Figure and Ground in SLA Theory

Michael Harrington

University of Queensland

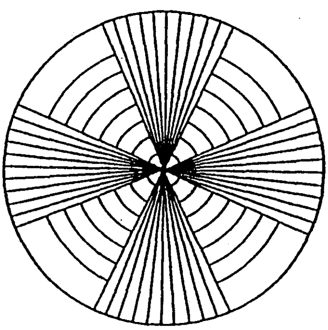
Abstract

Input-matching models of second language (L2) development assign a primary role to the environment in shaping learning outcomes. The input-matching perspective characterizes language development as an *emergent* process resulting from the interaction of the learning environment and general learning capacities. In this talk distinguishing features of the input-matching perspective are described and contrasted with other approaches to L2 development. Two kinds of input-matching models in the SLA literature, the Competition Model and connectionist simulations, are then examined. A key distinction is made between the approach as a theory of learning and as an approach to modeling. The scope of the models in accounting for L2 development is evaluated on the basis of this distinction.

Keywords: input-matching models, associative learning, connectionism, the Competition Model, eliminativist models, SLA theory

INTRODUCTION

The interplay of figure and ground is integral to visual perception. Consider the overlaid patterns below. The two patterns differ according to whether they are perceived as the figure or the ground.



The concentric lines are perceived as uninterrupted rings when they serve as the background for the pattern of radiating spokes. When the radiating spokes are backgrounded and the circles brought forward, the circles lose their continuity and take on the appearance of separate, pie-shaped pieces. The point is that we cannot make sense of the figure without also making sense of the ground.

By the same token, we cannot understand how a learner develops knowledge of a second language (L2) without also taking into account the learning environment in which this process unfolds. In the same way that there is perceptual dependency between the figure and ground, there is also a conceptual dependency in SLA research between the learner and learning environment. We cannot examine one without considering the other. This analogy will serve as the point of departure for my talk today, in which I will be discussing a class of models that assign a central role to the learning environment. These models provide the means to investigate the relationship between "figure" and "ground" in L2 development.

The focus here is on *input-matching* models of L2 development, which are represented in the SLA literature by the *Competition Model* (MacWhinney & Bates, 1989) and *connectionist simulations* of L2 development (Ellis & Schmidt, 1998). The terms "Competition Model" and "connectionism" are familiar to many, but it is important to note that they are actually used in three different ways. They can be used to describe a set of modeling formalisms, as a perspective on modeling and research, and as theory of language learning (which in turn is informed by a compatible theory of mind). Each use has its own set of concerns and criteria for success. As a modeling formalism each account is judged in terms of economy, computational efficiency and mathematical elegance. As a perspective on research into language development, the accounts are assessed in terms of descriptive and explanatory adequacy, as well as domain-specific considerations like ecological validity. As a theory of learning and mind, the issues of psychological and neurospsychological reality become paramount. Success at one level does not entail success at another, nor does failure at one level necessarily invalidate the account at the other levels.

All three aspects are important, but I will be focusing on the last two senses, that is, the tenability of the models as an approach to modeling L2 development and as a theory of learning. To anticipate my conclusion somewhat, I will be suggesting that current applications of these models in the SLA literature ignore the distinction between these two levels. They argue implicitly -- and sometimes explicitly -- that a successful demonstration of modeling a specific linguistic domain (e.g., plural morphology) reflects the validity of the theory of learning. I will show that this is not the case, that several claims concerning the input-matching perspective as a theory of learning and the mind have not been supported.

However, showing that current input-matching models are, in some key respects, inadequate as a theory of learning does not mean that they have nothing to offer the SLA researcher. A central point to be developed here is that input-matching models are of potential importance to SLA research because they provide a framework for modeling input to the learner. This is irrespective of the tenability of the claims made by their proponents concerning that nature of language and the mind. The nature of the input available to the learner is central to any theoretical perspective on language development. However, most treatments of input have been based on assumptions as to what shape that such input takes. Input-matching models provide, in principle, the means to systematically examine the role of input in development.

The talk is organized as follows. I will first describe the input-matching perspective as a theory of (adult) SLA. I will then present the two types of input-matching models in the SLA literature, the Competition Model and connectionist simulations. Particular attention will be given to the research logic and nature of the evidence used to support the respective accounts. Several major criticisms of the models as theories of learning will be raised, particularly concerning the adequacy of distributional statistics alone as the basis for syntactic development. I will conclude that current input-matching models have failed to support the strong version of the input-matching account, and argue instead for a mixed-model approach. In this alternative version pre-existing structural knowledge plays a central role in guiding the associative input-matching process. Three domains of L2 development where a mixed-model approach might be applied will then be identified. The talk will conclude with a brief discussion of SLA as a cognitive science.

Input-matching models of L2 development

The input-matching perspective attaches a primary role to the linguistic environment in explaining learning outcomes. Other terms are also available that denote the importance of the environment in learning, including *data-driven* (McLaughlin, 1987), *input-driven* (Plunkett, 1998), and *exposure-based* (Mitchell, Cuetos, & Corley, 1995). *Input-matching* is used because it underscores the simultaneous contribution of the environment *and* the individual to the learning process (Crain & Thornton, 1998). The input-matching perspective incorporates learner goals, the state of (and changes in) internal representations, and the learner environment in seeking to account for language development.

As a theory of language learning and representation, the input-matching perspective characterizes language development as the result of the interaction between the learning environment and domain-general learning capacities of the individual. It is a constructivist account of language development in the Piagetian sense, in that it assumes the individual can develop new knowledge representations on the basis of the input alone (MacWhinney, 1998).

The following characteristics serve to distinguish the input-matching perspective from other approaches to language and the mind.

Characteristics of the input-matching approach

• *Language is learned.* Input-matching assumes that language is *learned* through the active engagement of learning mechanisms that operate on the available linguistic input. The input can be characterized as *cues* that serve as multiple sources of information that are integrated in real-time processing to yield the response, judgment, etc., appropriate to the setting. The relationships between the surface phonological or orthographic cue and its underlying functions are expressed as relative statistical strengths, with learning a process of developing cue strengths. The cognitive mechanisms responsible for the development of cue strengths are not specific to language learning, but rather are shared by other higher order cognitive processes.[i] Language learning is thus the result of the interaction of competing cues, which include cognitive, linguistic, and interactional information. Language knowledge is thus described as an *emergent* property of the system, a term that is gaining increasing currency in language development literature (MacWhinney, 1998). Input-matching is a prime exemplar of *emergentist* (Ellis, 1998) views of language development.

The input-matching perspective is thus in sharp contrast to nativist approaches to language development like universal grammar (UG). In UG accounts, the development of syntactic knowledge is assumed to be the result of processes that are innate and largely autonomous from other types of knowledge. Language learning, or in more technical terms, *acquisition,* is assumed to be guided by biologically determined, language-specific capacities. In this view, input serves only as a trigger for the appearance of the innately specified grammar (Epstein, Flynn, & Martohardjono, 1996; Schwartz & Sprouse, 1996). The extent to which these capacities, first proposed to account for child language, are also available to the adult L2 learner has generated a sizeable literature (White, 1996).

• *Language learning is associative.* The essential mechanism responsible for development in the input-matching approach is associative learning. Associative learning reflects the individual's capacity to learn from the co-occurrence, or associations, among events in the environment (Shanks, 1995). Associative links can be of two types: *causal* links that serve to link two events; and *structural* links that consist of associations between features of a single entity (McLaren, Leevers, & Mackintosh, 1989). Language learning is viewed as an ongoing development of multiple, simple associations that are combined and incorporated into increasingly complex structures in the course of development. Sounds are associated with other sounds to form phonological units, these units are then associated with word meanings to form morphemes, and these individual morphemes are then associated in phrases, etc. Domain-general associative learning processes are assumed to be responsible for the development of all linguistic knowledge. This is again in contrast to UG and other formalist accounts of language learning, in which modular, domain-specific mechanisms are assumed to be responsible (e.g., Pinker & Prince, 1994).

