions acquire equivalence, becoming less distinguishable. Irrelevant variation is deemphasized.
- Transfer which requires complete reassignment of feature salience values is difficult.
- Stimulus features that are important in classification, vary independently of other parts, or occur frequently, may become represented by a specialized detector for that part.
- Whether a functional detector develops depends on both the objective frequency and subjective importance of the physical feature.

Language learning

The sound categories and categorical perception of L1 are subject to perceptual learning (Lively *et al.* 1994). Whether categorical perception effects are found at particular physical boundaries depends on the listener's native language. In general, a sound difference that crosses the boundary between phonemes in a language is more discriminable to speakers of that language than to speakers of a language in which the sound difference does not cross a phonemic boundary (Repp and Liberman 1987). Speech representations are not at the outset organized around individual speech sounds or phonemic segments; instead, according to the 'lexical restructuring hypothesis' only gradually, in early through middle childhood, do they become more fully specified and undergo segmental restructuring (Garlock *et al*. 2001; Metsala and Walley 1998; Storkel 2001). This emergent view has it that the words in a young child's lexicon may be relatively distinct with fewer neighbors than the same words in the fully developed lexicon. As a result, children may be able to rely on more holistic representations to uniquely differentiate each word from every other, and these representations may only become more detailed as words are acquired and density increases. So, as more and more similar words are acquired in the child's vocabulary, this drives an increasingly well-specified representation of these words, initially in terms of subunits like onset and rime, and this effect occurs first in dense phonological neighborhoods. It is the learner's knowledge of individual lexical items which drives the abstraction process, with the mental representation of known words only slowly changing to resemble the lexical structure of an adult.

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 Lalonde 1988; Werker and Tees 1984). In contrast to the newborn infant, the starting disposition of the neural stuff for L2 acquisition is already tuned to the L1 and is set in its ways, it is a *tabula repleta* with L1 entrenchment determining strong negative transfer (Sebastián-Gallés and Bosch 2005). The L2 learner's neocortex has already been tuned to the L1, incremental learning has slowly committed it to a particular configuration, and it has reached a point at which the network can no longer revert to its original plasticity (Elman *et al*. 1996). English learners of Chinese have difficulty with tones, and Japanese learners of English with the article system, both problems resulting from zero use in the L1. But, as described above, transfer which requires restructuring of existing categories is especially difficult. This is the essence of 'perceptual magnet theory' (Kuhl and Iverson 1995) in which the phonetic prototypes of one's native language act like magnets, or, in neural network terms, attractors (Van Geert 1993, 1994), distorting the perception of items in their vicinity to make them seem more similar to the prototype. What are 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. Iverson *et al*. (2003) present a detailed analysis of how early language experience alters relatively low-level perceptual processing, and how these changes interfere with the formation and adaptability of higher-level linguistic representations, presenting evidence concerning the perception of English /r/ and /l/ by Japanese, German, and American adults. The underlying perceptual spaces for these phonemes were mapped using multidimensional scaling and compared to native-language categorization judgments. The results demonstrate that Japanese adults are most sensitive to an acoustic cue, F2, that is irrelevant to the English /r/–/l/ categorization. German adults, in contrast, have relatively high sensitivity to more critical acoustic cues. Thus L1-specific perceptual processing can alter the relative salience of within- and between-category acoustic variation and thereby interfere with subsequent L2 acquisition. 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 created and subsequently entrenched by their L1.

McClelland (2001) presents a connectionist simulation of such effects. A Kohonen self-organizing map network was taught the mappings between phonological input patterns and phonetic representation space. When the model was trained with exemplars from two relatively distinct neighborhoods (representing /r/ and /l/), it learned separate representations and could correctly classify examples into these categories. If, however, the network had previously been trained with exemplars from one wide neighborhood representing the single Japanese alveolar liquid, thereafter it learned to treat the two /r/ and /l/ classes of input as the same and 'diabolically maintain[ed] this tendency, even when faced with input that would at first have caused it to represent the classes separately' (McClelland 2001).

Simultaneous bilingual L1 acquisition has been compared to successive L2 acquisition in recent simulations by Hernandez *et al.* (2005) and Li *et al.* (2004). Their DevLex model combines perceptual learning and the learning of mappings between phonological forms and semantics. The system is exposed either simultaneously or successively to two bilingual speech systems, the exemplars causing perceptual learning in self-organizing modules that are linked by Hebbian learning to simultaneously active conceptual representations. Early, balanced simultaneous bilingual simulations showed the clearest evidence of language separation: through self-organization, the network comes to separate the Chinese lexicon from the English lexicon, implicating distinct lexical representations for the two languages. In the DevLex model, less balanced or later L2 input produce representations whereby L2 is partially parasitic on L1, and adult second-language simulations show relatively little L1–L2 separation at a local level and maximal transfer and interference. The DevLex model instantiates the motto formulated by Elizabeth Bates (Bates *et al.* 1988: 284) that 'modules are made, not born' and emphasizes, as here, that the difficulties of adult L2 acquisition are a result of prior L1 learning, entrenchment, and transfer, rather than of a fixed neurologically-given critical period.

