Bilingual Lexical Representation in a Self-Organizing Neural Network Model

Xiaowei Zhao (xzhao2@richmond.edu)
Ping Li (pli@richmond.edu)
Department of Psychology, University of Richmond
Richmond, VA 23173 USA

Abstract

In this paper we present a self-organizing neural network model of bilingual lexical development. We focus on how the representational structure of the bilingual lexicon can emerge, develop, and change as a function of the learning history. Our results show that (1) distinct representations for the two lexicons can develop in our network during simultaneous acquisition, (2) the representational structure is highly dependent on the onset time of L2 learning if the two languages are learned sequentially, and (3) L2 representation becomes parasitic on L1 representation when L2 learning 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 limited plasticity as a function of the timing of learning.

Keywords: SOM; DevLex; Bilingual Lexicon.

Introduction

Mechanisms underlying early bilingual lexical acquisition are so far poorly understood. This lack of knowledge may be partly due to the methodological limitations associated with studying young bilingual children at early stages of language development (e.g., Bialystok, 2001). Work in the monolingual context has shown that neural network models are ideally suited for identifying mechanisms of early lexical acquisition (e.g., Li, Farkas & MacWhinney, 2004; Regier, 2005). 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 Li & Farkas, 2002; French & Jacquet, 2004; Thomas & van Heuven, 2005). Furthermore, no neural network model has been devoted to capture the impact of developmental time on bilingual children's lexical representations. Our study here attempts to bridge the gap by examining bilingual lexical representations with a self-organizing neural network.

An issue of enduring interest in bilingualism has been whether bilingual representation takes the form of a single, shared lexical storage or a separate, distinct storage in the mental lexicon (see French & Jacquet, 2004 and Kroll & Tokowicz, 2005 for recent reviews). The issue has been highly controversial, and has recently been further complicated by conflicting neuroimaging data (see Hernandez & Li, 2007), but researchers have come to recognize that a host of variables must be taken into consideration in dealing with this issue, such as bilingual proficiency, learning history (including age of acquisition), modality (comprehension vs. production), and word types (cognates vs. noncognates, abstract vs. concrete words).

The DevLex and DevLex-II models have been developed to capture the interactive developmental dynamics in language acquisition. These models rely on simple but powerful computational principles of self-organization and Hebbian learning. We have applied them successfully to account for a variety of empirical phenomena in early monolingual lexical development (see Li et al., 2004; Li, Zhao & Macwhiney, 2007). Here we apply a variant of the DevLex-II model to the bilingual context and focus on how the representational structure of the bilingual lexicon can emerge, develop, and change as a function of the learning history. In particular, we manipulate the onset time of L2 lexical learning, in three scenarios: simultaneous – onset time of L2 co-occurs with that of 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 structure for the two lexicons in our model would differ as a function of the learning history defined by L2 onset time. In addition, through analyzing the model's comprehension and production errors, we hope to show how the two developing lexicons compete and interact with each other.

The Model

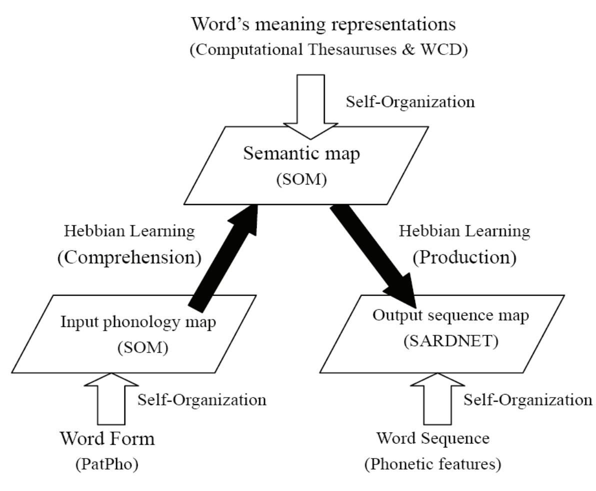

Figure 1: DevLex-II (Li, Farkas, & MacWhinney, 2007)