• *Language knowledge is distributed.* Linguistic knowledge is assumed to be represented in the mind as a complex pattern of associative links between units. These units, which are also referred to as *microfeatures* (Clark, 1993), are smaller than the traditional units of linguistic analysis (e.g., phoneme, word, phrase, rule, etc.). Representing knowledge in a distributed manner allows the system more flexibility, as these features can be re-used to process different representations. Models that attempt to account for language entirely in terms of these microfeatures (as in Rumelhart & McClelland, 1986) are often referred to as *subsymbolic* models of language and cognition (Stillings, Weisler, Chase, Feinstein, Garfield, & Rissland, 1995). These distributed knowledge representations produce rule-like behavior, but do not assume *symbolic* representations, which is the basic stuff of symbolic models of language and cognition. Symbolic approaches assume a compositional semantics and a combinatorial syntax as intrinsic to language and cognition, and thus differ fundamentally from the input-matching perspective.[ii] We will see that the sufficiency of subsymbolic accounts in representing complex syntactic relations is a major issue in cognitive science, and the debate has major implications for the input-matching models under discussion.

• *Language knowledge is graded.* Language knowledge is the result of the learner's experience with the language. The effect of this experience is incremental, or *graded*, with the mental representations for sounds, words, grammatical rules, etc., undergoing constant modification as a result of experience. Repeated exposure strengthens a given representation, while less exposure will serve to weaken and, in certain cases, extinguish it. Conceptualizing language knowledge as a network of graded representations makes the system more robust, in that it allows the processing of partial knowledge representations. It also allows the developmental process to be modelled more readily in quantitative terms. This is particularly amenable to capturing crosslinguistic transfer processes in L2 development. The graded nature of the representations contrasts with traditional symbolic approaches, in which these representations are assumed to be learned in an all-or-none manner (Hintzman, 1993).

However, it should be noted that recent symbolic models have also incorporated graded representations (Carpenter, Miyake, & Just, 1995).

• *Language knowledge is probabilistic.* The distributed, graded nature of knowledge representations means that language is readily describable in *probabilistic* terms. Knowledge of a given form can range on a continuum from near certitude, where the structure is always used correctly, to an intermediate state where appropriate usage varies, to a random state where the correct use of the structure doesn't rise above chance. Probabilistic models have an advantage over rule-based models in capturing the variable nature of behavior, which is important in contexts where variation itself is of theoretical interest.[iii] These models have been applied to a range of cognitive domains, including visual perception and speech perception (Massaro, 1987), decision-making (Kahneman, 1982), and sociolinguistic models of discourse variation (Sankoff, 1978). The probabilistic, statistical nature of L2 processing also contrasts with the traditional formal models where rule knowledge is represented and processed in a discrete fashion (Fodor & McLaughlin, 1990).

• *The unity of processing and learning.* The last characteristic of the input-matching perspective to be considered is the relationship between processing and learning. Input-matching assumes that the same mechanisms involved in processing input are also responsible for learning new knowledge (Sharkey, 1996). The graded and distributed nature of language knowledge means that new input is processed on the basis of previously stored experience, with the act of processing itself changing the strength of the knowledge representations. The input-matching perspective thus takes a more parsimonious view of processing and learning, unlike approaches like UG in which the learning and representation of linguistic knowledge are fundamentally distinct processes.

Input-matching, SLA, and cognitive science

Five characteristics of the input-matching perspective have just been introduced. The input-matching approach described here is domain-general in scope, assumes associative learning as the basic learning mechanisms responsible for the development of graded representations, which in turn can be described in probabilistic terms. The result is a system in which learning and processing are part of the same process. The input-matching perspective is in sharp contrast to nativist approaches to language development, in which learning is driven by innate, language-specific capacities. This approach ascribes only an incidental role to the learning environment, and makes a basic distinction between learning and processing.

The subsymbolic, domain-general nature of learning assumed by the input-matching approach as a theory of learning approach places it on one side of the current divide in cognitive science between "connectionist" and "classical" accounts of cognitive architecture (MacDonald & MacDonald, 1995). Over the past decade, the debate over whether human cognition requires symbolic computation has been a -- if not the -- major issue in cognitive science. The classical view of cognition (represented in the work of Fodor, Plyshyn and others), views the mind as a non-probabilistic computational machine that carries out discrete operations on symbols in a modular architecture. The input-matching perspective, in contrast, eschews a symbolic level of representation, characterizing language knowledge instead in distributed, probabilistic terms.

The debate between the two views centers around the ability of associative, subsymbolic models to learn novel representations and to carry out the structure sensitive operations needed for the processing of syntax, among other things. However, we will see that the ramifications of the larger debate are different for adult SLA than they are for child language and cognitive science.

Input-matching approaches to SLA (1): The Competition Model

The first example of an L2 input-matching model that we will consider is the Competition Model (CM). The CM is a model of sentence processing and learning developed by Elizabeth Bates and Brian MacWhinney, and applied with success to crosslinguistic, developmental and aphasic populations. The model has been applied to L2 as a framework for modeling transfer (Gass, 1987; Harrington, 1987; MacWhinney, 1987; Kilborn, 1989; Sasaki, 1991, 1994; Kempe & MacWhinney, 1998; Rounds & Kanagy, 1998). An overview of the model and representative research will be briefly presented.

Cue-based language knowledge: Direct mapping of form and function

The CM characterises linguistic knowledge as a complex set of mappings between available surface cues (form) and the underlying meanings (functions). Word order, for example, can provide an important cue as to the function of words in a sentence. In a canonical subject-verb-object (SVO) sentence like *Bill read a book*, the position of the noun before the verb is a strong cue as to which noun is doing the action. Word order cues are particularly important in English because they convey structural information (*e.g.*, thematic roles such as Agent, Theme or Patient). However, they are less important in other languages (e.g., Japanese) where this information is signaled by such means as case marking. The strength (or *weight*) of a given cue is developed through exposure. It reflects the frequency with which the cue occurs in the language, as well as the ways in which it interacts with other cues. Note that cues are not in a simple one-to-one relationship with the underlying function, rather, cue mappings are characterized by many-to-many links between multiple surface cues and underlying functions. It is the interaction among the cues that serves as the basis for processing and learning in the CM.

Cue-based language processing: Competition and convergence

In sentence comprehension cues can either compete or converge to yield a particular interpretation. This process of competition and convergence has been modeled in the CM research literature in the domain of thematic role assignment, in particular the interpretation of the sentential agent. The main experimental methodology used in CM studies is a sentence interpretation task in which a subject is asked to identify the agent in simple two-argument sentences like *Betty kissed Jack.* The experimental paradigm allows cue combinations to be systematically varied, thus allowing the effect of specific cues and cue combinations on agent assignment to be compared and contrasted across learners and groups.[iv] See Figure 1 for the set of cues used in Harrington (1987).

Figure 1. The sentence interpretation task: Assigning the Agent

<u>The boy</u> threw <u>the rock</u>.

NP_1 V NP_2

Word order cue: The preverbal noun is usually the Agent in English SVO sentences.

Animacy cue: Animate subjects typically *throw* things.

Cues used in Harrington (1987).

Sentence (Stress underlined)	Word order	Animacy favors	Contrastive stress favors
The <u>girl</u> kissed the rock	NVN	first noun	first noun
The man touched the <u>woman</u>	NVN	both nouns	second noun
The turtle the rock kicked.	NNV	first noun	neither

Cross-linguistic cue strength

The agent interpretation task has been applied to a wide range of languages, resulting in a crosslinguistic typology of language-specific cue strengths (Bates & MacWhinney, 1989; Kilborn, 1994, for reviews). A prototypical study is Bates, McNew, MacWhinney, Devescovi & Smith (1982), who investigated the use of sentence interpretation strategies by native Italian, German, and English speakers. The study used the word order, animacy and stress cues presented in Figure 1, and adapted to the languages in the study. Subjects in each language group did the sentence interpretation task in their respective languages and the findings were compared. The results revealed a reliable, language-specific difference in the use of the cues in interpreting the Agent of the sentence. Native English speakers made relatively greater use of word order cues than their Italian counterparts, who were more dependent on animacy and stress cues (Bates et al. 1982). The reliance on word order cues was evident both in responses to the canonical SVO (NVN) orders, where the first noun was almost always chosen as Agent regardless of animacy or stress contrasts involved. It was also apparent in the noncanonical (and rarely occurring) NNV and VNN orders, where the "second noun" (= preverbal noun) was reliably chosen as Agent. For the Italian subjects, word order cues alone were less reliable than animacy and stress cues in assigning the sentential Agent, reflecting the fact that Italian allows a variety of surface word orders.

