International Journal of Bilingual Education and Bilingualism
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t794297780

Bilingual lexical interactions in an unsupervised neural network model

Xiaowei Zhao[a]; Ping Li[b]
[a] Department of Psychology, Colgate University, Hamilton, NY, USA [b] Department of Psychology, Pennnsylvania State University, PA, USA

Online publication date: 16 August 2010

To cite this Article Zhao, Xiaowei and Li, Ping(2010) 'Bilingual lexical interactions in an unsupervised neural network model', International Journal of Bilingual Education and Bilingualism, 13: 5, 505 — 524
To link to this Article: DOI: 10.1080/13670050.2010.488284
URL: http://dx.doi.org/10.1080/13670050.2010.488284

International Journal of Bilingual Education and Bilingualism
Vol. 13, No. 5, September 2010, 505–524

Routledge
Taylor & Francis Group

Bilingual lexical interactions in an unsupervised neural network model

Xiaowei Zhao[a]* and Ping Li[b]

[a]*Department of Psychology, Colgate University, Hamilton, NY, USA;* [b]*Department of Psychology, Pennnsylvania State University, University Park, PA, USA*

(*Received 4 February 2009; final version received 24 December 2009*)

In this paper we present an unsupervised neural network model of bilingual lexical development and interaction. We focus on how the representational structures of the bilingual lexicons can emerge, develop, and interact with each other as a function of the learning history. The results show that: (1) distinct representations for the two lexicons can develop in our network when the two languages are learned simultaneously; (2) the representational structure is highly dependent on the onset time of the second language (L2) learning if the two languages are learned sequentially; and (3) L2 representation becomes parasitic on the representation of the first language when the learning of L2 occurs late. The results suggest a dynamic developmental picture for bilingual lexical acquisition: the acquisition of two languages entails strong competition in a highly interactive context and involves limited plasticity as a function of the timing of L2 learning.

Keywords: bilingual interaction; lexical development; neural network; DevLex

Introduction

An issue of enduring interest in bilingualism research has been how the two linguistic systems are represented, developed, and influenced by each other in the bilingual's mind. Despite significant progress in the field of bilingualism, the underlying computational mechanisms of early bilingual lexical acquisition are still poorly understood. In the empirical literature, there has been intense debate on whether bilingual representation takes the form of a single, shared lexical storage or a separate, distinct storage for the mental lexicon (see Dong, Gui, and MacWhinney 2005; French and Jacquet 2004; Kroll and Tokowicz 2005, for recent reviews). For example, in the study of early childhood bilingualism, some investigators have argued for the *unitary language system hypothesis* (e.g. Volterra and Taeschner 1978), according to which bilingual children who acquire two languages simultaneously often start with a single fusion system that combines the representations of both languages. This unitary representational system gradually differentiates into two systems that handle the two languages separately when the bilingual child grows older. In contrast, other researchers have argued for the *dual language system hypothesis* (Genesee 1989), according to which two separate linguistic systems are developed simultaneously in the child from the onset of their language acquisition. A new perspective to this debate is to ask, not whether there is a single or a dual system, but, rather, to what extent differences or separation

*Corresponding author. Email: xiaoweizhao@gmail.com

ISSN 1367-0050 print/ISSN 1747-7522 online

DOI: 10.1080/13670050.2010.488284
http://www.informaworld.com

exist and at what level they do so (such as the meaning or conceptual level vs. the phonological or morphosyntactic level; see Pavlenko 2009).

The issue of bilingual representation has recently been further complicated by conflicting neuroimaging data with regard to distinct or common neural substrates in bilingual language processing (see a review in Li 2009). The evidence so far seems to point to common neural systems for the processing of both first language (L1) and second language (L2), but the picture is muddied by the lack of careful consideration of the bilingual's learning history, the age effect, the proficiency and dominance of L1 vs. L2, and similarity distances between the bilingual's two languages. Given this situation, it is important to recognize that a host of variables must be taken into consideration in dealing with bilingual representation and the interaction between L1 and L2, such as bilingual proficiency, learning history, modality (comprehension vs. production), and word types (cognates vs. noncognates, abstract vs. concrete words; Van Hell and de Groot 1998).

Computational models offer particular advantages in dealing with complex interactions between variables by systematically bringing target variables under experimental control while holding other variables constant (see a recent review of computational cognitive models in McClelland 2009). Although the power of experimental research also lies in systematic control of variables, in natural language learning situations, especially in the bilingual case, it is often difficult to directly manipulate the learning environment in parametric ways (see a metaphor by Bialystok 2001 comparing bilingualism with the smorgasbord of food serving). In this paper, we plan to demonstrate the facility and utility of neural network models (or 'connectionist models') in the study of bilingualism.

A neural network is a computational model made of information processing units (neurons) that are connected in a network. The basic tenet of neural networks is that the brain works by co-ordinating the activation of large groups of neurons in response to particular cognitive or perceptual tasks, and that the activation is crucially determined by the specific connections that hold between the neurons. The connections can take on different degrees of strength (weights), while learning can change the strengths or eliminate the connections altogether depending on the task. In addition, a large number of neurons as information processing units may be activated in parallel to handle a specific problem. Neural network models argue for a high degree of interactivity between various levels of information processing, in contrast to classical cognitive theories that assume modularized, often serial, and 'informationally encapsulated' processes (see Fodor 1983, for the latter). Thus, neural network theories assume that knowledge representation and acquisition is distributed, parallel, and interactive in nature. To illustrate, neural networks learn through the adaptation of weights, the strengths of connections that hold between multiple and parallel working units, which serve as a simplification of the synaptic connections among real neurons.[1] There have been many algorithms developed for adjusting the weights to an optimal set of configurations, which may lead to the appropriate activation patterns of units that represent new knowledge (Haykin 1999). In the views of neural network models, linguistic representations can be best understood as properties that emerge out of learning (i.e. 'emergent properties') rather than as built in a priori; they emerge owing to the interaction of the learning system with the linguistic environment (Elman et al. 1996).

Researchers in neural network modeling have for quite some time been concerned with issues in language development and language processing. Work in the monolingual context has shown that neural network models are ideally suited for identifying mechanisms underlying phenomena in early lexical acquisition, including the U-shape learning of the English past tense (Plunkett and Marchman 1991), acquisition of lexical categories (Li, Farkas, and MacWhinney 2004), the vocabulary spurt (Li, Zhao, and MacWhinney 2007; Regier 2005), and acquisition of aspect (Zhao and Li 2009a). Unfortunately, the gap between neural networks and bilingualism is still wide open: to date, there have been only a handful of neural network models that are designed specifically to account for bilingual language processing and representation (see reviews in French and Jacquet 2004; Li and Farkas 2002; Thomas and van Heuven 2005). Furthermore, no neural network model has been devoted to capturing the impact of developmental time on bilingual children's lexical representations and the interaction between them. For example, popular existing models in bilingual studies have been developed based on the famous Interactive Activation (IA) network of McClelland and Rumelhart (1981), including the Bilingual Interactive Activation (BIA) model (Dijkstra and van Heuven 1998), BIA+ model (Dijkstra and van Heuven 2002), and the Bilingual Model of Lexical Access (BIMOLA, Grosjean 2008). These models can be said to be 'permanent' or 'stationary' models because mechanisms of learning and adaptation for representation are missing in these models. Specifically, the connections and their weights are fixed and manually coded to capture proficient bilingual adults' language processing, for example, in visual or spoken word recognition. Due to the lack of learning, these models have difficulties simulating the developmental dynamics of bilinguals' lexical representations and interactions.

