In: R. Heredia & J. Altarriba (Eds.), *Bilingual Sentence Processing*. North Holland: Elsevier Science Publisher.

A Self-Organizing Connectionist Model of Bilingual Processing

Ping Li and Igor Farkaš

University of Richmond

Abstract

Current connectionist models of bilingual language processing have been largely restricted to localist stationary models. To fully capture the dynamics of bilingual processing, we present SOMBIP, a self-organizing model of bilingual processing that has learning characteristics. SOMBIP consists of two interconnected self-organizing neural networks, coupled with a recurrent neural network that computes lexical co-occurrence constraints. Simulations with our model indicate that (1) the model can account for distinct patterns of the bilingual lexicon without the use of language nodes or language tags, (2) it can develop meaningful lexical-semantic categories through self-organizing processes, and (3) it can account for a variety of priming and interference effects based on associative pathways between phonology and semantics in the lexicon, and (4) it can explain lexical representation in bilinguals with different levels of proficiency and working memory capacity. These capabilities of our model are due to its design characteristics in that (a) it combines localist and distributed properties of processing, (b) it combines representation and learning, and (c) it combines lexicon and sentences in bilingual processing. Thus, SOMBIP serves as a new model of bilingual processing and provides a new perspective on connectionist bilingualism. It has the potential of explaining a wide variety of empirical and theoretical issues in bilingual research.

Introduction

Connectionism, parallel distributed processing (PDP) models, or neural networks have significantly influenced research in the cognitive sciences in the last fifteen years. Language, as one of the most important human cognitive components, has received in-depth treatments since the beginning of connectionist research. The acquisition of the English past tense (Rumelhart & McClleland, 1986), the recognition of speech (McClleland & Elman, 1986), and the processing of sentences (McClleland & Kawamoto, 1986) are among the earliest domains of connectionist research in the original PDP models. Unfortunately, connectionist models or modeling have had very limited impact on the field of bilingualism as a whole. To this date, there are only a handful of connectionist models that are implemented to account for the processing or representation of the bilingual mental lexicon (see our review below). This lack of interaction between connectionism and bilingualism is lamentable, and it gives us a good reason to pursue research that would fill this gap. In this chapter, we present a connectionist model, more specifically, a self-organizing neural network model of bilingual processing and representation.

Let us begin by reviewing a few important lines of research in connectionist bilingualism. In particular, we will examine the Bilingual Interactive Activation (BIA) model (Grainger, 1993; Dijkstra & van Heuven, 1998; van Heuven, 2000), the Bilingual Model of Lexical Access (BIMOLA; Grosjean, 1988, 1997; Léwy & Grosjean, in progress), and the simple recurrent network (SRN) model of bilingual memory (French, 1998). The first two belong to the so-called "localist" models, while the last one belongs to "distributed" models. In localist models, a word or a concept is represented by a single, unitary processing node in the network, whereas in distributed models, information about a word or a concept is distributed across several or many different units of processing.

The localist models, that is, BIA and BIMOLA, are based on earlier interactive models of word recognition, that is, the interactive activation model (IA) of McClelland and Rumelhart (1981) and the TRACE model of McClelland and Elman (1986), respectively. In the IA model, there are three levels of nodes, with ascending complexity: (1) features of a letter such as curves, straight lines, or crossbars, (2) individual letters, and (3) words. Information at all three levels can interact with each other during the word recognition process, which may flow both "bottom-up" (features to letters to words) and "top-down" (words to letters to features). Within levels, nodes compete for activation (thus inhibiting each other); across levels, nodes either inhibit or excite each other. According to IA, it is these inhibitory and excitatory connections that give rise to the appropriate activation of patterns that correspond to the perception of words.

The BIA model is a straightforward extension of the IA model. It consists of four levels of nodes: features, letters, words, and languages. As in IA, there are connections between nodes at each level and between nodes across levels. All nodes at the word level are interconnected with mutual inhibition. Feature units activate appropriate letters, and letter units activate appropriate words in the appropriate language. BIA uses the same parameters to regulate interactions within and across levels as in the original IA model. What is special to the BIA model (apart from the incorporation of two lexicons) are the language nodes (one for English and one for Dutch). Language nodes in BIA function as an important mechanism for the selection or inhibition of words in one or the other language, given that the model argues for and implements the language-independent access hypothesis, according to which words from different languages are represented in an integrated lexicon and are simultaneously contacted during word recognition. Results from BIA simulations suggest that the model is able to account for empirical results that directly support the language-independent access hypothesis, while at the same time is compatible with data previously thought to support the language-selective access hypothesis (Dijkstra & van Heuven, 1998; van Heuven, 2000).

While the BIA model focuses on visual word recognition, the BIMOLA model (Grosjean, 1988, 1997; Léwy & Grosjean, in progress) aims at handling the recognition of spoken words. It was partly inspired by the TRACE model (McClelland & Elman, 1986), an IA-based model for acoustic input. Like TRACE, it consists of three levels of nodes, corresponding to features, phonemes, and words. There is no separate level of language nodes in BIMOLA, unlike in BIA. The feature level is common to both languages, but the phoneme and word levels are organized in subsets according to languages (still in the same extended system). Features activate phonemes that, in turn, activate words. Connections (both excitatory and inhibitory) are unidirectional (ascending) between features and phonemes, but bidirectional between phonemes and words. Descending connections from top down (global language activation and higher linguistic information, especially the "bilingual speech mode"; Grosjean, 1997) serve to activate words that, in turn, can activate phonemes. Language activation (or selection) takes place through these descending connections but also through within-language connections at the phoneme and word levels. Compared to the BIA model, BIMOLA can account for language-specific activation without the use of language nodes – it is yet unclear whether language nodes or language tags are necessary components of bilingual processing (Li, 1998a; see also General Discussion). The BIMOLA model is currently being implemented and evaluated against empirical data (Léwy, in progress; Léwy & Grosjean, in progress).

Both BIA and BIMOLA can be said to be permanent (or stationary) models, despite their differences in input (visual vs. acoustic) and architecture (with or without language nodes). They differ from some connectionist models with a learning mechanism. In BIA and BIMOLA, the representations are fixed and manually coded, and are designed to capture proficient adult bilingual speakers' mental lexicon (but are not designed to evolve). In contrast, connectionist learning models dynamically develop representations from computing statistical characteristics of the input data (sentences). Along this direction, French (1998) presented a distributed model of bilingual memory based on SRN, a simple recurrent network (Elman, 1990) that learns representations online through sentence processing. As in the original SRN, the model takes in one input word at a time from a continuous stream of sentences and its task is to predict the next word of the current input in the sentence. The input stream is a series of artificially generated sentences of the N-V-N structure, with English and French sentences intermixed at the sentence level. Elman (1990) showed that distinct grammatical and semantic categories (e.g., nouns and verbs, animate and inanimate) can emerge in the SRN's hidden-unit representations once the network has learned the next-word prediction task with a reasonable size of sentences. This is because the prediction task involves detecting the context in which the current input occurs. With the SRN's exposure to mixed bilingual input, French showed that not only distinct categories emerge within each language in the hidden-unit representations (as in Elman 1990), but also there are distinct patterns of the two languages: words from the two languages are separated in space on a hierarchical cluster tree of the hidden-unit activations. Note that the network incorporates no mechanism (nodes or tags) to explicitly label words in the two languages. This model provides support to the hypothesis that bilingual memory is organized as a single distributed lexicon rather than two separately stored lexicons according to language. Thus, the model can display distinct behaviors of the two lexicons without invoking separate mechanisms for each language, unlike the BIA model that uses language nodes to separate the bilingual lexicon.

The connectionist model of bilingual processing that we present in this paper has the same spirit of the SRN model, but it differs from the above models in three important respects. First, our model combines both learning and representational properties. It is a learning model in the sense that lexical representations of both languages can emerge from the statistical learning of the input speech. This property is similar to that of the SRN, but is based on our network in explicitly modeling lexical co-occurrences in the acquisition of word meanings (Li, 1999, 2000, in press; Farkaš & Li, 2001, in press). On the other hand, our model also has some of the representational features of BIMOLA: lexical forms are encoded by articulatory features of the phonemes of words (see also Li & MacWhinney, 2001). The representational characteristics of the lexical forms and meanings can also become clearly discernible on a 2-dimensional topological space, given the self-organizing maps used in our model. In addition, given both word meanings and word forms in the model, learning can occur in the associative links between meanings and forms via Hebbian learning, a biologically plausible mechanism of co-occurrence learning (see later discussion).

