UC Riverside
Issues in Applied Linguistics

Title
Conditions on Transfer: A Connectionist Approach Yasuhiro Shirai

Permalink
https://escholarship.org/uc/item/1257n39j

Journal
Issues in Applied Linguistics, 3(1)

ISSN
1050-4273

Author
Shirai, Yasuhiro

Publication Date
1992-06-30

DOI
10.5070/L431005153

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Conditions on Transfer: A Connectionist Approach

Yasuhiro Shirai
Daito Bunka University

The purpose of this paper is twofold: (1) to comprehensively discuss conditions under which L1 transfer tends to occur, and (2) to explain these conditions in terms of the connectionist framework of second language representation, processing, and acquisition, primarily relying on the localized connectionist model (CLM = Connectionist Lexical Memory) of Gasser (1988). The conditions identified are: (1) interlingual mapping, (2) markedness, (3) language distance, (4) learner characteristics, (5) cognitive load, and (6) sociolinguistic context. It is argued that the connectionist framework explains L1 transfer effectively and that the interaction of these factors determines the degree of L1 transfer in interlanguage.

INTRODUCTION

The field of second language acquisition (SLA) has seen many changes in research trends in recent decades. This is particularly true with regard to the role of the first language (L1) in SLA. Contrastive Analysis (CA), which was the center of attention in applied linguistics in the 1950's and early 1960's, was neglected during the 70's, as research interest shifted from L1 transfer to universal aspects (e.g., the "natural order" of morpheme acquisition [Krashen, 1978]) in SLA. Since the 1980's, the pendulum seems to have swung back, and the present goals of research into L1 influence on SLA are "to state more precisely the conditions under which interference took place and the type of L1 knowledge that was utilized" (Ellis, 1985, p. 33; see also Felix, 1980 and Hatch, 1981 for similar points). This paper addresses this very problem--the conditions on transfer[1]--in a more comprehensive way. In so doing, I will attempt to apply a connectionist framework (Gasser, 1988, 1990) since it provides a new perspective on language transfer, posing new questions and reviving some old questions about L1

Issues in Applied Linguistics ISSN 1050-4273
© Regents of the University of California
Vol. 3 No. 1 1992 91-120

transfer. Connectionism is a radical departure away from the existing paradigm in cognitive science and is quickly gaining ground (e.g., MacWhinney, 1989). As Gasser (1990) states, the connectionist approach "has already reshaped the way many cognitive scientists think about mental representations, processing, and learning" (p. 179). Thus, if SLA is to be part of cognitive science, it should at least try to learn from the insights of that field and perhaps start research projects within a connectionist framework.

Paradigm Shift

Behind the criticism of the contrastive analysis hypothesis was, of course, a paradigm shift in the fields of linguistics and psychology; this shift became known as the "Chomskyan revolution" (Newmeyer, 1986). Chomsky's review of *Verbal Behavior* by B. F. Skinner (Chomsky, 1959) and his own *Syntactic Structures* (Chomsky, 1957) changed cognitive scientists' (psychologists' and linguists') view of human behavior, including language. As a result, structural linguistics, the audio-lingual approach, and the contrastive analysis hypothesis, all closely aligned with behaviorism and associationism, lost ground, and the universalist approach to L2 acquisition became predominant, a development which resulted in declining interest in language transfer.

Connectionism, on which this paper is based, is currently thought to have the potential to cause a new paradigm shift (e.g., Sampson, 1987; Schneider, 1987; Clark, 1989) in the field of cognitive science, which includes linguistics, psychology, neuroscience, philosophy, and artificial intelligence (AI). For decades, cognitive scientists have tried to understand the mind at a symbolic level, assuming that our cognitive processes involve using rules to manipulate symbols (Allman, 1989; Bates & Elman, 1992). Formal linguistics is a good example of such an approach. However, although much research has been done within the symbolic paradigm, its limitation has become apparent. Digital computers, which simulate human behavior using symbol manipulation, for instance, cannot cope with certain tasks that a child can easily handle, including natural language processing, even though they show human-like ability in other kinds of tasks, such as computation and chess. Connectionists, who propose a radical shift in the approach to human cognition, do not believe in symbolic rule systems; they hold that symbolic rule systems are actually the

manifestation of patterns of neural network activation and, as such, are *emergent*, that is, behavior which appears to be rule-driven has actually resulted from certain patterns of neural activation. The connectionist model has been simulated on computer programs, which have been able to learn some aspects of language.

Critics of connectionism (e.g., Fodor & Pylyshyn, 1988; Pinker & Prince, 1988)[2] say that it is only behaviorism in a new guise because connectionist learning is attained in a manner similar to the "stimulus-response" type of learning which characterized behaviorism. Although there seems to be some truth to this charge, the fundamental difference between connectionism and behaviorism is that the former does not disregard what is "inside the black box," i.e., what goes on between input (stimulus) and output (response). Indeed, connectionists try to simulate what goes on inside the black box by using computers, basing these simulations on a particular model of how the human brain works: that neural networks are connected with different weights and activated in response to stimuli and that this activation is massively parallel instead of serial. In other words, connectionism is concerned with the *internal representation* of knowledge and the architecture that supports it, whereas behaviorism is not.

If there is a similarity between connectionism and behaviorism, it is that connectionism is particularly relevant to the phenomenon of L1 transfer (Gasser, 1990). Therefore, as this paper will argue, the connectionist approach may provide new and more sophisticated interpretations of language transfer as well as new insights into the role of contrastive analysis in predicting language transfer.

One might wonder why it is necessary to discuss transfer in SLA using a connectionist model when it also can be explained by existing approaches. The advantages of the connectionist approach to existing approaches in SLA research are as follows:

> 1. Connectionism is concerned with what is going on inside the black box. Most of the present approaches in SLA try to explain "behaviors" by looking at input and output without much concern about what is going on inside the brain. Connectionism has "architecture"--a conception of the cognitve hardware that causes observable behavior. Moreover, to some extent, the way connectionist learning operates is constrained by neurobiological reality.[3]

2. Connectionism is a general cognitive model that handles human behavior in general, not only second language acquisition phenomena. Limited domain theories which have been proposed in SLA are important, but they should be reinterpreted within a more integrated framework (Hatch, Shirai & Fantuzzi, 1990) that would make research in SLA more relevant to general issues of cognitive science, thus creating a way for SLA to contribute to cognitive science in general.

3. As was stated earlier, connectionism is very different from the existing research paradigm (the symbolic approach) in cognitive science. SLA, which has also followed the traditional symbolic approach, should benefit from a connectionist approach, as other areas of cognitive science already have.

4. By applying the connectionist framework, it becomes possible to test certain hypotheses by using computer simulation, an important tool which has been used in cognitive science for understanding human cognition.

The Connectionist Framework

Connectionism can roughly be categorized into two types: localized connectionism and distributed connectionism. Localized connectionism (e.g., Cottrell & Small, 1983; Gasser, 1988; Waltz & Pollack, 1985) uses particular units to represent concepts, such as TREE, PUT, SUBJECT, etc. In this sense, it is similar to the symbolic approach; however, it is still radically different because it does not assume overt rule systems. Distributed approaches, often referred to as PDP models (e.g., McClelland, Rumelhart, & the PDP Research Group, 1986; Rumelhart, McClelland, & the PDP Research Group, 1986; Gasser, 1990; Sokolik & Smith, 1992; Henning & Roitblat, 1991), on the other hand, use a system of representation that is "distributed over many units, and each unit participates in the representation of many concepts" (Gasser, 1990, p. 181). For example, in Gasser (1990), concepts such as AGENT,

MARY, or SLEEP are represented over many units as patterns of activation.