Finally, an example of feature imprinting which has been clearly documented in the first and second learning of Chinese characters. Yeh *et al*. (2003) assessed the effect of learning experience upon the perceived graphemic similarity of Chinese characters by comparing results of shape-sorting tasks obtained from various groups of participants with different learning experiences and ages. Whereas both Taiwanese and Japanese undergraduates classified characters in relation to their configurational structures, American undergraduates, Taiwanese illiterate adults, and kindergartners categorized characters based on strokes or components. This trend of developmental changes from local details to more globally defined patterns which culminated in the identification of structure as consistently perceived by skilled readers is clearly the result of learning experience rather than simple maturation.

These various examples illustrate how a plastic, neural *tabula rasa* can become organized by early experience to optimally represent the phonological and orthographic perceptual input of the first language. Sufficient experience of L1 affords fluent accurate processing in this now-tuned and entrenched neural system, and subsequent second language learning is thus faced with maximal transfer and interference from L1, perceiving the L2 through the L1-tuned *tabula repleta*.

Transfer effects in L2 morpheme acquisition order

The first half of this paper culminated with a review of the morpheme acquisition studies which, averaging over L2 learners of different L1s, showed broadly similar orders of acquisition in L1 and L2 learners of English. This second half, therefore, parallels this organization by considering here more-focused L2 morpheme acquisition studies which demonstrate clear evidence of L1 transfer.

Hakuta and Cancino (1977) proposed that an L2 learner whose L1 does not make the same semantic discriminations as the L2 target with regard to particular morphemes experience more difficulty in learning to use these morphemes. There are various studies which support this claim. Hakuta (1976) reported the English language development of a Japanese L1-speaking child who showed particular difficulty with the definite/indefinite contrast, Japanese being a language that does not mark this distinction in the same way as English. Subsequent larger-scale investigations of ESL article use confirm these particular difficulties experienced by ESL speakers whose L1 does not include articles (Master 1997). Pak (1987), using the same BSM (Bilingual Syntax Measure) elicitation procedures as did Dulay and Burt (1974), showed that the order of English grammatical morpheme acquisition of a group of Korean-speaking children living in Texas was significantly different from that of Spanish and Chinese L1 speaking children, evidencing greater difficulty with the indefinite article and plural '-s', a finding confirmed by Shin and Milroy (1999) who showed that Korean children acquiring English as an L2 in New York City did very well on pronoun case and possessive '-s', but very poorly on articles, plural '-s', and third person singular '-s', a pattern also found in Japanese L1 children. Thus (1) there are identifiable differences in rank order of acquisition of morphemes between monolingual English-speaking children and second-language learners of English from particular L1 backgrounds, and (2) there is L1 influence on the course of L2 acquisition, with clear differences in rank order of acquisition of English morphemes between Spanish-speaking and Chinese-speaking children on the one hand (Dulay and Burt 1974) and Korean and Japanese speakers on the other (Shin and Milroy 1999). The fact that Japanese and Korean are morphosyntactically very similar confirms these language specific influences on L2 acquisition: L2 acquisition is clearly affected by the transfer of learners' knowledge of their first language.

Finally, though not directly a morpheme order study, the work of Taylor (1975a, 1975b) serves to both contextualize, as a useful reminder that these have been long-standing questions however much they drifted out of vogue, and to serialize, by quantifying the transition from L1 transfer induced errors to L2 overgeneralization errors in adult L2 acquisition. Taylor investigated the English of elementary and intermediate native Spanish speaking ESL students. He analyzed their errors in the Auxiliary and Verb Phrases of eighty sentences, categorizing them into those errors that resulted from L1 transfer, and those that resulted from overgeneralization of L2 patterns. The errors of the elementary and intermediate students were not strikingly qualitatively different—they were broadly of the same type. However, the rates of these errors were quantitatively different in the two groups, with transfer errors more prevalent among elementary students (40 percent) than intermediate students (23 percent), and overgeneralizations more prevalent in intermediate (77 percent) than elementary students (60 percent). As Taylor concludes, 'overgeneralization and transfer learning . . . appear to be two distinctive linguistic manifestations of *one psychological process*. That process is one *involving prior learning to facilitate new learning*. Whether transfer or overgeneralization will be . . . dominant . . . for a given learner will depend on his degree of proficiency in the target language' (Taylor 1975b: 87, emphasis added).