Second, our model displays both distributed and localist properties. Unlike SRN or other connectionist models that use back-propagation as the learning algorithm, our model employs principles of self-organization, a type of unsupervised learning (Kohonen, 1995). Although the inputs in our self-organizing network are in the form of distributed representations, the 2-D topological map of the network (the output) bears significant similarity to localist representations in that each unit on the map tends to represent one lexical item. The localized patterns on the map allow an "explicit" representation of the lexicon, instead of "implicit" representations as in the hidden-unit activations of an SRN (MacWhinney, 2000a). At the same time, each unit is surrounded by neighboring units on the map that can become co-activated, simulating a distributed lexicon in which similar words or word properties are grouped together (see details below).

Third, our model relies on the use of realistic linguistic data as input to the network, in particular, child-directed parental speech. In the SRN model of bilingual memory (French, 1998) as well as many current connectionist models, researchers have relied on the use of highly simplified, artificially generated input. Although such inputs are easy to construct and to control for, they are often isolated from realistic language uses, and run the risk of being successful just because of the use of certain properties in the input (see Lachter & Bever, 1988 for an earlier criticism of connectionist representations). For example, French (1998) structured the input data in such a way that the input has a fixed probability of 0.001 of switching from one language to another. In other words, on average, the network will first have learned 1000 sentences in one language before learning any sentences in the other language. We suspect that this artificially determined probability is what causes the network to display disparate behaviors for the two languages. To overcome potential problems associated with this approach to network modeling, we thus rely on corpus-based linguistic data that closely approximate the reality of language use (see also Li, in press, for discussion).

In what follows, we first present an overview of our model, SOMBIP, a connectionist self-organizing model of bilingual processing. We then discuss some preliminary simulation results from the model. We conclude by showing how our model can shed light on important issues in bilingual lexical and sentence processing.

The SOMBIP Model

Background

The design characteristics of the SOMBIP model are based on our self-organizing neural network model of language acquisition by young children (Li, 1999, 2000, in press; Farkaš & Li, 2001, in press). In recent years, self-organizing neural networks have become increasingly important for cognitive and perceptual studies (Hinton & Senjowski, 1999). Although significant progress has been made with models based on back-propagation, there are known limitations associated with these models, including catastrophic forgetting (inability to remember old information with new learning), scalability (inability to handle realistic, large-scale problems), and above all, its error-driven learning process, a procedure which propagates error signals from the discrepancy between desired and actual outputs to adjust weights. Some of these problems become most transparent when considered in the context of language acquisition (see Li, in press). Consideration of these problems lead us to look for models that bear more biological and psychological plausibility in the context of language learning, and we turn to the self-organizing models.

One of the most widely used self-organizing models is Kohonen's (1982, 1989, 1995) self-organizing map (SOM). SOM belongs to the class of "unsupervised" neural networks, because learning in these networks does not require the presence of a supervisor or an explicit teacher; learning is achieved by the system's self-organization in response to the input. During learning, the self-organizing process extracts an efficient and compressed internal representation from a high-dimensional input space and projects this new representation onto a 2-D topological structure (Kohonen, 1982, 1989, 1995). Several important properties of SOM and related features make it particularly well suited to the study of language.

(1) *Self-organization*. Self-organization in these networks typically occurs in a two-dimensional topological map, where each unit (or node, or neuron) is a location on the map that can uniquely represent one or several input patterns. At the beginning of learning, an input pattern randomly activates a group of the many units

on the map, according to how similar by chance the input pattern is to the weight vectors (codevectors) of the units. Once a unit becomes active in response to a given input, the weight vectors of the unit and its neighboring units are adjusted so that they become more similar to the input and will therefore respond to the same or similar inputs more strongly the next time. In this way, every time an input is presented, an area of units will become activated on the map (the so-called activity "bubbles"), and the maximally active units are taken to represent the input. Initially, activation occurs in large areas of the map, but gradually learning becomes more focused so that only the maximally responding unit or units are active. This process continues until all the inputs have found some maximally responding units.

(2) *Representation.* As a result of this self-organizing process, the statistical structures implicit in the high-dimensional input space are represented as topological structures on the 2-D space. In this new representation, similar inputs will end up activating the same units in nearby regions, yielding activity bubbles that become clearly visible on the map. Such a self-organizing process and its representation have clear implications for language acquisition: the formation of activity bubbles may capture critical processes for the emergence of semantic categories in the acquisition of the lexicon. In particular, the network organizes information first in large areas of the map and gradually zeros in onto smaller areas; this zeroing-in is a process from diffuse to focused patterns, as a function of the network's continuous adaptation to the input structure. This process allows us to model the emergence of semantic categories as a gradual process of lexical development. It naturally explains many generalization errors reported in the acquisition literature (e.g., substitutions of *put* for *give* or *fall* for *drop* as reported by Bowerman, 1978, 1982). It also explains language disorders that result from the breakdown of focused activation or the inability to form focused representations (Miikkulainen, 1997; Spitzer, 1999).

(3) *Hebbian learning*. Hebbian learning is not an intrinsic property of SOM, but several SOMs can be connected via Hebbian learning, such as in the multiple feature-map model of Miikkulainen (1993, 1997). Hebbian learning is a well-established biologically plausible learning principle, according to which the associative strength between two neurons is increased if the neurons are both active at the same time (Hebb, 1949). The amount of increase may be proportional to the level of activation of the two neurons. In the multiple SOM model developed by Miikkulainen, all units on one map are initially connected to all units on the other map. As self-organization takes place, the associations become more focused, such that in the end only the maximally active units on the corresponding maps are associated. Hebbian learning combined with SOM has strong implications for language learning: it can account for the process of how the learner abstracts relationships between word forms, meanings, and grammatical morphology, on the basis of how often they co-occur and how strongly they are co-activated in the representation.

Because of these properties, SOM (a) allows us to track the development of the lexicon clearly as an emergent property in the network's self-organization (from diffuse to focused patterns or from incomplete to complete associative links); (b) allows us to model one-to-many or many-to-many associations between forms and meanings in the development of the lexicon, and (c) provides us with a set of biologically plausible and computationally relevant principles to study bilingualism without relying on corrective feedback. It is ideally suited for our task also because the bilingual mental lexicon is constructed for the most part without supervision and undergoes continuous self-organization over time.

Architecture

Our SOMBIP model has been inspired by the multiple self-organizing feature-map model of Miikkulainen (1993, 1997). Miikkulainen proposed an integrated model of memory and natural language processing, in which multiple SOMs dedicated to different levels of information are connected. A sub-component of this model is DISLEX, a SOM model of the lexicon, in which different maps correspond to different linguistic information (orthography, phonology, or semantics) and are connected through associative links via Hebbian learning. Our model has also been inspired by the Hyperspace Analogue to Language (HAL) model (Burgess & Lund, 1997, 1999). In particular, we derive our meaning representations of the lexicon through a word co-occurrence detector (WCD), a mechanism similar to the principle of HAL computation. HAL attempts to capture meaning by reference to global lexical co-occurrences – how many words co-occur with the target word, and how often, in a moving window that runs through a large-scale language corpus (Usenet texts of 300 million words). A co-occurrence matrix for any number of words in a given window is derived, and weighted by the frequency of co-occurrences among words. Thus, a word is represented in HAL as a vector of the column and the row in the matrix that encodes the co-occurrence constraints in a high-dimensional space of language use.

Figure 1 presents a diagrammatic sketch of the different components of SOMBIP. Figure 1a depicts the two SOMs used in our model and their interconnections. During learning, a lexical form (phonological input) activates a unit or a group of units on SOM1, and simultaneously, its word meaning (semantic input) activates a unit or a group of units on SOM2. Note that in our current simulations we have used only phonological input in SOM1 to simulate the bilingual spoken lexicon; one can easily use orthographic input in SOM1 to simulate the bilingual visual lexicon. SOM1 and SOM2 are connected via associative links, such that the activation on one map can cause an activity to form in the other map. If the direction of the associative activity is from phonology to semantics, *comprehension* is modeled; if it goes from semantics to phonology, *production* is modeled. The associative links are trained by Hebbian learning, and the strengths of the connections are adjusted according to the form-meaning pairings in the input, which leads to adaptive associations between the two SOMs.