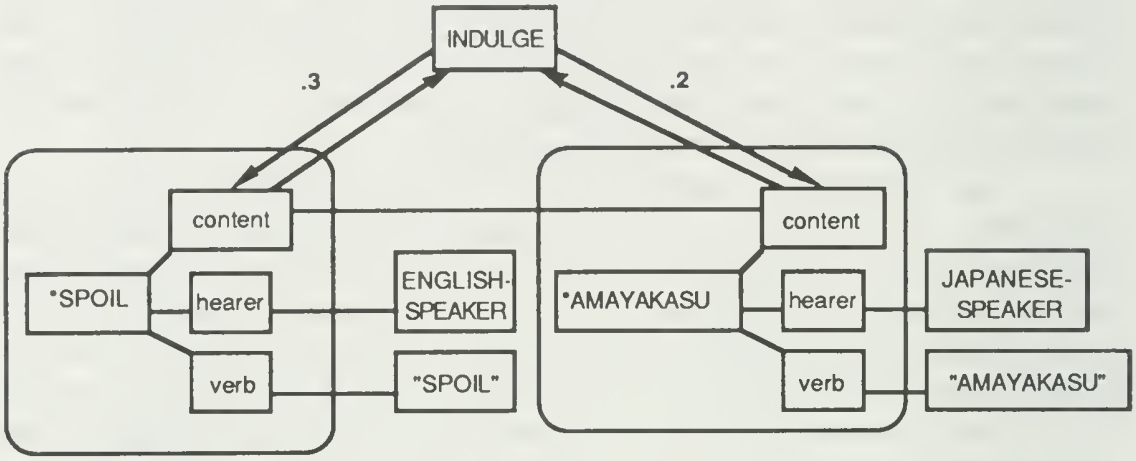
In this paper, a localized connectionist approach is used to illustrate the mechanism of transfer, simply because it is much easier to use a localized model to explain how connectionist models work in relation to language transfer. No claims are being made about the superiority of localized connectionist models vis-a-vis distributed models in terms of their neural plausibility.[4]

As a cognitive model, connectionism must address three problems: representation (the way knowledge is represented in the brain), process[5] (the way knowledge is activated/fired for use), and learning. Linguistic knowledge in localized connectionist models is represented over many nodes and in the strength of connection/association between nodes. For example, the words "bread" and "butter," both represented on different nodes, may have a strong connection because of their strong linguistic/conceptual association. In bilingual lexical knowledge, an L1 word is assumed to be strongly connected with its L2 equivalent (this point will be discussed in greater detail momentarily). In short, different kinds of knowledge (i.e., lexical, morphological, syntactic, pragmatic, and world knowledge) are represented over many nodes and are organized by means of connection weights.[6]

Processing in a connectionist model is a particular pattern of activation and firing of nodes. Each node has a threshold level (as does a neuron), and if its activation level reaches threshold, it fires. Moreover, if there is competition between nodes, the node that receives the greatest activation is the one which fires. For example, when a speaker asks, "Are you coming to the party tomorrow?" there should be competition between COME and GO, since they share very many features in common, and therefore are both activated. The only reason COME would be chosen (i.e., fired) is deictic information, e.g., whether the speaker is going to be at the party or not. It is therefore the greater activation from nodes relevant to this deictic information which makes the COME node fire the word "come" over "go."

An example from Gasser (1988) illustrates how such competition between nodes works. Figure 1 is a schematic depiction of an English-Japanese bilingual speaker's representation of what is cognitively involved in talking about "spoiling" someone:

Figure 1. Nodes and Connections in a Bilingual Context*



*From Gasser (1988, Figure 1.9, p. 16). Gasser's title for this figure was "Competing GUs in a Multilingual Context."

As can be seen in Figure 1, the nodes, shown as squares, represent semantic content, linguistic forms, hearer roles, and so forth. Here, the content of INDULGE is shown as being more strongly associated with English knowledge than with Japanese knowledge and therefore has a stronger connection weight to English (.3) than to Japanese (.2). Other things being equal, this bilingual speaker would thus say "spoil" instead of "*amayakasu*," the Japanese equivalent. In the real world, however, the presence of a monolingual Japanese interlocutor would activate a hearer node in the Japanese speaker, resulting in "amayakasu" being uttered by the speaker, while the presence of monolingual English interlocutor would activate a hearer node in the English speaker, resulting in the speaker uttering "spoil." If many bilingual speakers were present, both the Japanese and English hearer nodes would be activated in the speaker, perhaps resulting in language mixing, a situation which Gasser (1988) reports.

According to the connectionist model, learning is considered to be the result of this kind of processing. Essentially, the more often particular nodes at the ends of connections are activated and/or fired, the stronger the connections become; consequently, stronger connections become more easily activated, and this greater ease of activation constitutes learning. In the case represented by Figure 1,

if a Japanese-speaking learner of English keeps saying "spoil" instead of "*amayakasu*," the association will be stronger, and it will be easier for him to say the word when speaking in English. This process would constitute learning of the word "spoil" for this learner. To summarize, "representation" in the connectionist framework is equivalent to connection weights between nodes, "processing" refers to patterns of activation and firing at a given point in time, and "learning" refers to the change of weights (or strengths of connections) over time.

It was pointed out earlier that a connectionist approach to language acquisition is effective in dealing with language transfer. As Gasser (1990) states:

> Transfer is precisely what connectionist models are good at. Once a network has learned an association of a pattern P1 with pattern P2, when it is presented with a new pattern P3, this will tend to activate a pattern that is similar to P2 just to the extent that P3 is similar to P1. Thus the connectionist framework provides an excellent means of testing various notions about the operation of transfer in SLA . . . . The claim then would be that overlap of any type between L1 and L2 should be the basis of transfer. (pp. 189-190)

In a nutshell, because in the processing of information the existing representation is always activated, the L1 knowledge representation (i.e., the existing representation) always reshapes incoming L2 information, thus affecting subsequent learning of new L2 patterns.

The challenge for connectionists is how to represent L1 and L2 in a realistic way, and it should be admitted that so far the model has been limited to a small number of sentence patterns (Gasser, 1988). However, the model produces very interesting results and can even produce speech errors--those made by native speakers as well as L1 transfer errors often made by L2 learners. The following sections attempt to explain the mechanism of language transfer using the connectionist framework.[7]

Conditions on Transfer

In previous SLA research, the following factors (although the list may not necessarily be exhaustive) have been proposed as governing when L1 transfer tends or tends not to occur:

(1) Interlingual mapping
(2) Markedness
(3) Language distance (perceived/real)
(4) Learner characteristics
(5) Cognitive load
(6) Sociolinguistic context

Any one of these factors is of course by no means an absolute determiner of the presence or absence of language transfer. Instead, these factors are supposed to interact with one another, and the particular interactions are thought to affect the degree to which the transfer of L1 is observed in L2 processing/learning. I will discuss each of these factors below in greater detail.

(1) Interlingual Mapping

"Interlingual mapping" refers to linguistic mapping between L2 and L1. Any linguistic item of L1 has a particular mapping on L2. The basic claim which has been made is that the easier or more straightforward a particular mapping is, the more likely it is that transfer will occur. Tanaka & Abe (1989) state that "it can be hypothesized that language transfer is likely to occur in the linguistic areas where it is *easy to find interlingual equivalents*" (p.79) (emphasis mine).

Previous research has identified to some degree the linguistic levels where transfer tends or tends not to occur. Tanaka & Abe (1989) summarize these findings and state that language transfer is almost non-existent when learning morphemes,[8] not very strong in the learning of syntax (or becomes weaker as the learner's proficiency increases), and is strong in the area of semantics (lexicon), pronunciation, and discourse.[9] They further state that in the area of morphology, for example, since it is difficult for learners to find interlingual equivalents, transfer is weak in this area.

lexico-semantic transfer

Since Tanaka & Abe (1985) claim that transfer is persistent in semantics, let us first discuss this claim using the connectionist model of language processing/acquisition. Gasser (1988) provides an example of negative transfer produced by a Japanese learner of English. The learner keeps saying "water" when she is supposed to say "cold water." The reason for this persistent error is assumed by Gasser to be L1 transfer because in Japanese the concepts HOT

WATER and COLD WATER are expressed using separate words--"*yu*" and "*mizu*," respectively. In this case, she consistently makes the same error, although she knows it is an error in English. I would argue that from a connectionist point of view there are two possibilities for explaining this phenomenon. One is that this person first activates and fires the node for the Japanese word for "*mizu*," which then activates and fires the English equivalent "water" with which it is strongly associated. Another possibility is that her representation of the concept NOT-HOT-WATER is strongly associated with the word "water" and it is fired without any activation from Japanese word "*mizu*." Gasser (1988) argues for the latter possibility, but it is not really clear which is correct, and this uncertainty continues to be an issue in bilingual lexical representation (e.g., Kroll & Curley, 1988). Nevertheless, what is important is that in both cases a strong association between the L1 word and its L2 equivalent exists. In the case of the first possibility there is still a strong association between the two; in the second, there used to be a strong association between the two, which has resulted in an incorrect pattern of association (i.e., NOT-HOT-WATER = "water"), even though the Japanese word "*mizu*" is no longer fired.

