15

USAGE-BASED AND FORM-FOCUSED LANGUAGE ACQUISITION

The associative learning of constructions, learned attention, and the limited L2 endstate

Nick C. Ellis

1 Introduction

Cognitive Linguistics proposes that First Language Acquisition (L1A) involves the acquisition from language usage of constructions that map linguistic form and function. In this view, competence and performance both emerge from the dynamic system that is the frequency-tuned conspiracy of memorized exemplars of use of these constructions, with competence being the rationally integrated sum of prior usage and performance being its dynamic contextualized activation. L1A tunes the ways in which learners attend to language.

This chapter gives a psychological slant on the associative learning of linguistic constructions. The first part describes the aspects of associative learning that affect usage-based L1A and L2A both: frequency, 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 illustrating the combined operation of these factors in first and second language acquisition of English grammatical morphemes, a particular illustration of a broader claim that they control the acquisition of all linguistic constructions.

However, usage-based Second Language Acquisition (L2A) is typically much less successful that L1A, with naturalistic or communicatively-based L2A stabilizing at end-states far short of nativelike ability. Why? What is it that limits construction learning in L2A in comparison to L1A? The second half of the chapter considers the apparent irrationalities of L2A, the

shortcomings where input fails to become intake. It describes how "learned attention" explains these effects. The fragile features of L2A, those aspects of the second language that are not typically acquired, are those which, however available in the input, fall short of intake because of one of the factors of contingency, cue competition, salience, interference, overshadowing, blocking, or perceptual learning, all shaped by L1 entrenchment. Each phenomenon is explained within associative learning theory and exemplified in language learning. 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."

That the successes of L1A and the limitations of L2A both, paradoxically, derive from the same basic learning principles provides a non age-invoked biological explanation for why usage-based L2A stops short while L1A does not. These processes also explain why form-focused instruction is a necessary component of L2A, and why successful L2A necessitates a greater level of explicit awareness of the L2 constructions, a dialectic tension between the conflicting forces of the learner's current stable states of interlanguage and the evidence of explicit form-focused feedback, either linguistic, pragmatic, or metalinguistic, that allows socially scaffolded development.

2 Factors affecting usage-based L1A and L2A

2.1 Frequency

The past 50 years of research in psycholinguistics demonstrates that language processing is rational in that it is exquisitely sensitive to prior usage. Fluent language processing reflects frequency of usage at all sizes of grain: phonology and phonotactics, reading, spelling, lexis, morpho-syntax, formulaic language, language comprehension, grammaticality, sentence production, and syntax (N. C. Ellis, 2002a, 2002b). The words that we are likely to hear next, their most likely senses, the linguistic constructions we are most likely to utter next, the syllables we are likely to hear next, the graphemes we are likely to read next, and the rest of what's coming next across all levels of language representation, are made more readily available to us by our language-processing systems. Consider, for example, that while you are conscious of words in your attentional focus, you certainly did not consciously label the word "focus" just now as a noun; yet this sentence would be incomprehensible if your unconscious language analyzers did not treat "focus" as a noun rather than as a verb or an adjective. Nor, on reading "focus," were you aware of its nine alternative meanings or of their rankings in overall likelihood, or of their rankings in this particular context, rather than in different sentences where you would instantly bring a different meaning to mind. A wealth

of psycholinguistic evidence suggests that this information is available unconsciously for a few tenths of a second before your brain plumps for the most appropriate one in this context. Most words have multiple meanings, but only one at a time becomes conscious. This is a fundamental fact about consciousness (Baars, 1988, 1997). In these ways our unconscious language mechanisms present up to consciousness the constructions that are most likely to be relevant next.

Consider the particular constructions "Wonderful!," "One, two, three," "Once upon a time," and "Won the battle, lost the war." We have come to learn these sequential patterns of sound simply as a result of repeated usage. All perception is fundamentally probabilistic: every stimulus is ambiguous, as is any utterance or piece of language. Each of these formulaic constructions begins with the sound "wʌn." At the point of hearing this initial sound, what should the appropriate interpretation be? We perceive the most probable thing. Psycholinguistic analyses demonstrate that fluent language users are sensitive to the relative probabilities of occurrence of different constructions in the speech stream. Since we have experienced many more tokens (particular examples) of "one" than they have "won," in the absence of any further information, we favor the unitary interpretation over that involving gain or advantage.