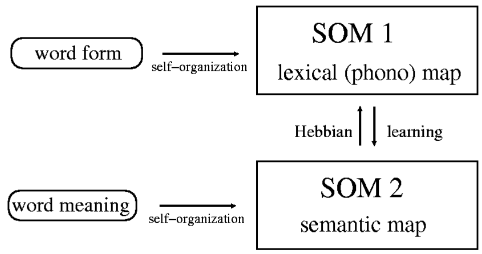

Figure 1a. SOM1 and SOM2 that self-organize on word forms and word meanings, respectively. They are interconnected via associative pathways, trained by Hebbian learning.

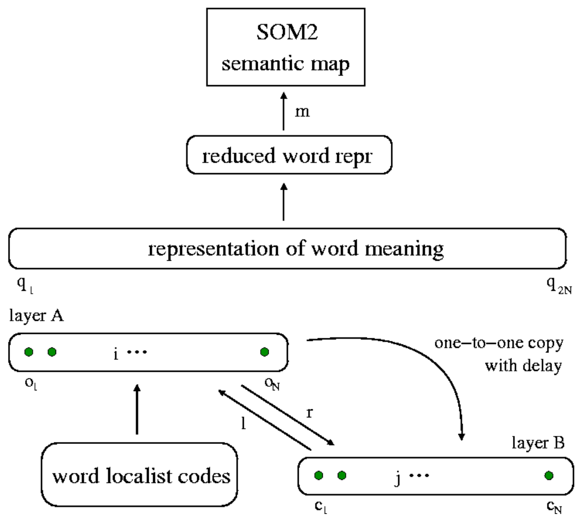

Figure 1b. WCD (lower panel) that computes lexical co-occurrence statistics. Layers A and B have full connectivity via modifiable **l** and **r** links. Other, one-to-one links serve to feed the unit activity from A to B with discrete single time-step delay. SOM2 receives distributed word representations extracted from the modifiable links. The representations are reduced to vectors of uniform length (100 dimensions) as inputs to SOM2 (see Farkaš & Li, in press, 2001, for details).

Figure 1b depicts WCD (lower panel), a special recurrent neural network that learns the lexical co-occurrence constraints of words. The WCD reads through a stream of input sentences (one word at a time), and given a lexicon sized N, it computes the transitional probabilities of words in the lexicon (see Farkaš & Li, in press, for details). Assume that at time t the current word is i, and is represented by a localist vector $\mathbf{o} = [o_1, \ldots, o_N]$ in layer A. Previous word j is represented by the vector $\mathbf{c} = [c_1, \ldots, c_N]$ in layer B. The adaptable connections (**l** and **r** links) between layers A and B serve to approximate the transitional probabilities between successive words, and as such, they are trained by Hebbian learning with weight decay so that they become normalized. Specifically, the link l_{ij} is updated to approximate $P(j^{t-1}|i^{t})$, i.e., the probability that the word j precedes the word i. At the same time, the link r_{ji} is updated to approximate $P(i^{t}|j^{t-1})$, that is, the probability that i follows j. Word i is characterized by a concatenation of vectors $\mathbf{l}_i = [l_{i1}, \ldots, l_{iN}]$, and $\mathbf{r}_i = [r_{1i}, \ldots, r_{Ni}]$, where $\mathbf{l}_i$ approximates the probability distribution of words preceding i (left context), and $\mathbf{r}_i$ the probability distribution of words following i (right context). The concatenated vectors, $\mathbf{q}_i = [\mathbf{l}_i,\mathbf{r}_i]$, then serve as distributed word representations to SOM2 (upper panel in Figure 1b). Because the dimensions of the vectors are determined by the size of the lexicon ($2N$ for any given vector), the vectors, before they are read by SOM2, are also submitted to a dimension-reduction mapping mechanism, which reduces the vectors to lower, fixed dimensions (e.g., 100 units, see Farkaš & Li, 2001).

As SOM2 takes representations from the WCD vectors $\mathbf{q}_i$, SOM1 also takes as its input the phonological representations of words. To represent the phonology of the bilingual lexicon, we have followed a syllable-based template coding originally developed by MacWhinney and Leinbach (1991) and recently by Li and MacWhinney (2001). This coding scheme has the advantage over traditional phonemic representations in that it can accurately capture the phonological similarities of multisyllabic words (most previous connectionist models have used only monosyllabic words as input). In this scheme, the phonology of a word is made up by combinations of syllables in a metrical grid, and the slots in each grid are made up by bundles of features that correspond to phonemes, C's (consonants) and V's (vowel).

To code our Chinese-English bilingual lexicon, we used 4 C-slots and 4 V-slots that allowed for the representation of words of one and two syllables, in the template of CVVCCVVC (each CVVC is a syllable; the first C represents the initial consonant, and the last C the final consonant). Thus, the Chinese (Cantonese) word *jat* (one) is represented in the slots as *ja*V*t*CVVC and the English word *about* is represented as C@VC*baUt*. Each C or V is represented by a set of 5 feature units, and the feature values (scaled between 0 and 1) are determined according to the articulatory features outlined by Ladefoged (1982) for English and by the Hong Kong Linguistic Society (1997) for Chinese (the 5 articulatory features are: Sound, Place, Manner, Chromacity, Sonority). For example, the phoneme /i/ is represented in both languages as [.1 .0 .0 .2 .3], indicating [vowel none none high front] for the 5 features. A separate set of 12 units are used to represent lexical tones in Cantonese (6 tones for each syllable), whereas these units are left empty for English. Thus, each word in the bilingual lexicon is represented by a vector of 52 units in the phonological representation (5 feature units for 8 phoneme-slots plus 12 tonal units).[1] Note that in neither the phonological nor the semantic representations described above is there a label or tag that tells which lexicon (English or Chinese) a given word should belong to.

Learning in the two SOMs is standard (Kohonen, 1989). Every SOM unit k has an array of connections in the form of a codevector $\mathbf{m}_k = [m_{k1}, \ldots, m_{k,2N}]$, which learns to approximate the inputs (semantic or phonological vectors) in such a way that every SOM unit becomes "specialized" for a given word, and that the neighboring units will become specialists ("winners") to similar words. During learning, both neighborhood radius and learning rate decrease over time.

Task and Procedure

Upon training of the network, a phonological input representation of a word is presented to SOM1, and simultaneously, the semantic representation of the same word is also presented to SOM2. By way of self-organization, SOM1 forms an activity in response to the phonological input, and SOM2 an activity in response to the semantic input. As the two SOMs receive input and continue to self-organize, they also simultaneously form associations through Hebbian learning for all the active units in the two maps that respond to the inputs. The network's task is to create an ordered layout for all the input words in the bilingual lexicon and be able to make the appropriate form-meaning associations.

Because our SOMBIP handles a bilingual lexicon, translation equivalents in the two languages are associated with each other in the following way: if the phonology of an English word is presented to SOM1, the semantics of the English word and that of its cross-language translation equivalent in Chinese are also presented to SOM2. Similarly, if the semantics of an English word is presented to SOM2, the phonology of the English word and that of its translation equivalent in Chinese are also presented to SOM1. For example, the word *boat* and *syun* are associated by the phonology of *boat* or *syun* co-occurring with the semantics of *boat* and *syun*, and vice versa. This procedure works in the same way for words in both languages, ensuring that translation equivalents have a chance to be associated. Because of the difficulty in determining what words should be counted as translation equivalents (especially with regard to English and Chinese), this procedure applied only to the nouns and verbs in our simulations (i.e., disregarding adjectives, adverbs, pronouns, prepositions, etc.). Multiple translation equivalents (e.g., English *tell* and Chinese *gong* and *waa*) were also associated through co-occurrences in the network, although such cases were rare in our data.

As discussed earlier, artificially generated input data are often problematic in matching up with realistic language use. In this study, we used a realistic bilingual data set, the Hong Kong Bilingual Corpus from the CHILDES database (Yip & Matthews, 2000; MacWhinney, 2000b). This corpus contains transcripts of conversations between a child and his native English-speaking father and native Cantonese-speaking mother. The recordings were made when the child was between ages 1 and 3. The parents followed the one parent-one language principle when addressing the child. The language between the parents was mainly Cantonese with a great deal of English mixed in, as is characteristic of the speech of Hong Kong middle class families. Despite the "one parent-one language" principle, the quantity of input from the two languages was not all balanced. On the whole, the child received more Cantonese than English input in his first three years. Because of the relatively young age at which the recordings were made, there was not enough productive speech from the child. However, there was plenty of parental speech as input. We therefore extracted all of the parental speech plus the speech of the student assistants who made the recordings during the home visits, forming the bilingual input speech corpus that we used for our simulations. These speech data also effectively allow us to simulate what the learning system (human or network) receives in a concurrent bilingual environment and how the system can, on the basis of the input, develop lexical representations from sentence processing (we used a similar procedure in modeling first language acquisition; see Li, in press; Farkaš & Li, 2001, in press).