Finally, many previous neural network models rely on the use of artificially generated input representations, rather than training sets derived from actual speech input to the learners. Moreover, the size of the bilingual lexicon that the models can handle is often very small. The use of synthetic or highly simplified lexicons provides certain modeling conveniences in terms of analysis of the linkage between input and output. However, the highly idealized models often do not directly speak to developmental and interactive patterns in realistic learning situations.

Our current study takes these gaps as starting points for building computational models of developmental bilingualism, in particular, by simulating bilingual lexical representations and interactions with an unsupervised neural network model. The backbone of our simulation is an unsupervised neural network model, DevLex-II. This model and its predecessor, DevLex, have been developed to capture interactive developmental dynamics in language acquisition. The models rely on simple but powerful computational principles of unsupervised and Hebbian learning. We have applied them successfully to account for a variety of empirical phenomena in early monolingual lexical development (see Li, Farkas, and MacWhinney 2004; Li, Zhao, and MacWhinney 2007; Zhao and Li 2009a).[2] In this study we implement a variant of the DevLex-II model for the bilingual context and focus on how the representational structure of the bilingual lexicon can develop and change as a function of the learning history. In particular, we manipulate the onset time of lexical learning of the L2, in three scenarios: simultaneous – onset time of L2 co-occurs with that of the L1; early learning – onset time of L2 is slightly delayed relative to that of L1; and late learning – onset time of L2 lags significantly behind that of L1. We hypothesize that the representational structures for the two lexicons in our model would differ as a function of the learning history defined by L2 onset time. In addition, we hope to show how in each scenario the two developing lexicons compete and interact with each other, by analyzing the output of the model for comprehension and production errors and by examining the distances between the lexical representations of translation equivalents in the model's behavior.

Below, we first provide details on the construction of the model, followed by a discussion of the behavioral phenomena we wish to capture and explain. The success of the model and what it reveals are then explored.

The model

Self-organizing feature map (SOM): an unsupervised neural network

Unlike many popular neural network models of language that have been previously used (see review in Elman et al. 1996), our model relies on unsupervised learning algorithms that require no explicit 'teachers' to provide constant error corrections to the network. The model presented in this study is a variant of the so-called self-organizing feature map (SOM; Kohonen 2001).

To construct a SOM, a group of nodes (or neurons) are arranged on a two-dimensional lattice (i.e. a topographic map) for the organization of input representations, where each node on the map has input connections to receive external stimulus patterns. On the map, a node k has a vector $\vec{m}_k$ associated with it to represent the weights of the input connections to it. At each training step of SOM, an external input pattern (e.g. the phonological or semantic representation of a word in our study) is randomly chosen and presented to all the nodes on the map. This activates many nodes, according to how similar, by chance, the input pattern is to the weight vectors of the nodes; and the node that has the highest activation is declared the 'winner' (the Best Matching Unit, BMU). Once a node becomes active in response to a given input, the weight vectors of that node and its neighboring nodes (the neighbors) are adjusted, so that they become more similar to the input, and the nodes will respond to the same or similar inputs more strongly the next time. In this way, every time an input is presented, an area of nodes will become activated on the map (the 'activity bubbles') and the maximally active nodes are taken to represent the input. Initially activation occurs in large areas of the map, that is, large neighborhoods, but gradually learning becomes focused and the size of the neighborhood reduces. This process continues until all the inputs have found some maximally responding nodes as their BMUs. As a result of this self-organizing process, the map falls into a topography-preserving state, which means that the inputs with similar features will end up activating nodes in nearby regions, yielding meaningful activity bubbles that can be visualized on the map. This property is consistent with the observation of topographic maps in the sensory and motor areas in our brain, given that the signals from adjacent body regions are often projected to and processed by neighboring cortical areas (Spitzer 1999). For our purposes, this topography-preserving property allows us to model the emergence of semantic categories as a gradual process of lexical learning.

A sketch of the model

DevLex-II is a multi-layer, unsupervised, SOM-based neural network model, as diagrammatically depicted in Figure 1 (see Li, Zhao, and MacWhinney 2007, for details). It includes three basic levels for the representation and organization of

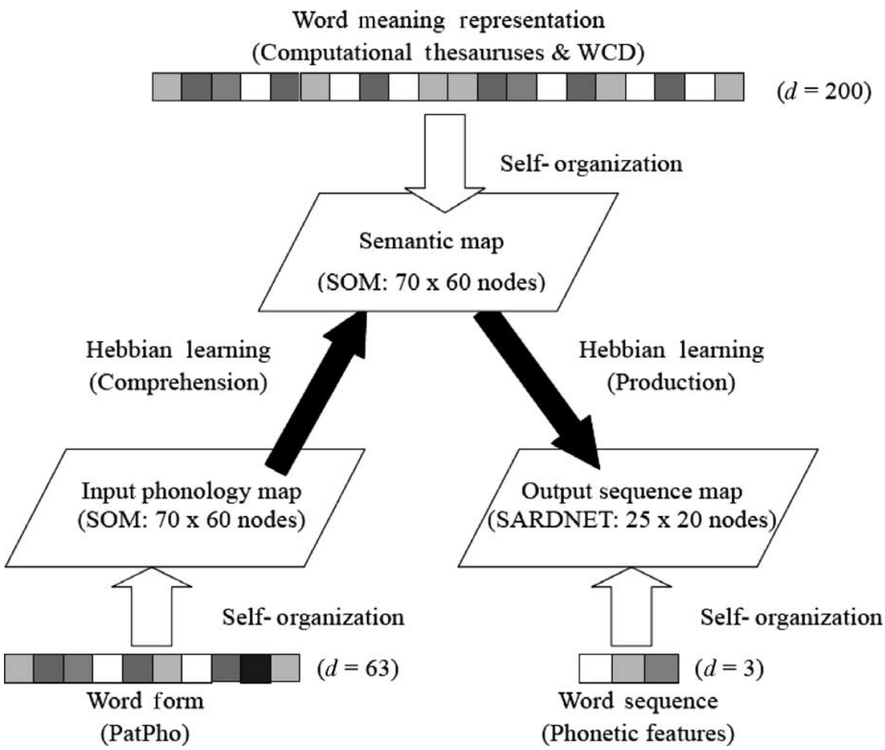

Figure 1. The architecture of the DevLex-II model. Each of the three self-organizing maps (SOM) takes input from the lexicon and organizes phonology, semantics, and phonemic sequence information of the vocabulary, respectively. The number of nodes in each map is indicated in parentheses. The dimension of the input vector for each map is indicated by '$d =$ ' in parentheses next to the input representation symbols. The maps are connected via associative links updated by Hebbian learning.

linguistic information: phonological content, semantic content, and the output sequence of the lexicon. The core of the model is a SOM that handles lexical–semantic representation. This SOM is connected to two other SOMs, one for input (auditory) phonology, and another for articulatory sequences of output phonology. Upon training of the network, the meaning representation, input phonology, and output phonemic sequence of a word are presented to and processed by the network. This process can be analogous to a child's analysis of a word's semantic, phonological, and phonemic information upon hearing a word.