The implication of the above example in general connectionist terms is this: when learners find an L1-L2 equivalent (either consciously or unconsciously), they tend to use it in production and comprehension.[10] In the case of lexical semantics, it is very easy to find interlingual equivalents (i.e., straightforward mapping), and in limited situations such a strategy works and makes the connection/association stronger, further resulting in an increase in the firing of the same nodes. Until negative evidence comes up (or even after that), this process is repeated and the connection becomes very solid. In previous research, it has also been claimed that lexico-semantic transfer is strong (e.g., Tanaka & Abe, 1985; Ijaz, 1986; Shirai, 1989), and this claim makes sense in view of the above explanation of transfer. It is not surprising, therefore, that lexico-semantic transfer is "pervasive and persistent."

morphology

The area of morphology, especially bound morphology, has been assumed to be relatively free from L1 constraints (e.g., Dulay & Burt, 1973; Krashen, 1981). However, the principle of "interlingual mapping" also holds in morphology. Indeed, although a "natural" order of morpheme acquisition has been claimed, there

still is cross-lingusitic influence. Andersen (1983), citing Cancino (1979), exemplified the difference in the acquisition orders of an L1 Spanish learner of English and an L1 Japanese learner of English, a difference which Andersen attributed to L1 transfer. For example, Japanese has a form similar to English for possessive (e.g., John *no* (possessive marker) pen = John's pen), but no plural markers, while Spanish exhibits the opposite situation. This contrast facilitated the Japanese learner's acquisition of the English possessive marker relative to the Spanish learner's. Sasaki (1987) makes a similar claim, pointing out that five Japanese-speaking learners (Hakuta, 1974, three subjects from Koike, 1981, and her own subject) of English in naturalistic settings all acquired the possessive *-s* earlier than in the proposed "natural order." Sasaki further states that the five Japanese subjects showed a high correlation among their own acquisition orders but a low correlation with the "natural order." Sasaki thus criticizes the Natural Order Hypothesis (e.g., Krashen, 1981), which treated Hakuta's Japanese subject (Uguisu), a counter-example to the natural order, as an idiosyncratic variation. Lightbown (1983) also reports a case of a group of French learners of English whose different accuracy order she attributes to the influence of L1.

Clearly, therefore, even though a natural order has been claimed in the literature based on correlational or implicational studies, we still need to resort to an L1-transfer explanation even in the area of morpheme acquisition. Further, the pattern of development can be explained by the principle of "interlingual mapping," which claims that if L1-L2 mapping is simple, transfer tends to occur. In other words, when L1 and L2 are similar in structure, not only positive transfer (e.g., the early acquisition of possessive *-s*) but also negative transfer (e.g., the overuse of possessive *-s*) may occur.

Hatch (1983) suggests that the acquisition of morphemes can be explained by "naturalness" factors such as perceptual saliency, frequency, and invariance of forms, as well as by the "L1" factor. In connectionist terms, such a claim can be interpreted as follows: the naturalness factor makes it easy for a particular form to be connected to a particular meaning/function. It will be easy to identify and easy to match; there will be many opportunities to strengthen connections. This results in the Natural Order. In addition, L1 factors facilitate the recognition and connection formation between an L2 form and its meaning/function when the learner has equivalent (similar) form/function relationships in his or her L1 (e.g., possessive in Japanese).

pragmatic/discourse transfer

As claimed by Tanaka & Abe (1989), it seems that transfer is also apparent in the area of pragmatics/discourse. Examples are abundant in the SLA literature (e.g., Richards, 1971; Scarcella, 1979, 1983; Schmidt & Richards, 1980; Blum-Kulka, 1982; Olshtain, 1983; Beebe, Takahashi, & Uliss-Weltz, 1990). Though the entire area of discourse/pragmatic transfer is too broad to be given proper treatment here, it is important to consider it briefly in two subsections--micro-level and macro-level transfer--because although both these categories are viewed as discourse/pragmatic transfer in SLA, they are different in terms of processing, i.e., in terms of the mechanism by which each takes place.

micro-level transfer: Micro-level transfer is surface-level transfer of L1 forms, which typically results in negative transfer in the realm of pragmatics/discourse. Kato (1989), for instance, notes that many English speakers who learn Japanese as an L2 use a direct translation of "Do you want to . . . ?" for invitations/suggestions, though this kind of construction is inappropriate as an invitation in Japanese. The function of "Do you want to . . . " in Japanese is restricted to purely literal inquiries, unlike in English. Similarly, as cited in Odlin (1989), Kasper's (1981) example of a German L1 speaker using "should" in cases where native speakers of English would use more indirect forms is also a case in point. In both of these examples, the result (or product) is an unintended illocutionary force. This kind of surface-level transfer of L1 forms can of course occur at the lexical, phrasal, or sentential level. In terms of processing, the cases of micro-level pragmatic transfer are not at the pragmatic/discourse level, although they are usually treated as such in SLA literature because the product results in a pragmatic error. In fact, the process (the firing of L1 nodes) which results in this kind of transfer is almost identical to L1 lexical transfer, as discussed above.

macro-level transfer: Macro-level transfer, in contrast, is truly pragmatic/discourse transfer in that the learner activates his or her L1 knowledge about pragmatics/discourse as part of "world knowledge." In other words, unlike micro-level transfer, what is transferred in macro-level transfer is not an L1 form, but L1 knowledge about discourse. For example, it has often been pointed out that L1 rhetorical structure is transferred to L2 and can result in either negative or positive transfer in comprehension and production

(e.g., Kaplan, 1966; Maccoun, 1983; Oi, 1986; Connor & Kaplan, 1987). As Hatch (1989) notes, in L2 comprehension/production we must rely on "world knowledge" even though we do not know which parts of world knowledge are universal and which are culture-specific. Therefore, in using L2, we activate relevant world knowledge when we perform in contexts similar to those for L1, and knowledge about L1 rhetorical structure is part of our world knowledge.

The same type of "world knowledge" transfer can occur in conversational interaction at the level of the speech event/act. For example, "How's it going?" as a casual greeting in American English is problematic for many ESL learners because, depending on the situation, it is not necessary to actually answer the question. In other languages, however, it may be the case that you need to give a verbal response (Schmidt & Richards, 1980). This difference sometimes results in international students feeling negatively about American people, whom they are perceive as "cold" or "superficial" because of failing to give a response in such a conversational environment.

Gasser's (1988) CLM includes three kinds of knowledge necessary for attaining a speaker's goal: (1) context-specific conventions (e.g., "I'll have . . . " in a restaurant context), (2) cross-contextual conventions (e.g., politeness), and (3) general knowledge about illocutionary acts, (e.g., "let the hearer know what is desired but convey pessimism about the chance of success" (p. 46) when making a polite request). Although knowledge about rhetorical structure has not yet been represented in connectionist models (though there are some symbolic AI models which do look at rhetorical structure), it seems possible that rhetorical knowledge could be incorporated as part of (3), general knowledge about illocutionary acts. In the CLM, then, the knowledge thus represented would be activated when a speaker or writer tries to attain a goal; thus, in L2 performance, it would be natural for L2 learners to activate the same L1 knowledge structure as they would when trying to attain the same goal. Moreover, knowledge types (2) and (3) are probably more likely to be activated because they appear to be more universal than (1).

(2) Markedness

In SLA literature, markedness[11] is assumed to be a strong constraint on transfer. The basic assumption is that when there are two structures which differ in markedness, the less marked (or

unmarked) one is more likely to be transferred (e.g., Gass, 1981). Although there can be problems in determining what a marked form is, the principle seems to hold true for the most part. For our purposes, we must consider markedness in the learner's L1 as well as in the L2. Other things being equal, we can assume that what is unmarked in L1 and what is unmarked in L2 are more transferable than what is marked, since Kellerman (1977, 1978, 1979) and Jordens (1977) have established that what is unmarked in L1 is more transferable to L2 than what is marked, in the area of semantic transfer. Moreover, Zobl (1983, 1984) supports this notion in the area of grammar (but see Eckman, 1977 for a different view in the domain of phonology).