A total of 5,154 word types and 185,279 word tokens are found in our bilingual parental corpus, according to the *freq* (frequency count) output of the CLAN (Child Language Analysis) program (MacWhinney, 2000b). For our purposes we trained our model on the 400 most frequent word types in this corpus, which effectively covers 56% of the entire data. These 400 most frequent words happened to contain 184 Chinese words, and 216 English words. During training, SOM1 received the phonological representations of the 400 words and self-organized on them. The WCD network of our model received the words in the input sentences one at a time, and built semantic representations from the lexical co-occurrence statistics. These representations were submitted to our dimension-reduction mapping so that all vectors had a uniform length of 100 units, and subsequently sent to SOM2 for self-organizing learning. SOM1 and SOM2 were accordingly linked by associative links. Note that training in SOM1 and SOM2 was simultaneous, and therefore learning in SOM1 and SOM2 did not begin until the WCD network and the dimension reduction had completed their job. In principle, however, SOM2 need not wait until WCD is done, given that it can self-organize on early stages of semantic representations as WCD learns online. This method has also been implemented in our model as an incremental learning process (see Farkaš & Li, 2001).

Results and Discussion

In this section, we report results from several simulations with SOMBIP, and our analyses will focus on the network’s performance with respect to distinct behaviors in the two languages, the formation of lexical categories, the interlingual priming and interference effects, and effects of proficiency and resource limitation.

Language Separation without Language Nodes

As discussed earlier, an important difference between different connectionist models of bilingualism is whether the model explicitly includes a separate level of language nodes. The BIA model does, whereas the BIMOLA and the bilingual SRN models do not. In BIMOLA, the feature level is common to both languages, and the phoneme and word levels contain subsets of units for each language. The differences between these subsets, coupled with global language activation and higher linguistic information from top down, allow the system to separate the two languages. The SRN approach of French (1998) is more radical, in that it makes no *a priori* assumptions about the differences between the two languages in the bilingual lexicon, but simply lets the system learn the bilingual (artificial) sentence data. However, the way input was structured in the SRN was problematic, as we discussed earlier. In our model, there is no place for explicit marking of languages, as in the SRN model. We trained our model on a realistic parental input corpus with bilingual speech intermixed between Chinese and English. Figure 2 presents a sketch of the phonological and semantic organizations of the bilingual lexicon in the two SOMs, after the network has been trained on the 400 target words for 500 epochs (the WCD network was trained on the input sentences for 3 epochs, i.e., 555,837 word tokens).[2]

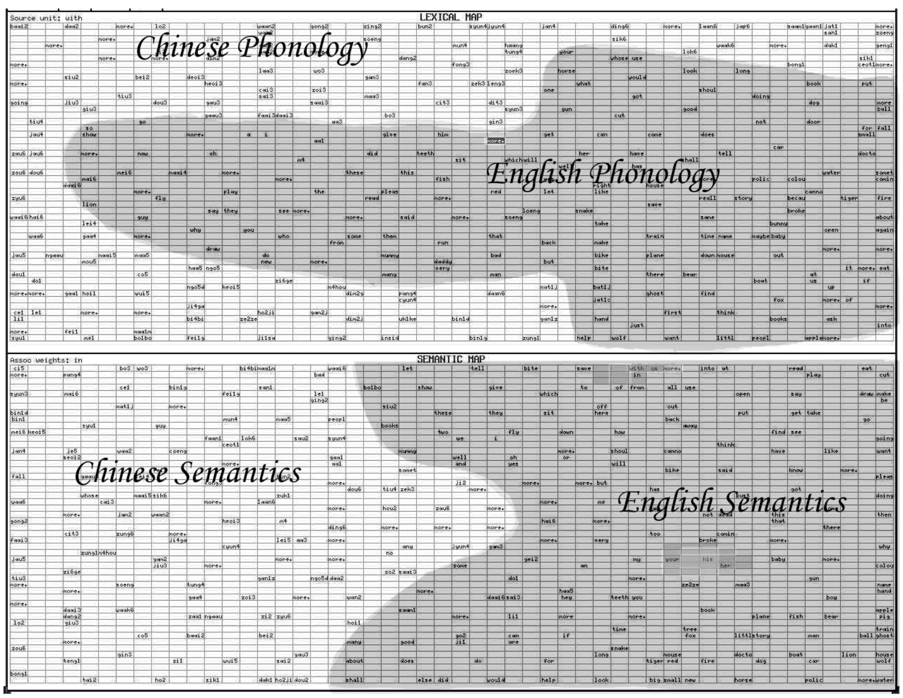

Figure 2. A sketch of the phonological and semantic organizations of the bilingual lexicon in SOM1 and SOM2 after the network has been trained on the representations for 500 epochs. The upper panel represents the lexical form map (SOM1), and the lower panel the semantic map (SOM2). Shaded areas indicate the English phonological and semantic representations, in contrast to the Chinese representations.

As can be seen, our network clearly displays distinct clusters of lexical representations of Chinese from those of English, on both the form (SOM1) and the meaning level (SOM2). Note that because of the topological nature of the maps and the “bubble-filling” nature of SOM learning (Kohonen, 1995), the border between the two languages is not a straight line in either case. Because the sketch leaves out the details, cases where the English and Chinese representations are intermingled are not shown here (but see our discussion in interlingual interference). In general, the intermingled cases are uncommon, and the overall separation of the two lexicons is clear on the 2-D maps.

Our network received no labels or tags for which language a given word should belong to, yet the self-organizing process was able to separate the two lexicons. On the phonological level (SOM1), it is clear that English and Cantonese differ greatly, even though some vowels may share the same quality. In our phonological representation scheme, we have tried to adapt the template in such a way that it could fit both languages (CVVCCVVC), cutting some corners in English (e.g., eliminating the representation of initial consonant clusters). This adaptation does not seem to have much effect on the cross-language distinction. Overall, the syllabic templates, the phonemic inventory, and the lexical tones have given the learning system a great deal of differences to explore for the identification of the two lexicons.

With respect to semantics, SOM2 shows an even clearer separation between the two languages. Note that our network performed essentially a sentence processing task on the speech corpus, in which the WCD reads through the streams of sentences, computes the co-occurrence statistics for each of the 400 words, disregarding words not in the target lexicon (treating them as one unknown type). The words that precede and follow each of the 400 words in the lexicon form the contextual history for that given word, and it is this contextual history that captures the use and meaning of the word. Thus, the words in the two lexicons will have very different contextual histories, because words in English are typically associated with English words, and words in Chinese are typically associated with Chinese words. This is true even with a mixed bilingual corpus like ours, in which English and Chinese sentences are intermixed. From the learner’s perspective, the input contains both English and Chinese sentences, but it is relatively consistent whether a given sentence is English or Chinese, because the intermixing occurs at the sentence level, but rarely within the sentence level. Although it is common for Hong Kong Cantonese speakers to code switch, the most common code-switching situation is one in which the speaker uses an English guest word in a Chinese sentence (Li, 1996a). Given that the coded-switched guest words are often isolated instances and their associations with the lexicon of the borrowing language are relatively weak, our network did not get confused about the identity of the two lexicons.

Overall, our simulation results match well with empirical studies of the bilingual lexicon that argue for the representation of language-specific (or distinct) lexicons, but in an integrated network (e.g., Kirsner et al., 1984).