On the semantic and phonological levels, the network forms representational patterns of activation according to the standard SOM algorithm as discussed above. Here, given a stimulus **x** (the phonological or semantic information of a word), a winner (i.e. BMU) on the corresponding SOM is found. After that, the weights of the nodes surrounding the winner in the neighborhood are updated. Unlike the SOMBIP model (Li and Farkas 2002), DevLex-II has a separate output sequence level. This level is slightly different from the other two levels where standard SOM is used. The addition of this level in the model is inspired by models of word learning based on temporal sequence acquisition. It is designed to simulate the challenge that language learners face when they need to develop better articulatory control of the phonemic sequences of words. Here, the activation pattern corresponding to the phonemic sequence information of a word is formed according to the algorithms of SARDNET (James and Miikkulainen 1995), a type of temporal or sequential SOM network (see Li, Zhao, and MacWhinney 2007, for further technical descriptions).

To better simulate language behaviors, previous researchers have linked together multiple SOMs to handle different linguistic aspects via cross-layer connections (see Li, Farkas, and MacWhinney 2004; Miikkulainen 1997). In DevLex-II, concurrent with the training of the three maps, the associative connections between maps are trained via Hebbian rule, a neurally inspired and biologically plausible mechanism of associative learning and memory (Cooper 2005), with a constant learning rate (see technical details in Li, Zhao, and MacWhinney 2007). The central idea here is that the weights of the associative connections between the frequently and concurrently activated nodes on two maps will become increasingly strong with training. To control the maximum weight value such a cross-map connection could have, we implemented a normalization process for the associative weights. This process allows for the gradual decrement of the weights of some associative links that are accidentally established at the early, and usually disorderly, stage of training (when the size of the 'activity bubble' is still large). As training progresses, the nodes connected by these links do not co-activate as frequently as before, therefore their weights gradually decay to zero. Meanwhile, there are links whose weights increasingly approach the maximum value ($=1$) because they connect those BMUs that consistently co-activate with each other at the late stage of training (when the size of the 'activity bubble' becomes small). After the cross-map connections are stabilized, the activation of a word form can evoke the activation of a word meaning via form-to-meaning links (to model word comprehension). If the activated unit on the semantic map is the BMU of the correct word meaning, we say that our network correctly comprehends this word; otherwise the network makes a comprehension error. Similarly, the activation of a word meaning can trigger the activation of an output sequence via meaning-to-sequence links (to model word production). If the activated units on the phonemic map are the BMUs of the phonemes making up the word in the correct order, we determine that our network correctly produces this word; otherwise the network makes an error in production.

Plasticity and stability in the model

To realistically simulate bilingual lexical development (especially L1 and L2 interaction) we must consider a fundamental problem called 'catastrophic interference' (see French 1999; Li, Farkas, and MacWhinney 2004). For example, if we train a network to acquire an L1 lexicon with 500 words and then train it on another 500 words in L2, in many traditional networks, the addition of L2 words may disrupt the network's knowledge of L1. In other words, the network will lose its representation of L1 words because of learning new words, which of course, is unlike human learning. This problem has been a 'plasticity–stability' dilemma in neural networks. Keeping the network's plasticity for new words often causes it to lose its stability for old knowledge; conversely, a network that is too stable often cannot adapt itself very well to the new learning task. To resolve this problem for our bilingual study, we introduced two new features into DevLex-II.

The first is a self-adjustable *neighborhood function*. For the standard algorithm of SOM, the radius of the neighborhood usually decreases according to a fixed training timetable. This type of development in the network, though practically useful, is subject to the criticisms that: (1) learning is tied directly (and only) to time (amount) of training, and is rather independent of the input-driven self-organizing process, and (2) the network often loses its plasticity for new inputs when the neighborhood radius becomes very small. In DevLex-II, we attempt to forestall these criticisms by using a learning process in which the neighborhood size is not totally locked with time, but is adjusted according to the network's learning outcome (experience). In particular, neighborhood function depends on the network's error level on each layer averaged across all the input patterns. Here, a 'quantization error' of an input pattern (as named in Kohonen 2001) is defined as the Euclidean distances (i.e. how similar) of the input pattern to the input weight vector of its BMU.[3] A second way in which we attempt to solve the plasticity–stability problem is to manage the training process as follows: for the input phonology map and the semantic map, during each training epoch, once a unit is activated as a BMU, it will become ineligible to respond to other inputs in the current training epoch. In this way, the old words are kept untouched in the training; the new words can be represented by novel units (new resources) on the maps. A similar procedure is also used for the output sequence map on the word level, where the same phoneme in different locations of a word will be mapped to different (but adjacent) nodes on the map. This mechanism resembles a process in which new neurons are recruited for novel inputs as computational resources become scarce (see Li, Farkas, and MacWhinney 2004, for an algorithm in new node recruitment). The use of a different but adjacent unit for new input is also empirically plausible: psycholinguistic research suggests that when young children encounter a novel word they tend to map it to a different category or meaning for which the child has not yet acquired a name (see Markman 1984; Mervis and Bertrand 1994).

Bilingual lexicons and input representations

To control for a host of extraneous variables in the study of bilingual lexicons, we used as our basis vocabulary from the CDI (Dale and Fenson 1996) for two languages, English and Chinese. The English lexicon was identical to that of Li, Farkas, and MacWhinney (2004). The Chinese lexicon was derived from the Chinese version of the CDI (Tardif, Gelman, and Xu 1999; Wu 1997). Each lexicon included 500 words chosen from the Toddler list of the corresponding CDI. The words were extracted roughly according to their order of acquisition by the toddlers, excluding homographs, word phrases, game words, words about time, words about place, and onomatopoeias.[4] The English lexicon included 286 nouns, 98 verbs, 51 adjectives, and 65 words in other categories; and the Chinese lexicon included 242 nouns, 145 verbs, 47 adjectives, and 66 words in other categories.

During the training of the model, the linguistics information was explicitly coded in the input representations. First, the sound pattern and phonemic makeup of a word were coded as the basic phonological input to the model. Second, the articulatory features of the 55 phonemes from the two languages (36 in Chinese; 38 in English) were coded and represented on the output sequence map. Third, the semantic information of words included two parts: (1) the co-occurrence probabilities of words in the discourse of the two languages, and (2) the 'semantics' of words, in the guise of word association, synonymy, and hyponymy, as represented in a normal thesaurus. The purpose of such a combination was to enhance the accuracy of our lexical representation (see Li, Farkas, and MacWhinney 2004, for rationale).

For the phonological input, we aligned the basic phonological patterns of English and Chinese words into a trisyllabic template with 18 phonemic slots according to PatPho, a generic phonological pattern generator for neural networks (Li and MacWhinney 2002; Zhao and Li 2009b). The template was CCCVVCCCVV CCCVVCCC, with Cs representing consonants and Vs representing vowels. Each phonemic slot consisted of three units which roughly represented the articulatory features of a phoneme (such a phonemic representation was also adopted for the output sequence map). In addition, a separate group of nine units was used to represent lexical tones in Chinese, and the values of these units were left empty for English.

Technically, the semantic information was entered into our model as a combination of two parts. The first set was computed using the parental input from the CHILDES corpus (MacWhinney 2000). We used WCD, a special recurrent neural network that learns the lexical co-occurrence constraints of words, to read a stream of input sentences one word at a time, and learn the adjacent transitional probabilities between words, which it represents as a matrix of weights (see Li, Farkas, and MacWhinney 2004, for details). WCD computes two vectors that correspond to the left and the right context, respectively; it then transforms these probabilities into normalized vector representations for word meanings.