Not only do we know the constructions that are most likely to be of overall relevance (i.e., first-order probabilities of occurrence), but we also predict the ones that are going to pertain in any particular context (sequential dependencies), and the particular interpretations of cues that are most likely to be correct (contingency statistics). Thus, in the context of "Alice in . . .," "wonderland" comes to the fore well ahead of "one"; we stop counting and sense wonder instead. These predictions are rational and normative in that they accurately represent the statistical covariation between events (N. C. Ellis, 2006). There is good evidence that human implicit cognition, acquired over natural ecological sampling as natural frequencies on an observation by observation basis, is rational in this sense (J. R. Anderson, 1990, 1991a, 1991b; Gigerenzer & Hoffrage, 1995; Sedlmeier & Betsc, 2002; Sedlmeier & Gigerenzer, 2001).

This evidence of rational language processing implies that language learning, too, is an intuitive statistical learning problem, one that involves the associative learning of representations that reflect the probabilities of occurrence of form–function mappings. Learners have to figure language out: their task is, in essence, to learn the probability distribution P(interpretation|cue, context), the probability of an interpretation given a formal cue in a particular context, a mapping from form to meaning conditioned by context (Manning, 2003). In order to achieve optimal processing, acquisition mechanisms must have gathered the normative evidence that is the necessary foundation for rationality. To accurately predict what is going to happen next, we require a representative sample of experience of

similar circumstances upon which to base our judgments, and the best sample we could possibly have is the totality of our linguistic experience to date. The systematicities of language competence, at all levels of analysis from phonology, through syntax, to discourse, emerge from learners' implicit tallying of the constructions in their usage history and the implicit distributional analysis of this lifetime sample of language input.

2.2 Contingency and DP

2.2.1 Learning theory

But it is not just the frequency of encounter of a construction that determines its acquisition. The degree to which animals, human and other alike, learn associations between cues and outcomes depends upon the contingency of the relationship as well. In classical conditioning it is the reliability of the bell as a predictor of food that determines the ease of acquisition of this association (Rescorla, 1968). In language learning it is the reliability of the form as a predictor of an interpretation that determines its acquisition (MacWhinney, 1987). The last thirty years of psychological investigation into human sensitivity to the contingency between cues and outcomes (Shanks, 1995) demonstrates that when given sufficient exposure to a relationship, people's judgments match quite closely the contingency specified by DP (the one-way dependency statistic, Allan, 1980) which measures the directional association between a cue and an outcome, as illustrated in Table 15.1.

Table 15.1 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

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.

DP = P(O|C) − P(O|-C) = a/(a+b)-c/(c+d).

DP is the probability of the outcome given the cue (P(O|C) minus the probability of the outcome in the absence of the cue (P(O|-C). When these are the same, when the outcome is just as likely when the cue is present as when it is not, there is no covariation between the two events

and DP = 0. DP approaches 1.0 as the presence of the cue increases the likelihood of the outcome and approaches −1.0 as the cue decreases the chance of the outcome—a negative association.

2.2.2 Language learning

Consider for example the acquisition of English grammatical morphemes in these terms. In the 25 years following Brown's (1973) descriptions of child L1A, the "morpheme order studies," classic milestones in the history of SLA theory, investigated the L2A of the grammatical functors, progressive -ing, plural -s, possessive -s, articles a, an, the, third person singular present -s, and past tense. These studies showed a remarkable commonality in the L2 and L1 orders of acquisition of these grammatical functors[1] (Bailey, Madden, & Krashen, 1974; Brown, 1973; Dulay & Burt, 1973; Pica, 1983), broadly:

1 plural "-s"   "Book s"
2 progressive "-ing"   "John go ing"
3 copula "be"   "John is here" / "John 's here"
4 auxiliary "be"   "John is going" / "John 's going"
5 articles "the/a"   "The books"
6 irregular past tense   "John went"
7 third person "-s"   "John like s books"
8 possessive " 's"   "John 's book"

Clearly, there are no 1:1 mappings between these cues and their grammatical interpretations. Plural -s, third person singular present -s, and possessive -s, are all 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, there are many instances of the cue being present but that outcome not pertaining, b in Table 15.1 is of high frequency, and DP accordingly low. View the mappings from the other direction too: plural -s, third person singular present -s, and possessive -s all have variant expression as the allomorphs [s, z, əz]. Therefore, if we evaluate any one of these, say [s], as a cue for a particular outcome, say plurality, there are many instances of that outcome in the absence of the cue, c in Table 15.1 is inflated, and DP 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. 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), as evidenced by the many pages of

explanation given to *the* in a grammar of English (Biber, Johansson, Leech, Conrad, & Finegan, 1999; Celce-Murcia & 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 DP 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 1:1 form to meaning mappings.