A Sketch of the Model

DevLex-II is a multi-layer self-organizing neural network model as diagrammatically depicted in Figure 1 (see Li, et al. 2007 for details). It includes three basic levels for the representation and organization of linguistic information:

phonological content, semantic content, and output sequence of the lexicon. The core of the model is a two-dimensional self-organizing, topography-preserving, feature map (SOM; Kohonen, 2001), which handles lexical-semantic representations. This feature map is connected to two other feature maps, one for input (auditory) phonology, and another for articulatory sequence of output phonology. Upon training of the network, the word meaning representations, input phonology, and output phonemic sequence of a word are presented to and processed by the network. This process can be analogous to the 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 standard SOM algorithm.

Here, given a stimulus **x** (the phonological or semantic information of a word), a winner node (or BMU, best matching unit) on the SOM is found if its weight vector has the smallest Euclidean distances to **x.** After a winner is identified, the weights of the nodes surrounding the winner in a given area (the *neighborhood*) are updated proportional to a constant learning rate α. Unlike the SOMBIP model (Li & 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 phonemic sequence information of a word is formed according to the algorithms of SARDNET (James & Miikkulainen, 1995), a type of temporal or sequential SOM network (see Li et al., 2007 for further details). In DevLex-II, the activation of a word form can evoke the activation of a word meaning via form-to-meaning links (to model word comprehension) and the activation of a word meaning can trigger the activation of an output sequence via meaning-to-sequence links (word production). Concurrent with the training of the three maps, the associative connections between maps are trained via Hebbian learning with a constant learning rate β.

Plasticity and Stability in the Model

To realistically model bilingual lexical development (especially the L2 acquisition) we must consider a core issue called "catastrophic interference" (see French, 1999; Li et al., 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 additional L2 words may disrupt the network's knowledge of L1. 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 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. In standard 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 neighborhood radius becomes very small. In DevLex-II, we attempt to correct these problems 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 will depend on the network's average quantization error on each layer, with quantization errors defined as the Euclidean distances between an input pattern and the input weight vector of its BMU (Kohonen, 2001). We implement this process as follows: (1) at each epoch (training with all available words), the network checks the quantization errors on each layer responding to input patterns and calculates their average errors for each layer; (2) the average errors from the current epoch are compared with those from previous epochs, and the neighborhood sizes on each layer are adjusted accordingly (either increase by 1 if the current error is larger than the previous average error, or decrease by 1 if it is smaller); (3) the neighborhood size should not be negative, and not larger than the final neighborhood size of the previous training stage; we split the training process into several stages to gradually present the network with new words. This method gives DevLex-II certain plasticity by increasing the neighborhood size a little when facing new patterns (due to the increment of error), but there is also certain degree of stability due to the restrictions in step (3). The learning process will thus no longer be fixed *a prior*, but be dependent on the experience level of the network.

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 response 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) in the maps. A similar procedure is also used in 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 in the map. This mechanism resembles a process in which new neurons are recruited for novel inputs as computational resources become scarce (see Li et al., 2004). The use of a different but adjacent unit for new input is also empirically motivated: 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 acquired a name yet (see Markman, 1984, *Principle of Mutual Exclusivity*; Mervis & Bertrand, 1994, *Principle of Novel-Name-Nameless-Category*).

Bilingual Lexicons and Input Representations

To control for a host of extraneous variables in the study of bilingual lexicon, we used as our basis the vocabulary from CDI, the MacArthur-Bates Communicative Development Inventories (Dale & Fenson, 1996). Each lexicon included 500 words and was ordered roughly according to their order of acquisition. The English lexicon was identical to that of Li et al. (2004). The Chinese lexicon was derived from the Chinese version of the CDI (Wu, 1997; Tardif et al., 1999).

