

Lexical Organization and Competition in First and Second Languages: Computational and Neural Mechanisms

Ping Li

Pennsylvania State University and National Science Foundation

Received 4 March 2008; received in revised form 8 October 2008; accepted 26 January 2009

Abstract

How does a child rapidly acquire and develop a structured mental organization for the vast number of words in the first years of life? How does a bilingual individual deal with the even more complicated task of learning and organizing two lexicons? It is only until recently have we started to examine the lexicon as a dynamical system with regard to its acquisition, representation, and organization. In this article, I outline a proposal based on our research that takes the dynamical approach to the lexicon, and I discuss how this proposal can be applied to account for lexical organization, structural representation, and competition within and between languages. In particular, I provide computational evidence based on the DevLex model, a self-organizing neural network model, and neuroimaging evidence based on functional magnetic resonance imaging (fMRI) studies, to illustrate how children and adults learn and represent the lexicon in their first and second languages. In the computational research, our goal has been to identify, through linguistically and developmentally realistic models, detailed cognitive mechanisms underlying the dynamic self-organizing processes in monolingual and bilingual lexical development; in the neuroimaging research, our goal has been to identify the neural substrates that subserve lexical organization and competition in the monolingual and the bilingual brain. In both cases, our findings lead to a better understanding of the interactive dynamics involved in the acquisition and representation of one or multiple languages.

Keywords: Language acquisition; Lexical development; DevLex; Dynamic self-organization; Bilingual representation; Neural representation of lexicon

1. Introduction

Language as a dynamical system, a proposal championed by Liz Bates, Jeff Elman, and other colleagues (e.g., Bates & Elman, 1993; Elman, 1990, 1995; Elman et al., 1996; Smith & Thelen, 1994; Van Geert, 1991), has had a profound impact on our thinking of the

Correspondence should be sent to Ping Li, Department of Psychology, Pennsylvania State University, University Park, PA 16802. E-mail: pul8@psu.edu

relationship between language and cognition. This perspective distinguishes itself from the view of cognition based on basic building blocks in the form of symbols and rules. Recent advances in developmental science and cognitive neuroscience have provided further empirical support for the dynamical perspective, and further neural and computational evidence for the dynamic changes that occur in the language learner and the learning environment. For example, a great deal of recent research has been devoted to identifying mechanisms of infant statistical learning (e.g., Gliozzi, Mayor, Hu, & Plunkett, 2009; Saffran, Aslin, & Newport, 1996; Yu & Smith, 2007), temporal dynamics in the processing and representation of linguistic and nonlinguistic materials (e.g., Altmann, 2009; Hare, McRae, & Elman, 2003, 2009; McRae, 2009), and computational and neural architectures of bilingual language competition and organization (Abutalebi & Green, 2007; Hernandez & Li, 2007; Hernandez, Li, & MacWhinney, 2005; Li & Green, 2007; Thomas & van Heuven, 2005). An integral picture of these lines of research looms on the horizon, offering new ways of thinking about the relationship among language, cognition, culture, and the brain. In this paper, I discuss data and theory from our research that contributes to this perspective, with specific reference to the dynamic self-organization and competition in language acquisition, particularly in the acquisition of the lexicon and its representation by monolingual and bilingual individuals.

One of the most influential papers in cognitive science, Elman's (1990) "Finding Structure in Time," provides a central premise of the language-as-dynamical-system perspective. Structured representations in human cognition such as those of linguistic categories, taxonomies, hierarchies, or even recursion, need not be the results of explicit rule learning or symbol manipulation, but could emerge from the dynamic interaction of multiple components or agents during the learning or discovery process. For example, learning in the simple recurrent network (SRN) leads to the representation of linguistic categories such as nouns and verbs, as a result of the SRN's discovering and clustering of the input space according to the input item's history of occurrence (that is, co-occurrence constraints in the linguistic environment, learned by the SRN's next-word prediction task). The different clusters formed in the SRN's similarity space correspond squarely with what linguists would call animate nouns, inanimate nouns, transitive verbs, intransitive verbs, and so on. In other words, linguistic categories need not be defined a priori as discrete entities in the mental representation, but could emerge as such from the interaction between the learner and the learning environment.

This way of looking at mental representation and linguistic structure has inspired many, including my own studies discussed in this article. More recently, Elman (2004; see also Elman, 2009) pushed this idea further, arguing that it is the "mental states," the contexts in which the lexical entries occur and interact, that define what words really are. This idea contrasts further with traditional linguistic views that a mental lexical entry contains a fixed representation of phonological, semantic, and grammatical information relevant to the construction of phrases and sentences, in and of themselves, stored in the long-term memory of the speaker (see, e.g., Jackendoff, 2002). This dynamic view of the lexicon coincides with other contemporary proposals based on large-scale, data-inspired computational models, in particular, the argument that words represent the aggregate of multiple, global,

co-occurrence constraints in high-dimensional spaces of language use (e.g., the latent semantic analysis [LSA] model, Landauer & Dumais, 1997; the hyperspace analog to language [HAL] model, Burgess & Lund, 1997, 1999). In this perspective, the functions of words in the mental lexicon are not statically defined, but dynamically provided and enriched by the context, that is, all the other items with which the word co-occurs in a sentence or text. This aggregated context provides the total linguistic environment for the usage history of the target word, which in turn defines the possible linguistic experience of the word for the listener and the speaker. This view, apparently a much broader and more inclusive perspective on what we can call “lexical” or “lexicon,” necessitates the inclusion in the representation of various dynamic properties of not just the lexical entries themselves (e.g., verbs), but their co-occurring items and contexts (e.g., verb arguments; see Li, Shu, Liu, & Li, 2006, for behavioral and electrophysiological evidence on this; see Elman, 2009, for a more extended view). In this perspective, we say that the language user has acquired the meaning of a word if he or she has learned the relevant total context in which the word occurs in speech (Burgess & Lund, 1999; Li, Burgess, & Lund, 2000; Li, Farkas, & MacWhinney, 2004a).

While the exact mechanisms for implementing this new view of the lexicon are subject to debate (e.g., whether we need placeholders for individual lexical entries in the representation), the need to understand lexical organization and competition through the study of dynamic interaction of learning and learner variables is clear and has been a major focus of our research in the last few years. Toward this end, we have developed the DevLex model, a self-organizing neural network model of the development of the lexicon, to account for a variety of empirical and theoretical issues in language acquisition. The model has also been extended to explain crosslinguistic discrepancies in the acquisition of vocabulary in typologically different languages and to account for language representation in bilingual individuals. In this paper, I provide a synthesis of the major findings from this model, focusing on three themes, namely, lexical organization in development, lexical competition between languages, and lexical competition within languages: (a) for the first theme, our model provides a mechanistic account of the dynamic changes that occur in the representational structure in the mental lexicon, as a result of enriched learning experience with the linguistic input; (b) for the second theme, competition clearly occurs between languages in any individual who is faced with learning and representing two linguistic systems either simultaneously or sequentially, and the dynamics of competition are best captured by self-organizing processes within the learning system; additionally, neuroimaging techniques can provide an *in vivo* view of the representational structure as a consequence of the learner’s diverse linguistic experience in multiple languages, complementing information gained from (a); and finally (c) for the third theme, our research indicates that different types of linguistic information are competing during learning and processing, and due to language-specific constraints, they may be learned in different ways (e.g., nouns vs. verbs) or may take different priorities in representation (e.g., semantics vs. prosody), as reflected in our model and in neuroimaging data. In the remainder of the discussion, I aim at laying out some of the perspectives toward unraveling the computational mechanisms and neural correlates underlying lexical organization and competition, in both monolingual and bilingual contexts.