Although connectionism is often regarded as anti-nativist, it is not true that innateness has no place in connectionism. As Bates & Elman (1992) state, "there is no logical incompatibility between connectionism and nativism" (p. 17). Therefore, the problem of markedness in SLA can be explained by assuming some type of pre-wiring in the network (Gasser, 1990). Indeed one possibility is that if we suppose there are innate predispositions for humans in language acquisition (e.g., a universal acquisition order for some structure), it could be assumed that there are innate patterns of association between connections which are difficult to alter by subsequent learning. (It must be noted, however, that in most cases connectionists do not assume an innate component to explain the acquisition of knowledge, including language.)

Another possibility is that some markedness conditions are not innate (e.g., Kellerman's markedness in lexical semantics) but learned. The scenario in this case would be that since unmarked items are usually more frequent or perceptually salient (e.g., Ferguson, 1984), connections for these items would be made very easily. Such connections acquired after birth may also become too strong to alter later in life. This possibility is, of course, more congruent with connectionism, which emphasizes learning by data available in the input/environment. Although the issue of markedness and transferability is still not resolved (Ellis, 1985), it appears that the connectionist framework is flexible enough to handle it.

(3) Language Distance (Perceived/Real)

Kellerman (1977, 1978, 1979) suggests that transfer is more likely to occur when L1 is typologically similar to L2. What is important for Kellerman is not the actual similarity or difference

between two languages but the perceived distance. For example, even though Japanese is typologically very similar to Korean, if a learner is totally unaware of the similarity between the two languages, s/he may not use transfer readily. This perceived distance is called "psychotypology" by Kellerman (1983).

In relation to this notion, Gasser (1990) used a distributed connectionist network to simulate certain aspects of second language acquisition. One result was that when L1 words were similar to L2 words, there was more transfer in word order, or, in other words, when L1 words were similar to L2 words (e.g., cognates), the word order of L1 tended to be transferred to that of L2. As Gasser points out, the result is interesting because it supports the view of language distance as a constraint on transferability. However, it also goes against the notion of "perceived" distance, since a computer simulation program does not have any intention or thought and therefore could not believe, or psychologically perceive, L1 as being different from L2.

These seemingly conflicting observations could be interpreted as follows: first, it is probably the case that regardless of the "perceived" distance between L1 and L2, if actual distance is close, transfer is more likely to occur. This would be possible because, if L1 and L2 are typologically similar, the use of prior knowledge (in connectionist terms, the existing strength of associations between nodes) tends to work well which in turn facilitates the learner's use of the same learning mechanism (prior strength will be further strengthened, with new connections between L1 and L2 nodes added on). Therefore, even without any perception/belief about language distance, transfer can be facilitated.

Second, it is plausible that "perceived language distance" may also have an effect on transferability. Knowledge about language distance can be either "meta-knowledge" about the distance between languages (i.e., conscious knowledge learned from some source or knowledge generalized through the learner's own struggle with the L2) or "unconscious knowledge" (i.e., typically, the kind of knowledge held by very young children acquiring two languages at the same time).[12]

Precisely how the "perception" of language distance would influence connectivity must yet be resolved. In any event, language distance (whether perceived or real) is an arguably important condition on language transfer.

(4) Learner characteristics

learning environment

It has been suggested in the literature that L1 transfer is strong when learners learn another language in acquisition/input-poor environments, i.e., in environments where the chances for naturalistic communication are limited, as is typically the case in foreign language (as opposed to second language) situations (Krashen 1978, 1981; Ervin-Tripp, 1974; McLaughlin, 1978), and this observation is easily explained in connectionist terms. When L2 learners learn the L2 in an acquisition-poor setting, they tend to use a "grammar translation approach," which necessitates that the learner "connect" L1 to L2. Hence, transfer errors, of the kind discussed in the section on lexical transfer, would be predicted. However, Tanaka & Abe (1985) revealed that Japanese learners learning English in English-speaking environments showed the same "rate" of lexical transfer as did classroom EFL learners, a finding which seemingly goes against the earlier notion that classroom EFL learning would tend to increase transfer. Nevertheless, Shirai (1990) has explained this surprising observation by the connectionist model: learners in ESL contexts need to communicate in the L2 and must initiate utterances by relying on their L1, to some degree, especially at early stages, which means they have to make L1-L2 connections between lexical concepts. As a result of this process, L1-L2 connections may become stronger and harder to eliminate later, and this might offset any advantages such learners have over EFL learners, who rely on grammar-translation but do not really have to communicate in the L2. This interpretation suggests a new hypothesis regarding a condition of transfer: when a necessity for production exists, there is a higher likelihood of creating L1-L2 connections, thus increasing transfer.

level of proficiency

It has been suggested in the literature that L2 learners rely on L1 knowledge when their L2 knowledge is not sufficient. This idea has been empirically supported by Taylor (1975) and Dommergues & Lane (1976) in the area of syntax. Major's (1986, 1987) Ontogeny Model also supports this notion in the area of phonology. The claim that L1 transfer is strong at the early stages of L2 development is of course not surprising. When there is a gap in what a learner can say and what s/he wants to say, it is necessary to

rely on whatever knowledge is available. When L2 syntax is not fully developed, a learner has to fall back on L1 syntax, for which the neural network is already fully developed.

This argument recalls the notion of "relexification" as being one of the characteristics of early L2 acquisition (e.g., Schumann, 1982). In this view, the learner, not knowing enough L2 syntax, substitutes L2 lexical items for L1 lexical items, using the same word order as in L1. Krashen (1981) also claims that learners use an L1 + Monitor mode when the knowledge of L2 is not sufficient.

In connectionist terms, L1 transfer at lower levels of L2 proficiency can be explained as follows: word order is determined in two ways in Gasser's (1988) CLM--one is by the lexicon (lexical items have syntactic information, too); the other is by the sequencing component. In early stages of L2 development, syntactic information has to be created from the facts of L2 input. However, while L2 syntactic knowledge is still meager, the learner has to utter something. Therefore, s/he must activate his or her readily available L1 word order knowledge to produce L2 utterances.

age

SLA literature has shown that adult learners tend to show more L1 transfer than younger learners, at least if a learner's performance is compared to the ultimate attainment achieved by adult and child learners. One area where L1 transfer has been shown to be especially evident for adults is phonology (e.g., Oyama, 1976). As the critical period hypothesis claims, it appears to be difficult to acquire native-like L2 pronunciation after puberty, and this difficulty may be due to neurobiological factors (e.g., brain plasticity, lateralization), to psychomotor factors (e.g., dexterity of speech muscles), or to affective factors (e.g., language ego, attitude) (Brown, 1987). Kennedy (1988), who has proposed an information-processing approach to explain child-adult differences in SLA, claims that once automatic knowledge in L1 is created (a process called "unitization"), it is difficult to alter it later in life. Although at this stage we do not know the precise reason for this difference, it does seem that adults are at least more prone to transfer than children.

Munro (1986) provides a connectionist explanation for a general critical/sensitive period in learning. His point is that a critical period can be accounted for by a "reduction in the modifiability of neuronal response characteristics as their synaptic connections are strengthened as a result of experience" (pp. 471-

472). In other words, experience would create a pattern of connections that would be difficult to modify or alter later in life. This reduction of modifiability corresponds to Kennedy's notion of unitization. When an L1 connection is formed and solidified as a system, it may indeed be the case that subsequent alteration of certain connections is difficult.

(5) Cognitive Load

attention and monitoring

It seems to be a common observation that when L2 learners speak in situations with high cognitive loads, it is difficult for them to maintain accuracy. For example, when I participate in English in a discussion at a graduate-level seminar, I find myself making many grammatical, morphological, and phonological mistakes. Specifically, I probably do not attend to the distinction between /l/ and /r/, which are allophones of a phoneme in my L1, and additionally make many errors in articles and 3rd person singular *-s*, even though I know all these rules and know how to produce correct forms. It has also been observed that, in interviews, L2 learners' accuracy and fluency drastically decline when topics become difficult (e.g., description of movies involving unfamiliar objects).