To derive the input representations of the bilingual lexicons for our model, first, we used PatPho, a generic phonological pattern generator for neural networks (Li & MacWhinney, 2002), to construct the basic input phonological patterns of the English and Chinese words. A left-justified template with 54 dimensions was adopted. In addition, a separate group of 9 units was used to represent lexical tones in Chinese, and the values of these units were left empty for English. Thus, each word in the bilingual lexicon was represented by a vector of 63 units for its input phonological representation. Second, there were 55 phonemes from the two languages, which we represented as vectors of articulatory features of the phonemes to the output sequence map (as in PatPho). Third, for each language, we constructed two sets of lexical semantic representations through two different methods, and then combined them to increase the accuracy of the lexical representation (see Li et al., 2004 for rationale). The first set was generated by WCD (the word co-occurrence detector, Li et al., 2004), a special recurrent network that learns the lexical co-occurrence constraints of words by reading through input speech in linguistic corpus (here it is the child-directed parental speech from the database of CHILDES: http://childes.psy.cmu.edu). The second set of semantic representations was generated from computational thesauruses available for each of the two languages (WordNet for English and HowNet for Chinese: http://www.keenage.com). Our method allows for a lexical representation with both semantic and syntactic information. It makes our semantic representation a kind of "language specific semantic representation" and closer to the "lemma component" of a lexical entry, which allows inter-language synonyms to have different representations.

Simulation Parameters

In the simulations reported below, the input phonology map and the semantic map each consisted of 70 x 60 nodes, and the output sequence map included 25 x 20 nodes. During training, both learning rate α and β were kept constant (0.25 and 0.1 respectively). The radii of a winner's neighborhood on each map were adjusted automatically according to the neighborhood function mentioned above. The initial radius on the SOM layer was set to be 20 and that on the SARDNET was 10. These numbers were chosen to be large enough to discriminate among the words and phonemes in the lexicon while keeping the computation process tractable.

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, learning of L2 is delayed relatively 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 (English) [1]. Then the L2 words 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 500 L2 words were 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 4 stages. Then the training continued for another 10 stages until all the 500 L2 words were seen by the network (so exposure to L2 words in all three scenarios was 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

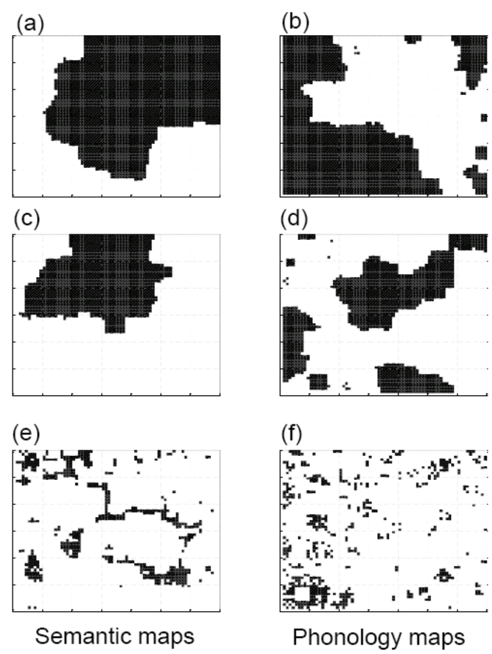

Figure 2: Examples of bilingual lexical representations on semantic map and phonological map. Dark areas correspond to L2 (Chinese) words. (a-b): simultaneously learning; (c-d): early L2 learning; (e-f) late L2 learning.

[1] In separate simulations (not reported here) we obtained similar results when Chinese was L1 and English was L2.

Bilingual Lexical Representations

First we examine the phonological and semantic organizations of the bilingual lexicon in the corresponding maps in our model. Figure 2 shows the examples of the distribution of the two lexicons on each map in 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 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 be best labeled by L2 words, whereas white regions indicate those neurons that best represent L1 words in the input space. [2]