2. Lexical organization in development

A fundamental puzzle for the language acquisition researcher is how the child manages to acquire a massive vocabulary with apparent ease and with rapid speed during the first few years of life. For some, this may be the beginning and the end of the “logical problem of language acquisition.”¹ Starting from identifying recurring patterns of rhythms, tones, phonemes, and other segments in the running speech, the child must come to grips with the meaningfulness of these patterns for the communicative purpose, that is, to solve the “form-meaning mapping” problem in early communicative interaction. Obviously, for a child learning two languages from the beginning, the problem becomes even more complicated and challenging.

While a great deal has been learned about how the child makes the initial form-meaning mapping (e.g., through statistical learning), relatively little is known about how the child mentally organizes various mapped items into a coherent whole and how the mental organization changes over time in the process of development. To fill this gap, our research attempts to understand the computational and neural mechanisms of organization in the context of the representation of the lexicon as a structured whole. The question to be addressed here is not how individual form-meaning pairings of words are acquired, but how the phonological and semantic representations of groups of words, at varying sizes, are formed and organized over the course of development, and how the representations compete to force structural changes. To illustrate this effort with an example, Fig. 1 shows how our DevLex model develops structured lexical representations across developmental stages. The model learns an early child lexicon of 500 words in an incremental fashion, as shown in the four snapshots, in which meaningful patterns emerge over time in the representational structure. At various points in time, the lexical categories are being shaped and consolidated, and boundaries established and shifted. By the end of the learning the network has formed the major lexical categories of what we can call nouns, verbs, adjectives, and close-class words. (Note that category labels are provided here by the modeler for discerning the meaningfulness of the output representations; the learning model did not receive such labels in the input nor did it provide them in the output).

2.1. *The DevLex model*

The scenario presented in Fig. 1 is an illustration of how our model can capture the dynamic expansion and change in the representation of lexical items and more important, their relationships in the representational structure (see Li et al., 2004a, for technical details). Traditionally, it was difficult to empirically probe into the dynamic structure of a large mental lexicon as the lexicon expands and the representation develops. Computational modeling has provided an ideal tool for us to study such changes and developments, due to its ability and flexibility in manipulating a large number of free parameters relevant to specific hypothesis testing (e.g., timing, amount, and rate of learning). In particular, our computational research has relied on the use of two simple but powerful learning principles, self-organization and Hebbian learning, in artificial neural networks (Hebb, 1949; Kohonen,

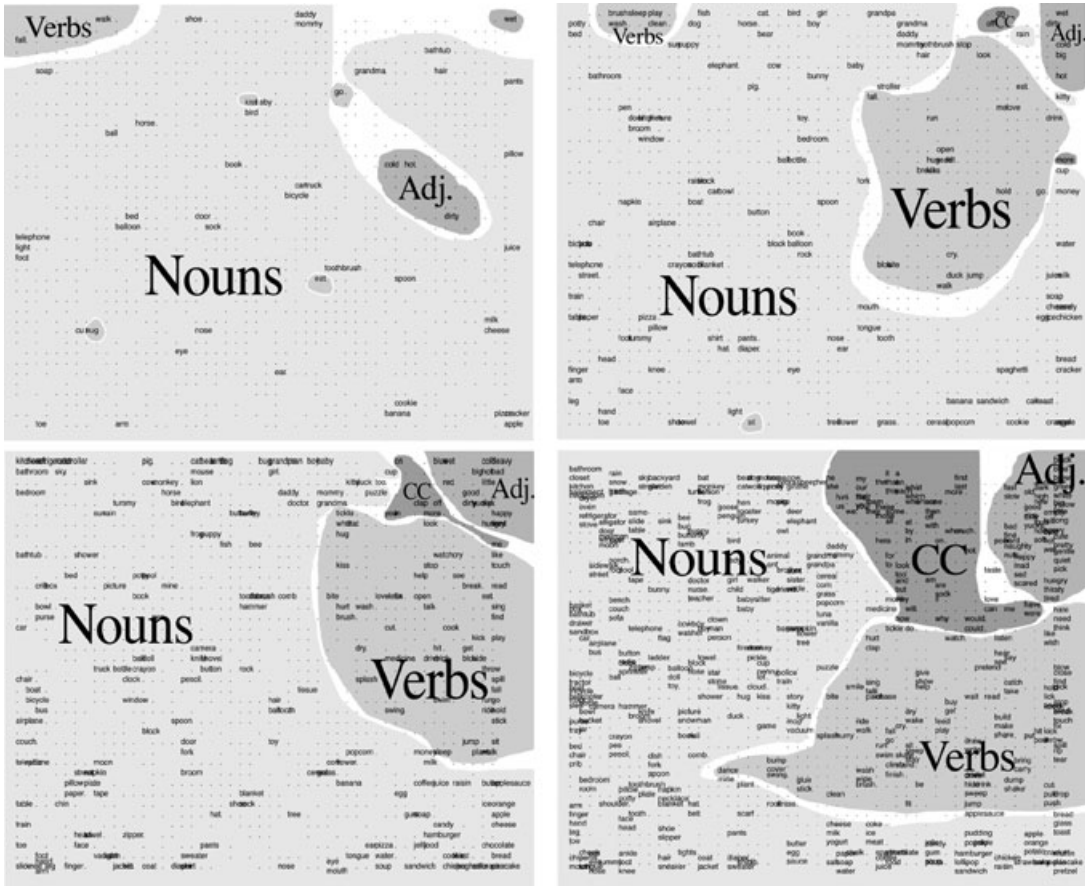


Fig. 1. Snapshots of DevLex across different developmental stages: Stage 1 (50 words—upper left), Stage 3 (150 words—upper right), Stage 5 (250 words—lower left), and Stage 10 (all 500 words—lower right). The sequence of images illustrates the nature of changes underlying the developmental process, as a result of the competition in an expanding vocabulary and the changing/enriching word representations. DevLex clearly separates the four major lexical categories and the semantic subcategories within each category toward the final stage. (Because of the large number of words involved in each map, the individual words are not legible in this figure. Reprinted from Hernandez et al., 2005, with permission from Elsevier.)

1982; Miikkulainen, 1993). Our neural network model, DevLex, has been designed to examine mechanisms of association, organization, competition, and plasticity in the development and representation of the lexicon. Fig. 2 presents a diagrammatic sketch of the revised model, DevLex-II. Compared with the original DevLex model (Farkas & Li, 2001; Li et al., 2004a), DevLex-II uses an additional output sequence map to better model comprehension and production as separate pathways (see brief discussion below; see details in Li, Zhao, & MacWhinney, 2007).