The second set of semantic representations was generated from computational thesauruses available for each of the two languages. For Chinese, it was derived from a Chinese computational database called HowNet (http://www.keenage.com). Through a program which calculated the similarity of Chinese words in the database (Liu and Li 2002), we derived a matrix that represents the similarity of all the 500 Chinese words. For English, as in Li, Farkas, and MacWhinney (2004), we used a feature generation system (Harm 2002) to derive semantic features from the WordNet database (Miller 1990), and the similarity of the 500 English words was further calculated according to these features.

A random mapping (Kohonen 2001) method was further used to reduce the size of each set of the semantic representations to a lower dimension (from $d = 500$ to $d = 100$), and the two sets were then combined together to form each word's semantic vector. Our method allows for a lexical representation with both semantic and syntactic information, which has the ability to introduce certain *language-specific information* into our representation.

Simulation scenarios

Our simulation included three learning scenarios: simultaneous, early, and late. In simultaneous training, the two lexicons were presented to the network gradually and in parallel. At the first stage, the training vocabulary included 50 English words and 50 Chinese words. Then at every new stage, 50 more English words along with 50 more Chinese words were added to the training pool until the final stage when the size of each lexicon reached 500. Here, a training stage included 10 epochs, which means that each available word was presented to the network 10 times at each stage. In the sequential learning situation, the learning of L2 was delayed relative to that of L1, either only slightly (early learning) or significantly (late learning). In the case of early L2 learning, the network was first trained on 100 L1 words (Chinese).[5] Then the L2 words (English) were presented to the network stage by stage (each stage with 50 more new L2 words) along with the corresponding increment of L1 words. The training would end 10 stages later, when the entire list of 500 L2 words was seen by the network. In the case of late learning, L2 words began to join the training only after 400 L1 words had been presented to the network during the first four stages. Then the training continued for another 10 stages until all the 500 L2 words were seen by the network (so that exposure to L2 words in all three scenarios was over 10 stages). Comparison of the three learning scenarios should allow us to see the effects that the consolidation of lexical organization in one language has on the lexical representation in the other language.

Results and discussion

Bilingual lexical representations

First we examine the phonological and semantic organizations of the bilingual lexicons in the corresponding maps in our model. Figure 2 shows the examples of the distribution of the two lexicons on each map in the different learning situations. Due to the large size of the lexicons and maps, only broad areas of the active neurons are displayed. In Figure 2, the boxes on the left represent the distributions of the bilingual lexicons in the semantic map, and the boxes on the right indicate the distributions in the phonological maps. Black regions represent those neurons that can best be labeled by L2 (English) words, whereas white regions indicate those neurons that best represent L1 (Chinese) words in the input space.

Here, Figures 2a and b represent the simultaneous acquisition situation. We can see that our network shows clear and distinct lexical representations of L1 and L2 on both the phonological and the semantic levels and within each language. The results are similar to Li and Farkas' (2002) SOMBIP, and the network's ability to develop distinct representations for each language shows that simultaneous learning of two languages allows the system to easily separate the lexicons during learning (see also similar findings from a different model by French and Janquet 2004). In the case of sequential acquisition, however, the results are not so clear-cut. If L2 was introduced into learning early on, then the lexical organization patterns were similar (though not identical) to those found in simultaneous acquisition, as shown in Figures 2c and d. The differences are reflected in terms of the slightly smaller spaces occupied by the L2 words (English, the dark areas on each map) as compared to the lexical space occupied by L1, and the more dispersed and fragmented distribution of L2 on the phonological map (Figure 2d) as compared to the distribution in the simultaneous learning case (Figure 2b). We can dub these as the 'L2 islands.' However, if L2 was introduced to learning late, the lexical organization patterns were significantly different from those found in simultaneous acquisition, as shown in Figures 2e and f. No large L2 islands appeared this time. In fact, we can say that the L2 representations were parasitic on or auxiliary to those of L1 words: compared with L1 words, the L2 words occupied only small and fragmented regions, and were dispersed throughout the map. There were small L2 chunks that were isolated from each other, and interspersed within L1 regions. A close inspection showed that the locations of the L2 words depended on how similar they were to the L1 words in meaning (for semantic map) or in sound (for phonological map). For example, in Figure 2f, the English words *boy*, *bee*, and *bear*

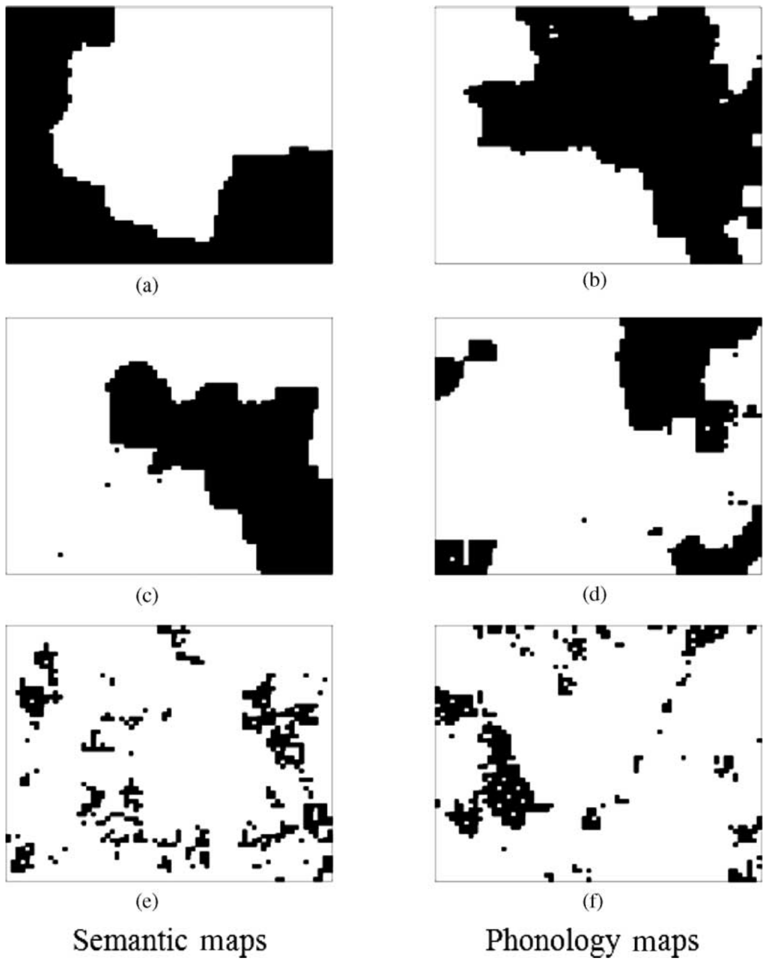

Figure 2. Examples of bilingual lexical representations on the semantic map and the phonological map. Dark areas correspond to L2 (English) words: (a–b), simultaneously learning; (c–d), early L2 learning; and (e–f), late L2 learning.

were located next to the Chinese word *bai2*(white) since they sound similar.[6] Other examples include *my* next to *mai3*(buy), *her* next to *he1*(drink), *ear* next to *ye2ye*(grandpa). Similar examples could also be found on the semantic map (Figure 2e): *girl* and *boy* were close to *mei4mei*(sister) and *nv3hai*(girl); *go* and *walk* close to verbs like *pao3*(run), *tiao4*(jump), and *pa2*(crawl); *chocolate, food,* and *cake* were projected to the location of Chinese words for food such as *qiao3ke4li4* (chocolate), *dan4gao1*(cake), *mian4bao1*(bread), and *tang2*(candy).