2.3 Multiple cues, the PCM, and Cue-Competition

2.3.1 Learning theory

Normative DP theory describes associative learning where learners have to acquire the relationship between a cue and an outcome and where the cue is the only obvious causal feature present. In such situations contingency is easy to specify and human learning is shown to be rational in that it accords with the normative DP rule. However, it is rarely if ever the case that predictive cues appear in isolation, and most utterances, like most other stimuli, present the learner with a set of cues which co-occur with one another, with the learner's task being to determine the ones that are truly predictive. In such cases of multiple cues to interpretation, the predictions of normative analysis using the DP rule are muddied by *selection effects*: learners selectively choose between potential causal factors. Thus, in some circumstances the cue may be selected for association with an interpretation while in other circumstances it may not, depending on the presence and status of other cues.

Cheng and Holyoak (1995) and Cheng and Novick (1990) have proposed an extended version of contingency theory, which they termed the Probabilistic Contrast Model (PCM), as a descriptive account of the use of statistical regularity in human causal induction. The model, which applies to events describable by discrete variables, assumes that potential causes are evaluated by contrasts computed over a "focal set." The focal set for a contrast is a contextually-determined set of events that the reasoner *selects* to use as input to the computation of that contrast. The focal set consists of all trials on which the target cue is present as well as all those trails that are identical to the target present trials except for the absence of the target, the PCM in this way approximating the logic of classical scientific method. Because the focal set is not the universal set of events, the results of this reasoning appear irrational when measured unconditionally against DP theory applied to the whole learning set. Shanks (1995) reviews the evidence of human reasoning in situations of multiple cues, concluding that the results are well accommodated by the PCM.

2.3.2 Language learning

There is considerable redundancy in language (Shannon, 1948), with the same meaning or intention potentially expressible in a wide variety of ways. So language is a prime example of a stimulus environment rich in multiple cues. The Competition Model (MacWhinney, this volume, MacWhinney & Bates, 1989; MacWhinney, Bates, & Kliegl, 1984) was explicitly formulated to deal with competition between multiple linguistic cues to interpretation. 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.

Experiments using miniature artificial languages have shown that, in the initial stages of acquisition, learners tend to focus on only one cue at a time (Blackwell, 1995; MacWhinney & Bates, 1989; Matessa & Anderson, 2000; McDonald, 1986; McDonald & MacWhinney, 1991). For example, when cues for determining the agent in sentences include word order, noun animacy and agreement of noun and verb, learners typically decide to focus attention on only one of these as the predictor of interpretation. MacWhinney, Pleh, and Bates (1985) demonstrated that the cue that children first focus upon is that which has the highest overall validity as measured by its availability (its frequency or probability of occurrence) times its reliability (its probability of correctly indicating the interpretation, broadly equivalent to its DP). The effect is that a cue with high availability but low reliability may initially be used over a cue that is of lower availability, even though it is in fact more reliable. Learners focus on one cue alone to begin with. Later on, after having tracked the use of this first cue, they will add a second cue to the mix and begin to use the two in combination, and, as development proceeds, so additional cues may be added if they significantly helped reduce errors of understanding, as measured by the statistic "conflict validity" which relates to how the cue affords extra predictive accuracy when its interpretation conflicts with that of a co-occurring cue. This variable by variable incremental sequence is as predicted by the probability contrast model and the Competition Model both.

2.4 Salience

2.4.1 Learning theory

The phenomena summarized by the PCM are qualified by other additional factors of associative learning, those relating to salience of cue and importance of outcome.[2] Experimental investigations of learning in

situations of multiple cues illustrate robust phenomena of *selective attention*. In such experiments two cues, C1 and C2, are always presented together during training and they jointly predict an outcome. In the test-phase, the strength of conditioning to C1 and C2 presented individually are measured. The typical outcome is that the strength of conditioning to each cue depends on their relative physical intensity. If C1 is a dim light and C2 a bright light then, after conditioning to the C1-C2 combination, the learned response to the bright light is very strong while the dim light alone produces little or no reaction (Kamin, 1969). The general perceived strength of stimuli is commonly referred to as their *salience*. Although it might in part be related to the physically measurable intensity of stimuli, salience refers to the intensity of the subjective experience of stimuli, not of the objective intensity of the stimuli themselves. Salience, as subjective experience, varies between individuals and between species. Rescorla and Wagner (1972) presented a formal model of conditioning which expresses the capacity any cue has to become associated with an outcome at any given stage of learning:

$$dV = ab\,(L - V)$$

The associative strength of the outcome to the cue is referred to by the letter V, the change in this strength which occurs on each trial of conditioning is called dV, a is the salience of the outcome, b is the salience of the cue and L is the amount of processing given to a completely unpredicted outcome. The more a cue is associated with an outcome, the less additional association the outcome can induce.