The model contains three layers of self-organizing feature maps: the input phonology map, the intermediate semantic map, and the output sequence map. These layers handle different levels of a word’s information: phonological content, semantic content, and

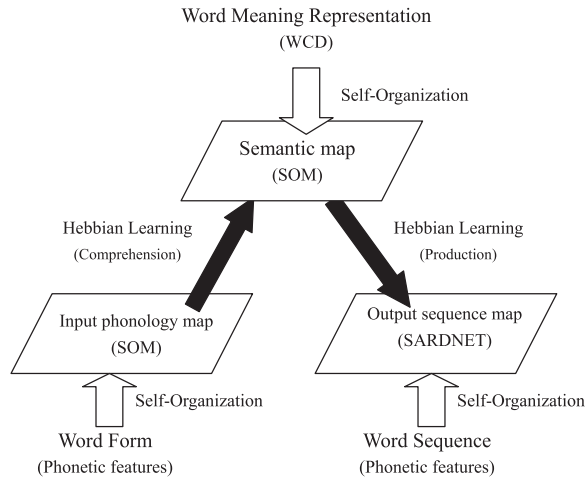


Fig. 2. A sketch of the DevLex-II model. Each of the three self-organizing maps (SOM) receives the lexical input and organizes the phonological, semantic, and phonemic sequence information of the lexicon, respectively. The associative connections between maps are trained by Hebbian learning (from Li et al., 2007; reproduced with permission from the Cognitive Science Society.)

articulatory sequence, respectively. Upon training of the network, word representations of these different levels are simultaneously presented to the network. This training process can be analogous to the child's analysis of a word's phonological, semantic, and phonemic sequence information upon hearing the word. For the phonological and semantic maps, the DevLex model forms activation patterns according to the standard algorithms of self-organizing maps (SOM; Kohonen, 1982, 2001).² For the output sequence map (which was absent from the original DevLex model), the DevLex-II model uses the algorithm of SARDNET (James & Miikkulainen, 1995), a type of temporal or sequential SOM network. The addition of the sequence map is inspired by models of word learning based on temporal sequence acquisition, and 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. Concurrent with the training of the three maps, the associative connections between maps are trained via Hebbian learning with a constant learning rate. Once the whole network is trained, the model's comprehension and production abilities can be probed into through phonology-to-semantics links and semantics-to-sequence links, respectively. (For technical description and other details of the model, see Li et al., 2007; Zhao & Li, 2007.)

The architecture of the model is built upon a number of considerations concerning the limitations of existing neural network models of language acquisition. First, most previous models have focused on highly simplified artificial patterns to approximate lexical items, while our model is designed to use realistic features of sound and meaning that make direct contact with the language input that the child receives (based on corpora of child-adult speech interactions). Second, most previous models have failed to extend the size of the input to a level that adequately simulates the actual growth of the child lexicon, while our

model is designed to scale up to a realistic level of early vocabulary such that details of organization and between-item competition could be studied. Third, our model is inspired by the idea that mental organization of the lexicon occurs without explicit instruction and corrective feedback (a self-organizing process), and thus it relies on unsupervised, self-organizing methods to learn (Li, 2003; MacWhinney, 2002).

The SOM-based self-organization has been motivated by topography-ordering features that can be found in many parts of the brain (Kohonen, 1982; Miikkulainen, Bednar, Choe, & Sirosh, 2005). The human cortex can be considered as consisting of multiple feature maps that handle auditory, visual, and other sensorimotor information, topographically ordered as a result of responding to input characteristics in the learning environment (Spitzer, 1999). The DevLex model, in addition to drawing on properties of SOM-based computation for within-modality self-organization, further relies on the training of Hebbian connections for between-modality interaction (see also Miikkulainen, 1997). In so doing it also achieves developmental realism with respect to the structure and characteristics of the input lexicon for the learning process.

2.2. *Nonlinear patterns of vocabulary growth: Insights from DevLex*

We can begin with two interesting observations about how children learn vocabularies. First, children clearly acquire a structured representation of the major linguistic categories, for example, of what we call nouns, verbs, and adjectives (see Fig. 1 for how this might be realized in a computational model over time). The structures in the representation tend to develop gradually, from a relatively random collection of items early on (e.g., mostly nouns that refer to objects), to a more organized group of words that distribute across the linguistic categories. Second, children clearly produce more and more words, but the size of the early vocabulary does not increase monotonically; rather, growth tends to follow a nonlinear pattern. Over the initial months, children add words to their productive inventory only slowly, which increases by a few words each month. After about 6 months of slow growth counting from the first productive words, children pick up the speed of learning when they can say about 50–100 words (roughly around 18–22 months), showing rapid vocabulary growth that is often referred to as the “vocabulary spurt” or “naming explosion” (Bates & Carnevale, 1993; Dromi, 1987; Goldfield & Reznick, 1990; McCarthy, 1946; Nelson, 1973; Van Geert, 1991). One of the goals of the DevLex model has been to link these two phenomena together, that is, to argue that structured representations lead to vocabulary spurt in development. This is possible with DevLex because the model, as described above, has the ability to learn the structure of the input through self-organization and to capture comprehension-production processes through trained Hebbian associative links.

A distinct advantage of DevLex is that it allows us to model the gradual emergence of structured lexical representations. This gradual emergence is the result of learning through self-organization in SOM, a process whereby similar inputs end up activating similar representational units in the network during pattern extraction and pattern formation. Over time, this process leads to topographically ordered structures in 2D maps, capturing the essential complexities of the input data. The basic algorithms of SOM give rise to localized

representations with soft boundaries while learning distributed patterns, which are ideally suited for understanding the structure of the early vocabularies that children build over time (see Ritter & Kohonen, 1989, for an earlier discussion of SOM's application to the study of words and meanings).

In a series of simulations (Li et al., 2007), we were able to connect when the network develops the basic structure of the lexicon with when it displays the ability of rapid vocabulary growth. The model gradually learns the representations of an early child lexicon (591 words from the CDI³; Dale & Fenson, 1996) and shows a rapid increase in the size of both receptive and expressive vocabularies after a slow start in early learning. This rapid increase, however, is prepared by the network's learning of the structured representations of phonemic sequence, phonology, and word meaning, as well as its learning of the mappings between these characteristics of the lexicon through Hebbian associative connections. Since the Hebbian connections are trained simultaneously with the network's self-organization in the individual maps, these two processes go hand in hand: Once patterns of representation have consolidated, the associative connections between the maps can also be reliably and rapidly strengthened. Recall that the associative Hebbian connections are used to model vocabulary comprehension and production processes, and hence a rapid association between the maps means a rapid acquisition of the relevant words for comprehension or production.

It is not difficult to see why this should be the case. When the network has formed basic structures in the corresponding maps, it means that the phonological structure and the semantic structure of some set of related words are identified. That is, meaningful clusters of words have formed and emerged in the SOM as concentrated patterns of activation, and these clusters or groups serve as catalysts or magnets for future learning. Imagine that a set of related nouns denoting the most familiar animals is learned and grouped together in the semantic map, and also that the corresponding associative connections between the semantic and phonological maps are established for these items. Next, a new word denoting another animal, "donkey," comes into learning. Because "donkey" shares significant similarities with already learned items such as "cow" and "horse," the system will quickly connect this new word with members of an existing category, due to their overlap in semantic or syntactic features. Once this overlap is identified, the system will readily map the new word through the already established associative pathways (e.g., from semantics to phonology), and relatively little effort will be involved in the new learning.

Li et al. (2007) showed that this process provides a mechanistic account of why vocabulary spurt tends to coincide with the system's building of a structured organization of the early lexicon (e.g., when about 100 words are learned). In this way the early structure provides a basic framework upon which later word learning can accelerate. In other words, earlier learned words help to form the initial links within and across the phonological and semantic levels so that future learning can capitalize more readily on the existing patterns and associations. At this point, word learning is no longer hampered by uncertainty or confusion on the maps, and the vocabulary spurt occurs.

How does this account of vocabulary spurt based on DevLex simulations square with other empirical and computational explanations of this phenomenon? In Li et al. (2007), we compared our modeling results with those from Plunkett, Sinha, Moller, and Strandsby

(1992), Siskind (1996), and Regier (2005) and suggested that our model provides a parsimonious but accurate account of the vocabulary spurt. In particular, Regier's (2005) LEX lexicon as exemplars model incorporated a mechanism of selective attention which allows the system to highlight relevant features for particular words while suppress irrelevant features, thus reducing memory interference. The net effect of this process is finer distinctions among word forms and word meanings, which is crucial for vocabulary spurt according to Regier (2005). On the one hand, there are clear parallels between LEX and DevLex, in that both models display high confusion rates in forms and meanings early on in learning and rapidly decreased confusion rates as vocabulary spurt begins. On the other hand, the two models differ in that while selective attention is required to drive successful learning in LEX, only more word learning itself is required for better representation in DevLex. This is because the self-organizing process continuously extracts more features for discrimination from the input space, which may be especially true in an incremental learning scenario, whereby lexical representation becomes enriched by incorporating more co-occurrence information over the course of learning (see e.g., Li et al., 2004a). Thus, while neither LEX nor DevLex makes recourse to factors external to learning, such as "naming insight" (McShane, 1979) or "communicative awakening" (Tomasello, 1999), DevLex further relies on characteristics inherent in the learning process itself for better and more efficient word learning at a later stage.