This parasitic feature of L2 on L1 representation in the late L2 learning situation was further tested by a quantitative measure. In particular, 32 pairs of translation equivalents were selected from the bilingual lexicons. These translation equivalents were all nouns with concrete meanings (see the Appendix). Furthermore, for each pair of translation equivalents, the Euclidean distance between their locations on the semantic map was calculated. The smaller the distance, the closer they were. The average of the 32 distances provides us with a general estimate of the relationship between the lexical representations of the two languages (particularly for words representing similar concepts in different languages). This measure was also applied to the early learning and simultaneous learning situations. Five networks were constructed for each learning situation, and the average distance based on the five networks (trials) was measured and shown in Figure 3. A one-way analysis of variance (ANOVA) was conducted to test the impact of the learning history (simultaneous, early L2, and late L2) on this distance measure. Learning history was significant, $F(2, 12) = 19.06$, $p < 0.001$, $\eta^2 = 0.76$, showing that the three different learning scenarios yielded different representational structures in terms of the distance of the lexical representations in the two languages. Post-hoc tests (pairwise differences, with Bonferroni correction) revealed that the late L2 learning condition had significantly shorter distances ($d = 25.13$) than the early L2 learning condition ($d = 36.52$) and the simultaneous learning condition ($d = 42.03$), $p < 0.01$. However, no significant difference between the early learning condition and the simultaneous learning condition was obtained. This result from the quantitative analysis is consistent with our qualitative analysis of Figure 2, in that the lexical representation of L2 in the late learning situation is fundamentally different from the other two conditions.

Why is late L2 learning so different from the other two situations? We believe that this is due to significant differences in our model's developmental changes as a function of learning history. In the late learning situation, L2 was introduced at a time when the learning system had already dedicated its resources and representational structure to L1, and L1 representations had been consolidated. So the L2 could only use existing structures and associative connections that were already established by the L1 lexicon. This is the sense in which we say that the L2 lexicon was parasitic on the L1 lexicon (see Hernandez, Li, and MacWhinney 2005). The network's re-organizational ability (plasticity) has been significantly weakened with the decrement of the neighborhood size on each map. Even though our model had a certain degree of plasticity in that it could recruit new resources into the computation when needed, this was still not sufficient to allow any radical restructuring or

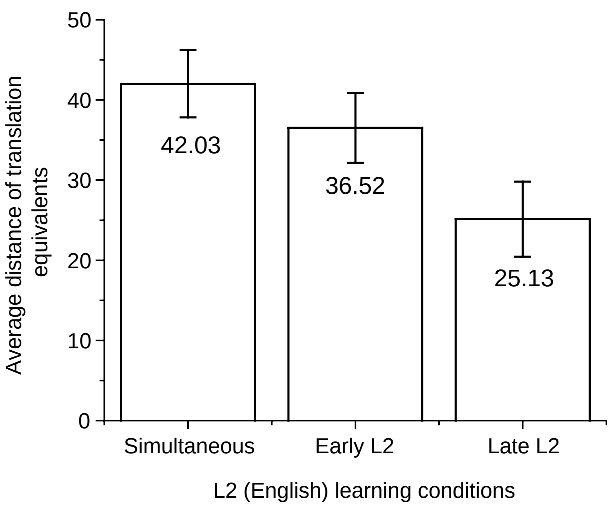

Figure 3. Average distance of the translation equivalents on the semantic map. The value changes as a function of the L2 learning history. Results are based on five trials for each condition and the error bar indicates standard deviation.

complete reorganization of the map's topology. Take the semantic layer as an example. Due to the weak plasticity of the network under the late L2 learning condition, the subtle differences in the meanings and the usage of L2 words (which were novel to the network) could not be fully represented. Thus, as demonstrated by our analyses of word density and lexical categories (see the next section), a clear structure of the L2 lexicon could not be developed on the map. In contrast, for the early L2 learning, the network still had significant plasticity and could continually reorganize the lexical space for the L2. The semantic and syntactic aspects of L2 words could then still be clearly captured by our network. Rather than becoming parasitic on the L1 lexicon, early learning allowed the entry of the L2 lexicon to present significant competition against the L1 lexicon. This reduced 'neural' plasticity for late L2 learning is consistent with what has been proposed in connectionist accounts of age of acquisition effects in the adult lexical processing literature, that is, that such effects are due to changes in the network's adaptive plasticity (Ellis and Lambon Ralph 2000; Elman 1993; Smith, Cottrell, and Anderson 2001; see Hernanderz and Li 2007, for a review).

Segalowitz and de Almeida (2002) have suggested that adolescent bilingual individuals are often slower and less accurate at judging the categories of words in their L2 (or less dominant language) than in their L1 (or more dominant language) in a simple semantic categorization task, but show an opposite pattern (L2 better than L1) in an arbitrary categorization task. Such a finding implies that the bilingual's mental representation of L2 words may be relatively 'impoverished' (i.e. bilinguals are aware of fewer senses and semantic associations of words) in comparison to the clear representation of L1. The latter facilitates classification in a simple semantic categorization task but causes interference in arbitrary categorization. Our simulation results are consistent with this interpretation, in that the L2 lexical representations are dispersed and fragmented.

An interesting prediction can be derived from the different L2 representations under the three different learning situations. In the late L2 learning condition, the location of many L2 words on the semantic map depended on how similar they were to the L1 words in meaning. They were often projected close to their translation equivalents (e.g. *nurse* and *hu4shi*), and therefore also became associated more closely to words in L1 which are semantically related to them (e.g. *nurse* and *dai4fu*(doctor), since *hu4shi* and *dai4fu* are closely related in L1). Such close distributions in the lexical representations across the two languages would make interactions across languages stronger in the late L2 bilinguals' language processing. This would explain the occurrence of the priming effects between translation equivalents or semantically related words in the late L2 bilingual's two languages. In contrast, in the early and simultaneous learning situations, due to the existence of two clear and distinct lexical representations, such cross-language priming effects should not be as apparent. This prediction has been partly supported by a recent behavioral study conducted by Kiran and Lebel (2007), in which they found that less balanced bilinguals often have stronger cross-linguistic semantic and translation priming effects than more balanced bilinguals. We have started to examine the empirical and computational bases of this prediction with a priming study and a computational model (see Zhao and Li 2009c, for preliminary computational findings).

Word density and learning history

Another way in which learning history may have an impact on bilingual representation in our model is the degree to which within-language lexical distributions are packaged. Inspecting the bilingual representations on the semantic and phonological maps, we found that the words were not evenly distributed in L1 and L2. Some areas were very dense while other areas were sparse. In some dense areas, the retrieval of the sound or the semantic content of a word could be difficult because the competition between words is strong and could thus result in a higher confusion rate.[7] To explore differences in density across the learning situations, we calculated mean densities for semantic and phonological neighborhoods. We defined the density of a word on a map as the number of words in its neighborhood (with radius of 1) divided by the total number of units in its neighborhood (usually nine, but could be six or four, depending on whether the tested word was on the border or at the corner of the map). The value of this density measure ranged from one-ninth (when only the word itself is in the neighborhood) to one (when all neighboring units of a word are occupied by other words). Then the average group densities of L1 and L2 words were calculated, respectively. Table 1 shows the average word densities for L1 and L2 in both the semantic and the phonological maps. Obviously, the larger the density measure, the more crowded the group members are on the map. We might expect to find more competition, confusion, and errors in a high-density group. We can see that under the late L2 learning situation, the density of the L2 words reaches a very high level (0.78 and 0.65 for the phonology and the semantic map, respectively), and this differs sharply from the other two learning situations.