These results were replicated across a number of typologically diverse languages, and have established the CM as a useful framework for crosslinguistic comparisons.

Cue-based language learning

The CM has been applied with some success in SLA as a model of interlanguage development and processing transfer. The characterisation of language knowledge as cue strength allows interlanguage development to be described as the quantitative development of the requisite cue strengths in the target L2 (MacWhinney, 1992). The individual learner develops the cue strengths for expressing a particular linguistic function, like assigning Agenthood, through experience with the language. This process can be facilitated or inhibited by the strength of the cue in the L1. For most linguistic domains, interlanguage development can be described as the position of the individual learner or group of learners along a cue-strength continuum from the L1 to the L2 (Harrington, 1987; forthcomingA).

For example, Harrington (1987) examined sentential Agent interpretations by groups of native English, native Japanese, and native Japanese ESL learners (the Interlanguage (IL) group), for evidence of L1-influenced processing strategies in L2 performance. Accuracy and reaction time results on sentences in which word order, animacy, and contrastive stress cues were varied indicated that the IL subjects did transfer L1 processing preferences in the interpretation of the L2 English sentences. In the the canonical NVN (=SVO) order, for example, the IL responses on the English sentences (68% of the first nouns were selected as Agent) were midway between the two L1 response means (81% for the L1 English, and 59% for L1 Japanese). This suggests that the IL participants were moving from the L1 to the L2 cue strengths in this domain. However, for the nonstandard word orders (NNV and VNN), the word order cues were not as strong, as the IL responses remained closer to the L1 Japanese. The latter finding was attributed to the interaction of Animacy cues with the two types of word order (Harrington, 1987, p. 368-9). The complex interaction of cues means that the patterns of convergence and competition among the various cues in yielding particular interpretations can be difficult to interpret.

A number of CM-based L2 studies have replicated the earlier studies and have established the model as a tool for understanding L1 transfer effects in the domain of agent assignment. Although the focus on Agent assignment has meant that the effect has been thoroughly replicated crosslinguistically, the challenge remains to extend the model to other domains.

Summary of the Competition Model

The first example of an input-matching model examined was the Competition Model. It has proven itself as a reliable means for modeling both crosslinguistic cue strengths and interlanguage development, although its domain of application remains limited. There are other specific criticisms of the individual studies that will not be addressed here. See Kilborn, 1994, for a thorough and balanced review. The purpose of this section has been to show how the Competition Model provides insight into L2 processing and development. How this insight relates to a cognitive theory of SLA will be discussed below, where the role of the input-matching perspective in SLA is evaluated.

Input-matching approaches to SLA (2): Connectionist models

The other class of input-matching models in SLA research we will consider are connectionist simulations, also called *neural networks*, *neural computing*, or *parallel distributed processing* (Sharkey, 1996). Connectionist models exploit a relatively simple set of learning algorithms and massive computational resources to extract patterns from input.

There are a wide variety of connectionist models currently in use, and they have been applied to issues in a number of fields, including economics, engineering and cognitive science. The approach has also captured the attention of researchers interested in L2 issues, who have used connectionist models to simulate a number of L2 domains (Gasser, 1990; Sokolik & Smith, 1992; Broeder & Plunkett, 1994; Ellis & Schmidt, 1998).

Model overview: connections and units.

As was the case with the Competition Model, connectionist models characterise language knowledge as the links between surface forms and underlying meanings. These links are represented as *input* and *output* units linked in a network. The various cues that contribute to the Agent interpretation in the CM framework can thus be formally represented in connectionist terms as input units that combine to yield an interpretation as an output. Learning consists of adjusting weights in the connections in network. This is based on the discrepancy between the current output of the network, and the desired output presented by the teacher or environment. Adjustments in weighting are made in proportion to the degree to which the particular weighted connection (= set of cues) is responsible for the given interpretation. At the outset of learning, adjustments are made on every trail across most of the connections. In later stages there are fewer discrepancies noted between actual and desired output, and the adjustment of cue weightings (i.e., learning) takes place only in the conflict situations. See Sharkey (1996) for an introduction to connectionist modeling in the L2 context.

The earliest models depended on a *Hebbian* learning rule in which input and output connections are directly linked. These models are limited in the kinds of knowledge they are able to learn, and as a result, most current connectionist models make use of some type of layer of *hidden units*[v]. The units are located between the input and output layers and permit the network to model a much greater range of knowledge.

L2 connectionist models

Ellis & Schmidt (1997,1998) is a good example of how connectionist modeling can be provide insight into basic learning processes. The studies investigated the frequency by regularity interaction that has been observed in the processing of past tense verbs in English. Prasada & Pinker (1993) have shown that the speed at which native speakers produce the past tense when presented with the present stem is sensitive to both the frequency and the regular/irregular nature of the form. In English the regular past tense is formed by affixing -*ed* to the present stem, *play-played* or *call-called*, while each irregular verb has a unique past tense form, as in *go-went*, or *run-ran*. The ability of native speakers to produce the various forms differs markedly by type frequency. For the regular past tense, the frequency of the present tense stem does not affect the speed with which the form is produced: high frequency past tense forms are produced just as fast as low frequency forms. The story is different for irregular verbs. High frequency irregulars are produced quickly, often faster than regular past tense verbs, while the production of low frequency irregular verbs is much slower. The irregular verbs are thus sensitive to differences in frequency of occurrence, while regular verbs are not.

This pattern of results led Pinker and others to posit two separate mechanisms as being responsible for the production of past tense verb. Irregular verbs, which are sensitive to frequency effects, are assumed to be stored as individual items in associative memory. Production of these forms is a matter of retrieval from memory, a process that is extremely sensitive to frequency effects. On the other hand, performance on the regular past tense verbs shows little sensitivity to differences in frequency, and are assumed to be rule-based. Regular past tense verbs are generated by a rule binding the past tense morpheme to the stem at the time of production. Beck (1997) presented similar results for advanced L2 learners for production on the regular verb forms, but did not find a frequency effect for irregular verbs.

Ellis & Schmidt point out two shortcomings with this research. First, the findings were based on the final state of learning. The data in the studies were obtained from fluent native or advanced L2 speakers, who had already attained complete or advanced levels of knowledge. The proposed interpretation is thus applicable to the processing of these forms, but not necessarily to the learning. Second, the assumptions about frequency of item occurrence were based on frequency counts obtained from published corpora analyses. The underlying assumption was that published frequency counts reflect the learner's actual exposure to these items.

To address these issues, Ellis & Schmidt examined how these forms were learned, both by human participants and in a connectionist simulation. The comparison of empirical data and simulation results is a standard research tactic in cognitive science, but remains unusual in SLA research. In the study a group of subjects was taught the plural morphology of an artificial language. The language preserved the frequency and regularity features of the past tense verb data, and provided the means to systematically control the frequency of presentation of both the regular and irregular forms. The study tracked the individual's progress over successive learning trials. High frequency items, whether regular or irregular, were learned faster. For the low frequency items, the regular forms behaved in the same way as their high frequency counterparts. In other words, there was no frequency effect for the regular forms. The low frequency irregular forms, on the other hand, were much slower than the other three types. The irregular forms, then, were sensitive to frequency, as the previous studies had shown.

Although the results were consistent with Pinker's dual mechanism account, Ellis & Schmidt argued that the pattern of findings did not *require* a symbolic, rule-based approach (Pinker & Prince, 1994).[vi] They suggested that a connectionist model, based on a very simple associative learning mechanism, could also account for the data without the necessity of positing two separate mechanisms.

In order to test this claim, a connectionist simulation was carried out using the same stimuli used with the human subjects. The network was trained to associate the correct stems and plural affixes across different frequency and regularity conditions.[vii] The simulation results replicated those of the human subjects, and by doing so demonstrated that the frequency by regularity interaction can be accounted for in terms of simple associative learning mechanisms within a connectionist network.