2.3. *Crosslinguistic patterns of vocabulary growth*

Our major goal in developing the DevLex model is to account for the dynamic interaction of the learner with the learning environment. While the discussion above can be taken to reveal how the system captures the learning process and the resulting representation, our data below indicate more clearly how the learning environment plays a crucial role in shaping the outcome of development. In particular, we apply our model to examine crosslinguistic similarities and differences in early lexical development and investigate how specific characteristics of the linguistic input can affect the developmental time course.

Empirical studies of children's vocabulary development often indicate distinct patterns of acquisition as a function of the specific type of language being acquired. For example, English-speaking children's early lexical repertoire shows a clear "noun bias" (Gentner, 1982), a preponderance of nouns compared to other categories of words. This noun bias has been recently found to be weak or nonexistent in some East Asian languages, for example, Chinese and Korean (Choi, 2000; Tardif, 1996, 2006), casting doubt on the generality of a universal, conceptually based, early lexical composition. Just what causes such differences is yet unclear. Some researchers suggest that factors such as parental input (and language-specific properties of the lexicon; see more discussion in section 3.2) might be key in accounting for such differences. For example, Tardif (2006) argued that there is a prevalence of verbs in adult Chinese as compared with adult English, and that verbs occur more frequently in child-directed parental speech in Chinese than in English. Thus, the linguistic input to the two groups of children might be different from the outset.

To realistically capture input characteristics of the child’s learning environment, DevLex has relied on the use of CDI, as discussed earlier. If input differences in English and Chinese hold the key for crosslinguistic differences in early lexicon, our model should be able to test this hypothesis by using identical parameters of learning while receiving different input to simulate early vocabulary learning in the two languages. To do this, we used the English CDI (Dale & Fenson, 1996) and the Chinese CDI (Tardif, Fletch, Liang, & Zhang, 2002) as the basis of the model’s input (500 words for each language). The corresponding CDI words were then treated in the standard DevLex procedure for phonological, phonemic, and semantic representations (see Li et al., 2007; Zhao & Li, 2007, 2008). To accurately represent word meanings, the model relied on a WCD, *word co-occurrence detector* (Farkas & Li, 2001; Li et al., 2004a) to extract the transitional probabilities of the input words from the CHILDES parental database (Li et al., 2000; MacWhinney, 2000), and then combined this information with semantic feature vectors based on computational thesauruses available for each of the two languages (WordNet and its Chinese equivalent HowNet; <http://www.keenage.com>). The use of CDI vocabularies and the CHILDES parental speech has been a consistent feature of our model in achieving both linguistic and developmental realism.

Two sets of networks with identical simulation parameters were run separately, one for each language. During a simulation, words from the training lexicon (Chinese or English) were presented to the network one by one, according to a word’s frequency of occurrence in the parental speech of the CHILDES database. Fig. 3 presents the average number of words correctly produced by the network for three major grammatical categories, nouns, verbs, and adjectives, as a function of the network’s expanding vocabulary size.

The results reveal a number of crosslinguistic similarities and differences. First, for both English and Chinese, our model, during most stages of vocabulary learning (represented by the total number of learned words), can correctly produce more nouns than verbs, and more verbs than adjectives. This overall noun advantage in our network’s productive vocabulary for both languages runs counter to the argument that verbs dominate over nouns in early

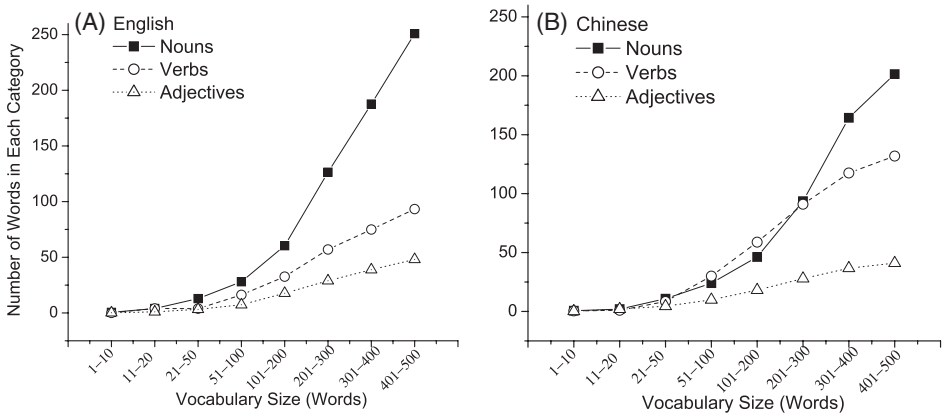


Fig. 3. Mean number of nouns, verbs, and adjectives learned by DevLex-II at different developmental stages for (A) English and (B) Chinese. Results are averaged across 10 simulation trials (from Zhao & Li, 2008).

Chinese vocabulary (Tardif, 1996), but is consistent with our recent corpus-based analyses of crosslinguistic vocabulary development (Liu, Zhao, & Li, 2008).⁴ Second, clear differences exist between the two languages, in that the network generally produced more nouns in English than in Chinese, and more verbs in Chinese than in English, across most stages. For Chinese, it produced comparable numbers of nouns and verbs at the early vocabulary stages. In two cases (i.e., when the total vocabulary size was between 51–100 and between 101–200 words), the model actually produced more verbs than nouns (by a small margin), but after 300 words it produced more nouns than verbs. In English, by contrast, nouns always dominated over verbs in number, starting from the very beginning. Moreover, the rate of increase for nouns in English was also more rapid than that for verbs. Thus, our network displayed a much stronger noun bias in English than in Chinese, consistent with empirical findings regarding crosslinguistic differences in vocabulary growth.

Why does our network show the different patterns in lexical composition for the two languages? We considered two important factors in the input to our network that could be the sources of the differences, namely, word length and word frequency in the input. Fig. 4A, B plot the word length distributions (in phonemes) of nouns and verbs in our Chinese and English lexicons, respectively. The distribution of Chinese words shown in Fig. 4A indicates that most verbs have their phonemic length in the range of two to four phonemes (left-skewed). The peak of the distribution is at three phonemes (corresponding to a typical monosyllable in Chinese), and more than 40% of all the verbs have such a word length. The nouns are more normally distributed around an average of four phonemes (from three to six phonemes, either monosyllables or disyllables), and nearly 20% of nouns have six phonemes. Fig. 4B shows a different picture for English. The distributions of nouns and verbs overlap more closely, and both follow a left-skewed pattern in English. Most of the words are in the range of three to five phonemes. The highly skewed distribution for Chinese verbs reflects the fact that a large percentage of words in this language are made up of monosyllables (represented by single characters),⁵ and this feature is more pronounced for verbs than for nouns. With regard to word frequency, we found that more verbs have higher