These different distributions and word density patterns appear to reflect the learning curve. Figure 4 presents the number of L2 words that can be successfully produced by our network as a function of the L2 words available to the network at different stages. Not surprisingly, the vocabulary sizes of the L2 words increased over time under all three learning situations. However, a regression analysis indicates more rapid learning for the early than the late learning situation. In fact, the pattern for early L2 learning is quite similar to that for simultaneous learning. The regression equation for early L2 learning has a slope of 0.95 (t (48) = 34.75, $p < 0.001$), which is close to the regression slope of 1.008 for the simultaneous learning situation (t (48) = 34.75, $p < 0.001$). The late L2 learning regression equation has a slope of 0.79 (t (48) = 25.86, $p < 0.001$), which indicates a slower learning speed than the early L2 and simultaneous learning situations. The empirical bases and implications of these learning trajectories, however, need to be further investigated.

Table 1. Word density and comprehension and production errors of L1 (Chinese) and L2 (English) in the phonology and semantic maps. Results are based on the average of five trials.

		Word density		Number of error	
		Phonology	Semantic	Comprehension	Production
Simultaneous	L1(Chinese)	0.385	0.369	36.4	37.8
	L2(English)	0.397	0.427	18.6	51.8
Early L2	L1(Chinese)	0.241	0.272	20.0	36.4
	L2(English)	0.583	0.566	30.4	75.2
Late L2	L1(Chinese)	0.283	0.255	20.6	11.2
	L2(English)	0.781	0.650	134.2	161.2

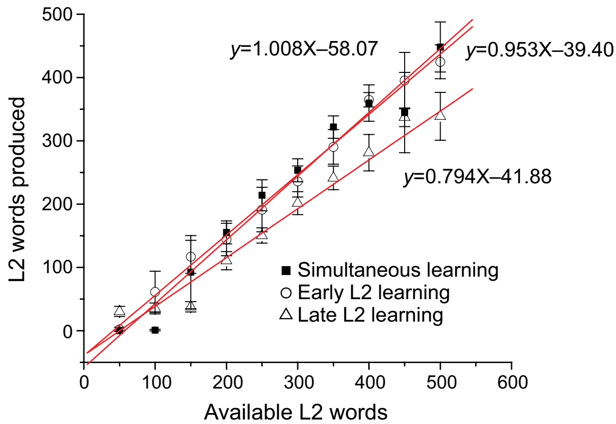

Figure 4. Correctly produced words as a function of available L2 words at different stages. Error bars indicate standard deviations, and the lines were fitted through regression analyses (the three regression equations show the different slopes of increase for the different learning situations). Each data point was calculated based on five trials.

Comprehension and production errors

L2 learners often have comprehension and production problems in the L2, particularly with pronunciations that they are unsure of. Recent empirical studies have shown that bilinguals, as compared with monolinguals, often have more difficulties generating fast and accurate names in picture naming or word naming tasks ('deficit' in lexical retrieval of L2, Craik and Bialystok 2006; Gollan et al. 2005). One possible source of such production difficulties, based on our simulation results, could be due to the nature of the L2 representation. Lexical items in L2 are represented in more dense neighborhoods on the map, and hence in a more likely confusable fashion. The bilingual speaker could have difficulties in retrieving the correct L2 items due to increased lexical competition from nearby items during speech production. As shown in Table 1, our model under the late L2 learning situation showed more comprehension and production errors for L2 words (134.2 and 161.2 on average in five trials) than under the other two learning situations. In addition, when L1 and L2 errors are considered together, most errors happened to the L2 words. Word density is relatively low for the L1 words in general. They are more robust than words in high-density areas and thus more resistant to competition or damage.

A correlation analysis was conducted to evaluate the relationship between the L2 word density and the number of production errors in L2. On the semantic map, the correlation is significant, $r(13) = 0.73$, $p < 0.01$. The more densely populated L2 words were on the semantic map, the more production errors our network made. Similarly, significant positive correlations are also found between our model's L2 comprehension errors and the word density measure of L2 on both the phonological map ($r(13) = 0.87$, $p < 0.01$) and the semantic map ($r(13) = 0.82$, $p < 0.01$). The results suggest that the intense competition of words in the densely packed L2 areas on the two linguistics levels may cause errors and confusions in our models' comprehension and production of L2 words.

In the current study, we also found interesting error patterns in our network's comprehension and production performance. DevLex-II has been shown to be able to capture children's error patterns in a monolingual environment (Li, Zhao, and MacWhinney 2007). These comprehension and production errors in bilingual situations, we believe, reflect the interaction and interference of the two lexicons in our model.

First, very strong within-language interference could be observed in the comprehension errors in all of the three bilingual learning scenarios. Such interference might be caused by the similarity either in sound or in meaning between two words in the same language. For example, an activation of the English word *she* on the input phonology map led to the activation of *see* on the semantic map. This is an example of incorrect comprehension (from phonology to semantics) due to within-language phonological interference. Other examples include: *stove-stone; bump-jump; glass-grass; pull-pool; qing3*(invite)-*qin1*(kiss); and *zang1*(dirty)-*zhang1* (piece). Semantic similarity may also lead to comprehension errors such as: *kick-drop; cut-tear; hei1*(black)-*lv4*(green); and *mi4feng1*(bee)-*ma3yi3*(ant).

Second, comprehension errors due to between-language interference were also found in our mode. Most of them were due to phonetic similarities (i.e. cross-language homophones): *a-e2* (goose)*; tongue-tang2*(sugar)*; hair-hei1*(black); *ear-ye2ye* (grand-pa); and *when-wan3*(bowl) (see Li and Farkas 2002, for similar errors); very few were due to semantic similarities: *Mao1*(cat)-*bear*; *shou3*(hand)-*toe* (Chinese as L2); and *kiss-qin1*(kiss) (English as L2).[8] However, these between-language interference errors were observed only in the late L2 learning situation. The absence of such interference errors in the simultaneous and early situations is probably due to the more distinct, less dense lexical representations in these situations as compared to the late learning situation. This analysis seems to be consistent with empirical reports as summarized by Francis (2005) that between-language interference errors are not as common as within-language interference errors.

Another interesting finding from our modeling is that the between-language interference was unidirectional, that is, the comprehension of L2 words was affected by L1 knowledge only. There was little evidence of interference from L2 to L1 in our simulations. This also supports our earlier analysis that L2 representations are often parasitic on L1 representations under late learning. In monolingual simulations (see Zhao and Li 2005), DevLex-II shows lexical confusions, omissions, replacements, or incorrect sequencing of phonemes in production. However, for the late L2 learning situation, many such errors were due to phonemes unique to L2. For example, [z] as in *zoo* and [ð] as in *then,* were two phonemes not found in Chinese and therefore they were often confused with each other on the map when English was learned as L2. Other examples include confusion of phonemes like [ɔ:] as in *born* and [ɒ] as in *pot*. Similar examples were obtained when Chinese was learned as L2 in seperate simulations. For instance, *c* ([ts']) and *ch* [tʂ'] were two phonemes not found in English and therefore they were often confused with each other on the map. Other examples include confusion of phonemes such as *j*, *q*, *x* ([tɕ], [tɕ'], [ɕ]), *z* and *zh* ([ts], [tʂ]), *s* and *sh* ([s], [ʂ]). In late L2 learning, the subtle differences between those phonemes are not highly distinguishable in a system that has already committed itself to the L1 phonemic inventory. These simulated patterns match up well with speech learning theories indicating that early learners can create new phonetic categories more easily than late learners, and that such differences are due to the stabilization of the phonetic representation of L1 vs. L2 over the lifespan of learning (see Flege 1995).