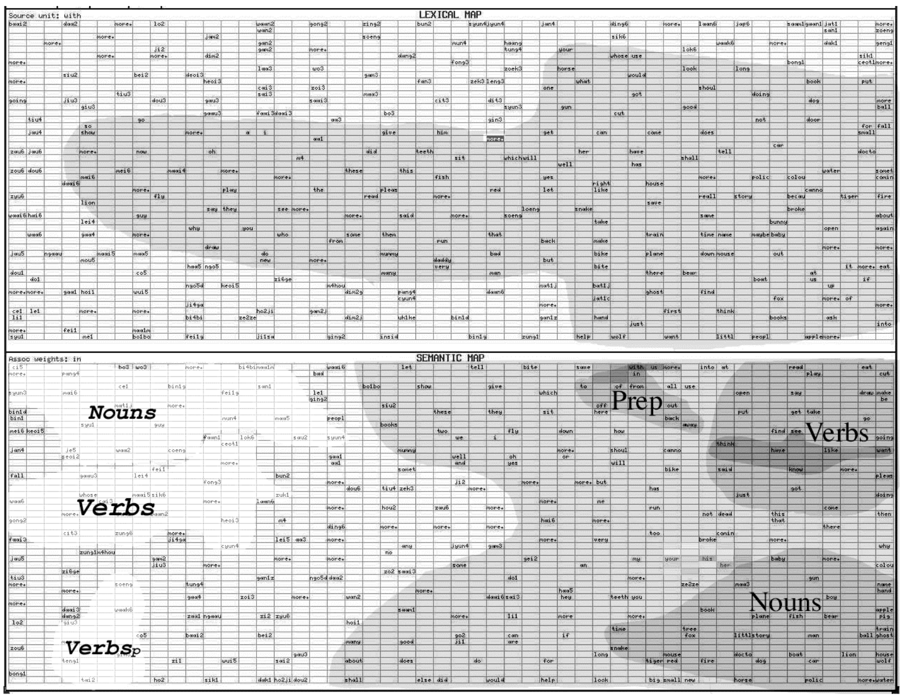

Figure 3. Emergence of lexical categories (e.g., nouns, verbs, and subcategories in Chinese and English) in SOM2.

Lexical Categories in the Bilingual Lexicon

Given that our model was able to clearly differentiate the two lexicons, does it show any ability in developing categories within each language? In other words, is our network able to organize the phonological and semantic features into structurally meaningful clusters? Our previous research shows that SOMs are good at deriving semantic categories of the lexicon, even in cases where linguists fail to identify the categories with clear labels (e.g., Whorf's cryptotypes; Li, 1999, in press). A further examination of Figure 2 reveals that the network developed a considerable amount of structure on its semantic organization.[3] Figure 3 presents the same results as in Figure 2, with illustrative labels on the lexical categories in SOM2.

In Figure 3, we can see that the SOM2 treated the nouns and verbs distinctly, in both Chinese and English. The network also grouped other words together as clusters that share the same grammatical categories (e.g., English prepositions and pronouns occurred in the same neighborhood). More interestingly, within each of the categories (assuming there are enough instances), semantically similar words also occurred together. For example, in English, the state verbs (e.g., *know*, *like*, *have*, *want*) were grouped together (the darkest shaded area under the label

"Verbs"), in contrast to other activity verbs;[4] in Chinese, words that are related to cooking/eating (within the verb category) were grouped together (e.g., *sik* 'eat', *jam* 'drink', *cit* 'cut', *gaau* 'gnaw'). One can also see a category of verbs indicating perceptual/mental activities in Chinese (marked as "Verbs-p", including *teng* 'listen', *tai* 'look', *gin* 'see', *seong* 'think'). These effects of categorical emergence from statistical learning match up with results from our previous monolingual simulations (Li, 1999, 2000, in press; Farkaš & Li, in press).

Interlingual Priming and Interference

In the empirical literature, there are heated debates on whether bilingual lexical representation is language-specific or language-independent, and whether the two lexicons are stored separately or together (see Smith, 1997, for a summary). These debates determine how researchers look at well-known effects of cross-language priming and interference in empirical studies (e.g., de Groot & Nas, 1991; Kirsner et al., 1984). In a way, our above simulation results already indicated that the two lexicons can reside in an integrated network, but separately in space. Given that the two lexicons exhibit distinct behaviors in our network, and that there is no language-independent conceptual representation in our network (as is in many other bilingual models, see Conclusions), how can we account for priming effects or interference that show that the two lexicons do interact with each other? Recall that our procedure involved training the mapping between word forms and meanings in such a way that the mapping is not one-to-one, but many-to-many (for nouns and verbs, at least). Thus, words that are considered as translation equivalents are linked together through associative pathways that are trained by Hebbian learning.

After training the network on the bilingual speech corpus for 500 epochs on the 400 words, we tested the network's performance on comprehension (from phonology to semantics) and production (from semantics to phonology). Probing into the network's associative pathways is laborious, which involves examining each of the 400 words on SOM1 or SOM2 to see what words it can activate on the other map. Our analyses indicate that there are clear effects of priming and interference. To illustrate with examples, the activation of the following words on SOM2 (left) led to the activation of words on SOM1 (right) via the associative links (items in parentheses are translations of Chinese words):

SOM2 SOM1
bear → *bear*, *there*, *fish*, *this*
jat (one) → *jat* (one), *loeng* (two), *soeng* (think), *saam* (three), *san* (new)
bibi (baby) → *bi4bi1* (baby), *maami* (mummy), *baby*, *maybe*, *mummy*
sik (eat) → *it*, *eat*, *sik* (eat), *maai* (buy)

These are some representative examples that indicate semantic as well as phonological priming effects in production. For example, *bear* activates *fish* because *fish* was semantically related to *bear* and SOM2 placed it next to *bear*. *There* and *this* were activated because of their phonological similarity to *bear* and *fish*, respectively (SOM1 placed *there* next to *bear*, and *this* next to *fish*). The second example in Chinese (*jat*) can also be similarly explained. These two examples show only within-language priming effects – no translation equivalents exist in our simulated lexicon. The next two examples show cross-language priming effects: *mummy* became activated because of its semantic similarity to *baby*, which in turn was activated by the Chinese word *bibi*. In addition, *maybe* was activated because of its phonological similarity with *baby* on SOM1. The last example, *sik* 'eat', is a similarly interesting case. Its strongest response on SOM1 was *it* and not *eat*, a phonologically similar word. Note that these priming effects represent possible interference in language production. Because of strong associative pathways, the speaker might have said *it* instead of *sik* or *eat*, especially in a strong bilingual speech mode (Grosjean, 1997).

Similarly, our model was also able to show cross-language interference effects in comprehension. For example:

SOM1 SOM2
look → *lok* (fall), *fall*, *look*, *tai* (look), *ceot* (go out).
boat → *boat*, *dog*, *car*, *syun* (boat), *sau* (hand)
waan (play) → *waanx* (play), *wan* (look for), *play*, *read*, *find*, *see*
sik (eat) → *eat*, *cut*, *sik* (eat), *maai* (buy)

Again, cross-language priming resulted from both semantic and phonological similarities as a function of the co-activation of forms (SOM1) and meanings (SOM2). For example, because of the phonological similarity of *look* and *lok* on SOM1, *look* activated *lok* in SOM2, which in turn activated *fall* as a result of semantic similarity (*lok* means "fall"). Li (1996a) found that bilingual Chinese-English speakers do not need more time to process cross-language homophones like *look*/*lok* than do monolinguals in processing monolingual materials, when the homophone occurs in a sentence context; in other words, there was no switching cost associated with the recognition of code-switched English words. This finding contradicts the notion of a language switch or monitor mechanism that says that switching languages takes time (Macnamara, 1967; Obler & Albert, 1978; see also Green, 1998). Li suggested that a parallel activation mechanism might be involved: multiple candidates in the bilingual lexicon are activated, and sentence context plays a crucial role in affecting the final activation of the target word (see also Altarriba et al., 1996, for a similar argument). The results shown in this simulation are consistent with the view that multiple candidates may be simultaneously activated and selected due to effects of cross-language priming.

Effects of Proficiency and Individual Differences

To extend the scope of the SOMBIP model discussed above, we modeled the effects of proficiency and individual differences in two separate simulations using the same bilingual lexicon. Proficiency was modeled by having SOM2 self-organize on "novice" word representations – representations that come out of the WCD network's learning of a smaller sample of English sentences (92,640 word tokens, as compared with 555,837 tokens in the above), simulating a beginning Chinese learner of English. Figure 4 presents a snapshot of the network's representation with the same learning parameters as in Figures 1 and 2.