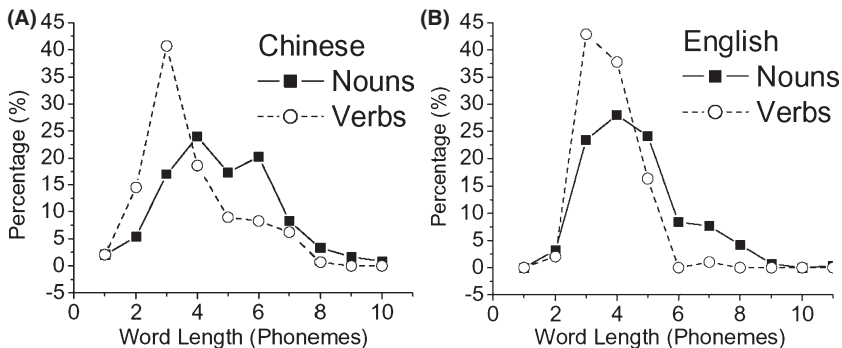


Fig. 4. Crosslinguistic differences in the distribution of nouns and verbs as a function of phonemic length of words: (A) Chinese; (B) English (from Zhao & Li, 2008).

frequencies in Chinese than in English when comparing the Chinese and the English parental speech in the CHILDES corpus (see Zhao & Li, 2008 for details). Consequently Chinese verbs may be generally easier to learn than nouns (but see Goodman, Dale, & Li, 2008, for a discussion of the role of parental input frequency in early vocabulary development). With these analyses we can suggest that children, as well as our network, capitalize on the linguistic characteristics of the lexicon in acquiring the early vocabulary, and in this process, verbs may be more advantaged in Chinese than in English (and other Indo-European languages) where nouns dominate in early learning.

Why is DevLex sensitive to language-specific properties such as word length and word frequency in vocabulary acquisition? The latter property has been extensively examined in connectionist language models since Plunkett and Marchman (1991), but the former has not been particularly investigated in this context. In DevLex, as in other connectionist models, word frequency is modeled by the number of times (tokens) the network is trained on for a given lexical item (type), and as such, items of higher frequency tend to lead to more robust representations in the system and are acquired earlier. In realistic learning, word frequency interacts with a host of other variables, including word length, in determining the time course of acquisition (see Goodman et al., 2008, for a recent discussion). With respect to word length, DevLex relies on a sequential phonemic output procedure that has a short-term temporal memory loop (see Li et al., 2007), and therefore the longer a word is, the more difficult the network will have in correctly outputting the phonemic sequence of that word. Over time, incremental learning of nouns versus verbs results in the different orders of acquisition as a function of different word lengths of items in the two categories.

In summary, our simulated results show clear crosslinguistic differences in early vocabulary development, and analyses of the nature of the learning environment, that is, the properties of the input to the model, shed light on the sources of such differences. Since our networks learning Chinese and English use the same architecture and the same learning parameters, we can safely conclude that the different input characteristics in the two lexicons must have contributed significantly to the cross-language differences observed. Elsewhere we have made similar arguments and analyses on how the model can be applied to capture crosslinguistic differences in other contexts, for example, in the acquisition of cryptotypes and tense-aspect morphology (e.g., Li, 1993, 2003; Li & MacWhinney, 1996; Li & Zhao, 2009; Zhao & Li, 2009). Much of our modeling work hinges on the notion that the mental lexicon develops as a system, and the study of its dynamic organization and structure facilitates the understanding of not only vocabulary acquisition but also the development of related linguistic systems (see General Discussion for more details; Li, 2006, Li, 2008). This contrasts with previous empirical focus on how children acquire lexical form-meaning mappings at the individual word level (e.g., on the initial fast-mapping process). In empirical studies it would be difficult to study large-scale lexical organizations in a systematic way as we do here, for example, by varying only the type of input and holding other learning parameters constant. This point brings us to the next section in which the power of computational modeling can be further attested and appreciated in the context in which not one, but two languages are acquired, either simultaneously or sequentially.

3. Lexical competition between languages

Our ability to acquire not only a first (L1) but also a second (L2) or more languages has long puzzled cognitive scientists. How does the learner, when faced with streams of multiple language signals, learn to identify, differentiate, represent, and retrieve words from a bilingual lexicon? Does the competition among words from two or more languages lead to mixed or confused representations during the course of learning? If so, how does the learner gradually arrive at a reasonably well-organized representation, if ever, for the bilingual lexicon? Does the representation differ as a function of when, how, and in which context L1 and L2 are acquired? These are only a few of the many important questions that naturally arise when we extend the view of lexical development from the monolingual to the bilingual context.

In studying bilingual lexical development, one might wish to consider a metaphor used by Bialystok (2001, pp. 224–225): If bilingualism is analogous to a smorgasbord at a dinner table, we will observe that some people eat several things at once, others eat one at a time (at varying speed), and still others come late and take only a bite of the dessert. While it is difficult if not impossible to study this type of volatile scenario empirically, that is, to systematically control for the great number of variables in the learning process, our research has taken advantage of the powerful modeling tools to parametrically manipulate the relevant variables (e.g., timing, rate, frequency, and quantity of L1 vs. L2 input). Our manipulations can be seen as a computational instantiation of how the linguistic smorgasbord is served and taken. In what follows, I report findings specifically from our studies of the timing (age of acquisition) and the amount of L2 input in the learning process. These findings lend us some insights into how multiple language systems interact and compete with each other, and how this competition affects the organization in the representation of bilingual lexicon.

3.1. *Competition in the acquisition of bilingual lexicon*

Two fundamental issues have occupied the minds of bilingual and second language acquisition researchers for decades. First, researchers have debated about the existence and the nature of a critical or sensitive period for second language acquisition, that is, the hypothesis that it is difficult or impossible to attain native proficiency in a second language after a given age period (e.g., Birdsong, 1999; Flege, Yeni-Komshian, & Liu, 1999; Johnson & Newport, 1989; Liu, Bates, & Li, 1992). Although there has been considerable dispute on this issue (see Hakuta, Bialystok, & Wiley, 2003; Harley & Wang, 1997; Long, 1990; Snow & Hoefnagel-Hohle, 1978), a general consensus has been that some form of age effects does exist, and it exists more strongly for certain aspects of the linguistic system than others (e.g., phonology and syntax, as compared with lexicon or semantics). Early explanations resorted to maturational accounts (Lenneberg, 1967), while more recent evidence points to the dynamic interaction of multiple language systems and early sensorimotor experience as potential sources of the age effects (Bates, 1999; Hernandez & Li, 2007; Hernandez et al., 2005; Kuhl, 2004). Second, an issue of enduring interest in bilingualism has been whether

bilingual representation takes the form of a single, shared lexical storage or separate, distinct storages in the mental lexicon (Kroll & de Groot, 1997; Kroll & Stewart, 1994; Potter, So, Von Eckhart, & Feldman, 1984; see Kroll & Tokowicz, 2005, for a review). This issue has been highly controversial and has recently been further complicated by seemingly relevant but conflicting neuroimaging data (see discussion below, and reviews in Abutalebi, Cappa, & Perani, 2001, 2005; Hernandez & Li, 2007). What seems to be certain from the debate is that a host of variables must be taken into consideration in resolving this issue, such as bilingual proficiency, age of acquisition, modality (comprehension vs. production), and type of linguistic stimuli in question (e.g., cognates vs. noncognates; de Groot, 1993; Dong, Gui, & MacWhinney, 2005; Kirsner, Smith, Lockhart, King, & Jain, 1984).

Traditionally, the age of acquisition issue has attracted more research attention in the study of second language acquisition (SLA), while the lexical representation issue has attracted more attention in the bilingual memory research. Our computational work has attempted to link these two issues together, that is, to argue that the nature of bilingual lexical representation depends on age of acquisition, or that age effects will directly impact the way in which words are acquired and represented. In making this connection we see SLA and bilingual processing as two faces of the same coin, especially when we take a developmental perspective toward the evolving representations during learning.