Conclusion

In this study we extended DevLex-II, an unsupervised neural network model, to the simulation of bilingual lexical acquisition, representation, and interaction. We highlight three key features of our model:

(1) In contrast to previous computational models of bilingualism, our model is a learning model based on unsupervised learning and Hebbian learning, two powerful and biologically plausible principles of computation. These principles have allowed us to simulate the dynamics underlying both monolingual and bilingual lexical representations and interactions.
(2) Our model relies on the use of large-scale realistic linguistic data as the input. In total, we were able to simulate the development of 1000 words that were based on children's early lexicons from the two languages. The input representations of these words were carefully coded to reflect the linguistic features of the target languages. Their semantic features were extracted from the parental speech in the CHILDES database. By simulating actual lexical forms and meanings, we were able to achieve developmental and lexical realism.
(3) We considered computational learning properties (e.g. self-adjustable neighborhood functions) against the context of realistic language learning so that the DevLex-II model has the ability to handle the plasticity–stability problem in L2 learning. These properties are crucial in providing our model with the flexibility to systematically simulate the impact of the learning history of L1 vs. L2 on the linguistic representations of the two languages.

Our findings suggest that the nature of bilingual representation is the result of a highly dynamic process in which mechanisms of learning interact with the timing and history of learning to determine developmental trajectories. The three scenarios of learning simulated in our model demonstrate how developmental patterns are shaped by the interactive dynamics inherent in learning. In particular, when the learning of L2 is significantly delayed relative to that of L1, the structural consolidation of the L1 lexicon will adversely impact the representation and retrieval of L2 words, which results in a parasitic L2 representation due to reduced plasticity in the system's structuring of an L2 (Hernandez and Li 2007; Hernandez, Li, and MacWhinney 2005). Thus, late L2 learning differs in fundamental ways from early L2 or simultaneous L1–L2 learning, and connectionist models such as those simulated in DevLex and DevLex-II can provide detailed computational and mechanistic specifications for unveiling such differences through interactive dynamics in development. Future research is needed to further link our simulation patterns with learning data and to evaluate the model against human performance.

Acknowledgements

We would like to thank two anonymous reviewers and the editor for their valuable comments and suggestions on the earlier versions of this article. Preparation of this article was made possible by a grant from the National Science Foundation (BCS-0642586) to Ping Li and the Colgate University faculty research grant to Xiaowei Zhao.

Notes

1. In some models, weights may be prewired and cannot be changed, which is the case, for example, in earlier models like the IA model of McClelland and Rumelhart (1981).
2. We have also obtained some preliminary results of applying our model to bilingual acquisition, on which the current study is based (see Zhao and Li 2007, 2008).
3. The detailed procedure of implementing this new neighborhood function is described in Li, Zhao, and MacWhinney (2007).
4. We excluded homographs in our simulations because the unique semantic representations for them are difficult to get; and excluded phrases because they include more than one word. See Bates et al. (1994) for reasons for excluding the other four types of words from a normal analysis of vocabulary development.
5. In separate simulations, we obtained similar results when English was L1 and Chinese was L2.
6. The number in the Chinese phonetic transcription indicates the tone of the corresponding word.
7. Initially, high density may bring a certain advantage to the learning of novel words in the dense areas, in that once a novel word is learned, its close neighbors may be more easily mapped to the semantic category to which they belong. However, the disadvantages caused by strong competition and high confusion could overwhelm the advantages eventually.
8. We constructed separate models in which either Chinese or English was the L2 (with the same modeling parameters). Given that the results from these models were very similar, we report here mainly the results from modeling Chinese as L2 and English as L1.

References

Bates, E., V. Marchman, D. Thal, L. Fenson, P.S. Dale, J.S. Reznick, J.S. Reilly, and J.P. Hartung. 1994. Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language* 21, no. 1: 85–123.

Bialystok, E. 2001. *Bilingualism in development*. Cambridge and New York: Cambridge University Press.

Cooper, S. 2005. Donald Hebb's synapse and learning rule: A history and commentary. *Neuroscience & Biobehavioral Reviews* 28, no. 8: 851–74.

Craik, F., and E. Bialystok. 2006. Positive and negative effects of bilingualism on cognitive aging. Paper presented at the 47th annual meeting of the Psychonomic Society, November 16–19, in Houston, TX.

Dale, P.S., and L. Fenson. 1996. Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers* 28, no. 1: 125–7.

Dijkstra, T., and W.J.B. van Heuven. 1998. The BIA model and bilingual word recognition. In *Localist connectionist approaches to human cognition*, ed. J. Grainger and A.M. Jacobs, 189–225. Mahwah, NJ: Erlbaum.

Dijkstra, T., and W.J.B. van Heuven. 2002. The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition* 5, no. 3: 175–97.

Dong, Y., S. Gui, and B. MacWhinney. 2005. Shared and separate meanings in the bilingual mental lexicon. *Bilingualism: Language and Cognition* 8, no. 3: 221–38.

Ellis, A., and M. Lambon Ralph. 2000. Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: Insights from connectionist networks. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26, no. 5: 1103–23.

Elman, J. 1993. Learning and development in neural networks: The importance of starting small. *Cognition* 48, no. 1: 71–99.

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

Flege, J.E. 1995. Second language speech learning: Theory, findings, and problems. In *Speech perception and linguistic experience*, ed. W. Strange, 233–77. Timonium, MD: York Press.

Fodor, J. 1983. *The modularity of mind*. Cambridge, MA: MIT Press.

Francis, W.S. 2005. Bilingual semantic and conceptual representation. In *Handbook of bilingualism: Psycholinguistic approaches*, ed. J.F. Kroll and A.M.B. de Groot, 251–67. New York: Oxford University Press.

French, R.M. 1999. Catastrophic forgetting in connectionist networks. *Trends in Cognitive Sciences* 3, no. 4: 128–35.

French, R.M., and M. Jacquet. 2004. Understanding bilingual memory. *Trends in Cognitive Science* 8, no. 2: 87–93.

Genesee, F. 1989. Early bilingual development: One language or two? *Journal of Child Language* 16, no. 1: 161–79.

Gollan, T.H., R.I. Montoya, C. Fennema-Notestine, and S.K. Morris. 2005. Bilingualism affects picture naming but not picture classification. *Memory & Cognition* 33, no. 7: 1220–34.

Grosjean, F. 2008. The Léwy and Grosjean BIMOLA model. In *Studying bilinguals*, ed. F. Grosjean, 201–12. New York: Oxford University Press.

Harm, M. 2002. *Building large scale distributed semantic feature sets with WordNet.* Technical Report PDP-CNS-02-1. Pittsburgh, PA: Carnegie Mellon University.

Haykin, S. 1999. *Neural networks: A comprehensive foundation.* 2nd ed. Upper Saddle River, NJ: Prentice Hall.

Hernandez, A., and P. Li. 2007. Age of acquisition: Its neural and computational mechanisms. *Psychological Bulletin* 133, no. 4: 638–50.

Hernandez, A., P. Li, and B. MacWhinney. 2005. The emergence of competing modules in bilingualism. *Trends in Cognitive Sciences* 9, no. 5: 220–5.