CONCLUSIONS

Corder (1967) proposed the 'error analysis' model in place of the CAH, introducing the notion of the system of interlanguage (IL) at the same time. Errors were no longer viewed simply as an indication of difficulty, but instead they illustrated a learner's *active* attempts at systematic development via *intake*, a *process* which involved the construction of an IL, a 'transitional competence' reflecting the *dynamic* nature of the learner's developing system, where every learner sentence should be regarded as being *idiosyncratic* until shown to be otherwise. Selinker's development of this concept of interlanguage emphasized the wide range of *cognitive* influences on this complex and often *conscious constructive* process: language transfer was indeed an integral part of the mix, but it was accompanied by a range of other factors including overgeneralization of L2 rules, transfer of training, strategies of L2 learning, and communication strategies (Selinker 1971, 1972). Indeed, every sentence is idiosyncratic, as indeed it is systematic too. Every sentence conspires in the system, but the system is more than the sum of the parts. Every new usage is created dynamically, influenced by interactions among the different parts of the complex system that are unique in time (N. C. Ellis 2005; Larsen-Freeman 1997; Larsen-Freeman and Ellis in press, December 2006). The morpheme acquisition studies that I have concentrated upon for illustration here, however comprehensive, provide little more than crude nomothetic summaries of highly variable dynamic systems. Bayley (1994, 1996) and Bayley and Preston (1996) describe detailed variation analyses of the use of past-tense morphology in advanced Chinese learners of English who overtly inflected in obligatory contexts anywhere between 26 percent and 80 percent of verbs, depending upon (1) the salience of the phonetic difference between inflected and base forms (e.g. suppletive, ablaut irregular, other irregular, regular syllabic, regular nonsyllabic, etc.); (2) the grammatical aspect (perfective aspect favors ($p_i = .68$) and imperfective aspect disfavors ($p_i = .32$) past tense marking); and (3) phonological factors involving the preceding and following phonetic segments. His studies clearly show how interlanguage is systematically conditioned by a range of linguistic, social, and developmental factors, and that acquisition of past tense marking may best be described as proceeding, not stepwise from unacquired to acquired, but along a continuum. This is not a simple continuum, however. It is not a 'uni-dimensional' or 'linear one'. Rather, in Bayley's (1994) words, 'it is multidimensional, with the perfective–imperfective aspectual opposition, phonetic saliency, phonological processes (such as '-t', 'd' deletion) that converge with particular morphological classes, and social and developmental factors constituting the different dimensions' (1994: 178). As explained in the first half of this paper and demonstrated in Goldschneider and DeKeyser (2001), add to this utterance-by-utterance variability the systematic influences of frequency, contingency, semantic complexity, and broader aspects of salience and syntactic category. Next, as explained in the second half of this paper and demonstrated in L2 transfer research such as that of Shin and Milroy (1999), add the ways first language usage induces interference, overshadowing and blocking, and perceptual learning, all biasing the ways in which learners selectively attend to their second language.

It does not matter how you see grammatical functors as operating in acquisition or processing, whether they serve as direct cues to interpretation in structuralist/functional/construction grammar accounts, whether they serve as cues to parameter resetting in UG-based accounts, whether acquisition of routines for their processing is what serves as the major determinant of acquisition as in Processability Theory (Pienemann 1998), or whatever your favorite flavor, or blend. *In all cases*, the functors have to be perceived as cues before they can partake in acquisition. And the factors of L1-driven, learned selective attention and perceptual tuning described in this paper modulate and attenuate their perception accordingly. In this pair of papers I have tried to outline several of the factors affecting the processing of each and every experience of language and the associative learning consequently accrued from each of these exemplars. Together, these experiences conspire dynamically in the emergence of interlanguage systems. An understanding of associative learning theory illuminates both the rationality of L1 fluency and the apparent irrationalities of fragile L2 acquisition and fossilization.

Final version accepted May 2005

ACKNOWLEDGEMENTS

I thank Gabi Kasper, Martin Bygate, Robert DeKeyser, anonymous *Applied Linguistics* Readers, and Diane Larsen-Freeman, Usha Lakshmanan, and University of Michigan ELI 692 students for insightful and constructive advice on prior versions of this paper.

REFERENCES

Anderson, M. C. and **R. A. Bjork.** 1994. 'Mechanisms of inhibition in long-term memory: A new taxonomy' in D. Dagenbach and T. Carr (eds): *Inhibitory Processes in Attention, Memory and Language*. San Diego: Academic Press, pp. 265–326.