In Li and Farkas (2002) we used a variant of the DevLex model, SOMBIP, to establish the validity of self-organizing neural networks in understanding the emergence of lexical representations in two languages. The network was trained on child-directed speech from a bilingual Chinese-English family (Yip & Matthews, 2000), and with learning the network developed distinct lexical representations for each language, consistent with findings from French and Jacquet (2004) based on a bilingual SRN network trained on English and French data. In what follows, I discuss simulations that focused on how the representational structure develops and changes as a function of the learning history. In particular, in our simulations we manipulated the onset time of L2 learning relative to that of L1 learning to investigate the locus of age of acquisition effects. We hypothesized that the representational structure for the two lexicons in our model would differ as a function of L2 onset time. In addition, through analyzing the model's comprehension and production errors, we would see how the two developing lexical systems compete and interact with each other.

As in the monolingual model reported earlier (see section 2.3), we used the Chinese and English CDI vocabularies as the basis of our model's learning target (500 words from each language). The phonological, phonemic, and semantic representations were coded in the standard DevLex procedure (Li et al., 2004a, 2007; Zhao & Li, 2007). Again, the lexical semantic representations were based on both lexical co-occurrence information of the input words from the CHILDES parental corpus and the semantic features extracted from computational thesauruses available for each of the two languages.⁶ Unlike the monolingual simulations that employed two sets of networks separately, one for each language, the bilingual models were trained on the input lexicon from both languages in the same network.

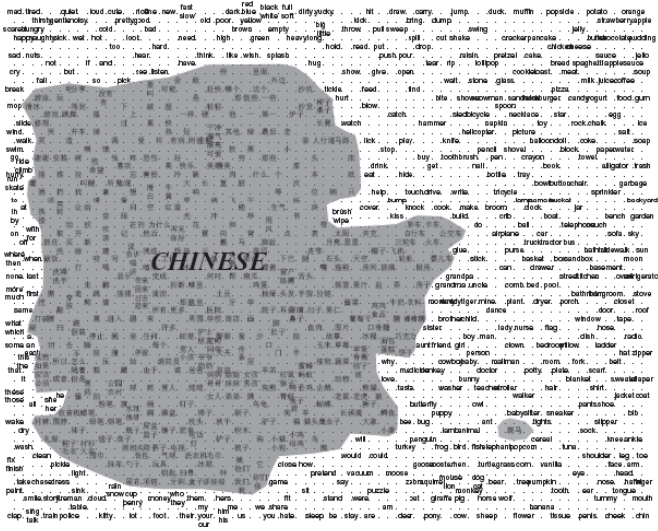
The training scenarios were manipulated in our model as follows. First, in the simultaneous training situation, the two lexicons were presented to the network gradually and in parallel. At the first stage, the training vocabulary included 50 English words and 50

Chinese words. Thereafter for every new stage, 50 more English words along with 50 more Chinese words were added to the training pool, until the size of each lexicon reached 500 at the final stage. Here, a training stage included 10 epochs so that each available word was presented to the network 10 times at each stage. Second, in the sequential training situation, 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 (English).⁷ Then the L2 words were presented to the network stage by stage (each stage with 50 more new L2 words) along with the corresponding increments 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 four stages. Then the training continued for another 10 stages until all the 500 L2 words were seen by the network (so that the exposure to L2 words in all three scenarios was constant, that is, involving 10 stages). By systematically manipulating the vocabulary onset time for L2 relative to that of L1, our model provides a principled way for us to identify the impact of lexical organization in one language on the lexical representation in the other language.

Fig. 5 presents a snapshot of the results from early versus late L2 learning in our simulations. These results contrast with the patterns from simultaneous learning situations (not shown here; see Fig. 2 of Li & Farkas, 2002), where the network is exposed to both languages from the beginning and could easily separate the two lexicons during learning and develop distinct representations for different languages. As seen in Fig. 5, however, when the network is exposed to the two languages sequentially, the results show a clear age effect. Specifically, if L2 was introduced into learning early on, the network could still restructure the lexical space, continually though slowly, to establish a lexical representation for the L2 independent of that for the L1, shown here as the “big island” surrounded by the L1 representation (Fig. 5A). If L2 was introduced to learning late in the process, then the network was unable to establish an independent lexical territory for L2 representation: Compared with L1 words, the L2 words occupied fragmented regions, often small and compact, seen as densely populated “small islands” (Fig. 5B). Table 1 presents a quantitative comparison of the averaged lexical space occupied by L1 and L2 lexicons, and the corresponding confusion rates (i.e., words that cannot be successfully distinguished by the network) associated with each language, under the three learning situations.

Why are the lexical organization patterns so different in the three learning situations? We believe that this is due to the significant competition that occurs between the two lexical systems and the associated developmental changes during learning. In the simultaneous learning situation, L1 and L2 lexicons can effectively compete with each other from the beginning.⁸ In the early learning situation, even though the initial 100 words may have set up the basic structure for the L1 lexicon, the network is still sufficiently “plastic,” when the functional units and their connections are not fully specified and are therefore still open to change. Because of this plasticity, the increased L2 vocabulary in the learning process presents a significant competition against the L1 lexicon, allowing for independent territories of the L2 representation to be established. Finally, in the late learning situation, L2 is introduced at a time when the learning system has been “entrenched,” when it has

(A) Early L2 learning



(B) Late L2 learning

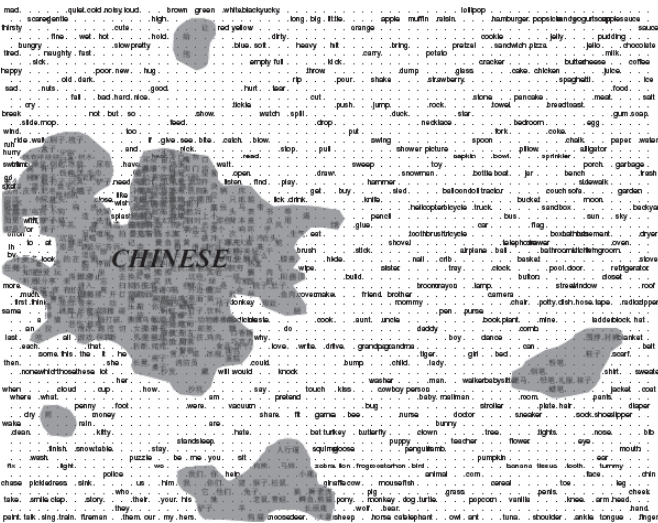


Fig. 5. Lexical organization as a function of early versus late L2 learning. In early learning (A), independent L2 representation can be established, while in late learning (B), only fragmented and compressed representations (“L2 islands”) can occur (from Zhao & Li, 2006).

Table 1
Size of lexical space and rate of confusion for English (L1) vs. Chinese (L2)

	Lexical Space ^a	Confusion Rate ^b
Simultaneous L1-L2 learning		
Chinese	2038	12.8%
English	2162	12.8%
L2:L1	0.94:1	1:1
Early L2 learning		
Chinese	1803	20.6%
English	2397	11%
L2:L1	0.75:1	1.87:1
Late L2 learning		
Chinese	956	64%
English	3244	2%
L2:L1	0.3:1	32:1

Note: ^aSize of lexical space is calculated as the number of nodes occupied by L1 or L2 words.