The Rescorla-Wagner model pulled together the findings of hundreds of experiments each designed with an empirical rigor unsurpassed outside animal learning research. Its generality of relevance makes it arguably the most influential formula in the history of conditioning theory. It encapsulates the phenomena that the salience of the cue and the importance of the functional outcome are essential factors in any associative learning. A language learner might never get round to noticing low salience cues, particularly when the interpretation accuracy afforded by the other more obvious cues does well enough for everyday communicative survival. The Rescorla and Wagner (1972) model predicts that for low salience cues whose redundancy denies them any more than low outcome importance, dV on any learning trial will be negligible, and thus they may never become integrated into a consolidated construction.

2.4.2 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 physical 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, in press, this volume; Jurafsky, Bell, Gregory, & Raymond, 2001; Zuraw, 2003). As Slobin (1992, p. 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've" 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 & Goodman, 1997) and in speech directed to children (Goodman, Nusbaum, Lee, & Broihier, 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 no more than 40% to 50% of the time (Herron & Bates, 1997). If fluent native speakers can only hear these grammatical functors from the bottom-up evidence of input 40%–50% 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 & Goodman, 1997, pp. 51–52). 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 second as in first language learners (Bailey, Madden, & Krashen, 1974; Brown, 1973; Dulay & 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

380

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.

2.5 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: each of the above explanations, low DP, 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 this 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," described above, that 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. Although each of the factors of input frequency, semantic complexity, grammatical complexity, phonological form, and perceptual salience has been historically considered within SLA theory 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, p. 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, əz] 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, while 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 DP by inflating c, homophony by inflating b (see Table 15.1). 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% 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-analyzed 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 SLA. 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.

3 Factors special to L2

3.1 The limited endstate of usage-based SLA

Children almost invariably eventually acquire nativelike grammatical competence in their first language. Although the acquisition of functional morphology takes many months of input analysis (Brown, 1973; Tomasello, 1992, 2003; Lieven & Tomasello, this volume) and a large critical mass of evidence is necessary for all the subtle generalizations to

emerge (Bates & Goodman, 1997), nevertheless, nativelike competence is the norm. In stark contrast, adults almost invariably fail to acquire nativelike competence in a second language from naturalistic exposure. Although second language learners, too, are surrounded by language, not all of it "goes in," and L2A is typically much less successful than L1A. This is Corder's distinction between input, the available target language, and intake, that subset of input that actually gets in and which the learner utilizes in some way (Corder, 1967). Associative L2 learning from naturalistic usage typically falls far short of a nativelike endstate, often stabilizing at a "Basic Variety" of interlanguage which, although sufficient for everyday communicative purposes, predominantly comprises just nouns, verbs, and adverbs, with little or no functional inflection and with closed-class items, in particular determiners, subordinating elements, and prepositions, being rare, if present at all (Klein, 1998). What are the additional associative learning factors which explain this paradox?

3.2 Interference

3.2.1 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 a 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 US, 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 (M. C. Anderson & Bjork, 1994). Quite simply, when multiple traces are associated to the same cue, they tend to compete for access to conscious awareness (M. C. Anderson & Neely, 1996; Baddeley, 1976, chapter 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).

3.2.2 Language learning

Prior proposals for understanding aspects of SLA 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, p. 1). PI underpins a variety of fundamental phenomena of language learning and language transfer, 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 are different will be difficult" (Lado, 1957, p. 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 chapter, 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

384

USAGE-BASED AND FORM-FOCUSED SLA

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, Mervis, & Hirsh-Pasek, 1994; Gopnik & 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 more likely get hats hung on them next. Kaminsky, Call, and Fischer (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 tested 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, i.e. 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 & 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.

3.3 Overshadowing and blocking

3.3.1 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 discussing selective attention above, I introduced 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 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 & 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 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 15.2, 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 rise or not 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

Table 15.2 The design and outcome of Chapman & 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 prediction of X	BD -> X	D	77

(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 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 extend 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 PCM (Cheng & Holyoak, 1995) introduced above, explains the deviations up and down from 57 as follows. The focal set for cue C is just the AC and A trials; 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 15.2, 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 & Blair, 2000; Kruschke & Johansen, 1999; Shanks, 1995)—they are not restricted to complex learning environments with dozens of cues and outcomes and intricate interactions.

Krushke 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 & 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 & 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.

3.3.2 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, p. 59).