Anderson, M. C. and **J. H. Neely.** 1996. 'Interference and inhibition in memory retrieval' in E. L. Bjork and R. A. Bjork (eds): *Memory. Handbook of Perception and Cognition*, 2nd edn. San Diego, CA: Academic Press, pp. 237–313.

Baddeley, A. D. 1976. *The Psychology of Memory*. New York: Harper and Row.

Baddeley, A. D. 1997. *Human Memory: Theory and Practice*, Rev edn. Hove: Psychology Press.

Bailey, N., C. Madden, and **S. Krashen.** 1974. 'Is there a ''natural sequence'' in adult second language learning?' *Language Learning*, 24: 235–43.

Bardovi-Harlig, K. 1992. 'The use of adverbials and natural order in the development of temporal expression.' *IRAL* XXX: 299–320.

Bardovi-Harlig, K. 2000. *Tense and Aspect in Second Language Acquisition: Form, meaning, and use*. Oxford: Blackwell.

Bates, E. and **J. C. Goodman.** 1997. 'On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia and real-time processing.' *Language and Cognitive Processes* 12: 507–86.

Bates, E., I. Bretherton, and **L. Snyder.** 1988. *From First Words to Grammar*. Cambridge: Cambridge University Press.

Bayley, R. 1994. 'Interlanguage variation and the quantitative paradigm: Past tense marking in Chinese-English' in E. Tarone, S. M. Gass, and A. D. Cohen (eds): *Research Methodology in Second-Language Acquisition*. Hillsdale, NJ: Lawrence Erlbaum, pp. 157–81.

Bayley, R. 1996. 'Competing constraints on variation in the speech of adult Chinese learners of English' in R. Bayley and D. R. Preston (eds): *Second Language Acquisition and Linguistic Variation*. Amsterdam: John Benjamins, pp. 97–120.

Bayley, R. and **D. R. Preston.** 1996. *Second Language Acquisition and Linguistic Variation*. Amsterdam: John Benjamins.

Biber, D., S. Johansson, G. Leech, S. Conrad, and **E. Finegan.** 1999. *Longman Grammar of*

Spoken and Written English. Harlow, UK: Pearson Education.

Bloom, P. 2000. *How Children Learn the Meanings of Words*. Harvard, MA: MIT Press.

Brown, R. 1973. *A First Language: The early stages*. Cambridge, MA: Harvard University Press.

Brunswik, E. 1955. 'Representative design and probabilistic theory in a functional psychology,' *Psychological Review* 62: 193–217.

Bybee, J. 2006. 'Mechanisms of change in grammaticalization: The role of frequency' in R. D. Janda and B. D. Joseph (eds): *Handbook of Historical Linguistics*. Oxford: Blackwell.

Celce-Murcia, M. and **D. Larsen-Freeman.** 1999. *The Grammar Book: An ESL/EFL teacher's course*, 2nd edn. New York: Heinle & Heinle.

Chapman, G. B. and **S. J. Robbins.** 1990. 'Cue interaction in human contingency judgment,' *Memory and Cognition* 18: 537–45.

Cheng, P. W. and **K. J. Holyoak.** 1995. 'Adaptive systems as intuitive statisticians: causality, contingency, and prediction' in J.-A. Meyer and H. Roitblat (eds): *Comparative Approaches to Cognition*. Cambridge MA: MIT Press, pp. 271–302.

Cheng, P. W. and **L. R. Novick.** 1990. 'A probabilistic contrast model of causal induction,' *Journal of Personality and Social Psychology* 58: 545–67.

Corder, S. P. 1967. 'The significance of learners' errors,' *International Review of Applied Linguistics* 5: 161–9.

Croft, W. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

Dale, P. S. and **L. Fenson.** 1996. 'Lexical development norms for young children,' *Behavioral Research Methods, Instruments, and Computers* 28: 125–7.

Diesing, M. 1992. *Indefinites*. Cambridge, MA: MIT Press.

Doughty, C. 2001. 'Cognitive underpinnings of focus on form' in P. Robinson (ed.): *Cognition and Second Language Instruction*. Cambridge: Cambridge University Press, pp. 206–57.

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

Dulay, H. C. and **M. K. Burt.** 1973. 'Should we teach children syntax?' *Language Learning* 23: 245–58.

Dulay, H. C. and **M. K. Burt.** 1974. 'Natural sequences in child second language acquisition,' *Language Learning* 24: 37–53.

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* 24/2: 143–88.

Ellis, N. C. 2002b. 'Reflections on frequency effects in language processing,' *Studies in Second Language Acquisition* 24/2: 297–339.

Ellis, N. C. 2005. 'At the interface: Dynamic interactions of explicit and implicit language knowledge,' *Studies in Second Language Acquisition* 27: 305–52.

Ellis, N. C. 2006. 'Language acquisition as rational contingency learning,' *Applied Linguistics* 27/1: 1–24.