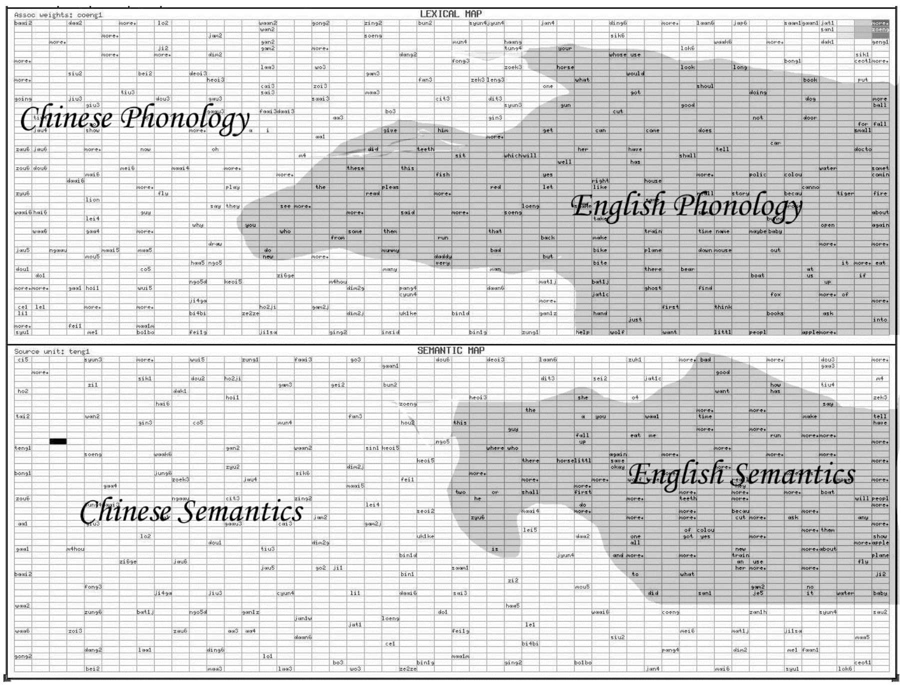

Figure 4. Effects of proficiency in lexical representations. English semantic representations in SOM2 were based on the WCD network's learning of a limited amount of input (92,640 word tokens).

As is illustrated in SOM2 of Figure 4, the semantic representations of the English words were squeezed into the upper right-hand corner, and occupy only a very small portion of the map. Although there was some rudimentary structure in these vectors (e.g., *good* and *bad* were nearby on the map, so were *say* and *tell*), the semantic vectors from the WCD network were still fuzzy and did not represent the word meanings accurately; thus the SOM could not fully differentiate them and accordingly organize them into meaningful structures. This picture shows that for novice bilingual speakers, the bilingual mental lexicon is still largely monolingual, and words in the second language are largely clamped together. It is also consistent with the view that the bilingual lexicon is a dynamically developing system, one that changes as a function of which language becomes the dominant language (Heredia, 1997). As the bilingual speaker becomes more proficient and fine-tunes the semantic system (amongst other developments, e.g., an expanding size of the L2 lexicon), this "dominance" situation may shift to an L2 instead of L1 dominance (Heredia & Altarriba, in press).

Finally, to simulate individual differences in the representation of bilingual lexicon, we constructed a smaller-in-size but otherwise identical network as in previous simulations. In the simulations reported above, we used 2,500 units each (a 50×50 grid) in SOM1 and SOM2. In the new simulation, we used 1,600 units (a 40×40 grid). This reduction in map size is meant to simulate a bilingual processor that has limited resource (e.g., working memory capacity) at disposal. There is evidence that individual differences in language processing are often due to working memory capacity (Just & Carpenter, 1992). Our network that used a smaller map size might thus show disrupted representational structure or processing abnormalities.

Analyses of this manipulation show that the network was able to display language separation, lexical categories, and priming and interference effects as in the bigger network. However, given the same amount of training (500 epochs), the network had difficulty in distinguishing between many words that are either phonologically or semantically similar. In other words, the same units on the SOM often represented two or more similar lexical items. Recognition rate dropped to 57% for SOM1 (as compared to 74% with the 50×50 map) and to 54% for SOM2 (as compared to 74% with the 50×50 maps).[5] As a result of the decrease in recognition fidelity, the network's associative links were much more diffuse than those in the larger networks. Thus, there were stronger within and cross-language priming and interference effects as compared with the bigger network, some of which could not be explained easily. For example, activation of *let* on SOM1 caused *bei* (let), *let*, *show*, *and*, *mouse*, and *red* to become activated on SOM2; activation of *go* on SOM2 caused *say*, *waa* (say), *waai* (bad), *hai* (be), *gong* (say), *hoei* (go), and *know* to become activated on SOM1, and so on. These results indicate that the unusual patterns in bilingual processing may be due to diffuse bilingual representations, which might in turn be due to limited processing capacity.

In sum, manipulations of size of the learning corpus and size of the network extend the capabilities of our model in accounting for a variety of bilingual phenomena.

General Discussion and Conclusions

In this chapter we provided an overview of connectionist bilingualism, and presented a sketch of SOMBIP, a self-organizing connectionist model of bilingual processing. A number of theoretical and architectural considerations of the model are discussed, and preliminary simulation results from the model are analyzed. Our SOMBIP model consists of two interconnected self-organizing neural networks, coupled with a recurrent neural network that computes lexical co-occurrence constraints. Simulations with our model indicate that (1) the model can account for distinct patterns of the bilingual lexicon without the use of language nodes or language tags, (2) it can develop meaningful lexical-semantic categories through self-organizing processes, and (3) it can account for a variety of priming and interference effects based on associative pathways between phonology and semantics in the lexicon, and (4) it can explain lexical representation in bilinguals with different levels of proficiency and working memory capacity. These capabilities of our model are due to its design characteristics in that (a) it combines localist and distributed properties of processing, (b) it combines representation (through self-organization) with learning (recurrent neural network WCD, Hebbian learning of associative pathways), and finally (c) it combines lexicon and sentences in bilingual processing. Most previous connectionist bilingual models (e.g., BIA and BIMOLA) are stationary models. To fully capture the dynamics in bilingual language processing, a model needs to incorporate the above mentioned properties, especially the learning characteristics such as in SOM and WCD.

Green (1998) proposed an inhibitory control (IC) model of bilingual lexical processing. Although the IC model is not a connectionist model – on the contrary, it has quite some symbolic AI flavor, for example, with goal-oriented decision boxes and control schemas – it attempts to integrate activation-based accounts (e.g., interactive activation mechanisms). According to Green, there are multiple levels of control, with each level associated with a specific schema, from high-level event scripts to low-level articulatory controls. The particular level at which the IC model operates is an intermediate level, the lemma level, whereby an inhibitory mechanism suppresses the activation of lemmas that are tagged as belonging to the language other than the intended one. Crucial to the functioning of this mechanism are the language tags, tags that are believed to be part of the conceptual system of the lexicon. Green likened the language tags to the language nodes in the BIA model, which function to reinforce lexical activation of the currently activated language, while simultaneously decreasing lexical activation in the other lexicon. Interestingly, in a brief note, Dijkstra and van Heuven (1998) mentioned that the BIA language nodes might be thought to represent abstractions of very different statistical regularities in the two languages due to lexical co-occurrence differences. This view differs from what Green assumed, but is completely consistent with our argument here (in SOMBIP, lexical co-occurrence statistics are explicitly incorporated through WCD).

In a commentary on the IC model, Li (1998a) questioned the necessity of language tags. One piece of evidence against the notion of language tags in IC is that some recent studies show that there need not be a cost associated with language switching, especially in natural speech situations (Grosjean, 1988, 1997; Grosjean & Miller, 1994; Li, 1996a) – according to IC, language switching takes time, since the process of switching to another language involves the inhibition of previous language tags. In the IC model, the inhibition of a particular stimulus shuts down the activation of all other related stimuli in the same language from top down. This assumption seems to contradict several activation-based accounts that the bilinguals' two languages may be always activated, though the strength of the activation differs in specific linguistic situation, depending on the frequency of the target words, the sentence context, the speaker's proficiency in the two languages, and the speech mode (Grosjean, 1988, 1997; Li, 1996a). Thus, Li (1998a) argued that it is unclear whether the language tags are ad hoc constructs or necessary components of bilingual processing, just as it is unclear whether language nodes are necessary. Li suggested that the seemingly separate representations of the two lexicons, and the related interlingual priming/interference effects, might arise as a result of lexical and grammatical learning in a self-organizing neural network, in which no distinct labels are given to items of the two or more languages. Specifically, words from both languages may exist in the same topological map, but over time the network can develop localized patterns of activity in learning the mappings between phonology, orthography, morphology, and semantics. These localized patterns of activity may then correspond to the learner's internalized, distinct representations of the two lexicons. In this study, we have implemented precisely such a model, the SOMBIP, which nicely accounts for the independent (yet interactive) representations of the bilingual lexicon. A supra-lexical level of language nodes or language tags is absent in SOMBIP, but the effects of the language nodes or tags are captured in the system.