• Summary of L2 connectionist models

Like the Competition Model, connectionist models provide a controlled means to investigate specific domains of L2 development. They achieve this by assuming single, unified mechanism for learning and processing, and provide an explicit characterization of the developmental process. This is promising for SLA theory, because it directly addresses a pressing need for current SLA theory, namely a psycholinguistically tenable account of the transition mechanisms responsible for the development of L2 knowledge (Gregg, 1996).

The brief survey of the two types of input-matching models current in the L2 literature leads us to the first main point of the talk.

Point #1: Input-matching models have been successful as empirical models of L2 learning and transfer, but...

I have tried to show how the input-matching perspective can contribute to our understanding of L2 development. The focus has been on the modeling process and accounts of specific domains of learning. My attention will now turn to the adequacy of the perspective as a cognitive learning account, an issue that will have a direct bearing on the role that these models might play in SLA research.

Before we examine the input-matching perspective as a cognitive learning theory, let's briefly consider what the proponents of the input-matching perspective see as its role in SLA theory. A reader might come away from reading the studies discussed here with the impression that the input-matching account is a sufficient basis for a theory of L2 development, at least in terms of syntactic development. In Ellis & Schmidt (1998), for example, they state that their model accounts for the learning of plural morphology, and strongly suggest that it can account for morphological development across the board. Likewise, they argue for a connectionist account for long-distance dependencies, which also seems to apply to general syntactic development. In these terms the input-matching account is presented as a *sufficient* account of L2 development.[viii]

I will try to show why this is not the case by identifying three significant problems with input-matching as a sufficient account of adult SLA. The problems are both empirical and theoretical in nature. This does not mean, however, that I will be writing off the input-matching perspective in SLA research. On the contrary, in the last part of the talk I will make the case that input-matching models, properly understood, can and will make an important contribution to SLA theory development. In particular, I will consider the role of these models in the development of an L2 transition account.

Problem #1: Input-matching models are unable to go beyond the input in extracting structural regularities.

The first problem with the input-matching models is that they have yet to demonstrate the ability to go beyond the input, that is, to learn language knowledge that is not manifestly present in the input. Clark & Thornton (1997) identify two ways in which statistical regularities such as grammar rules or collocations can be extracted from the input. The first way, which they term *basic*, is driven by the statistical distribution of forms. Collocations are an example of such statistical regularities, as are the cue validity counts in the Competition Model. These regularities can be extracted directly from the input, through the calculation of relative frequencies of occurrence of forms.

The second way in which regularities can be extracted from the input is indirect, or what Clark & Thornton called *derived*. These patterns represent higher order structural relations, and can only be extracted by some kind of systematic recoding of the input data.

The need for such recoding, or transformation, of the raw input has long been recognized (Chomsky, 1959). Regularities like the long-distance dependencies governing pronominal reference cannot be learned by reliance on cooccurence statistics alone. As is evident in the example sentences, there is nothing in the distribution statistics of the words that reflects the structural relationship between *Mary* and *she*.

a. *Mary* watched television before *she* had her dinner.
b. Before *Mary* had her dinner *she* watched television.
c. Before *she* had her dinner *Mary* watched television.
d. **She* watched television before *Mary* had her dinner.

The ability to learn these relations depends on a mechanism for recoding, or the *redescription*, of the raw input into more abstract representations (Clark & Karmiloff-Smith, 1993). The ultimate success of the input-matching perspective as an account of L1 development will depend on demonstrating how this recoding occurs. The most widely cited success story in this regard is Elman's model of subject-verb-subject agreement and cross-clausal dependencies (Elman, 1990, 1993). Elman used a simple recurrent network in which temporal order could be simulated. This allowed him to train simpler forms before more complex ones, as well as to incrementally increase the memory space available in learning. Both of these aspects are integral to child language learning. By starting low and slow Elman was able to train up the network to simulate agreement and long-distance dependencies. The network was first trained on simple distinctions, including whether a word was a noun, verb or relative pronoun, as well as the singular-plural distinction. The simple primitives thus trained were then used as the basis for more complex mappings, e.g., *The girl loves the boy* (Elman 1993).

By starting small, the system was able to first develop the primitives of noun and verb from the input and then use the knowledge to drive the redescription into higher order structures. Elman thus established the capacity of a connectionist model to capture, at least in principle, the recoding of the statistical input into higher level representations.

Problem #2: Input-matching models cannot account for the development of novel knowledge representations.

However, Elman's demonstration notwithstanding, there is little evidence that these models are able to develop the primitive representations needed for the higher order structures in the same manner as people do. In the input-matching models we have examined, the basic representations used to model the domain in question were all assumed. In the Competition Model studies, for example, all the cues were prewired. The cues of preverbal position, animacy contrasts, and singular-plural distinctions were all assumed in the sentence interpretation tasks. The same holds for the Ellis & Schmidt studies, as indeed it does for input-matching models in the developmental literature, particularly the studies discussed in the *Rethinking Innateness* volume by Elman et al. (Marcus, 1998).

In addition to the absence of research that actually tests the notion that these primitives can be learned directly from the input, there is also counterevidence to the evidence available. Elman's network "learned" the concept noun and verb by grouping the items on the basis of the similarity in contexts in which they appeared. The network was trained to predict the correct continuation of a string like "the girl loves the__". It learned to activate only nouns (dog, boy, cat) for the slot, instead of verbs. These basic concepts were then the basis for learning the higher order structures.

Marcus (1998) points out that the ability of the network to correctly assign an item to a category (as in identifying a noun as belonging to a collection of nouns), does not necessarily reflect the acquisition of the concept of "nounness" in a number of significant respects. Most notably, it does not enable the system to generalise basic grammatical relationships from one word to another. The network predicts how words that share a close resemblance will behave in particular contexts. However, it is unequipped to handle genuinely novel items, that is, items that have little overlap in features with the items in the training set. It is unable to go beyond the input in a simple but profound way.

Marcus tested the ability of Elman's network to generalize simple abstract relations to novel words. He trained the network on two different sentence frames to see if the noun concept developed on one set of items could generalise to another in the way it would with a human. Sentences trained in the first frame took the form of *"The bee sniffs the__"*, where the blank consisted of flower words, *rose, daisy, lilac*, etc. Sentences in the second frame took the form of a simple linking relationship expressing identity, as in "*A rose is a rose*". Nine of the ten flower words used in the first frame were also trained in the second frame, as in *"A rose is a rose"*, *"A lily is a lily"*, etc. One flower term, *lilac*, was not trained in the second frame. This item was used to test the ability of the network to generalize the relationship learned over the other items.

The network mastered all links on which it was trained in the second frame. However, when the term *lilac* was presented as a test *"A lilac is a ___"*, the network was unable to generalise to "*a lilac is a lilac*". In other words, the trained network was unable to use the distribution information about the lilac's appearance in the first frame to complete the task in the second frame. The network apparently lacks the abstract, but basic, notion of noun that would allow it to generalise a novel noun to the *X is an X* relationship. It is important to emphasise that the failure of the network to generalise was not due to any specific quirk of the Elman model. Rather, it was the result of the training regimes shared by most connectionist models, wherein training is done on local features of the input (Marcus, 1998).

In conclusion, the most widely cited test of the claim that input-matching models can develop novel representations on the basis of input regularities, has been shown to be inadequate. This suggests that a level of representation higher than associative links is needed in models of language development.

Problem #3: Input-matching models attempt to account for language processing and learning without recourse to an abstract level of syntactic representation.

The third problem with the input-matching models is actually a restatement of Problems #1 and #2, in terms of SLA theory development. As a non-symbolic model of cognition, the input-matching models have rejected the need for a distinct layer of syntactic representation. Grammar rules, rather than being a set of relations computed over discrete symbols, are characterized as patterns of activation based on associative patterns. Given the inability of these models to show that they can learn crucial language, any model that rejects an explicit syntactic component appears to face a difficult challenge. A cognitive model of SLA needs a grammar, and the models considered here have yet to show that they can provide it.