I believe that this redundancy is much more influential in second rather than first language acquisition. Children learning their native language only acquire the meanings of temporal adverbs quite late in development (Dale & 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 this nonredundant 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 C in Table 15.2 above, and they readily fit the requirements for the phenomena involving overshadowing. Thus, another pervasive reason for the non-acquisition of low-salience cues in SLA 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*" or, more properly, "stabilized" with a Basic Variety of IL devoid of functional inflections and closed-class items (Klein, 1998).

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 thus: "Since many features of L2 input are likely to be infrequent, non-salient, and communicatively redundant, intentionally focused attention may be a practical (though not theoretical) necessity for successful language learning" (Schmidt, 2001). Terrell characterized explicit grammar instruction as "the use of instructional strategies to draw the students' attention to, or focus on, form and/or structure" (Terrell, 1991), with instruction targeted at increasing the salience of inflections and other commonly ignored features by firstly pointing them out and explaining their structure, and secondly by providing meaningful input that contains many instances of the same grammatical meaning–form relationship. 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 & Williams, 1998, p. 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).

3.4 Perceptual learning

3.4.1 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 & Poggio, 2002; Goldstone, 1998; Seitz & Watanabe, 2003; Watanabe, Sasaki, & Nanez, 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 component 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). These psychological representation spaces can be charted using the statistical technique of multidimensional scaling (MDS) (Nosofsky, 1992) rather than factor analysis, and the emergence of structure can be simulated using connectionist techniques such as 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

390

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 & 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, Webb, and Cavedo (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

391

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's the way you see it (or that's the way you first see it), and those are the features whose strengths are incremented on each subsequent processing episode.

Goldstone (Goldstone, 1998; Goldstone & Steyvers, 2001; Kersten, Goldstone, & Schaffert, 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 & Tanne, 1994; Shiu & Pashler, 1992).

3.4.2 Language learning

The sound categories and categorical perception of L1 are subject to perceptual learning (Lively, Pisoni, & Goldinger, 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 & 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, Walley, & Metsala, 2001; Metsala & 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

392

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 one and four 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 & Lalonde, 1988; Werker & Tees, 1984). In contrast to the newborn infant, the starting disposition of the neural stuff for SLA 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 & 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, p. 369). 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 & 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, Kuhl, Akahane-Yamada, Diesch, Tohkura, Kettermann, 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 SLA. Under normal L1 circumstances, usage optimally tunes the language system to the input. A sad irony for an L2 speaker under such circumstances of transfer is that more input simply compounds their error; they dig themselves ever deeper into the hole begun and subsequently entrenched by their L1.

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, p. 112).

Feature imprinting has been clearly exemplified in the first and second learning of Chinese characters. Yeh, Li, Takeuchi, Sun, and Liu (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-entrenched *tabula repleta*.

3.5 *Transfer effects in L2 morpheme acquisition order*

The first half of this chapter 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 a second-language learner whose L1 does not make the same semantic discriminations as the L2 target with regard 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 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 SLA, with clear differences in rank order of acquisition of English morphemes between Spanish-speaking and Chinese-speaking children on the one hand (Dulay & Burt, 1974) and Korean and Japanese speakers on the other (Shin & Milroy, 1999). The fact that Japanese and Korean are morphosyntactically very similar confirms these language-specific influences on SLA: L2 acquisition is clearly affected by the transfer of learners' knowledge of, and attention to, the features and cues that are criterial for 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 longstanding 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 SLA. Taylor

investigated the English of elementary and intermediate native Spanish-speaking ESL students. He analyzed their errors in the auxiliary and verb phrases of 80 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%) than intermediate students (23%), and overgeneralizations more prevalent in intermediate (77%) than elementary students (60%). 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, p. 87, my emphasis).

4 Conclusions

Many of the chapters in this Handbook concern theories of Cognitive Grammar, cognitive semantics, and the attentional system of language. These analyses illuminate the content of what is learned and transferred. Their necessary complements are theories of associative learning and the ways that prior language knowledge tunes attention to language and cognition about language, thus to understand the processes of acquisition.

Corder (1967) proposed the "error analysis" model in place of Lado's 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 & Ellis, in press, December 2006).