In response to Li's commentary, Green agreed that the bilingual lexico-semantic system is composed of self-organized networks, as a presumption of the subset hypothesis (Paradis, 1989) adopted by the IC model. But he argued that language tags are necessary and that even in self-organizing networks such as the one used by Miikkulainen (1997) there is still room for language tags. Interestingly, Miikkulainen (1993) did mention a multiple SOM network that functions exactly as Green would argue: each language has its own separate lexicon, stored on different maps. A high-level control network would determine which language modules are active, switching between conversations in two languages. Such an approach was intended to model, according to Miikkulainen, distinct language behaviors in the representation of the bilingual lexicon in the brain. But Miikkulainen also speculated on a different (in our view a more plausible) model, one that is very similar in spirit to SOMBIP. This model would represent all word forms from the multilingual lexicon in a single lexical map, and all semantic representations in a single semantic map. These two maps could be associated with each other via the associative links. Cross-language priming and interference ("slips between languages" in his terms) could be modeled by noisy semantic mapping (like the diffuse map in our limited-resource network) or by associative propagation. In this version of the model, there would still be a high-level control network, but the task of the control mechanism merely modulates language output such that the same sentence parser could work on different language materials. In SOMBIP, however, this control network was unnecessary as the WCD network works equally efficiently in both languages.

The ability of our network to show distinct patterns of the bilingual lexicon without distinct mechanisms (e.g., separate language tags or nodes) is yet another case of the classical connectionist argument: modular linguistic behaviors that are different or dissociated need not arise from dual-mechanisms or dual-routes in the representational or processing system. They can naturally emerge from learning the input characteristics within a single system, a system that can be operationalized in terms of statistical or connectionist principles. This argument was advanced by the original PDP studies of the English past-tense acquisition (Rumelhart & McClelland, 1986), and has since been supported by research in many domains of language studies, including language acquisition, normal and impaired word reading, and language organization in the brain (e.g., Elman et al., 1996; Plaut et al., 1996; Miikkulainen, 1997; Small et al., 1995).

As a final important note, our model departs from the tradition that has been adopted by many bilingual researchers (see Kroll & de Groot, 1997, for discussion) in that it does not assume there is an integrated level of meaning-concept or meaning and concept (the "conceptual level"). The mixture of meaning with concept has created much confusion, in both first language and second language acquisition research (see Li & Shirai, 2000). This confusion is further complicated by the fact that many psycholinguists adopt a language-specific lexical-semantic level, the lemma level (Levelt, 1989). To avoid these confusions, thus, in our model there is no language-independent "conceptual" representation. Our semantic representation, or more accurately, lexical representation that contains both syntactic and semantic information, comes out the WCD network's analyses of global lexical co-occurrence information. This level of representation matches more closely with what linguists call "meaning", rather than "denotata" (Lyons, 1977) – entities of concepts that many bilingual researchers use in their picture naming tasks.

To conclude, the SOMBIP model outlined in this chapter serves as a new model of bilingual processing and provides a new perspective on connectionist bilingualism. It has the potential of explaining a wide variety of empirical and theoretical issues in bilingual research. So far, we have tested the SOMBIP model only on spoken (transcribed) bilingual Chinese-English data. How the model generalizes to other bilingual situations should be further tested. Chinese differs significantly from English and other Indo-European languages in phonological, syntactic, and lexical structures and these differences are bound to affect language processing (Li, 1996b, 1998b). When one examines bilingual pairs of languages that are more similar (e.g., Dutch and German), it is likely that the model will show more integrated phonological and semantic representations of the bilingual lexicon. Moreover, if the mixed bilingual data contain more intra-sentential code-switches (rather than inter-sentential as in our data), the network that receives such data as input might also produce more integrated semantic representations. In addition, the current SOMBIP design could easily be extended to simulate orthographic-semantic relationships, and be evaluated against empirical findings and results from the BIA model. Other interesting extensions of the model would include simulating impaired semantic, phonological, or orthographic structures by adding noise to the SOM maps and by damaging the associative pathways that connect word form and word meaning. Such modeling efforts could shed light on bilingual representation and processing in the brain (Fabbro, 1999).

Notes

[1] The current scheme represents a scaled-down version of the full tri-syllabic template representation of English (see Li & MacWhinney, 2001, for the full version). As such, it is limited in certain respects; for example, it is not able to represent consonant clusters as in *string* /strIN/, which would require additional C-slots (e.g., CCCVVC for one syllable).

[2] It is impossible to show all the details of the original figure here because each of the 2-D maps contains 400 words spread on a 50×50 grid of 2,500 units.

[3] The organizations of the phonological representations are mainly based on phonemic and template similarities, and are relatively less interesting for our purpose here.

[4] The distinction between state and activity verbs has attracted significant debates concerning the nature and origin of this distinction in language acquisition. Some researchers propose that this distinction is innately given (Bickerton, 1981, 1984). We have argued that this distinction can arise from children's statistical learning of the semantic properties of verbs, in the same way as it would occur distinctly in our connectionist networks (Li & Shirai, 2000).

[5] The recognition rate is calculated as the rate of correctly mapped words in the SOM over the total number of words in the lexicon. Whether a word is mapped correctly is determined by a function that calculates the nearest distance between the input and the codevectors in the map. Note that the recognition rate of 74% is not too impressive at first glance. This is because the network learns a many-to-many mapping function for the noun and verb translation equivalents. If the network learns a one-to-one mapping, the recognition rate would go up to 93% and 97% for SOM1 and SOM2, respectively.

References

Altarriba, J., Kroll, J.F, Sholl, A., & Rayner, K. (1996). The influence of lexical and conceptual constraints on reading mixed-language sentences: Evidence from eye fixations and naming times. *Memory & Cognition, 24,* 477-492.

Bickerton, D. (1981). *Roots of language*. Ann Arbor, MI: Karoma.

Bickerton, D. (1984). The language bioprogram hypothesis. *Behavioral and Brain Sciences, 7*, 173-188.

Bowerman, M. (1978). The acquisition of word meaning: An investigation into some current conflicts. In N. Waterson & C.E. Snow (Eds.), *Development of communication* (pp. 263-287). New York: John Wiley and Sons.

Bowerman, M. (1982). Reorganizational processes in lexical and syntactic development. In E. Wanner & L.R. Gleitman (Eds.), *Language acquisition: The state of the art* (pp.319-346). Cambridge, UK: Cambridge University Press.

Burgess, C. & Lund, K. (1997). Modelling parsing constraints with high-dimensional context space. *Language and Cognitive Processes, 12*, 1-34.

Burgess, C. & Lund, K. (1999). The dynamics of meaning in memory. In E. Dietrich & A. Markman, (Eds.), *Cognitive dynamics: Conceptual and representational change in humans and machines* (pp. 17-56). Mahwah, NJ: Erlbaum.

Dijkstra, T. & van Heuven, W. (1998). The BIA model and bilingual word recognition. In J. Grainger & A.M. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 189-225). Mahwah, NJ: Erlbaum.

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

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.

Fabbro, F. (1999). *The neurolinguistics of bilingualism: An introduction*. East Sussex, UK: Psychology Press.

Farkaš, I. & Li, P. (2001). Modeling the development of lexicon with a growing self-organizing map. Manuscript submitted.

Farkaš, I. & Li, P. (in press). A self-organizing neural network model of the acquisition of word meaning. In E. Altmann, A. Cleeremans, C. Schunn, & W. Gray (Eds.), *Proceedings of the 4th International Conference on Cognitive Modeling*.

French, R.M. (1998). A simple recurrent network model of bilingual memory. In M.A. Gernsbacher & S.J. Derry (Eds.), *Proceedings of the 20th Annual Conference of the Cognitive Science Society*, (pp. 368-373). Mahwah, NJ: Erlbaum.