Point #2: Key foundation assumptions of input-matching models concerning the nonsymbolic and constructivist nature of language development have not been supported by research to date.

The key foundation assumptions of the input-matching perspective concerning the nonsymbolic and constructivist nature of language development have not been supported by research to date. There is little evidence to sustain the notion that associative learning alone is sufficient to account for the development of L2 knowledge. Likewise, it does not seem likely that we can account for several crucial structural aspects of language without invoking structural representations that have symbolic content.

What, then, does this mean for the input-matching studies that we have examined? If the theory of learning that underpins these models is not tenable, what contribution can the Competition Model and connectionist simulations make to SLA theory? Let's consider the alternatives.

• Implications for a model of L2 development based on the input-matching framework.

1. The approach is valid, but the research is still in its infancy. The input-matching models we have considered are scarcely a decade old, and the fact that they cannot provide a comprehensive account of development is not surprising. Elman (1990,1993) has already shown that these models can, in principle, account for long-distance dependencies and other structure sensitive processes, which had hitherto been considered beyond simple associative learning processes. As more powerful tools become available (e.g., more accurate frequency statistics via corpus analysis), the models will provide us with an increasingly refined picture of the statistical nature of linguistic input, and the means to test the role of this information in learning outcomes.

2. The approach is fundamentally misguided. The opposite alternative is that the input-matching approach is wrong. Fodor, Pylshyn and others have argued cogently that the input-matching model of language and cognition suffers fatal flaws, several of which I identified above. Elman's demonstration aside, the models are still subject to the limitations of statistical-based learning (Fodor & Plyshyn, 1988; Pinker & Prince, 1988; Fodor & McLaughlin, 1990). More specifically, these researchers reject an *eliminativist* view, in which a subsymbolic, capacity-general architecture replaces, *in its entirety*, the classical symbolic model of computation. The eliminativist view is embodied in the list of features that I set out at the beginning of the talk. It is presented in explicit terms in Rumelhart & McClelland (1986), and the more recent volume *Rethinking Innateness (*Elman et al. 1996).

3. *The strong version is too strong*. Although the eliminativist view of cognition as described here may not be tenable, it is not the only alternative available. It represents the strong version of the approach; and other accounts in the connectionist literature have incorporated key elements of the eliminativist model, while still retaining features of traditional symbolic processing models. (Smolensky, 1988; Bever, 1992; Shastri & Ajjanagadde, 1993; Hummel & Holyoak, 1997.[ix] These approaches are referred to as *mixed* or *hybrid* models, and their potential for SLA theory has yet to be explored.

Mixed models

Mixed models take many shapes and forms, but all assume some kind of pre-existing structure in which the associative learning mechanisms are embedded. A number of accounts stress the importance of domain-specific structure in learning (Burgess & Hitch, 1992; Regier, 1995; Shallice, Glasspool & Houghton, 1995; Hartley & Houghton, 1996), and others emphasise the importance of initial structural biases or categorical knowledge (Seidenberg & McClelland, 1989; Anderson, 1995; Clark & Thornton, 1997).

An example of a mixed model is Hummel & Holyoak's model of analogical reasoning. The model combines the structural sensitivity of symbolic computational models with the flexibility of distributed representations. Analogical knowledge is represented in terms of symbolic, propositional structures that bind to distributed associative representations containing specific semantic content. The distributed representations are used repeatedly across different contexts by the model in the process of mapping analogic relationships from a source analog (e.g., the sun) to a novel recipient concept (an atom). The distributed representations allow the model to carry out the mapping with a flexibility similar to that of a person (Hummel & Holyoak, 1997).

Mixed models assume a role for symbolic language representations in the more traditional sense, and thus represent more continuity with the symbolic approach than the eliminativist model. Whether mixed models will make an significant contribution to a cognitive theory of SLA remains to be seen. As a research approach they avoid several of the significant problems that beset the eliminativist perspective, and may be particularly suited to investigating adult L2 development.

The mixed-model approach provides a means to investigate the interaction of structural and associative bases of L2 development.

There are two apriori reasons why the mixed model approach deserves attention as a tool for investigating adult L2 development. In the first instance, the adult learner brings a vast amount of pre-existing knowledge to the L2 learning experience. Unlike the child, the adult learner begins with a set of structural representations in place, which means the ability of the model to learn novel representations is less relevant to adult SLA. Secondly, associative learning processes play a relatively greater role in adult learning outcomes than they do in child language. The incorporation of associative learning mechanisms with explicit symbolic representations should permit the relative contribution of associative and structural bases to L2 development to be investigated systematically.

The role of associative learning in L2 development

The effect of associative learning in L2 development is readily evident across a range of processes and learning outcomes. Associative learning is a central mechanism in vocabulary learning, in the development of fluent production (Schmidt, 1992), and is strongly implicated in differences in learner aptitude (Pimsleur, Sundland & McIntyre, 1966). The dependence on formulae and patterns for "productive" language use, particularly at the early stages of development, is another way in which associative learning drives the development of the L2 system. Despite the recognised importance of these processes, the extent to which differences in associative learning contribute to individual differences across adult learners, as well as between child and adult learning, remains largely unexplored.

A better understanding of the associative bases of L2 knowledge also has implications for our understanding of the development of structural knowledge. The use of associative-based rote forms allows the learner to extend linguistic performance beyond current computational capacity (Bever, 1992). By using memorizing and using chunks containing grammatical structures beyond her current level of competence, the learner has the chance to use, analyze and possibly integrate these structures into her developing grammar. Associative learning can thus serve as an important device for the learner to bootstrap syntactic development. At the same time, rote formulae and patterns often obscure the structural knowledge that the learner possesses. The researcher must ascertain whether the use of a particular grammatical structure is the result of productive knowledge of the rule, or due to rote learning. This thick habitual overlay in performance requires the use of methods that allow the researchers to tap the underlying knowledge more directly as, for example, in the use of grammaticality judgments.

All this suggests that associative learning will be an integral part of any cognitive theory of SLA, regardless of the cognitive architecture and theory of mind assumed. Mixed models provide a means to investigate the interface between associative and structural knowledge, although it remains to be seen what such models will look like. The mixed model of analogical reasoning proposed by Hummel and Holyoak (1997) may be particularly useful for understanding L2 processes. The model allows the presentation of input over time, and under different structural assumptions, and provides a principled means to assess the contribution of novel learning and existing knowledge to ongoing development.

Although it remains to be seen whether these models can be applied productively to L2 development, I would now like to suggest two domains where mixed models hold promise. Both domains represent the direct interaction of abstract structural knowledge and associative-based learning processes, and thus serve as a potential testing ground for the approach.

Associative and structural knowledge bases of L2 sentence processing.

Sentence processing research focuses on the real-time production and comprehension of sentences. A major enterprise in psycholinguistics, work in the area has yielded important insights into language and the mind. These insights are of great potential importance for SLA theory, as the successful access and integration of linguistic knowledge in real-time performance are critical elements of L2 proficiency. As a result, researchers are beginning to look more closely at L2 sentence processing (Harrington, forthcomingB). For example, Juffs & Harrington (1995, 1996) examined the ability of advanced L2 learners to process complex syntax. The focus of the study was the interpretation of grammatical structures involving *Wh*-extractions from Object and Subject clauses, as in *Who did Jane say her friend likes_?* and *Who did Jane say_likes her friend?*, respectively. The two structures differ in the difficulty they present for L2 learners. Results from grammaticality judgments suggest that Object extraction structures (e.g., *Who did Jane say her friend likes_?)* are easier for L2 learners than their Subject extraction counterparts (*Who did Jane say_likes her friend?*). The aim of the study was to identify the source of the processing asymmetry. This was done by attempting to identify the relative contribution of grammatical knowledge of the structures, and the learners' ability to use that knowledge in real time processing (Juffs & Harrington 1995). The pattern of word-by-word reading latencies on the two types of structure led us to conclude that observed differential in difficulty was the result of a processing deficit rather than a knowledge deficit.