This inability to attend to accuracy in situations of high cognitive load can be explained within an information-processing framework (e.g., McLaughlin, 1978, 1987; McLaughlin, Rossman, & McLeod, 1983; Færch & Kasper, 1986; Bialystok, 1988; Kennedy, 1988). Such an argument maintains that in any kind of skill learning, we tend to automatize our knowledge. When a skill is automatized (i.e., overlearned), it is no longer necessary to pay attention for the behavior to be successful. When we must pay attention to our behavior, before the knowledge or skill is automatized, we must use our working (short-term) memory, which has a limited capacity. Therefore, when L2 learners confront tasks with a high cognitive load (e.g., discussions of complex matters, descriptions of unfamiliar objects), their attention is diverted and their accuracy of L2 production breaks down.

When it is difficult to pay attention to their behavior, L2 learners tend to rely on automatized knowledge, which does not require any space in working memory. The relevant automatized knowledge available for L2 learners is interlanguage knowledge which has become automatic and L1 knowledge. Therefore, when a learner's automatized interlanguage knowledge is unstable or

insufficient for communication needs, s/he tends to fall back on L1 knowledge, which is already automatic and readily available. Hence, the conclusion which has been documented in SLA research: when cognitive load is high, L1 transfer tends to occur. Now, how might this concept of cognitive load relate to connectionism?

conscious knowledge and connectionism

It must be admitted that conscious knowledge seems to be a problematic area both for Gasser's (1988) CLM and for connectionism in general. In the CLM, conscious knowledge, or what Krashen calls the Monitor, is not represented. The problem is that if conscious knowledge is to be represented and built into the system, as Gasser (1988) states, it "must operate on the output of generation as it is currently handled in the model. That is, they [i.e., rules of grammar] would have to check words that are about to be uttered" (p. 202).

This problem of incorporating conscious knowledge is not a trivial matter in modelling an L2 learner's knowledge. In the case of L1 knowledge, it may not be a serious problem because the knowledge is unifaceted; that is, the native speaker's knowledge of language can be accessed basically in an automatic fashion in comprehension/production. On the other hand, L2 learners (especially L2 learners in foreign language settings) have knowledge (e.g., 3rd person singular *-s*) which may not be readily accessible in a communicative setting.[13] Furthermore, even if a conscious knowledge component is added to the CLM, it may be difficult to specify how it will affect the change in unconscious knowledge (i.e., learning) represented in the network.

How, then, can connectionist models handle this problem? It may be possible, as Norman (1986) suggests, that a kind of conscious knowledge is closely mapped onto the network of unconscious knowledge, activating and inhibiting the connections between relevant nodes. Further, Clark (1989) claims that the connectionist system can handle such conscious rule-learning. However, the mechanism by which these might be done is not clear. This issue remains an area where future work is necessary.

Another important issue in relation to consciousness is how initial stages of language acquisition can be characterized. In L1 acquisition, conscious strategies may be minimal.[14] Gasser (1988) states that L1 acquisition "would begin with knowledge of concepts and some of the goals that are realized by linguistic means" (p. 198) and that, later, patterns of association would develop between

concepts and phrases (e.g., BATH TAKING and "take a bath"). Gasser's assumption, in line with Peters (1983), that "syntax would arise out of the learning of lexical patterns" is a view of how representation of language presumably develops in the early stages of L1 acquisition.

Adult L2 acquisition, on the other hand, may be more strongly guided by conscious strategies (Schmidt, 1990). Certainly in classroom settings, it is teaching which primarily develops a learner's conscious component. Yet even in naturalistic settings, a learner's observations will basically develop a conscious component. In both cases, the conscious knowledge developed will gradually be automatized, eventually becoming available for production and comprehension.

(6) Sociolinguistic Context

hearer-role

It is Speech Accommodation Theory (e.g., Beebe & Giles, 1984; Giles & Byrne, 1982) which explains style-shifting in interlanguage within a social psychological framework: that the speech of L2 learners tends to be similar to that of interlocutors when they identify with the interlocutor group (a process called "convergence"), but tends to be dissimilar when they do not identify with the interlocutor group ("divergence"). In Beebe's (1988) example of speech accommodation by Chinese-Thai children, when the children speak with a Chinese speaker, their L2 (Thai) reflected greater Chinese influence (i.e., L1 transfer). In my experience in foreign language classrooms in Japan, when, as is often the case, students intentionally speak English (especially in oral reading) in a heavily Japanese-accented English, this would be a case of convergence to their peer group and divergence from the target-language group. It thus seems that L1 transfer is stronger when the learner is "converging" to the L1 group, while it is weaker when the learner is "diverging" from the L1 group.

This influence of the hearer can certainly be implemented by Gasser's CLM connectionist model, which has "hearer-role" nodes. In Gaser's terms, when the learner is speaking with someone from the same culture, the hearer-role (represented as a node) is specified as such. Whether the learner likes it or not (i.e., is goal-driven or not), the hearer-role is activated, which leads to a spreading activation of the nodes connected to it. Thus, the model would be

able to show the kinds of adjustment which are called "accommodation."[15]

subject matter

Beebe (1988) also suggests that even when the hearer is the same, Chinese-Thai children had a greater L1 (Chinese) influence in their L2 (Thai) when they discussed Chinese holidays than when they discussed Thai holidays. This finding suggests that when a learner discusses something about his or her own country, the L1 schema may be strongly activated, thus facilitating L1 transfer. Such an observation recalls the results of research using word association tests, in which subjects showed different responses to L1 and L2 stimuli and produced responses which were often culturally-loaded. Ervin-Tripp (1964, 1967) found, for example, that Japanese-English bilinguals' responses to English stimulus words were similar to those of monolingual Americans, while their responses to Japanese cues were similar to those of monolingual Japanese. These findings point to the strong connection between L1 lexicon and L1 conceptual structure as well as between L2 lexicon and L2 conceptual structure. Consequently, if the L1 conceptual structure is activated, L2 performance will be influenced by L1.

The Contrastive Analysis Hypothesis Revisited

One of the weaknesses of the CA hypothesis was that its predictions were based only on the linguistic analysis of two languages. When the hypothesis was tested against actual learner data, it was not consistently supported (Briere, 1968, for phonology; Whitman & Jackson, 1972, for syntax). As Tanaka & Abe (1988) state, the difference-difficulty hypothesis may not hold at the level of individual linguistic items, though it may hold true at the level of typological differences between the two languages in question. Also, as Wardhaugh (1970) has claimed, a priori predictions of difficulty at linguistic levels seem problematic.

On the other hand, it should be pointed out that there is supporting evidence for the CA hypothesis. Recent studies (e.g., Tanaka, 1983; Takahashi, 1985) have tried to test a somewhat modified version of the hierarchy of difficulty based on the contrastive analysis hypothesis (e.g., Prator, 1967, as cited in Brown, 1987) in the area of lexico-semantics, and the results have given some support for the hierarchy. It has also been claimed that learners whose L1 has an article system (Master, 1987) or plural

morphemes (Young, 1990) learn the corresponding system more easily than learners whose L1 does not have these structures. We thus cannot deny the possibility that the CA hypothesis may still have some validity in a priori predictions of difficulty at the level of certain linguistic items, if other variables can be controlled.

In this connection, it should also be pointed out that in the present framework contrastive analysis is understood to be merely one of many factors influencing transfer. It is seen as interacting with all the other factors discussed in this paper, not operating entirely on its own. It is therefore not reasonable to expect too much from contrastive analysis; it may have some a priori predictive power in terms of transfer, but only on the condition of "other things being equal."

Another consideration with regard to the CA hypothesis is the general mechanism of transfer. One condition proposed in this paper is interlingual mapping, which predicts that straightforward mapping helps form connections, thus facilitating transfer. It was also argued that in the acquisition of morphemes, when L1 and L2 have similar structures, (positive) transfer helps early acquisition. This view recalls Oller & Ziahosseiny's (1970) study, in which learners whose native languages used Roman scripts made more spelling errors due to transfer than learners whose L1 did not use Roman alphabets. These findings also suggest that similarity leads to transfer, whether negative or positive (see also Weinreich, 1953; Wode, 1978; Long & Sato, 1984). Indeed, the "difference-difficulty hypothesis" of the CA hypothesis (Lado, 1957) may be reinterpreted in a more neutral way, because difference does not necessarily lead to transfer errors. Such a reinterpretation could be called a "similarity-transfer hypothesis" and would be totally congruent with a connectionist approach, since when a new pattern is encountered which is similar to another existing pattern in the learner's representation, the new pattern would activate the existing pattern.