The morpheme acquisition studies that I've concentrated upon for illustration here, however comprehensive, provide little more than crude nomothetic summaries of highly variable dynamic systems. Bayley (1994, 1996; Bayley & Preston, 1996) describes 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%–80% of verbs, depending upon (i) the salience of the phonetic difference between inflected and base forms (e.g. suppletive, ablaut irregular, other irregular, regular syllabic, regular nonsyllabic, etc.), (ii) the grammatical aspect (perfective aspect favors (p_i = .68) and imperfective aspect disfavors (p_i = .32) past tense marking), and (iii) phonological factors involving the preceding and following phonetic segments. His studies clearly show how interlanguage is systematically conditioned by a range of usage factors, linguistic, social, and developmental, and that acquisition of past tense marking may best be described as proceeding, not stepwise from unacquired to acquired, but along a continuum. Not a simple continuum though, "not a uni-dimensional or linear one. Rather, 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" (Bayley, 1994, p. 178). As explained in the first half of this chapter 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 chapter and demonstrated in L2 transfer research like 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. These are the associative mechanisms that underlie learned attention as it affects "thinking for speaking" and "thinking for listening," the usage that underpins language learning itself.

Notes

1 Later studies did call for some qualification of this conclusion by demonstrating some variability across L1/L2 that resulted from language transfer. We will consider this in more detail towards the end of this chapter, but for the moment concentrate on the commonalities.
2 In constructional terms, outcome importance is the degree to which successful interpretation of the construction is essential to successful interpretation of the message as a whole.

Bibliography

Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society*, *15*, 147–149.

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

Anderson, J. R. (1991a). Is human cognition adaptive? *Behavioral and Brain Sciences*, *14*(3), 471–517.

Anderson, J. R. (1991b). The adaptive nature of human categorization. *Psychological Review*, *98*(3), 409–429.

Anderson, M. C., & Bjork, R. A. (1994). Mechanisms of inhibition in long-term memory: A new taxonomy. In D. Dagenbach & T. Carr (Eds.), *Inhibitory processes in attention, memory and language* (pp. 265–326). San Diego: Academic Press.

Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In E. L. Bjork & R. A. Bjork (Eds.), *Memory. Handbook of perception and cognition* (2nd ed., pp. 237–313). San Diego, CA: Academic Press.

Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge: Cambridge University Press.

Baars, B. J. (1997). *In the theater of consciousness: The workspace of the mind*. Oxford: Oxford University Press.

Baddeley, A. D. (1976). *The psychology of memory*. New York: Harper and Row.

Baddeley, A. D. (1997). *Human memory: Theory and practice*. (Revised ed.). Hove: Psychology Press.

Bailey, N., Madden, C., & Krashen, S. (1974). Is there a "natural sequence" in adult second language learning? *Language Learning*, *24*, 235–243.

Bardovi-Harlig, K. (1992). The use of adverbials and natural order in the development of temporal expression. *International Review of Applied Linguistics in Language Teaching* (*IRAL*), *30*, 299–320.

Bardovi-Harlig, K. (2000). *Tense and aspect in second language acquisition: Form, meaning, and use*. Oxford: Blackwell.

Bates, E., & Goodman, J. C. (1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia and real-time processing. *Language and Cognitive Processes*, *12*, 507–586.

Bayley, R. (1994). Interlanguage variation and the quantitative paradigm: Past tense marking in Chinese-English. In E. Tarone, S. M. Gass, & A. D. Cohen (Eds.), *Research methodology in second-language acquisition* (pp. 157–181). Hillsdale, NJ: Lawrence Erlbaum.

Bayley, R. (1996). Competing constraints on variation in the speech of adult Chinese learners of English. In R. Bayley & D. R. Preston (Eds.), *Second language acquisition and linguistic variation* (pp. 97–120). Amsterdam: John Benjamins.

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

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

Blackwell, A. (1995). *Artificial languages, virtual brains*. Unpublished doctoral dissertation, University of California at San Diego.

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.

Bybee, J. (in press). Mechanisms of change in grammaticalization: The role of frequency. In R. D. Janda & B. D. Joseph (Eds.), *Handbook of historical linguistics*. Oxford: Blackwell.

Celce-Murcia, M., & Larsen-Freeman, D. (1999). *The grammar book: An ESL/EFL teacher's course* (2nd ed.). New York: Heinle & Heinle.

Chapman, G. B., & Robbins, S. J. (1990). Cue interaction in human contingency judgment. *Memory and Cognition*, *18*, 537–545.

Cheng, P. W., & Holyoak, K. J. (1995). Adaptive systems as intuitive statisticians: Causality, contingency, and prediction. In J.-A. Meyer & H. Roitblat (Eds.), *Comparative approaches to cognition* (pp. 271–302). Cambridge MA: MIT Press.

Cheng, P. W., & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology*, *58*, 545–567.

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

Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavioral Research Methods, Instruments, and Computers*, *28*, 125–127.

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* (pp. 206–257). Cambridge: Cambridge University Press.

Dulay, H. C., & Burt, M. K. (1973). Should we teach children syntax? *Language Learning*, *23*, 245–258.