The study marked an advance over previous off-line studies, as it provided insight into the interaction of processing and structural knowledge in real time. However, there were two shortcomings to the study as a window on L2 processing and, more importantly, on development. The first shortcoming concerns the population studied. The research focused on the ability of advanced L2 learners to process these forms, with all participants selected for having what, in relative terms, was considered a "final state" L2 grammar. As was the case in the Prasada & Pinker (1993) study, the effects observed reflected the outcome of the learners' experience, and provide little insight into how that knowledge developed, or the course of development that took place. The second limitation is that all the processing variation evident in the word-by-word reading times, both within and across individuals, was interpreted in terms of structural complexity demands as predicted by a GB-based parsing model (see Juffs

& Harrington, 1995, for details). The source of the proposed processing deficit itself was unspecified, and it is likely that the observed processing outcomes are the result of multiple processes, some of which are based on associative memory. Lexical access and the effect of cooccurence frequency are two prominent candidates in this regard.

These limitations can be overcome by using a model of processing that combines explicit structural assumptions with an account of the associative learning mechanisms responsible for the strength of the specific lexical items and speed of processing. Such a model would allow the relative contribution of multiple sources of knowledge contributing to performance to be assessed. It would also provide a framework for assessing variation across learners at a given point, and within individual learners across time. In terms of the study in question, if associative learning effects are demonstrated (e.g., through manipulation of the frequency of presentation), then it would provide support for a processing deficit account by identifying, in part, where the processing difficulties reside.

Associative and structural contributions to vocabulary development

The interaction of structure and associative learning is also manifested in L2 lexical development. Associative learning is widely recognised as an integral factor in vocabulary learning. Paired-associate learning has long been a basic vocabulary learning technique, and vocabulary knowledge more generally reflects the key properties of statistical learning, most notably the importance of frequency (Schmidt, 1992). The effect of higher ("top-down") structure in vocabulary development is also a readily evident, if often discounted, factor. The organization of lexical material at the time of presentation can play an important role in how well items are learned. For example, one line of research has compared the effect of thematic organization and semantic category organization on L2 vocabulary presentation (Svenconis & Kerst, 1994; Harrington, 1994; Tinkham, 1997; Harrington & Park, 1997). The studies suggest that the type and degree of item relatedness can have a marked effect on vocabulary learning outcomes. The research has been driven by concerns about pedagogic practice, particularly computer-based vocabulary presentation, but it also addresses theoretical issues concerning the way in which conceptual organization interacts with bottom-up associative processes in the course of L2 vocabulary development (Harrington, 1994). A mixed model provides a means to systematically consider the effect of conceptual organization on the learning of individual items (e.g., by identifying the limits on the number of items to be presented at a time for a given structure type).

I have sketched out two domains where I believe mixed models hold promise for SLA theory development. This leads me to my third, and concluding, point.

Point #3: As the ground constrains how the figure is perceived, the learning environment shapes L2 learning outcomes. Input-matching models provide an important tool for modeling that process, and thus can play an important role in SLA research.

Three general conclusions about the input-matching perspective in SLA research can be drawn from today's talk.

• *The associative bases of L2 performance suggest a central role for input-matching models of L2 development.* The effects of associative learning vary across the domains of SLA development, as well as across the individual's development. The ability of input-matching models to capture these learning processes means that the models will play a central role in the development of SLA theory.

• *Input-matching models complement other sources of knowledge in L2 development.* The modified input-matching approach is part of a larger theory of SLA. The applications of these models to domains of L2 development can play a particularly important role in the development of a transition theory of SLA (Gregg, 1996). Powerful associative learning mechanisms are embedded within a base of pre-existing structural knowledge, and together they work to extract patterns from the input.

• *The "big issues" are of limited relevance to adult SLA research.*

The stakes in the debate between eliminativist and mixed models are significant for cognitive science and a theory of mind, which is why the debate has engendered so much intensity. However, the main force of the debate is irrelevant to SLA research. A key issue in the debate concerns whether primitives like noun or verb are innate, or emerge in the course of development. The adult L2 learner brings to the task of learning another language a vast store of first language knowledge, both linguistic and metalinguistic. This experience provides a number of initial primitives and structural biases for the L2 system. Input-matching models of L2 development, particularly adult L2 development, to some degree <u>have</u>

to be mixed models, given the knowledge of at least one previous language that the adult learner brings to the task of learning another language.

Conclusion.

In this talk I have described the input-matching perspective on language development and examined two types of input-matching models in the SLA literature, the Competition Model and connectionist models. I noted that both terms are used in at least three different ways. The terms are used to refer to specific modeling formalisms, to a perspective on the modeling process, and as a theory of learning. Three points were made about input-matching research and its relevance to SLA. The first point concerned the models' capacity to formally model domains of L2 development. This can take the form of developmental processes, as in the Ellis & Schmidt study, or transfer processes, as was the case with the Competition Model. However, the success as a demonstration of a particular domain is independent of the claims about the nature of learning made by these input-matching models, which was the second point. There is little support for an eliminativist version of the input-matching perspective.

The value of input-matching models of L2 development rests in their capacity to capture associative learning processes, and to model how the outcomes of these processes interact with overall development. In particular, this capacity can help us to better understand the interface between the associative and structural bases of L2 knowledge, and I identified two domains where this might be achieved.

NOTES

* Michael Harrington, CLTR, University of Queensland, Brisbane 4072, Qld, Australia, mharr@lingua.uq.edu.au

[i] Learning outcomes reflect the individual's experience with the environment. Thus the input-matching perspective is not a *tabula rasa* view of language learning, in which the learner brings an empty and unstructured set of cognitive capacities to the task. Rather, the point of departure in attempting to account for learning outcomes is that innate predispositions, such as those proposed in UG theory, make only a minimal contribution to the learning task. The emergent view sees development not in terms of either nurture or nature, but as a rich and complex interaction between the two dimensions (Elman, et al, 1996). A central issue in the input-matching research program is identifying the nature of those predispositions for the second language learner and the role that they play in the developmental process.

[ii] The "classical" view of cognition characterizes the mind as a general purpose computational system that is best understood as a symbol manipulation process (Newell, Rosenbloom, & Laird, 1989). A symbolic architecture assumes that knowledge is represented directly in symbols, and that computations are carried out on these representations. In natural language computation these symbols include phonemes, morphemes, grammar rules, *etc.*, and the processor works directly on these elements to yield an interpretation. The comprehension process can be described as a predicate logic in which rules are applied to a string of input representations to

yield an output. Thus the string *the man runs quickly* is understood as (is *transformed* into) the structured symbolic proposition RUNS(MAN,QUICKLY). These arguments and predicates are generated at the level of syntactic representation that is independent of the semantics of the specific items involved. However, the symbolic-subsymbolic distinction is not rigid, as some models combine distributed representation with symbolic structural knowledge (Smolensky, 1988; Hummel & Holyoak, 1997).

[iii] The treatment of second language learning in terms of probabilistic processes has major implications for research methodology and design, particularly in terms of the statistical tests used to establish reliable relationships between data. Null-hypothesis-testing, which has long been the statistical method of choice in experimental social science, assumes a discrete criterion value at which the null hypothesis is either rejected or retained and statistically "significant" difference inferred. Probabilistic processes, in contrast, are best described with correlational statistics, where the relative strengths (correlations) of multiple cue associations are of central theoretical interest (Howson & Urbach, 1989). Real-time language use wherein multiple cues naturally occur and interact is a prime example of multiple cue interaction. Probabilistic statistical techniques, notably *Bayesian modeling*, replaces the two-valued statistical decision of the *t-test* or ANOVA with a *goodness-of-fit* approach in which a range of values are estimated and theoretical decisions made based on the degree of agreement. Correlational statistics have traditionally been considered problematic for establishing causal relations between variables, with a resulting (pronounced) bias toward null hypothesis testing. However, the use and abuse of null hypothesis-testing as the default statistical tool in cognitive studies has come under increasing fire in behavioural research (Loftus, 1997).