The discussion in this section is merely a first attempt at reinterpreting the CA hypothesis in a more sensible way; obviously, further consideration of this problem is necessary. Indeed, SLA research, I believe, must address the issues regarding the CA hypothesis, which was "laid to rest" (Brown, 1987) without rigorous empirical testing because it lost its theoretical foundation. Today, with a new framework (connectionism), we ought to take another look at this important question of transfer.

CONCLUSION

In this paper, I have tried to specify conditions where L1 transfer can be understood to tend to occur, using a connectionist model of human information processing.[16] The paper tries to be comprehensive in identifying the factors behind transfer: I have tried not only to identify factors but also to explore the "lower-order" (i.e., closer to the neural level) of processing. Although preliminary and speculative at this stage, this essay is an attempt to look inside the "black box" of second language acquisition/processing.

Among the issues that ought to be explored further within a connectionist framework are (1) which factors are strong and (2) which factor(s) interact with which other factor(s). There are at least two possible ways to investigate these questions. One, as used in the connectionist framework of SLA research proposed by Gasser (1990), is computer modeling. Though it may be premature to claim that computer modeling is identical to human L2 processing/acquisition, it is an interesting method for investigating SLA mechanisms because it can control for intervening variables and has no sampling problem as far as subjects are concerned. Therefore, it does not necessitate statistical testing; the experiment conducted on a machine is always replicable, and the results can be compared with what is known about L2 acquisition in humans. Henning & Roitblat's (1991) study, which simulates a Spanish-speaker's acquisition of English negation, is a project along these lines. Secondly, another possible method is to use actual SLA data, though it goes without saying that we would need a large data base and sound operational definitions of the factors discussed in this paper. One such study, Puolisse & Schils (1989), an investigation of communication strategies which effectively controls for several independent variables, suggests that time pressure (which is closely related to cognitive load factor) is stronger than proficiency level in determining the degree of "transfer strategy" used by a learner. Continuing this line of research would certainly be valuable in addressing some of the issues I have raised.

Apparently, most language teachers and learners assume that the learner's mother tongue has an important role in second language learning. It seems that after the confusion created when universals in acquisition were overemphasized and variability studies including L1 transfer did not get much attention, studies on transfer are enjoying a comeback. What we now need is a sound theory and

sound methodology to test the theory. The connectionist framework, as presented in this paper, may contribute to the further specification of L1 transfer: which factors condition transfer and the role transfer plays in second language processing and acquisition.

ACKNOWLEDGEMENTS

I thank Evelyn Hatch, John Schumann, Mike Gasser, Miyuki Sasaki, Cheryl Fantuzzi, Bob Jacobs, Mitsuru and Kurumi Hiki, Barbara Baptista, Foong-Ha Yap, and several anonymous *IAL* reviewers for their helpful comments on an earlier version of this paper. Any remaining deficiencies are of course mine.

NOTES

[1] In this paper, transfer is defined as the effect of the knowledge of one language on the learning/processing of another. With this definition, not only the effect of L1 on L2 learning but also the effect of L2 on L1 can be included. This definition also includes the role of a third language in language transfer.

[2] See also the criticisms of Fodor & Pylyshyn (1988) and of Pinker & Prince (1988) by Clark (1989, chapters 8 and 9).

[3] See Schumann (1990, 1991) for why it is necessary to have architecture and neurobiological constraints for theory building in SLA. See also the upcoming special issue of *IAL* on neurobiology and language acquisition, to be guest-edited by John Schumann.

[4] This is not a trivial question because some of the strongest points to be made in support of the connectionist approach are realizable only through distributed representation (e.g., graceful degradation; see Clark, 1989, chapter 5; Gasser, 1990; Hinton, McClelland, & Rumelhart, 1986). In this paper, however, I will not go into this issue, but will discuss L1 transfer in such a way as to be congruent with both localized and distributed approaches. Simply put, locally represented concepts/nodes can also be represented in a distributed fashion (Gasser, 1988). For further details concerning localized vs. distributed representation, see Gasser (1988, chapter 11) and Feldman (1986).

[5] In this paper, the word "process" is synonymous with "processing" but is used in a sense which is different from that current in SLA literature. For example, "acquisition process" in SLA refers to the process of change that extends over time. It should also be noted that in most SLA studies, the distinction between representation and process(ing) is not at all clear (Carroll, 1989).

[6] Gasser's (1988) model of connectionist lexical memory (CLM) does not have a phonological component, but he claims it would not be difficult to add such a component to the model.

[7] A comprehensive review of connectionism is unfortunately beyond the scope of this paper. For more comprehensive reviews of connectionism in relation to SLA, see Schmidt (1988) and Gasser (1988, 1990).

[8] "Morphemes" here should be interpreted as inflectional morphemes.

[9] "Discourse" here means the linguistic level which cannot be handled by the knowledge of one sentence only. This involves the realms of what Canale (1983) called "discourse competence" and "sociolinguistic competence," domains which roughly correspond to Bachman's (1990) "textual competence" and "pragmatic competence," respectively.

[10] Although the examples discussed are cases of "negative transfer," there are obviously many cases of "positive transfer." Otherwise, L2 learners would not use the strategy of searching interlingual equivalents. Evidence of such searches is pervasive in the literature of child L2 acquisition research, where it is pointed out that learners constantly request translation equivalents or offer comments on the lack of equivalents in the native and target languages (see Hatch, 1978).

[11] See, for example, Ellis (1985, chapter 8) for the role of markedness in second language acquisition.

[12] The distinction between "conscious" and "unconscious" should be viewed not as a dichotomy but as a continuum (e.g., Takala, 1984). Moreover, even young children use conscious strategies in learning L2 (e.g., Hatch, 1978).

[13] For example, even though I (an advanced L2 learner of English) know the rule of third person singular *-s*, I often fail to put *-s* in obligatory contexts when speaking communicatively.

[14] This notion is congruent with the fact that babies have been shown to have limited development of those cortical areas which are responsible for declarative knowledge (i.e., the medial temporal region) (Schumann, personal communication).

[15] The idea of a "hearer-role" can also be applied to "foreigner talk," which is a case of convergence to the nonnative speaker by the native speaker. It also partially explains language switching/mixing.

[16] I have argued strongly in this paper for a connectionist approach because it provides a new perspective from which to interpret SLA phenomena. However, the debate between the two camps--symbolic and connectionist--is by no means settled (for a critical evaluation of connectionism from the symbolic perspective, see Pinker & Mehler, 1988). It has also been suggested that a promising approach would be a hybrid of the two paradigms (e.g., Clark, 1989; Holyoak, 1991).

REFERENCES

Allman, W. F. (1989). *Apprentices of wonder: Inside the neural network revolution.* New York: Bantam.

Andersen, R. W. (1983). Transfer to somewhere. In S. M. Gass & L. Selinker (Eds.), *Language transfer in language learning* (pp. 177-201). Rowley, MA: Newbury House.

Bates, E. A. & Elman, J. L. (1992). *Connectionism and the study of change* (CRL Technical Report No. 9202). San Diego: University of California, Center for Research in Language.

Bachman, L. F. (1990). *Fundamental considerations in language testing.* Oxford: Oxford University Press.

Beebe, L. M. (1988). Five sociolinguistic approaches to second language acquisition. In L. M. Beebe (Ed.), *Issues in second language acquisition: Multiple perspectives* (pp. 43-77). New York: Newbury House.

Beebe, L. M. & Giles, H. (1984). Speech accommodation theories: A discussion in terms of second-language acquisition. *International Journal of the Sociology of Language, 46,* 5-32.

Beebe, L. M., Takahashi, T., & Uliss-Weltz, R. (1990). Pragmatic transfer in ESL refusals. In R. C. Scarcella, E. Andersen, & S. D. Krashen (Eds.), *Developing communicative competence in a second language* (pp. 55-73). Cambridge, MA: Newbury House.