^bConfusion rate is defined as the percent of words that shared the same nodes with other words on the map (i.e., words that cannot be distinguished by the model).

significantly consolidated the L1 representation and dedicated its resources to L1 (learning of 400 words, 80% of the total target L1 vocabulary). In this case, new learning of L2 words can only use existing structures and the associative connections that have already been established by L1. We could say that at this point the L2 lexicon can only be parasitic on the L1 lexicon (see Hernandez et al., 2005). With regard to the dynamics of network learning, it is clear that along with the functional specification of the L1 representational structure comes the reduced plasticity of the network for radical restructuring or complete reorganization of the self-organizing map's topography. This reduced "neural" plasticity 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 & Lambon Ralph, 2000; Elman, 1993; Smith, Cottrell, & Anderson, 2001; see Hernandez & Li, 2007, for a review).

What implications do different L1 and L2 lexical representations have for bilingual language processing and production? Recent studies have suggested that bilingual individuals, as compared with monolinguals, often have more difficulty generating fast and accurate names in picture naming tasks (Gollan, Montoya, Fennema-Notestine, & Morris, 2005). One possible source of such production difficulties, based on our simulation results, could be due to the nature of the L2 representation, that is, that because lexical items in L2 are represented in more dense neighborhoods and hence in a more confusable fashion, the bilingual speaker has difficulty in retrieving the correct L2 items due to increased lexical competition from nearby items during speech production. In contrast to production, however, within-language semantic associations between lexical items in L2 comprehension might actually occur faster than between lexical items in L1 comprehension, due to the more compact L2 representations in close proximity. Such predictions can be tested empirically with further behavioral, computational, and neural studies.

3.2. *Neural signatures of bilingual lexical competition and representation*

If our modeling results above can be interpreted as suggesting an age of acquisition effect, what neural correlates will we find with bilingual individuals who have two lexical systems that are in constant competition? One way to answer this question is to directly probe into the representation by testing the same bilingual individual with comparable lexical items from both languages. This procedure can now be conducted effectively with the functional neuroimaging method. Researchers in the past 15 years have made significant progress in identifying the neural substrates of bilingual language representation and acquisition, using various noninvasive neuroimaging techniques, particularly the event-related potentials and the fMRI methods (see reviews and discussions in Abutalebi et al., 2001, 2005; Abutalebi & Green, 2007; Chee, 2006; Grosjean, Li, Münte, & Rodriguez-Fornells, 2003; Hernandez & Li, 2007; Li & Green, 2007; Vaid & Hull, 2002). A major question in this line of inquiry is how variables such as age of acquisition and bilingual proficiency jointly or separately modulate patterns of neural response in the bilingual brain (e.g., Chee, Tan, & Thiel, 1999; Kim, Relkin, Lee, & Hirsh, 1997; Klein, Milner, Zatorre, Meyer, & Evans, 1995; Mahendra, Plante, Magloire, Milman, & Trouard, 2003; Mechelli et al., 2004; Perani et al., 1998; Wartenburger et al., 2003). While our computational modeling clearly provides evidence for representational differences in kind as a function of age of acquisition (e.g., more fragmented and compressed for late L2, as compared with early L2 representations), the neuroimaging data to date have been unclear in this regard. Below I briefly discuss the implications of our recent fMRI study that complement our computational findings with respect to the nature of lexical representation in the bilingual brain.

A central issue in the cognitive neuroscience of language has been how the brain represents linguistic categories such as nouns and verbs. This issue is directly related to one of our major themes regarding lexical organization in development and the emergence of lexical categories (see section 2). In English, it has been shown that nouns and verbs elicit distinct cortical responses, in that nouns activate the posterior brain systems encompassing temporal-occipital regions while verbs activate the prefrontal and frontal-temporal regions (e.g., Damasio & Tranel, 1993; Martin, Haxby, Lalonde, Wiggs, & Ungerleider, 1995; Pulvermuller, 1999). Neuropsychological data from patients with brain injuries first showed this verb-frontal noun-posterior double dissociation (e.g., Bates, Chen, Tzeng, Li, & Opie, 1991; Caramazza & Hillis, 1991; Miceli, Silveri, Nocentini, & Caramazza, 1988). However, more recent evidence suggests that this view might be overly simplistic (see, e.g., Pulvermuller, 1999; Tyler, Russell, Fadili, & Moss, 2001). Li et al. (2004b) examined the representations of nouns and verbs in Chinese in an fMRI study and found that unlike in English, nouns and verbs in Chinese activate a wide range of overlapping brain areas as distributed neural networks, in both the left and the right hemisphere. We attributed the cortical response differences between Chinese and English to the language user's experience with and their sensitivity to language-specific properties of the lexicon and grammar: In English and other Indo-European languages, nouns and verbs are explicitly marked by grammatical morphology in a sentence context, whereas in Chinese, most of these grammatical markers for nouns and verbs are absent (with the exception of aspect markers); in addition, there is a

large set of class-ambiguous words (up to about 30% according to some estimate) that can be used as nouns and verbs and unlike their counterparts in English and other languages, these ambiguous words involve no morphological changes when used in the sentence as either nouns or verbs. Experiences with such specific properties of the target language may have helped to shape the different neural response patterns found with Chinese versus English speakers.

An interesting question arises as to how speakers of *both* Chinese and English might represent the bilingual lexicon in the same brain. Would nouns and verbs from Chinese and English be represented and processed in the same way as in one language, or would they exhibit specific cortical patterns of response according to the specific language? The question of how the bilingual brain represents two languages has been hotly debated. In particular, researchers debate whether bilingual representation involves a common or overlapping neural system for both languages or whether it is supported through distinct neural mechanisms that are weighted by properties of the particular languages (e.g., Abutalebi & Green, 2007; Chee et al., 1999; Kim et al., 1997; Klein et al., 1995; Perani et al., 1998). While the evidence so far seems to favor the common neural systems hypothesis for bilingual representation, no bilingual neuroimaging study has yet systematically examined the same types of words in typologically distinct languages. If we adopt the common neural systems hypothesis, we might predict that bilinguals will show the same patterns of brain activity when processing nouns and verbs from different languages.

To answer this question, we conducted an fMRI study, in which 11 early bilinguals were asked to read Chinese and English words (see Chan et al., 2008 for details). These bilinguals grew up as native speakers of Cantonese in Hong Kong, a bilingual environment, and started to learn English between ages 3 and 5. Anatomical and functional brain images were acquired while the participants performed a lexical decision task on nouns, verbs, and filler words from the two languages. Functional images were grouped into four sets: English nouns, English verbs, Chinese nouns, and Chinese verbs. Time course of the hemodynamic responses for selected regions of interest (ROI) was charted. Fig. 6 presents the results from the ROI analysis in two selected regions to contrast patterns of activation between Chinese words and English words.