[iv] The test sentences consist of two nouns and an active verb; one noun representing the Agent of the sentence and the other noun the object. A typical sentence used in a CM experiment is *The cow kicked the tree*, where the identification of the *cow* as the sentential Agent by English participants is based on word order and semantic cues. The preverbal noun is interpreted as Agent in an active SVO sentence, and the animacy contrast between the two nouns (a cow is more likely to kick a tree than vice versa) also strongly biases the interpretation of cow toward *cow* as the Agent for English-speaking respondents. The structure of the task allows for the straightforward examination of cue effects. As an illustration, animacy cues are contrasted in *The tree kicks the cow* versus *The cow kicks the tree.* Word order contrasts are compared in *The cow kicks the tree* versus *The cow the tree kicks.* The joint effect of animacy and word order on interpretation is reflected in pairs like *The tree the cow kicks* versus the *cow the tree kicks.* The artificial nature of task has led some to question the validity of such interpretative tasks for real-time processing which is usually context-dependent. However, the experimental paradigm allows the researcher to study the effect of cues that are not readily available or, more often, confounded in natural language. (Bates & MacWhinney, 1989; McLaughlin & Harrington, 1989).

[v] In general terms they can only learn knowledge representations that manifest linear separability. The either/or problem as represented by the Boolean relationship *Either A or B*, is an example of knowledge that is not linearly separable.

[vi] This is an oversimplification, as there are other systematic differences between regular and irregular verbs that also led Pinker to adopt a rule-based model for the regular verbs (Pinker, 1991).

[vii] The singular word stems were the input units and the plural prefixes were the output units. The output units were either the regular plural or irregular forms. The network was trained by presenting input the singulars, which would then activate one of the plural output units. At the outset this process was random, with a given input stem activating any of the plural output units. Each time a given input unit activated a particular output unit, the model would compare the activation of that mapping with the activation weight of the correct output unit, given that input unit. The backpropagation learning mechanism was used to calculate the difference between the mapping produced and the desired mapping, and to then make an incremental adjustment in weights. As a result, the next time the stem input was presented it was closer to the correct output unit in terms of activation strength. All the input-output mappings were thus trained separately over many blocks of trials. In addition, Ellis and Schmidt also had a singular stem that was used to test how well the network could generalize to novel items. They found that there was a small tendency for the untrained singular to activate the regular plural (e.g., as *wugs* is given as the plural for *wug).*

[viii] Under this view SLA becomes, in essence, a theory of implicit learning. The characteristics of the input-matching approach that I set out at the beginning are identical to most accounts of implicit learning (Cleeremans, Destrebecqz, & Boyer, 1998), except for the greater role for conscious control of learning that, in principle, is part of the input-matching perspective.

[ix] For example, Smolensky names his approach Integrated Connectionist/Symbolic architecture (ICS) Smolensky, (1988).

REFERENCES

Anderson, J. A. (1995). *An Introduction to Neural Networks*. Cambridge, MA: MIT Press.

Beck, M. (1997). Regular verbs, past tense and frequency: tracking down a potential source of NS/NSS competence differences. *Second Language Research, 13*(2), 93-115.

Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? In S. Gass & J. Schachter (Eds.), *Linguistic Perspectives on Second Language Acquisition* (pp. 41-68). Cambridge: Cambridge University Press.

Bates, E., McNew, S., MacWhinney, B., Devescovi, A., & Smith, S. (1982). Functional constraints on sentence processing. *Cognition, 11*, 245-299.

Bever, T. (1992). The demons and the beast -- Modular and nodular kinds of knowledge. In R. G. Reilly & N. E. Sharkey (Eds.), *Connectionist Approaches to Natural Language Processing* (pp. 213-252). Hove: Lawrence Erlbaum.

Broeder, P., & Plunkett, K. (1994). Connectionism and second language acquisition. In N. Ellis (Eds.), *Implicit and Explicit Learning of Languages* (pp. 421-454). San Diego, CA: Academic Press.

Burgess, N., & Hitch, G. J. (1992). Toward a network model of the articulatory loop. *Journal of Memory and Language, 31*, 429-460.

Carroll, S. E. (1995). The hidden dangers of computer modelling: remarks on Sokolik & Smith's connectionist learning model of French gender. *Second Language Research, 11*(3), 193-205.

Carpenter, P. A., Miyake, A., & Just, M. A. (1995). Language comprehension: Sentence and discourse processing. *Annual Review of Psychology, 46*, 91-120.

Chalmers, D. (1990). Syntactic transformations on distributed representations. *Connection Science, 2*, 33-63.

Chomsky, N. (1959). Review of B.F. Skinner *Verbal Behavior. Language, 35*, 25-58.

Clark, A. (1993). *Associative Engines*. Cambridge, MA: MIT Press.

Clark, A., & Karmiloff-Smith, A. (1993). The cognizer's innards: A psychological and philosophical perspective on the development of thought. *Mind and Language, 8*, 487-519.

Clark, A. & Thornton, C. (1997). Trading spaces: computation, representation, and the limits of uninformed learning. *Behavioral and Brain Sciences 20*, 57-90.

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Science, 2*(10), 406-416.

Crain, S., & Thornton, R. (1998). *Investigations in Universal Grammar.* Cambridge, MA: MIT Press.

Ellis, N., & Schmidt, R. (1998). Rules or associations in the acquisition of morphology? The frequency by regularity interaction in human and PDP learning of morphosyntax. *Language Acquisition and Connectionism. Special Issue of Language and Cognitive Processes., 13*(2/3), 307-336.

Ellis, N. C., & Schmidt, R. (1997). Morphology and long-distance dependencies: Laboratory research illuminating the A in SLA. *Studies in Second Language Acquisition, 19*(2), 145-171.

Elman, J. (1995). Language as a dynamical system. In R. F. Port & T. Van Gelder (Eds.), *Mind as Motion. Explorations in the Dynamics of Cognition.* (pp. 195-225). Cambridge, MA: MIT Press.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14,* 179-211.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition, 48,* 71-99.

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.

Epstein, S. D., Flynn, S., & Martohardjono, G. (1996). Second language research: Theoretical and experimental issues in contemporary research. *Behavioral and Brain Sciences, 19*(4), 677-749.

Fodor, J., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition, 28,* 3-71.

Fodor, J. A., & McLaughlin, B. P. (1990). Connectionism and the problem of systematicity: Why Smolensky's solution doesn't work. *Cognition, 35.*

Gass, S. (1987). The resolution of conflicts among competing systems: A bidirectional perspective. *Applied Psycholinguistics, 8,* 329-350.

Gasser, M. (1990). Connectionism and universals of second language acquisition. *Studies in Second Language Acquisition, 12*(2), 179-199.

Gregg, K. (1996). The logical and developmental problems of second language acquisition. In W. C. Ritchie & T. Bhatia (Eds.), *Handbook of Second Language Acquisition* (pp. 49-81). San Diego: Academic Press.

Harrington, M. (1987). Processing strategies as a source of interlanguage variation. *Applied Psycholinguistics, 8*, 351-378.

Harrington, M. (1992) Review of MacWhinney, B. & Bates, E. (eds.) *The crosslinguistic study of sentence processing. American Journal of Psychology*, 105, 3, 484-492

Harrington, M. (1992). Working memory capacity as a constraint on L2 development. In R. J. Harris (Ed.), *Cognitive Processing in Bilinguals* (pp. 123-135). Amsterdam: North Holland.

Harrington, M. (1994). CompLex: A tool for the development of L2 vocabulary. *Journal of Artificial Intelligence in Education, 15*(4), 481-498.

Harrington, M., & Park, S. (1997). Link-based vocabulary learning in Korean. *On CALL, 11*, 3, 16-30.

Harrington, M. (forthcoming A). *Input-matching Models of Second Language Learning.* Matwah, NJ: Lawrence Erlbaum Associates.

Harrington, M. (forthcoming B). Sentence processing. In P. J. Robinson (Ed.), *Cognition and Second Language Instruction* . Cambridge: Cambridge University Press.

Hartley, T., & Houghton, G. (1996). A linguistically constrained model of short term memory for nonwords. *Journal of Memory and Language, 35*, 1-31.

Hintzman, D., L. (1993). Twenty-five years of learning and memory: Was the cognitive revolution a mistake? In D. E. Meyer & S. Kornblum (Eds.), *Attention and Performance XIV.* (pp. 359-391). Cambridge, MA: MIT Press.