Ellis, R. 2001. 'Introduction: Investigating form-focused instruction,' *Language Learning* 51(Supplement 1): 1–46.

Elman, J. L., E. A. Bates, M. H. Johnson, A. Karmiloff-Smith, D. Parisi, and **K. Plunkett.** 1996. *Rethinking Innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Fahle, M. and **T. Poggio.** (eds) 2002. *Perceptual Learning*. Cambridge, MA: MIT Press.

Faurud, K. 1990. 'Definiteness and the processing of noun phrases in natural discourse,' *Journal of Semantics* 7: 395–433.

Garlock, V. M., A. C. Walley, and **J. L. Metsala.** 2001. 'Age-of-acquisition, word frequency, and neighborhood density effects on spoken word recognition by children and adults.' *Journal of Memory and Language* 45/3: 468–92.

Goldschneider, J. M. and **R. DeKeyser.** 2001. 'Explaining the ''natural order of L2 morpheme acquisition'' in English: A meta-analysis of multiple determinants,' *Language Learning* 51: 1–50.

Goldstone, R. L. 1994. 'Influences of categorization on perceptual discrimination,' *Journal of Experimental Psychology: General* 123: 178–200.

Goldstone, R. L. 1995. 'Effects of categorization on color perception,' *Psychological Science* 6: 298–304.

Goldstone, R. L. 1998. 'Perceptual learning,' *Annual Review of Psychology* 49: 585–612.

Goldstone, R. L. 2000. 'A neural network model of concept-influenced segmentation' in Proceedings of the Twenty-second Annual Conference of the Cognitive Science Society. Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 172–7.

Goldstone, R. L. and **M. Steyvers.** 2001. 'The sensitization and differentiation of dimensions during category learning,' *Journal of Experimental Psychology: General* 130: 116–39.

Golinkoff, R. 1992. 'Young children and adults use lexical principles to learn new nouns,' *Developmental Psychology* 28: 99–108.

Golinkoff, R., C. B. Mervis, and **K. Hirsh-Pasek.** 1994. 'Early object labels: The case for a developmental lexical principles framework,' *Journal of Child Language* 21: 125–56.

Goodman, J. C., H. C. Nusbaum, L. Lee, and **K. Broihier.** 1990. 'The effects of syntactic and discourse variables on the segmental intelligibility of speech.' Paper presented at the Proceedings of the 1990 International Conference on Spoken Language Processing, Kobe, Japan.

Gopnik, A. and **A. N. Meltzoff.** 1997. *Words, Thoughts, and Theories*. Cambridge, MA: MIT Press.

Hakuta, K. 1976. 'A case study of a Japanese child learning ESL,' *Language Learning* 26: 321–52.

Hakuta, K. and **H. Cancino.** 1977. 'Trends in second language acquisition research,' *Harvard Educational Review* 47: 294–316.

Hawkins, J. 1978. *Definiteness and Indefiniteness: A Study of Reference and Grammaticality Prediction*. London: Croon Helm.

Hawkins, J. 1991. 'On (in)definite articles: Implicatures and (un)grammaticality prediction,' *Journal of Linguistics* 27: 405–42.

Hernandez, A., P. Li, and **B. MacWhinney.** 2005. 'The emergence of competing modules in bilingualism,' *TRENDS in Cognitive Sciences* 9: 220–5.

Herron, D. and **E. Bates.** 1997. 'Sentential and acoustic factors in the recognition of open- and closed-class words,' *Journal of Memory and Language* 37: 217–39.

Hock, H. S., E. Webb, and **L. C. Cavedo.** 1987. 'Perceptual learning in visual category acquisition,' *Memory and Cognition* 15: 544–56.

Iverson, P., P. K. Kuhl, R. Akahane-Yamada, E. Diesch, Y. Tohkura, A. Kettermann, *et al.* 2003. 'A perceptual interference account of acquisition difficulties for non-native phonemes.' *Cognition* 87: B47–B57.

James, C. 1980. *Contrastive Analysis*. London: Longman.

James, W. 1890. *The Principles of Psychology*, Vol. 1. New York: Holt.

Jurafsky, D., A. Bell, M. Gregory, and **W. D. Raymond.** 2001. 'Probabilistic relations between words: Evidence from reduction in lexical production' in J. Bybee and P. Hopper (eds): *Frequency and the Emergence of Linguistic Structure*. Amsterdam: Benjamins, pp. 229–54.

Kamin, L. J. 1969. 'Predictability, surprise, attention, and conditioning' in B. A. Campbell and R. M. Church (eds): *Punishment and Aversive Behavior*. New York: Appleton-Century-Crofts, pp. 276–96.