Here, Figures 2a & 2b represent the simultaneous acquisition situation. We can see that our network shows clear distinct lexical representations of L1 and L2 on both the input phonological and the semantic level and within each language. The results are similar to Li and Farkas's (2002) previous work, 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 lexicon during learning (See also French & 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 2c and 2d. The differences are reflected as the slightly smaller spaces occupied by the L2 words (Chinese, the dark areas on each map) as compared to the lexical space occupied by L1, and more dispersed and fragmental distributions of L2 on the phonological map (Figure 2d) as compared to simultaneous learning results (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 2f. No 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 with L1 regions. A close investigation shows 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, Chinese words *lang2* (wolf), *leng3*(cold) and *tang1* (soup) were located close to the English words *long* and *leg* since they sound similar. [3] Other examples: *tou2* (head) is close to *toe*, and *gou3* (dog) close to *go*. Examples like these could also be found in the semantic map (Figure 2e): *mei4mei* (sister) and *nv3hai* (girl) were close to *girl* and *boy; qiao3ke4li4* (chocolate), *dan4gao1*(cake) and *pu2tao2gan1* (raisin) were projected to the location of English words for food like *coffee, chocolate, milk* and *egg*.

Why is the late L2 learning so different from the other two situations? We believe that this is due to the significant difference in developmental changes as a function of learning history. In the late learning situation, L2 is introduced at a time when the learning system has dedicated its resources and representational structure to L1, and L1 representations are consolidated such that L2 can only use existing structures and associative connections that are already established by the L1 lexicon. In this sense we say that the L2 lexicon is parasitic on the L1 lexicon (see Hernandez, Li, & MacWhinney, 2005). This is because the network's re-organizational ability (plasticity) has been significantly weakened with the decrement of the neighborhood sizes on each map. Even though our model has certain degree of plasticity by recruiting new resources into the computation when needed, it is still not enough to make the radical restructuring or complete reorganization of the map's topology. In contrast, for the early L2 learning, the network still has significant plasticity and can continually reorganize the lexical space for L2. Rather than becoming parasitic on the L1 lexicon, early learning allows the increase of the L2 lexicon to present a significant competition against the L1 lexicon.

Word Density and Learning History

Another way in which learning history has impacted bilingual representation in our model is the degree to which within–language lexical distributions are packaged. Inspecting the bilingual representations in the semantic and phonological map, we found that the words were not evenly distributed in L1 and L2. Some areas were very dense while other areas were sparse. It was obvious that in some dense areas, the retrieval of the sound or the semantic content of a word could become difficult. In densely populated areas, the competition between words was often strong and thus might result in a higher confusion rate. Here we wanted to see if L2 words acquired during late L2 learning were distributed as a group in high density. For this purpose, we defined the average word density of a group represented on a map as the vocabulary size of the group divided by the total number of units on the map which can be best labeled by the members of the same group. Obviously, if the vocabulary size for the group is fixed, then the larger the density measure is, the more crowded the group members will be in the map. We may expect to find more competitions, confusions, and errors in a highly dense group. Table 1 shows the average word densities of L1 and L2 words in both the semantic and the phonological map. We can see that under the late L2 learning situation, the density of the L2 words reached a very high level (0.99 on a 0-1 scale).

Our density analysis is consistent with our previous representation analysis. Moreover, high density and small islands (i.e., the fragmental representations) may cause a

[2] Just as finding BMU for an input pattern, we can find the "best matching word" in the input space for each unit on the map. Then the unit can be marked by the label of its best matching word.

[3] The number in the Chinese phonetic transcription indicates the tone of the corresponding word.

Table 1: Word density, and comprehension (Com) and production (Pro) errors of L1 & L2 in the phonology (Phon) and semantic (Sem) maps. Results are based on the average of 5 trials.

		Word Density		Error #	
		Phon	*Sem*	*Com*	*Pro*
Simultaneous	*L1*	0.244	0.263	18.6	51.8
	L2	0.236	0.218	36.4	29.8
Early L2	*L1*	0.174	0.168	10.0	4.4
	L2	0.382	0.430	46.8	34.8
Late L2	*L1*	0.135	0.135	10.4	4.2
	L2	0.999	0.998	140.2	186.4

high level of competition and a high rate of lexical confusion between lexical items during the speaker's word retrieval for production. These patterns may account for the empirically observed 'deficit' in lexical retrieval during word naming in L2 (Craik & Bialystok, 2006). As seen in Table 1, our model under the late L2 learning situation showed more comprehension and production errors for L2 words (140.2 and 186.4 on averages in 5 trials) than under the other two learning situations. In addition, when L1 and L2 errors were considered together, most errors happened to the L2 words. Word density was quite 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.