Dulay, H. C., & Burt, M. K. (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), 143–188.

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*, 305–352.

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., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Fahle, M., & Poggio, T. (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., Walley, A. C., & Metsala, J. L. (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), 468–492.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*, 684–704.

Goldschneider, J. M., & DeKeyser, R. (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* (pp. 172–177). Hillsdale, NJ: Lawrence Erlbaum Associates.

Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General*, *130*, 116–139.

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

Golinkoff, R., Mervis, C. B., & Hirsh-Pasek, K. (1994). Early object labels: The case for a developmental lexical principles framework. *Journal of Child Language*, *21*, 125–156.

Goodman, J. C., Nusbaum, H. C., Lee, L., & Broihier, K. (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., & Meltzoff, A. N. (1997). *Words, thoughts, and theories*. Cambridge, MA: MIT Press.

Hakuta, K. (1976). A case study of a Japanese child learning ESL. *Language Learning*, *26*, 321–352.

Hakuta, K., & Cancino, H. (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: Croom Helm.

Hawkins, J. (1991). On (in)definite articles: Implicatures and (un)grammaticality prediction. *Journal of Linguistics*, *27*, 405–442.

Herron, D., & Bates, E. (1997). Sentential and acoustic factors in the recognition of open- and closed-class words. *Journal of Memory and Language*, *37*, 217–239.

Hock, H. S., Webb, E., & Cavedo, L. C. (1987). Perceptual learning in visual category acquisition. *Memory and Cognition*, *15*, 544–556.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., 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., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In

J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229–254). Amsterdam: Benjamins.

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior* (pp. 276–296). New York: Appleton-Century-Crofts.

Kaminsky, J., Call, J., & Fischer, J. (2004). Word learning in a domestic dog: Evidence for "Fast Mapping". *Science, 304*, 1,682–1,683.

Kersten, A. W., Goldstone, R. L., & Schaffert, A. (1998). Two competing attentional mechanisms in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1,437–1,458.

Klein, W. (1998). The contribution of second language acquisition research. *Language Learning, 48*, 527–550.

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?* (pp. 207–236). Mahwah, NJ: Lawrence Erlbaum Associates.

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*, 812–863.

Kruschke, J. K., & Blair, N. J. (2000). Blocking and backward blocking involve learned inattention. *Psychonomic Bulletin and Review, 7*, 636–645.

Kruschke, J. K., & Johansen, M. K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 5*, 1,083–1,119.

Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the "perceptual magnet effect". In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 121–154). Timonium, MD: York Press.

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*, 409–419.

Larsen-Freeman, D. (1976). An explanation for the morpheme acquisition order of second language learners. *Language Learning, 26*, 125–134.

Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied Linguistics, 18*, 141–165.

Larsen-Freeman, D., & Ellis, N. C. (Eds.) (in press, December 2006). *Language Emergence: Implications for Applied Linguistics* (Vol. 27).

Lively, S. E., Pisoni, D. B., & Goldinger, S. D. (1994). Spoken word recognition. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 265–318). San Diego, CA: Academic Press.

Livingston, K. R., & Andrews, J. K. (1995). On the interaction of prior knowledge and stimulus structure in category learning. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 48A*, 208–236.

Lubow, R. E. (1973). Latent inhibition. *Psychological Bulletin, 79*, 398–407.

Lubow, R. E., & Kaplan, O. (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, U.K: Cambridge University Press.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, *82*, 276–298.

MacWhinney, B. (1987). The Competition Model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 249–308).

MacWhinney, B. (1989). Competition and lexical categorization. In R. Corrigan (Ed.), *Linguistic categorization* (pp. 195–242). Amsterdam: Benjamins.

MacWhinney, B., & Bates, E. (1989). *The crosslinguistic study of sentence processing*. Cambridge: Cambridge University Press.

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior*, *23*(2), 127–150.

MacWhinney, B., Pleh, C., & Bates, E. (1985). The development of sentence interpretation in Hungarian. *Cognitive Psychology*, *17*(2), 178–209.

Manning, C. D. (2003). Probabilistic syntax. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic linguistics* (pp. 289–341). Cambridge, MA: MIT Press.

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*, 215–232.

Matessa, M., & Anderson, J. R. (2000). Modeling-focused learning in role assignment. *Language and Cognitive Processes*, *15*(3), 263–292.

McClelland, J. L. (2001). Failures to learn and their remediation: A Hebbian account. In J. L. McClelland & R. S. Siegler (Eds.), *Mechanisms of cognitive development: Behavioral and neural perspectives. Carnegie Mellon symposia on cognition* (pp. 97–121). Mahwah, NJ: Erlbaum.