Kaminsky, J., J. Call, and **J. Fischer.** 2004. 'Word learning in a domestic dog: Evidence for ''Fast Mapping'',' *Science* 304: 1682–3.

Kersten, A. W., R. L. Goldstone, and **A. Schaffert.** 1998. 'Two competing attentional mechanisms in category learning,' *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24: 1437–58.

Kohonen, T. 1998. 'The self-organization map, a possible model of brain maps' in K. H. Pribram (ed.): *Brain and Values: Is a Biological Science of Values Possible?* Mahwah, NJ: Lawrence Erlbaum Associates, pp. 207–36.

Kruschke, J. K. 1992. 'ALCOVE: an exemplar-based connectionist model of category learning,' *Psychological Review* 99: 22–44.

Kruschke, J. K. 1993. 'Human category learning: Implications for backpropagation models,' *Connection Science* 5: 22–44.

Kruschke, J. K. 1996. 'Base rates in category learning,' *Journal of Experimental Psychology: Learning, Memory and Cognition* 22: 3–26.

Kruschke, J. K. 2001. 'Toward a unified model of attention in associative learning,' *Journal of Mathematical Psychology* 45: 812–63.

Kruschke, J. K. and **N. J. Blair.** 2000. 'Blocking and backward blocking involve learned inattention,' *Psychonomic Bulletin and Review* 7: 636–45.

Kruschke, J. K. and **M. K. Johansen.** 1999. 'A model of probabilistic category learning,' *Journal of Experimental Psychology: Learning, Memory, and Cognition* 5: 1083–119.

Kuhl, P. K. and **P. Iverson.** 1995. 'Linguistic experience and the ''perceptual magnet effect'' ' in W. Strange (ed.): *Speech Perception and Linguistic Experience: Issues in Cross-language Research*. Timonium, MD: York Press, pp. 121–54.

Lado, R. 1957. *Linguistics across Cultures: Applied Linguistics for Language Teachers*. Ann Arbor: University of Michigan Press.

Larsen-Freeman, D. 1975. 'The acquisition of grammatical morphemes by adult ESL students,' *TESOL Quarterly* 9: 409–19.

Larsen-Freeman, D. 1976. 'An explanation for the morpheme acquisition order of second language learners,' *Language Learning* 26: 125–34.

Larsen-Freeman, D. 1997. 'Chaos/complexity science and second language acquisition,' *Applied Linguistics* 18: 141–65.

Larsen-Freeman, D. and **N. C. Ellis.** (eds) in press, December 2006. *Language Emergence: Implications for Applied Linguistics*.

Li, P., I. Farkas, and **B. MacWhinney.** 2004. 'Early lexical development in a self-organizing neural networks,' *Neural Networks* 17: 1345–62.

Lively, S. E., D. B. Pisoni, and **S. D. Goldinger.** 1994. 'Spoken word recognition' in M. A. Gernsbacher (ed.): *Handbook of Psycholinguistics*. San Diego, CA: Academic Press, pp. 265–318.

Livingston, K. R. and **J. K. Andrews.** 1995. 'On the interaction of prior knowledge and stimulus structure in category learning,' *Quarterly Journal of Experimental Psychology: Human Experimental Psychology* 48A: 208–36.

Lubow, R. E. 1973. 'Latent inhibition,' *Psychological Bulletin* 79: 398–407.

Lubow, R. E. and **O. Kaplan.** 1997. 'Visual search as a function of type of prior experience with target and distractor,' *Journal of Experimental Psychology: Human Perception and Performance* 23: 14–24.

Lyons, J. 1999. *Definiteness*. Cambridge: Cambridge University Press.

Mackintosh, N. J. 1975. 'A theory of attention: Variations in the associability of stimuli with reinforcement,' *Psychological Review* 82: 276–98.

MacWhinney, B. (ed.). 1987. *Mechanisms of Language Aquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

MacWhinney, B. 1989. 'Competition and lexical categorization' in R. Corrigan (ed.): *Linguistic Categorization*. Amsterdam: Benjamins, pp. 195–242.

MacWhinney, B. 1997. 'Second language acquisition and the Competition Model' in A. M. B. De Groot and J. F. Kroll (eds): *Tutorials in Bilingualism: Psycholinguistic Perspectives.* Mahwah, NJ: Lawrence Erlbaum Associates, pp. 113–142.

MacWhinney, B. 2001. 'The competition model: The input, the context, and the brain' in P. Robinson (ed.): *Cognition and Second Language Instruction*. New York: Cambridge University Press, pp. 69–90.

MacWhinney, B. and **E. Bates.** 1989. *The Crosslinguistic Study of Sentence Processing*. Cambridge: Cambridge University Press.