Grainger, J. (1993). Visual word recognition in bilinguals. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 11-26). Amsterdam/Philadelphia: John Benjamins.

Green, C.D. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism, 1*, 67-82.

de Groot, A.M.B. & Nas, G.L.J. (1991). Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory & Language, 30,* 90-123.

Grosjean, F. (1988). Exploring the recognition of guest words in bilingual speech. *Language and Cognitive Processes, 3*, 233-274.

Grosjean, F. (1997). Processing mixed languages: Issues, findings and models. In A.M.B. de Groot & J.F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 225-254). Mahwah, NJ: Erlbaum.

Grosjean, F. & Miller, J. (1994). Going in and out of languages: An example of bilingual flexibility. *Psychological Science, 5,* 201-206.

Hebb, D. (1949). *The organization of behavior: A neuropsychological theory.* New York: John Wiley.

Heredia, R.R. (1997). Bilingual memory and hierarchical models: A case for language dominance. *Current Directions in Psychological Science, 6,* 34-39.

Heredia, R.R. & Altarriba, J. (in press). Bilingual language mixing: Why do bilinguals code-switch? *Current Directions in Psychological Science*.

van Heuven, W.J.B. (2000). Visual word recognition in monolingual and bilingual readers: Experiments and computational modeling. Unpublished doctoral dissertation, NICI, University of Nijmegen, Netherlands.

Hinton, G.E. & Sejnowski, T. (1999). *Unsupervised learning: Foundations of neural computation.* Cambridge, MA: MIT Press.

Hong Kong Linguistic Society. (1997). *Yueyu pinyin zibiao* (Cantonese romanization scheme). Hong Kong Linguistic Society.

Just, M.A. & Carpenter, P.A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122-149.

Kirsner, K, Smith, M.C., Lockhart, R.S., King, M-L., & Jain, M. (1984). The bilingual lexicon: Language-specific effects in an integrated network. *Journal of Verbal Learning and Verbal Behavior, 23,* 519-539.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, *43*, 59-69.

Kohonen, T. (1989). *Self-organization and associative memory* (3rd ed.). Berlin: Springer-Verlag.

Kohonen, T. (1995). *Self-organizing maps.* Heidelberg: Springer.

Kroll, J.F. & de Groot, A.M.B. (1997). Lexical and conceptual memory in the bilingual: Mapping from meaning in two languages. In A.M.B. de Groot, & J.F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 169-200). Mahwah, NJ: Erlbaum.

Lachter, J. & Bever. T. (1988). The relation between linguistic structure and associative theories of language learning: A constructive critique of some connectionist learning models. *Cognition*, *28,* 195-247.

Ladefoged, P. (1982). *A course in phonetics* (2nd ed). San Diego, CA: Harcourt Brace.

Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Léwy, N. (in progress). Computational psycholinguistics and spoken word recognition in the bilingual and the monolingual. Unpublished doctoral dissertation, University of Neuchâtel, Switzerland.

Léwy, N. & Grosjean, F. (in progress). The computerized version of BIMOLA: A bilingual model of lexical access. Manuscript, University of Neuchâtel, Switzerland.

Li, P. (1996a). Spoken word recognition of code-switched words by Chinese-English bilinguals. *Journal of Memory and Language, 35,* 757-774.

Li, P. (1996b). The temporal structure of spoken sentence comprehension in Chinese. *Perception and Psychophysics, 58*, 571-586.

Li, P. (1998a). Mental control, language tags, and language nodes in bilingual lexical processing. *Bilingualism, 1,* 92-93.

Li, P. (1998b). Crosslinguistic variation and sentence processing: The case of Chinese. In D. Hillert (Ed.), *Sentence processing: A crosslinguistic perspective* (pp. 33-51). San Diego, CA: Academic Press.

Li, P. (1999). Generalization, representation, and recovery in a self-organizing feature-map model of language acquisition. In M. Hahn & S.C. Stoness (Eds.), *Proceedings of the 21st Annual Conference of the Cognitive Science Society* (pp. 308-313). Mahwah, NJ: Erlbaum.

Li, P. (2000). The acquisition of lexical and grammatical aspect in a self-organizing feature-map model. In L. Gleitman, & A. Joshi (Eds.), *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 304-309). Mahwah, NJ: Erlbaum.

Li, P. (in press). Language acquisition in a self-organizing neural network. In: P. Quinlan (Ed.), *Connectionism and developmental theory*. Philadelphia/Briton: Psychology Press.

Li, P. & MacWhinney, B. (2001). PatPho: A phonological pattern generator for neural networks. Manuscript submitted for publication.

Li, P. & Shirai, Y. (2000). *The acquisition of lexical and grammatical aspect.* Berlin/New York: Mouton de Gruyter.

Lyons, J. (1977). *Semantics* (Vol. 2). Cambridge, UK: Cambridge University Press.

Macnamara, J. (1967). The bilingual's linguistic performance: A psychological overview. *Journal of Social Issues*, *23*, 59-77.

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

MacWhinney, B. (2000a). Lexicalist connectionism. In P. Broeder & J.M. Murre (Eds.), *Models of language acquisition: Inductive and deductive approaches* (pp. 9-32). Oxford, UK: Oxford University Press.

MacWhinney, B. (2000b). *The CHILDES project: Tools for analyzing talk* (3rd ed). Mahwah, NJ: Erlbaum.

McClelland, J.L. & Elman, J. (1986). Interactive processes in speech perception: The TRACE model. In J.L. McClelland, D.E. Rumelhart, & the PDP research group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2, pp. 58-121). Cambridge, MA: MIT Press.

McClelland, J.L. & Kawamoto, A.H. (1986). Mechanisms of sentence processing: Assigning roles to constituents. In J.L. McClelland, D.E. Rumelhart, & the PDP research group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2, pp. 272-325). Cambridge, MA: MIT Press.

McClelland, J.L. & Rumelhart, D.E. (1981). An interactive activation model of context effects in letter perception. Part 1: An account of basic findings. *Psychological Review*, *88*, 375-405.

Miikkulainen, R. (1993). *Subsymbolic natural language processing: An integrated model of scripts, lexicon, and memory*. Cambridge, MA: MIT Press.

Miikkulainen, R. (1997). Dyslexic and category-specific aphasic impairments in a self-organizing feature map model of the lexicon. *Brain and Language, 59*, 334-366.

Miikkulainen, R. (1999). *The DISLEX simulator* (new version) [on-line]. Available at ftp://ftp.cs.utexas.edu/pub/neural-nets/software/dislex.2.1.tar.Z.

Obler, L. & Albert, M. (1978). A monitor system for bilingual language processing. In M. Paradis (Ed.), *Aspects of bilingualism* (pp. 105-113). Columbia, SC: Hornbeam Press.

Paradis, M. (1989). Bilingual and polyglot aphasia. In F. Boller, & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 2, pp. 117-140). Amsterdam: Elsevier.

Plaut, D.C., McClelland, J.L., Seidenberg, M.S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*, 56-115.

Rumelhart, D. & McClelland, J. (1986). On learning the past tenses of English verbs. In J. McClelland, D. Rumelhart, & the PDP research group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2, pp. 216-271). Cambridge, MA: MIT Press.

Small, S.L., Hart, J., Nguyen, T., & Gordon, B. (1995). Distributed representations of semantic knowledge in the brain. *Brain, 118*, 441-453.

Smith, M.C. (1997). How do bilinguals access lexical information. In A.M.B. de Groot & J.F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 145-169). Mahwah, NJ: Erlbaum.

Spitzer, M. (1999). *The mind within the net.* Cambridge, MA: MIT Press.

Yip, V. & Matthews, S. (2000). Syntactic transfer in a Cantonese-English bilingual child. *Bilingualism, 3*, 193-208.

Acknowledgments

This research was supported by a grant from the National Science Foundation (#BCS-9975249) to the first author. We would like to thank François Grosjean and Nicolas Léwy for comments and suggestions on an earlier draft of the paper, and Kathy Reid and Clint Schlenker for their editorial assistance and data preparation. The simulations were run on a Pentium PC with Linux OS. We used the DISLEX code (Miikkulainen, 1999) to run the SOM1 and SOM2 and their associative links, and the WCD code was written by the second author. Correspondence concerning this paper should be addressed to Ping Li, Department of Psychology, University of Richmond, Richmond, VA 23173. E-mail: pli@richmond.edu.