Due to the influence of these different distribution and word density patterns, lexical development may also be impacted by different L2 learning history. In Figure 3, we present 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 the three learning situations. A regression analysis indicated more rapid learning for the early than the late learning situation (see the slope function of the fitting line). In fact, the pattern for early L2 learning is quite similar to that for simultaneous learning. The empirical bases and implications of these findings, however, need to be further investigated.

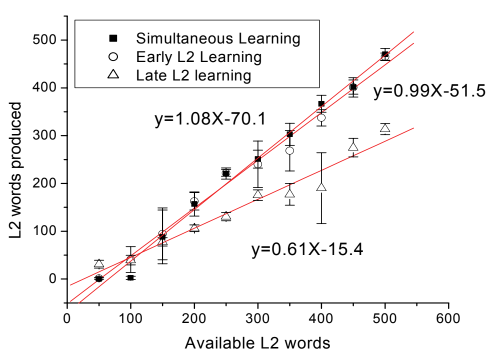

Figure 3: Correctly produced words as a function of available L2 words at different stages. Error bars indicate standard deviations, and the lines were fitted through regression analyses.

Comprehension and Production errors

Novice learners of L2 will often encounter problems when they use their second language. They may misunderstand unfamiliar L2 words, or may not get their words understood due to particular pronunciations. DevLex-II has been shown to be able to capture children's error patterns in a monolingual environment (Li et al., 2007). In the current study, we also found interesting patterns especially in our network's comprehension errors.

First, very strong within-language interferences could be found in the comprehension errors in all of the three bilingual learning scenarios. Such interference could 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 caused the responding of *see* on the semantic map. This is an example of within-language interference due to sound similarity. Other examples include *bump - jump*; *glass-grass*; *pull-pool; qing3* (invite)-*qin1* (kiss); *zang1* (dirty) - *zhang1* (piece). Semantic similarity may also cause comprehension errors such as: *kick-drop; cut-tear;* and *hei1*(black)-*lv4*(green); *mi4feng1*(bee)-*ma3yi3*(ant).

Second, comprehension errors caused by between-language interferences could also be found. Most of them were caused by phonetic similarities (i.e., cross-language homophones): *e2*(goose)-*a; tang2*(sugar)-*tongue; ye2ye* (grandpa)-*ear* (see Li & Farkas, 2002, for similar errors); fewer were caused by semantic similarities, *Mao1*(cat)-*bear*; *shou3*(hand)-*toe*. However, as in empirical studies summarized by Francis (2005), such interferences were not as common as within-language interferences, and in our model it could be found only in the late L2 learning situation. The absence of such interferences in the simultaneous and early situations is probably due to the distinct, less dense lexical representations in these situations as compared to the late learning situation.

Another interesting finding is that the between-language interference is unidirectional, that is, the comprehension of L2 words was interfered by L1 knowledge only. There was rare evidence of a reversed 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. Under this situation, L1 representations have been consolidated such that the processing of L2 word tends to use existing structures and associative connections that are established by the L1 lexicon. This sometimes causes unfamiliar L2 words to be wrongly projected to the regions of L1 words.

As in monolingual simulations (see Li et al., 2007), DevLex-II also showed lexical confusions, omissions, replacements, or incorrect sequencing of phonemes in bilingual production. However, for the late L2 learning situation, many errors were caused by phonemes that are unique to L2. For example, *c* ([tsʻ]) and *ch* [tʂʻ] are two phonemes not found in L1 (English) and therefore they are often confused in the map. Other examples include confusions of phonemes among *j, q, x* ([tɕ], [tɕʻ], [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 L1 phonemes. These simulated patterns match up well with empirically based hypotheses 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, a self-organizing neural network model, to the simulation of bilingual lexical representation and development. The model has been modified to handle the plasticity-stability problem in L2 learning. Our findings suggest that the nature of bilingual representations will depend on important developmental factors such as timing and history of learning. Comparing across three scenarios of learning, our model demonstrates how developmental patterns are determined by learning dynamics. In particular, when the learning of L2 is early relative to that of L1, functionally distinct lexical representations may be established for both languages; 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, resulting in parasitic L2 representation due to reduced plasticity in the structuring of a second language (Hernandez et al., 2005; Hernandez & Li, 2007). In this latter case, we can see how early learning leads to dedicated cognitive and neural structures that affect the shape and outcome of later development. These findings point to a highly dynamic process in which mechanisms of learning interact with the timing and history of learning to determine developmental trajectories. Connectionist models such as DevLex and DevLex-II provide excellent computational accounts and mechanistic specifications for such interactive dynamics in development.