MacWhinney, B., E. Bates, and **R. Kliegl.** 1984. 'Cue validity and sentence interpretation in English, German, and Italian,' *Journal of Verbal Learning and Verbal Behavior* 23/2: 127–50.

MacWhinney, B., C. Pleh, and **E. Bates.** 1985. 'The development of sentence interpretation in Hungarian,' *Cognitive Psychology* 17/2: 178–209.

Markman, E. 1989. *Categorization and Naming in Children: Problems of Induction*. Boston, MA: MIT Press.

Master, P. 1997. 'The English article system: acquisition, Function, and Pedagogy,' *System* 25: 215–32.

Matessa, M. and **J. R. Anderson.** 2000. 'Modeling focused learning in role assignment,' *Language and Cognitive Processes* 15/3: 263–92.

McClelland, J. L. 2001. 'Failures to learn and their remediation: A Hebbian account' in J. L. McClelland and R. S. Siegler (eds): *Mechanisms of Cognitive Development: Behavioral and Neural Perspectives*. Carnegie Mellon symposia on cognition. Mahwah, NJ: Erlbaum, pp. 97–121.

McDonald, J. L. 1986. 'The development of sentence comprehension strategies in English and Dutch,' *Journal of Experimental Child Psychology* 41: 317–35.

McDonald, J. L. 1987. 'Sentence interpretation in bilingual speakers of English and Dutch,' *Applied Psycholinguistics* 8: 379–413.

McGeoch, J. A. 1942. *The Psychology of Human Learning: An Introduction*. New York: Longmans.

McWhorter, J. 2002. *The Power of Babel: A Natural History of Language*. San Francisco, CA: W. H. Freeman & Co.

Meisel, J. 1987. 'Reference to past events and actions in the development of natural second language acquisition' in C. Pfaff (ed.): *First and Second Language Acquisition*. New York, NY: Newbury House.

Merriman, W. 1999. 'Competition, attention, and young children's lexical processing' in B. MacWhinney (ed.): *The Emergence of Language*. Hillsdale, NJ: Erlbaum, pp. 331–58.

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

Müller, G. E. and **A. Pilzecker.** 1900. 'Experimentalle Beitrage zur Lehre com Gedachtnis,' *Zeitschrift für Psychologie* 1: 1–300.

Nosofsky, R. M. 1986. 'Attention, similarity, and the identification-categorization relationship,' *Journal of Experimental Psychology: General* 115: 39–57.

Nosofsky, R. M. 1987. 'Attention and learning processes in the identification and categorization of integral stimuli,' *Journal of Experimental Psychology: Learning, Memory, and Cognition* 13: 87–108.

Nosofsky, R. M. 1992. 'Similarity scaling and cognitive process models,' *Annual Review of Psychology* 43: 25–53.

Odlin, T. 1989. *Language Transfer*. New York: Cambridge University Press.

Osgood, C. E. 1949. 'The similarity paradox in human learning: A resolution,' *Psychological Review* 56: 132–43.

Pak, Y. 1987. Age difference in morpheme acquisition among Korean ESL learners: Acquisition order and acquisition rate. Unpublished PhD dissertation, University of Texas, Austin.

Pevtzow, R. and **R. L. Goldstone.** 1994. 'Categorization and the parsing of objects' in *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 717–22.

Pica, T. 1983. 'Adult acquisition of English as a second language under different conditions of exposure.' *Language Learning* 33: 465–97.

Pienemann, M. 1998. *Language Processing and Second Language Development: Processability Theory*. Amsterdam: John Benjamins.

Postman, L. 1971. 'Transfer, interference, and forgetting' in J. W. Kling and L. A. Riggs (eds): *Woodworth and Schlosberg's: Experimental Psychology*, 3rd edn. New York: Holt, Rinehart & Winston, pp. 1019–1132.

Quine, W. V. O. 1960. *Word and Object*. Cambridge, MA: MIT Press.

Regier, T. 2003. 'Emergent constraints on word-learning: a computational perspective,' *Trends in Cognitive Science* 7: 263–268.

Repp, B. H. and **A. M. Liberman.** 1987. Phonetic category boundaries are flexible' in S. Harnad (ed.): *Categorical Perception: The Groundwork of Cognition*. Cambridge: Cambridge University Press, pp. 89–112.

Rescorla, R. A. and **A. R. Wagner.** 1972. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement' in A. H. Black and W. F. Prokasy (eds): *Classical Conditioning II: Current Theory and Research* New York: Appleton-Century-Crofts, pp. 64–99.

Robinson, P. (Ed.). 2001. *Cognition and Second Language Instruction*. Cambridge: Cambridge University Press.

Sagi, D. and **D. Tanne.** 1994. 'Perceptual learning: learning to see,' *Current Opinion in Neurobiology* 4: 195–9.