James, D., and R. Miikkulainen. 1995. SARDNET: A self-organizing feature map for sequences. In *Advances in neural information processing systems 7*, ed. G. Tesauro, D.S. Touretzky, and T.K. Leen, 577–84. Cambridge, MA: MIT Press.

Kiran, S., and K.R. Lebel. 2007. Crosslinguistic semantic and translation priming in normal bilingual individuals and bilingual aphasia. *Clinical Linguistics & Phonetics* 21, no. 4: 277–303.

Kohonen, T. 2001. *The self-organizing maps.* 3rd ed. Berlin: Springer.

Kroll, J.F., and N. Tokowicz. 2005. Models of bilingual representation and processing: Looking back and to the future. In *Handbook of bilingualism*, ed. J.F. Kroll and A.M.B. de Groot, 531–53. New York: Oxford University Press.

Li, P. 2009. Lexical organization and competition in first and second languages: Computational and neural mechanisms. *Cognitive Science* 33, no. 4: 629–64.

Li, P., and I. Farkas. 2002. A self-organizing connectionist model of bilingual processing. In *Bilingual sentence processing*, ed. R. Heredia and J. Altarriba, 59–85. North-Holland: Elsevier Science.

Li, P., I. Farkas, and B. MacWhinney. 2004. Early lexical development in a self-organizing neural network. *Neural Networks* 17, no. 8: 1345–62.

Li, P., and B. MacWhinney. 2002. PatPho: A phonological pattern generator for neural networks. *Behavior Research Methods, Instruments & Computers* 34, no. 3: 408–15.

Li, P., X. Zhao, and B. MacWhinney. 2007. Dynamic self-organization and early lexical development in children. *Cognitive Science: A Multidisciplinary Journal* 31, no. 4: 581–612.

Liu, Q., and S. Li. 2002. Word similarity computing based on How-net. *Computational Linguistics and Chinese Language Processing* 7, no. 2: 59–76.

MacWhinney, B. 2000. *The CHILDES project: Tools for analyzing talk*. Hillsdale, NJ: Lawrence Erlbaum.

Markman, E. 1984. The acquisition and hierarchical organization of categories by children. In *The 18th annual Carnegie symposium on cognition*, ed. C. Sophian, 376–406. Hillsdale, NJ: Lawrence Erlbaum.

McClelland, J.L. 2009. The place of modeling in cognitive science. *Topics in Cognitive Science* 1, no. 1: 11–28.

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

Mervis, C.B., and J. Bertrand. 1994. Acquisition of the novel name-nameless category (N3C) principle. *Child Development* 65, no. 6: 1646–63.

Miikkulainen, R. 1997. Dyslexic and category-specific aphasic impairments in a self-organizing feature map model of the lexicon. *Brain and Language* 59, no. 2: 334–66.

Miller, G.A. 1990. WordNet: An on-line lexical database. *International Journal of Lexicography* 3, no. 4: 235–312.

Pavlenko, A. 2009. Conceptual representation in the bilingual lexicon and second language vocabulary learning. In *The bilingual mental lexicon: Interdisciplinary approaches*, ed. A. Pavlenko, 125–60. Tonawanda, NY: Multilingual Matters.

Plunkett, K., and V. Marchman. 1991. U-shaped learning and frequency effects in a multi-layered perceptron: Implications for child language acquisition. *Cognition* 38, no. 1: 43–102.

Regier, T. 2005. The emergence of words: Attentional learning in form and meaning. *Cognitive Science: A Multidisciplinary Journal* 29, no. 6: 819–65.

Segalowitz, N., and R. de Almeida. 2002. Conceptual representation of verbs in bilinguals: Semantic field effects and a second-language performance paradox. *Brain and Language* 81, nos. 1–3: 517–31.

Smith, M.A., G.W. Cottrell, and K. Anderson. 2001. The early word catches the weights. In *Advances in neural information processing systems 13*, ed. T.K. Leen, T.G. Dietterich, and V. Tresp, 52–8. Cambridge, MA: MIT Press.

Spitzer, M. 1999. *The mind within the net: Models of learning, thinking, and acting*. Cambridge, MA: MIT Press.

Tardif, T., S.A. Gelman, and F. Xu. 1999. Putting the "noun bias" in context: A comparison of English and Mandarin. *Child Development* 70, no. 3: 620–35.

Thomas, M.S.C., and W.J.B. van Heuven. 2005. Computational models of bilingual comprehension. In *Handbook of bilingualism: Psycholinguistic approaches*, ed. J.F. Kroll and A.M.B. de Groot, 202–25. New York: Oxford University Press.

Van Hell, J.G., and A.M.B. de Groot. 1998. Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition* 1, no. 3: 193–211.

Volterra, V., and T. Taeschner. 1978. The acquisition and development of language by bilingual children. *Journal of Child Language* 5, no. 2: 311–26.

Wu, J. 1997. Language, play and general development for Chinese infant-toddlers. PhD diss., University of Colorado, Boulder.

Zhao, X., and P. Li. 2005. A self-organizing connectionist model of early word production. In *Proceedings of the twenty-seventh annual conference of the Cognitive Science Society*, ed. B.G. Bara, L. Barsalou, and M. Bucciarelli, 2434–9. Mahwah, NJ: Lawrence Erlbaum.

Zhao, X., and P. Li. 2007. Bilingual lexical representation in a self-organizing neural network. In *Proceedings of the 29th annual conference of the Cognitive Science Society*, ed. D.S. McNamara and J.G. Trafton, 759–60. Austin, TX: Cognitive Science Society.

Zhao, X., and P. Li. 2008. Vocabulary development in English and Chinese: A comparative study with self-organizing neural networks. In *Proceedings of the 30th annual conference of the Cognitive Science Society*, ed. B.C. Love, K. McRae, and V.M. Sloutsky, 1900–5. Austin, TX: Cognitive Science Society.

Zhao, X., and P. Li. 2009a. Acquisition of aspect in self-organizing connectionist models. *Linguistics: An Interdisciplinary Journal of the Language Sciences* 47: 1075–112.

Zhao, X., and P. Li. 2009b. An online database of phonological representation for Mandarin Chinese monosyllables. *Behavioral Research Methods* 41, no. 2: 575–83.

Zhao, X., and P. Li. 2009c. Cross-language priming in L1 and L2: A computational study. Paper presented at the 39th annual meeting of the Society for Computers in Psychology, November 19, in Boston, MA.

Appendix

The 32 pairs of English–Chinese translation equivalents used in our study.

English	Chinese
Sock	Wa4zi
Ant	Ma3yi3
Pillow	Zhen3tou
Bug	Chong2zi
Star	Xing1xing
Cup	Bei1zi
Ear	Er3duo
Hair	Tou2fa
Spoon	Shao2zi
Monkey	Hou2zi
Bird	Xiao3niao3
Button	Kou4zi
Train	Huo3che1
Bottle	Ping2zi
Coat	Da4yi1
Nail	Ding1zi
Duck	Ya1zi
Rock	Shi2tou
Hat	Mao4zi
Tongue	She2tou
Mouse	Lao3shu3
Toy	Wan2ju4
Basket	Lan1zi
Flower	Hua1duo3
Sky	Tian1kong1
Sweater	Mao2yi1
Shoe	Xie2zi
Animal	Dong4wu4
Drawer	Chou1ti
Tiger	Lao3hu3
Corn	Yu4mi3
Nose	Bi2zi