McDonald, J. L. (1986). The development of sentence comprehension strategies in English and Dutch. *Journal of Experimental Child Psychology*, *41*, 317–335.

McDonald, J. L., & MacWhinney, B. (1991). Levels of learning: A comparison of concept formation and language acquisition. *Journal of Memory and Language*, *30*(4), 407–430.

McGeoch, J. A. (1942). *The psychology of human learning: An introduction*. New York: Longman.

McWhorter, J. (2002). *The power of Babel: A natural history of language*. San Franciso, 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* (pp. 331–358). Hillsdale, NJ: Erlbaum.

Metsala, J. L., & Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 89–120).

Müller, G. E., & Pilzecker, A. (1900). Experimentalle Beitrage zur Lehre com Gedachtnis. *Zeitschrift fur 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*, 132–143.

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., & Goldstone, R. L. (1994). Categorization and the parsing of objects. In *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society* (pp. 717–722). Hillsdale, NJ: Lawrence Erlbaum Associates.

Pica, T. (1983). Adult acquisition of English as a second language under different conditions of exposure. *Language Learning*, *33*, 465–497.

Postman, L. (1971). Transfer, interference, and forgetting. In J. W. Kling & L. A. Riggs (Eds.), *Woodworth and Schlosberg's: Experimental psychology* (3rd ed., pp. 1,019–1,132). New York: Holt, Rinehart & Winston.

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., & Liberman, A. M. (1987). Phonetic category boundaries are flexible. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 89–112). Cambridge: Cambridge University Press.

Rescorla, R. A. (1968). Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative and Physiological Psychology*, *66*, 1–5.

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

Robinson, P. (Ed.). (2001). *Cognition and second language instruction*. Cambridge: Cambridge University Press.

Sagi, D., & Tanne, D. (1994). Perceptual learning: learning to see. *Current Opinion in Neurobiology*, *4*, 195–199.

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

Schumann, J. H. (1978). *The pidginization process: A model for second language acquisition*. Rowley, MA: Newbury House.

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

Schyns, P., & Rodet, L. (1997). Categorization creates functional features. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 681–696.

Sebastián-Gallés, N., & Bosch, L. (2005). Phonology and bilingualism. In J. F. Kroll

& A. M. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches*. Oxford: Oxford University Press.

Sedlmeier, P., & Betsc, T. (2002). *Etc.—Frequency processing and cognition*. Oxford: Oxford University Press.

Sedlmeier, P., & Gigerenzer, G. (2001). Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General*, *130*, 380–400.

Seitz, A. R., & Watanabe, T. (2003). Is subliminal learning really passive? *Nature*, *422*, 36–37.

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*, 209–231.

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*, 623–656.

Sharwood Smith, M. (1981). Consciousness raising and the second-language learner. *Applied Linguistics*, *2*, 159–168.

Shin, S. J., & Milroy, L. (1999). Bilingual language acquisition by Korean school children in New York City. *Bilingualism: Language and Cognition*, *2*, 147–167.

Shiu, L., & Pashler, H. (1992). Improvement in line orientation discrimination is retinally local but dependent on cognitive set. *Perception and Psychophysics*, *52*, 582–588.

Slobin, D. I. (1992). *Psycholinguistics* (2nd ed.). 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*, 1,321–1,337.

Taylor, B. P. (1975a). Adult language learning strategies and their pedagogical implications. *TESOL Quarterly*, *9*, 391–399.

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., & Xu, F. (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. (1992). *First verbs: A case study of early grammatical development*. New York: Cambridge University Press.

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 & E. Thelen (Eds.), *A dynamic systems approach to development: Applications* (pp. 265–331). Cambridge, MA: MIT Press.

van Geert, P. (1994). *Dynamic systems of development: Change between complexity and chaos*. New York: Harvester Wheatsheaf.

VanPatten, B. (1996). *Input processing and grammar instruction in second language acquisition*. New York: Ablex.

Watanabe, T., Sasaki, Y., & Nanez, J. (2001). Perceptual learning without perception. *Nature*, *413*, 844–848.

Weinreich, U. (1953). *Languages in contact*. The Hague: Mouton.

Werker, J. E., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology*, *24*, 672–683.

Werker, J. E., & Tees, R. C. (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., Li, J.-L., Takeuchi, T., Sun, V. C., & Liu, W.-R. (2003). The role of learning experience on the perceptual organization of Chinese characters. *Visual Cognition*, *10*, 729–764.

Zuraw, K. (2003). Probability in language change. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic linguistics* (pp. 139–176). Cambridge, MA: MIT Press.