Bialystok, E. (1988). Psycholinguistic dimensions of second language proficiency. In W. Rutherford & M. Sharwood Smith (Eds.), *Grammar and second language teaching* (pp. 31-50). New York: Newbury House.

Blum-Kulka, S. (1982). Learning to say what you mean: A study of speech act performance of learners of Hebrew as a second language. *Applied Linguistics, 3,* 29-59.

Briere, E. J. (1968). *A psycholinguistic study of phonological interference.* The Hague: Mouton.

Brown, H. D. (1987). *Principles of language learning and teaching* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Canale, M. (1983). From communicative competence to communicative language pedagogy. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 2-27). London: Longman.

Cancino, H. (1979). *Grammatical morphemes in second language acquisition: Marta.* Unpublished qualifying paper, Harvard University, Cambridge, MA. Cited in Andersen (1983).

Carroll, S. (1989). Second language acquisition and the computational paradigm. *Language Learning, 39,* 535-594.

Chomsky, N. (1957). *Syntactic structures.* The Hague: Mouton.

Chomsky, N. (1959). A review of B. F. Skinner's *Verbal behavior. Language, 35,* 26-58.

Clark, A. (1989). *Microcognition: Philosophy, cognitive science, and parallel distributed processing.* Cambridge, MA: MIT Press.

Connor, U. & Kaplan, R. (Eds.). (1987). *Writing across languages: Analysis of L2 written text.* Reading, MA: Addison-Wesley.

Cottrell, G. & Small, S. (1983). A connectionist scheme for modelling word sense disambiguation. *Cognition and Brain Theory, 6,* 89-120.

Dommergues, J. & Lane, H. (1976). On two independent sources of error in learning the syntax of a second language. *Language Learning, 26,* 111-123.

Dulay, H. & Burt, M. (1973). Should we teach children syntax? *Language Learning, 23,* 245-58.

Eckman, F. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning, 27,* 315-30.

Ellis, R. (1985). *Understanding second language acquisition.* Oxford: Oxford University Press.

Ervin-Tripp, S. M. (1964). An analysis of interaction of language, topic and listener. *American Anthropologist, 66* (part 2), 86-102.

Ervin-Tripp, S. M. (1967). An Issei learns English. *Journal of Social Issues, 23,* 78-90.

Ervin-Tripp, S. M. (1974). Is second language learning like the first? *TESOL Quarterly, 8,* 111-27.

Færch, K. & Kasper, G. (1986). Cognitive dimensions of language transfer. In E. Kellerman & M. Sharwood Smith (Eds.), *Crosslinguistic influence in second language acquisition* (pp. 49-65). New York: Pergamon.

Feldman, J. A. (1986). *Neural representation of conceptual knowledge* (Technical Report No. 189). Rochester, NY: University of Rochester, Department of Cognitive Science.

Felix, S. (1980). Interference, interlanguage, and related issues. In S. Felix (Ed.), *Second language development: Trends and issues* (pp. 93-107). Tubingen: Gunter Narr Verlag.

Ferguson, C. A. (1984). Repertoire universals, markedness, and second language acquisition. In W. E. Rutherford (Ed.), *Language universals and second language acquisition* (pp. 247-258). Amsterdam: John Benjamins.

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

Gass, S. M. (1981). An investigation of syntactic transfer in adult second language learners. In R. C. Scarcella & S. D. Krashen (Eds.), *Research in second language acquisition* (pp. 132-141). Rowley, MA: Newbury House.

Gasser, M. E. (1988). *A connectionist model of sentence generation in a first and second language.* Unpublished doctoral dissertation, University of California, Los Angeles.

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

Giles, H. & Byrne, J. (1982). An intergroup approach to second language acquisition. *Journal of Multilingual and Multicultural Development, 3*, 17-40.

Hakuta, K. (1974). A preliminary report on the development of grammatical morphemes in a Japanese girl learning English as a second language. *Working Papers on Bilingualism, 3*, 18-43.

Hatch, E. M. (1978). Introduction. In E. M. Hatch (Ed.), *Second language acquisition: A book of readings* (pp. 1-18). Rowley, MA: Newbury House.

Hatch, E. M. (1981). Discussion of "Input from the inside: The role of a child's prior linguistic experience in second language learning." In R. W. Andersen (Ed.), *New dimensions in second language acquisition research* (pp. 104-108). Rowley, MA: Newbury House.

Hatch, E. M. (1983). *Psycholinguistics: A second language perspective.* Rowley, MA: Newbury House.

Hatch, E. (1989). *Scope of SLA research: Promises and problems.* Paper presented at ETS Invitational Symposium on Language Acquisition, Princeton University, Princeton, NJ.

Hatch, E., Shirai, Y., & Fantuzzi, C. (1990). The need for an integrated theory: Connecting modules. *TESOL Quarterly, 24*, 697-716.

Henning, M. & Roitblat, H. (1991). Developmental stages of negation: A connectionist model. Paper presented at the Second Language Research Forum, University of Southern California, Los Angeles.

Hinton, G. E., McClelland, J. L., & Rumelhart, D. E. (1986). Distributed representations. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (vol. 1: *Foundations*, pp. 77-109). Cambridge, MA: MIT Press.

Holyoak, K. J. (1991). Symbolic connectionism: Toward third-generation theories of expertise. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 301-335). Cambridge, UK: Cambridge University Press.

Ijaz, I. H. (1986). Linguistic and cognitive determinants of lexical acquisition in a second language. *Language Learning, 36,* 401-451.

Jordens, P. (1977). Rules, grammatical intuitions and strategies in foreign language learning. *Interlanguage Studies Bulletin, 2,* 5-76.

Kasper, G. (1981). *Pragmatische aspecte in der interimsprache* [Pragmatic aspects of interlanguage]. Tubingen: Gunter Narr Verlag. Cited in Odlin (1989).

Kato, A. (1989). *A discourse analysis of "want" as an indirect directive in comparison to Japanese directives.* Unpublished master's thesis, University of California, Los Angeles.

Kaplan, R. (1966). Cultural thought patterns in inter-cultural education. *Language Learning, 16,* 1-20.

Kellerman, E. (1977). Towards a characterisation of the strategy of transfer in second language learning. *Interlanguage Studies Bulletin, 2,* 58-145.

Kellerman, E. (1978). Giving learners a break: Native speaker intuition as a source of predictions about transferability. *Working Papers on Bilingualism, 15,* 59-92.

Kellerman, E. (1979). Transfer and non-transfer: Where we are now. *Studies in Second Language Acquisition, 1,* 37-57.

Kellerman, E. (1983). Now you see it, now you don't. In S. M. Gass & L. Selinker (Eds.), *Language transfer in language learning* (pp. 112-134). Rowley, MA: Newbury House.

Kennedy, B. L. (1988). Adult versus child L2 acquisition: An information-processing approach. *Language Learning, 38,* 477-495.

Koike, I. (1981). *Second language acquisition of grammatical structures and relevant verbal strategies.* Unpublished doctoral dissertation, Georgetown University, Washington, DC. Cited in Sasaki (1987).

Krashen, S. D. (1978). The monitor model for second language acquisition. In R. Gingras (Ed.), *Second language acquisition and foreign language teaching.* Arlington, VA: Center for Applied Linguistics.

Krashen, S. D. (1981). *Second language acquisition and second language learning.* New York: Pergamon.

Kroll, J. F. & Curley, J. (1988). Lexical memory in novice bilinguals: The role of concepts in retrieving second language words. In M. Gruneberg, P. Morris, & R. Sykes (Eds.), *Practical aspects of memory* (vol. 2, pp. 389-395). London: John Wiley & Sons.

Lado, R. (1957). *Linguistics across cultures.* Ann Arbor: University of Michigan Press.

Long, M. H. & Sato, C. J. (1984). Methodological issues in interlanguage studies: An interactionist perspective. In A. Davies, C. Criper, & A. P. R. Howatt (Eds.), *Interlanguage* (pp. 253-279). Edinburgh, UK: Edinburgh University Press.

Lightbown, P. M. (1983). Exploring relationships between developmental and instructional sequences in L2 acquisition. In H. W. Seliger & M. H. Long (Eds.), *Classroom oriented research in second language acquisition* (pp. 217-245). Rowley, MA: Newbury House.