Hintzman, D. L. (1991). Why are formal models useful in psychology? In W. E. Hockley & S. Lewandowsky (Eds.), *Relating Theory and Data: Essays on Human Memory in honor of B. B. Murdock.* Hillsdale, NJ: Lawrence Erlbaum.

Howson, C., & Urbach, P. (1989). *Scientific Reasoning: The Bayesian Approach.* La Salle, IL: Open Court.

Hummel, J. E. & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review 104*, 3, 427-466.

Juffs, A., & Harrington, M. (1995). Parsing effects in L2 sentence processing: Subject and Object asymmetries in WH-extraction. *Studies in Second Language Acquisition, 17*, 483-512.

Juffs, A., & Harrington, M. (1996). Garden path sentences and error data in second language sentence processing. *Language Learning, 46*(2), 283-326.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press.

Kempe, V. & MacWhinney, B. (1998). The acquisition of case-marking by adult learners of Russian and German. *Studies in Second Language Acquisition 20*, 4, 543-587

Kilborn, K. (1994). Learning a language late: Second language acquisition in adults. In M. Gernsbacher (Eds.), *Handbook of Psycholinguistics*. New York: Academic Press.

Kilborn, K. (1989). Sentence processing in a second language: The timing of transfer. *Language and Speech, 32*(1), 1-23.

Kirsner, K., Lalor, E., & Hird, K. (1993). The bilingual lexicon: Exercise, meaning & morphology. In R. Schreuder & B. Weltens (Eds.), *The Bilingual Lexicon*. Philadelphia: John Benjamins.

MacDonald, C., & MacDonald, G. (Eds.). (1995). *Connectionism: Debates on Psychological Explanation*. Oxford: Basil Blackwell.

MacWhinney, B. (1987). Applying the competition model to bilingualism. *Applied Psycholinguistics, 8*, 315-327.

MacWhinney, B. (1992). Competition and transfer in second language learning. In R. J. Harris (Ed.), *Cognitive Processing in Bilinguals*. Amsterdam: Elsevier Science Publishers.

MacWhinney, B. (1987). The competition model. In B. MacWhinney (Eds.), *Mechanisms of Language Acquisition* (pp. 249-308). Hillsdale, NJ: Lawrence Erlbaum..

MacWhinney, B. (1998). Models of the emergence of language. *Annual Review of Psychology, 49*, 199-227.

MacWhinney, B., & Bates, E. (Ed.). (1989). *The Crosslinguistic Study of Sentence Processing*. New York: Cambridge University Press.

Marcus, G. (1998) Can connectionism save constructivism? *Cognition* 66, 2, 153-182.

Massaro, D. W. (1987). *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry*. Hillsdale, N. J.: Lawrence Erlbaum.

McLaren, I. P. L., Leevers, H. J., & Mackintosh, N. J. (1989). An associative theory of the representation of stimuli: Applications to perceptual learning and latent inhibition. In R.

G. M. Morris (Eds.), *Parallel Distributed Processing: Implications for Psychology and Neurobiology* (pp. 102-130). Oxford: Oxford University Press.

McLaughlin, B. (1987). *Theories of Second Language Learning*. London: Edwin Arnold.

McLaughlin, B., & Harrington, M. (1989). Second language acquisition. *Annual Review of Applied linguistics, 10*, 122-134.

Mitchell, D. C., Cuetos, F., & Corley, M. M. B. B. (1995). Exposure-based models of human parsing: Evidence for the use of coarse-grained (nonlexical) statistical records. *Journal of Psycholinguistic Research, 24*, 469-488.

Pinker, S. (1991). Rules of language. *Science, 253*, 530-536.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition, 28*, 73-193.

Pinker, S., & Prince, A. (1994). Regular and irregular morphology and the psychological status of rules. In S. D. Lima, R. L. Corrigan, & G. K. Iverson (Eds.), *The Reality of Linguistic Rules* (pp. 321-351). Philadelphia, PA: John Benjamins.

Plunkett, K. (1998). Connectionism and language learning. *Language and Cognitive Processes, 13*(2-3), 97-104.

Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered preceptron: Implications for child language acquisition. *Cognition, 38*, 43-108.

Plunkett, K., & Marchman, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition, 48*, 21-69.

Pulvermüller, F., & Schumann, J. H. (1994). Neurobiological mechanisms of language acquisition. *Language Learning, 44*, 681-734.

Redington, M., & Chater, N. (1998). Connectionist and statistical approaches to language acquisition: A distributional perspective. *Language and Cognitive Processes, 13*(2/3), 97-424.

Regier, T. (1995). A model of the human capacity for categorizing spatial relations. *Cognitive Linguistics, 6*, 63-88.

Rounds, P. L. & Kanagy, R. (1998). Acquiring linguistic cues to identifying AGENT: Evidence from children using Japanese as a second language. *Studies in Second Language Acquisition, 20, 4 509-542.*

Rumelhart, D. E. (1989). The architecture of mind: A connectionist approach. In K. Posner (Eds.), *Foundations of Cognitive Science* (pp. 133-159). Cambridge, MA: MIT Press.

Rumelhart, D. E., & McClelland, J. L. (1986b). On learning the past tense of English verbs. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel Distributed processing: Explorations in the Microstructure of Cognition. Psychological and Biological Models* (pp. 216-271). Cambridge MA: MIT Press.

Sasaki, Y. (1991). English and Japanese interlanguage comprehension strategies: An analysis based on the competition model. *Applied Psycholinguistics, 12*, 47 - 73.

Sasaki, Y. (1994). Paths of processing strategy transfers in learning Japanese and English as foreign languages: A Competition Model approach. *Studies in Second Language Acquisition, 16*(1), 43-72.

Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research, 12*(1), 40-72.

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition, 14*(4), 357-385.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed developmental model of visual word recognition and naming. *Psychological Review, 96*, 523-568.

Shallice, T., Glasspool, D. W., & Houghton, G. (1995). Can neuropsychological evidence inform connectionist modeling? Analyses of spelling. *Language and Cognitive Processes, 10*, 195-225.

Shanks, D. (1995). *The Psychology of Associative Learning*. Cambridge: Cambridge University Press.

Sharkey, N. E. (1996). Fundamental issues in connectionist processing. In G. Brown, K. Malmkjær, & J. Williams (Eds.), *Performance and Competence in Second Language Acquisition* . Cambridge: Cambridge University Press.

Shastri, L., & Ajjanagadde, V. (1993). From simple associations to systematic reasoning: a connectionist representation of rules, variables, and dynamic bindings using temporal synchrony. *Behavioral and Brain Sciences, 16*, 417-496.

Sokolik, M. E., & Smith, M. E. (1992). Assignment of gender to French nouns in primary and secondary language: A connectionist model. *Second Language Research, 8*, 39-58.

Stillings, N. A., Weisler, S. E., Chase, C. H., Feinstein, M. H., Garfield, J. L., & Rissland, E. L. (1995). *Cognitive Science. An Introduction*. Cambridge, MA: MIT Press.

Svenconis, D. J. & Kerst, S. (1994) Investigating the teaching of second-language vocabulary through semantic mapping in a hypertext environment. *CALICO Journal 12*, 2&3, 33 - 57.

Tinkham, T. (1997). The effect of semantic and thematic clustering on the learning of second language vocabulary. *Second Language Research, 13*, 2 138-163.

Tomlin, R. S. (1990). Functionalism in second language acquisition. *Studies in Second Language Acquisition, 12*, 155-177.

Tomlin, R., & Gernsbacher, M. (1994). Cognitive foundations of SLA: Introduction. *Studies in Second Language Acquisition, 16*(2).

Van Gelder, T. (1998). Cognitive Architecture: What choice do we have? In Pylyshyn, Z. (Ed.), *Constraining Cognitive Theories*, (pp191-204). Stamford, Ct.: Ablex Publishing.

White, L. (1996). Universal grammar in SLA. In R. W. R. & K. Bhatia (Eds.), *Handbook of Second Language Acquisition*. San Diego: Academic Press.