Saussure, F. d. 1916. Cours de linguistique générale (Roy Harris, Trans.). London: Duckworth.

Schmidt, R. 2001. 'Attention' in P. Robinson (ed.): *Cognition and Second Language Instruction*. Cambridge: Cambridge University Press, pp. 3–32.

Schumann, J. 1987. 'The expression of temporality in basilang speech,' *Studies in Second Language Acquisition* 9: 21–41.

Schumann, J. H. 1978. *The Pidginisation Process: A Model for Second Language Acquisition*. Rowley, MA: Newbury House.

Schyns, P. and **L. Rodet.** 1997. 'Categorization creates functional features,' *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23: 681–96.

Sebastián-Gallés, N. and **L. Bosch.** 2005. Phonology and bilingualism' in J. F. Kroll and A. M. de Groot (eds): *Handbook of Bilingualism: Psycholinguistic Approaches*. Oxford: Oxford University Press.

Seitz, A. R. and **T. Watanabe.** 2003. 'Is subliminal learning really passive?' *Nature* 422: 36–7.

Selinker, L. 1971. 'A brief reappraisal of contrastive linguistics,' *University of Hawaii Working Papers in Linguistics* 3: 1–10.

Selinker, L. 1972. 'Interlanguage,' *IRAL, International Review of Applied Linguistics in Language Teaching* 10: 209–31.

Shanks, D. R. 1995. *The Psychology of Associative Learning*. New York: Cambridge University Press.

Shannon, C. E. 1948. 'A mathematical theory of communication,' *Bell Systems Technological Journal* 27: 623–56.

Sharwood Smith, M. 1981. 'Consciousness raising and the second-language learner,' *Applied Linguistics* 2: 159–68.

Shin, S. J. and **L. Milroy.** 1999. 'Bilingual language acquisition by Korean school children in New York City,' *Bilingualism: Language and Cognition* 2: 147–67.

Shiu, L. and **H. Pashler.** 1992. 'Improvement in line orientation discrimination is retinally local but dependent on cognitive set,' *Perception and Psychophysics* 52: 582–8.

Simon, H. A. 1957. *Models of Man: Social and Rational*. New York: Wiley and Sons.

Slobin, D. I. 1992. *Psycholinguistics*, 2nd edn. Glenview, IL: Scott, Foresman and Company.

Storkel, H. L. 2001. 'Learning new words: Phonotactic probability in language development,' *Journal of Speech, Language, and Hearing Research* 44: 1321–37.

Taylor, B. P. 1975a. 'Adult language learning strategies and their pedagogical implications,' *TESOL Quarterly* 9: 391–9.

Taylor, B. P. 1975b. 'The use of overgeneralization and transfer learning strategies by elementary and intermediate students in ESL,' *Language Learning* 25: 73–108.

Tenenbaum, J. and **F. Xu.** 2000. 'Word learning as Bayesian inference.' Paper presented at the 22nd Annual Conference of the Cognitive Science Society.

Terrell, T. 1991. 'The role of grammar instruction in a communicative approach,' *The Modern Language Journal* 75: 52–63.

Tomasello, M. 2003. *Constructing a Language*. Boston, MA: Harvard University Press.

Van Geert, P. 1993. 'A dynamic systems model of cognitive growth: Competition and support under limited resource conditions' in L. B. Smith and E. Thelen (eds): *A Dynamic Systems Approach to Development: Applications*. Cambridge, MA: MIT Press, pp. 265–331.

Van Geert, P. 1994. *Dynamic Systems of Development: Change Between Complexity and Chaos*. New York: Harvester Wheatsheaf.

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

Watanabe, T., Y. Sasaki, and **J. Nanez.** 2001. 'Perceptual learning without perception,' *Nature* 413: 844–8.

Weinreich, U. 1953. *Languages in Contact*. The Hague: Mouton.

Werker, J. E. and **C. E. Lalonde.** 1988. 'Cross-language speech perception: Initial capabilities and developmental change,' *Developmental Psychology* 24: 672–83.

Werker, J. E. and **R. C. Tees.** 1984. 'Cross-language speech perception: Evidence for perceptual reorganization during the first year of life,' *Infant Behavior and Development* 7: 49–63.

Yeh, S.-L., J.-L. Li, T. Takeuchi, V. C. Sun, and **W.-R. Liu.** 2003. 'The role of learning experience on the perceptual organization of Chinese characters,' *Visual Cognition* 10: 729–64.

Zuraw, K. 2003. 'Probability in language change' in R. Bod, J. Hay and S. Jannedy (eds): *Probabilistic Linguistics*. Cambridge, MA: MIT Press, pp. 139–76.