Maccoun, W. (1983). *On the acquisition of argumentative discourse from the comprehension point of view.* Unpublished master's thesis, University of California, Los Angeles.

MacWhinney, B. (1989). Competition and connectionism. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 423-457). New York: Cambridge University Press.

Major, R. C. (1986). The ontogeny model: Evidence from L2 acquisition of Spanish *r*. *Language Learning, 36,* 453-503.

Major, R. C. (1987). A model for interlanguage phonology. In G. Ioup & S. H. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 101-124). Cambridge, MA: Newbury House.

Master, P. A. (1987). *A cross-linguistic interlanguage analysis of the acquisition of the English article system.* Unpublished doctoral dissertation, University of California, Los Angeles.

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

McLaughlin, B. (1978). The monitor model: Some methodological considerations. *Language Learning, 28,* 309-32.

McLaughlin, B. (1987). *Theories of second language learning.* London: Edward Arnold.

McLaughlin, B., Rossman, T., & McLeod, B. (1983). Second language learning: An information-processing perspective. *Language Learning, 33,* 135-158.

Munro, P. W. (1986). State-dependent factors influencing neural plasticity: A partial account of the critical period. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructures of cognition* (vol. 2: *Psychological and biological models*, pp. 471-502). Cambridge, MA: MIT Press.

Newmeyer, F. J. (1986). Has there been a Chomskyan revolution in linguistics? *Language, 62,* 1-18.

Norman, D. A. (1986). Reflections on cognition and parallel distributed processing. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructures of cognition* (vol. 2: *Psychological and biological models*, pp. 531-546). Cambridge, MA: MIT Press.

Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning.* Cambridge, UK: Cambridge University Press.

Oi, K. (1986). Cross-cultural difference in rhetorical patterning: A study of Japanese and English. *JACET Bulletin, 17,* 23-48.

Oller, J. W. & Ziahosseiny, S. M. (1970). The contrastive analysis hypothesis and spelling errors. *Language Learning, 20,* 183-189.

Olshtain, E. (1983). Sociocultural competence and language transfer: The case of apology. In S. Gass & L. Selinker (Eds.), *Language transfer in language learning* (pp. 232-249). Rowley, MA: Newbury House.

Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of psycholinguistic research, 5,* 261-258.

Peters, A. M. (1983). *The units of language acquisition.* Cambridge, UK: Cambridge University Press.

Pinker, S. & Mehler, J. (Eds.). (1988). *Connections and symbols.* Cambridge, MA: MIT Press.

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

Prator, C. H. (1967). Hierarchy of difficulty. Unpublished classroom lecture, University of California, Los Angeles. Cited in Brown (1987).

Puolisse, N. & Schils, E. (1989). The influence of task- and proficiency-related factors on the use of compensatory strategies: A quantitative analysis. *Language Learning, 39*, 15-48.

Richards, J. (1971). Error analysis and second language strategies. *Language Sciences, 17*, 12-22.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group (Eds.) (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (vol. 1: *Foundations*). Cambridge, MA: MIT Press.

Sampson, G. (1987). Review of *Parallel distributed processing* by D. E. Rumelhart, J. L. McClelland, & the PDP research group. *Language, 63*, 871-886.

Sasaki, M. (1987). Is Uguisu an exceptional case of "idiosyncratic variation"?: Another counterexample to the "natural order." *Chugoku-Shikoku Academic Society of Education Research Bulletin, 32*, 170-174.

Scarcella, R. (1979). On speaking politely in a second language. In C. A. Yorio, K. Perkins, & J. Schachter (Eds.), *On TESOL '79: The learner in focus* (pp. 275-287). Washington, DC: TESOL.

Scarcella, R. (1983). Discourse accent in second language performance. In S. M. Gass & L. Selinker (Eds.), *Language transfer in language learning* (pp. 306-326). Rowley, MA: Newbury House.

Schneider, W. (1987). Connectionism: Is it a paradigm shift for psychology? *Behavior Research Methods, Instruments, & Computers, 19*, 73-83.

Schumann, J. H. (1982). Simplification, transfer, and relexification as aspects of pidginization and early second language acquisition. *Language Learning, 32*, 337-366.

Schumann, J. H. (1990). Extending the scope of the acculturation/pidginization model to include cognition. *TESOL Quarterly, 24*, 667-684.

Schumann, J. H. (1991). *Falsification: Has it ever happened in SLA?* Paper presented at the Conference on Theory Construction and Methodology in Second Language Research, Michigan State University, East Lansing, MI.

Schmidt, R. (1988). The potential of parallel distributed processing for SLA theory and research. *University of Hawaii Working Papers in ESL, 7*, 55-66.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 129-158.

Schmidt, R. & Richards, J. (1980). Speech acts and second language learning. *Applied Linguistics, 1*, 129-157.

Shirai, Y. (1989). *The acquisition of the basic verb PUT: Prototype and transfer.* Unpublished master's thesis, University of California, Los Angeles.

Shirai, Y. (1990). U-shaped behavior in L2 acquisition. In H. Burmeister & P. L. Rounds (Eds.), *Variability in second language acquisition: Proceedings of the Tenth Meeting of the Second Language Research Forum* (vol. 2, pp. 685-700). Eugene, OR: Department of Linguistics, University of Oregon.

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.

Takahashi, T. (1985). Language transfer and the acquisition of patterns of lexico-semantic correspondence. Paper presented at the Second Language Research Forum, University of Southern California, Los Angeles.

Takala, S. (1984). Review of *Language two. Language Learning, 34*, 157-174.

Tanaka, S. (1983). *Language transfer as a constraint on lexico-semantic development in adults learning a second language in acquisition-poor environments*. Unpublished doctoral dissertation, Teachers College, Columbia University, New York.

Tanaka, S. & Abe, H. (1985). Conditions on interlingual semantic transfer. In P. Larson, E. L. Judd, & D. S. Messerschmitt (Eds.), *On TESOL '84: A brave new world for TESOL* (pp. 101-120). Washington, DC: TESOL.

Tanaka, S. & Abe, H. (1988). Gaikokugo gakushu ni okeru gengo ten-i no mondai [The problem of language transfer in foreign language learning], 1. *Eigo Kyoiku* [The English Teacher's Magazine], November, 32-35.

Tanaka, S. & Abe, H. (1989). Gaikokugo gakushu ni okeru gengo ten-i no mondai [The problem of language transfer in foreign language learning], 3. *Eigo Kyoiku* [The English Teacher's Magazine], January, 78-80.

Taylor, B. (1975). The use of overgeneralization and transfer learning strategies by elementary and intermediate students of ESL. *Language Learning, 25*, 73-107.

Wardhaugh, R. (1970). The contrastive analysis hypothesis. *TESOL Quarterly, 4*, 123-130.

Waltz, D. L. & Pollack, J. B. (1985). Massively parallel parsing: A strongly interactive model of natural language interpretation. *Cognitive Science, 9*, 51-74.

Weinreich, U. (1953). *Languages in contact*. New York: Linguistic Circles of New York.

Whitman, R. & Jackson, K. L. (1972). The unpredictability of contrastive analysis. *Language Learning, 22*, 29-41.

Wode, H. (1978). Developmental sequences in naturalistic L2 acquisition. In E. M. Hatch (Ed.), *Second language acquisition: A book of readings* (pp. 101-117). Rowley, MA: Newbury House.

Young, R. (1990). Functional interpretations of variation in interlanguage morphology. Paper presented at the 10th Second Language Research Forum, University of Oregon, Eugene.

Zobl, H. (1983). Markedness and the projection problem. *Language Learning, 33*, 293-313.

Zobl, H. (1984). Cross-language generalisations and the contrastive dimension of the interlanguage hypothesis. In A. Davies, C. Criper, & A. P. R. Howatt (Eds.), *Interlanguage* (pp. 79-97). Edinburgh, UK: Edinburgh University Press.

Yasuhiro Shirai has a Ph.D. in applied linguistics from UCLA and is now an assistant professor in the Department of Foreign Languages at Daito Bunka University in Tokyo, Japan, where he teaches EFL and linguistics. His research interests include L1/L2 acquisition of tense-aspect morphology and cognitive models of L1/L2 processing and acquisition.