Acknowledgments

This research was supported by a grant from the National Science Foundation (BCS-0131829).

References

Bialystok, E. (2001). *Bilingualism in development.* Cambridge, NY: Cambridge University Press.

Craik, F., & Bialystok, E. (2006). *Positive and negative effects of bilingualism on cognitive aging.* 47th Annual Meeting of the Psychonomic Society, Houston, TX.

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

Francis, W. S. (2005). Bilingual semantic and conceptual representation. In J.F. Kroll & A.M.B. De Groot (Eds.), *Handbook of bilingualism* (pp. 251-267). New York: Oxford University Press.

French, R.M. (1999). Catastrophic forgetting in connectionist networks. *Trends in Cognitive Sciences, 3,* 128–135.

French, R.M., & Jacquet, M. (2004). Understanding bilingual memory. *Trends in Cognitive Science, 8*, 87-93.

Flege, J.E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange, (Ed.), *Speech perception and linguistic experience* (pp.233-277). Timonium, MD: York Press Inc.

Hernandez, A., & Li, P. (2007). Age of acquisition: Its neural and computational mechanisms. *Psychological Bulletin*, *133*.

Hernandez, A., Li, P., & MacWhinney, B. (2005). The emergence of competing modules in bilingualism. *Trends in Cognitive Sciences, 9,* 220-225.

James, D., & Miikkulainen, R. (1995). SARDNET: A self-organizing feature map for sequences. In G. Tesauro et al., (Eds.), *Advances in neural information processing systems 7* (pp.577-584). Cambridge, MA: MIT Press.

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

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

Li, P., & Farkas, I. (2002). A self-organizing connectionist model of bilingual processing. In R. Heredia & J. Altarriba (Eds.), *Bilingual sentence processing* (pp.59-85). North-Holland: Elsevier Science Publisher.

Li, P., Farkas, I., & MacWhinney (2004). Early lexical development in a self-organizing neural network. *Neural Networks. 17*, 1345-1362.

Li, P., & MacWhinney, B. (2002). PatPho: A phonological pattern generator for neural networks. *Behavior Research Methods, Instruments, and Computers, 34,* 408-415.

Li, P., Zhao, X., & MacWhinney, B. (2007). Dynamic self-organization and early lexical development in children. *Cognitive Science*. (in press)

Markman, E. (1984) The acquisition and hierarchical organization of categories by children. In C. Sophian (Ed.), *The 18th annual Carneigie symposium on cognition* (pp.376-406). Hillsdale, HJ: Lawrence Erlbaum.

Mervis, C.B., & Bertrand, J. (1994). Acquisition of the novel name-nameless category (N3C) principle. *Child Development, 65*, 1646-1663.

Regier, T. (2005). The emergence of words: Attentional learning in form and meaning. *Cognitive Science, 29*, 819.

Tardif, T., Gelman, S.A., & Xu, F. (1999) Putting the "noun bias" in context: a comparison of English and Mandarin. *Child Development, 70*, 620-635.

Thomas, M.S.C. & Van Heuven, W. (2005). Computational models of bilingual comprehension. In J.F. Kroll & A.M.B. deGroot (Eds.) *Handbook of bilingualism*. New York: Oxford University Press.

Wu, J. (1997). *Language, play and general development for Chinese infant-toddlers.* Unpublished Ph.D. dissertation, University of Colorado at Boulder.