Clear differences emerged between languages in the neural patterns due to the bilingual speakers' processing of nouns and verbs in the two languages. First, for Chinese, nouns and verbs activated a wide range of overlapping regions, distributed in frontal, parietal, and occipital areas as well as the cerebellum. This pattern is consistent with findings from our study of monolingual speakers' processing of nouns and verbs in Chinese (Li et al., 2004b). However, unlike the monolingual results, there were two regions that showed stronger activations when Chinese nouns were directly compared with Chinese verbs, the left fusiform (Fig. 6A) and the right middle frontal gyrus (Fig. 6B). Second, for English, nouns and verbs evoked a more focused activation of brain regions, mostly in the left frontal and parietal gyri. When directly compared with nouns, English verbs elicited significant activations in several additional regions, including the left putamen (Fig. 6C), cerebellum (Fig. 6D), supplementary motor area, and the right visual cortex. These additional regions are implicated

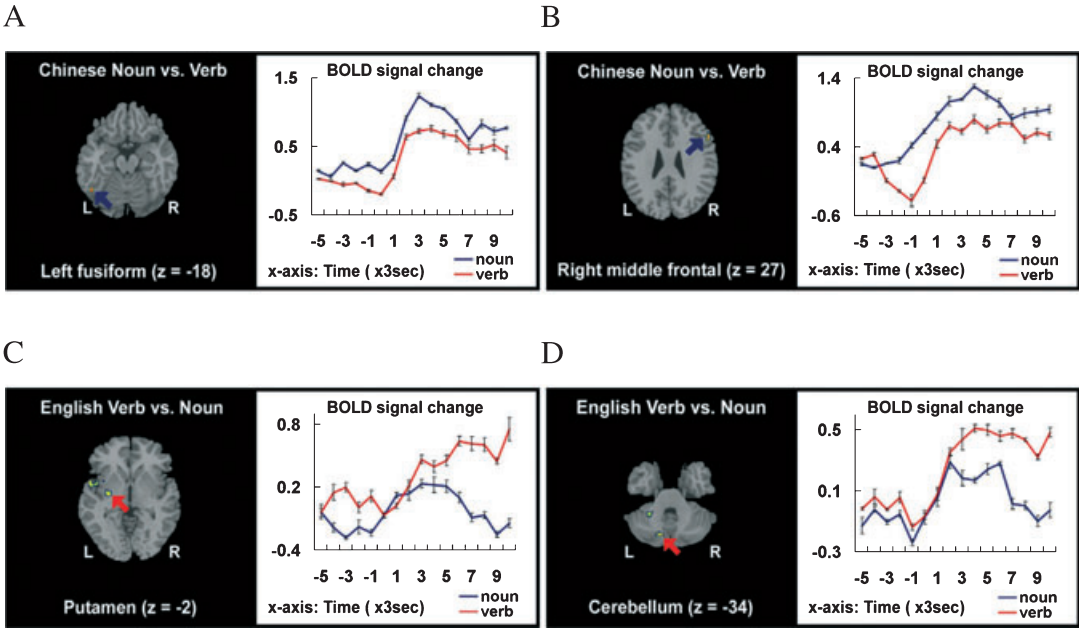


Fig. 6. Selected brain regions showing significant activation differences between Chinese and English in noun versus verb processing. Activation maps and time course results indicate that Chinese nouns elicited stronger activations than verbs in (A) the left fusiform and (B) the right middle frontal areas, while English verbs elicited stronger activations than nouns in (C) the left putamen and (D) the cerebellum. Error bars indicate standard errors of the mean (based on Chan et al., 2008; reproduced with permission from John Wiley & Sons).

in motor and sensory functions (e.g., formulating motor programming and coordinating speech articulation), and activation in these areas may have been due to the action and movement valence denoted by verbs.

These neuroimaging data show that our proficient early bilinguals display distinct neural patterns in processing nouns and verbs in the two languages, suggesting that lexical representation may take very different forms in the bilingual brain. Although our bilinguals do not behave as two monolinguals in the respective languages (hence the bilingual patterns are not the aggregate of the corresponding monolingual patterns; c.f. Grosjean, 1989), their neural responses in L1 and L2 are fundamentally distinct, suggesting that the bilingual brain is highly plastic early on and responsive to the specific linguistic properties of the two target languages. This is consistent with the argument that linguistic experience with language-specific properties shapes the neural systems of language representation (e.g., Li, 2006; Li et al., 2004a, 2004b; Perfetti et al., 2007; Tan et al., 2003): More specifically, the bilingual neural circuits are highly dependent on and responsive to specific phonological, morphological, or grammatical characteristics in each of the target languages. By contrast, our findings are at odds with the hypothesis that there is a common neural system for bilingual language representation. In our view, the common neural systems hypothesis has been based on cognitive performance tasks that draw on the language user's general linguistic experience

(e.g., verb generation, word repetition), tasks that do not tap into the representational system directly.

Our neuroimaging results are also consistent in general with our computational analyses discussed earlier. First, in the monolingual case (section 2.3), our crosslinguistic modeling results indicate that given the same learning parameters, the model produces different developmental profiles for nouns and verbs, depending on which language provides the initial input. Second, in the bilingual case (section 3.1), the structure of lexical representation may be fundamentally different depending on whether L2 is introduced early or late into the learning process. In our neuroimaging study, the bilingual speakers' L1 is Chinese, but they all have learned English at an early age, and therefore are able to build an English lexical representation system strong enough to compete with their L1 lexicon from early on. Thus, their distinct patterns of neural response accord well with the modeling patterns due to age of acquisition effects (see section 3.1). Future neuroimaging studies are needed to examine the neural correlates of noun and verb processing in late Chinese-English bilinguals in order to corroborate the analyses and hypotheses based on our computational modeling.⁹

4. Lexical competition within language

Our discussion so far indicates that the linguistic brain is highly plastic and hence highly sensitive to distinctive features of particular languages. This plasticity, however, holds only within certain time windows, as mentioned earlier. The question that remains is, If late language learning is so fundamentally different from early language learning, what dimensions of learning might be responsible for such differences? One hypothesis, based on the available neural and computational evidence so far, is the sensorimotor integration hypothesis (see Hernandez & Li, 2007, for discussion), which stipulates that sensorimotor learning is privileged early on for linguistic and musical skills, as this type of learning taps more directly into the acoustic, auditory, and phonological, as opposed to semantic and conceptual codes of language during early stages. This earlier-acoustic and later-semantic sequence may turn out to be significant for explaining age-related effects in language acquisition. In this section I provide neural evidence that indicates how the brain relies differentially on acoustic versus semantic cues during normal language processing, which we believe has strong implications for understanding mechanisms of language acquisition.

As is the case with nouns and verbs, typologically different languages tend to highlight different aspects of the linguistic system, leading to the idea that the same linguistic cues (e.g., morphology, syntax, semantics) may have different validities for speakers of different language. This idea has been captured elegantly by the Competition Model of Bates and MacWhinney (1982, 1987, 1989), according to which different linguistic cues compete with each other during language processing and language acquisition, and the cues that have higher validities tend to win the competition, causing faster processing and earlier acquisition.

Within the lexicon, there are at least the following cues that learners can attend to: phonemes (phonetic repertoire of a language), phonotactics (regularities underlying

phonemic co-occurrences), prosody (e.g., rhythm, stress, tone), and the semantic content of words. Ramus and Mehler (1999) considered phonetic repertoire, phonotactics, and prosody as the “prelexical” cues that listeners can use in discriminating between different languages, which differ from “lexical” knowledge that contains semantic and conceptual representations, which listeners also undoubtedly use in language processing. How different prelexical and lexical cues compete and interact during early language perception, in both the monolingual and the bilingual context, has been a topic of much empirical research. From a developmental point of view, prosodic and phonological processing abilities clearly develop very early, as young children have to use multiple prelexical cues for early word segmentation and speech perception (e.g., Jusczyk, Houston, & Newsome, 1999; Morgan & Demuth, 1996), and only gradually would they acquire the lexical semantic contents of words as their linguistic experience accrues.

To understand how prelexical and lexical cues compete in language perception, we used a paradigm called “language discrimination,” in which listeners are provided with natural or synthesized speech stimuli that contain either one or several prelexical cues, with or without lexical cues. The listeners’ task is to identify the language membership of the stimuli. This task is analogous to a situation in which we hear several groups of people talking in a restaurant and want to determine if they all speak the same language. Previous studies in language discrimination, however, have focused on prelexical cues and have not examined how both prelexical and lexical cues compete during language processing. We are interested in identifying the competition of prelexical and lexical cues, and the underlying neural mechanisms of how patterns of cue competition are reflected in cortical activities.

It has become increasingly clear that prosodic and phonological processing versus lexical semantic processing are subserved by separate brain regions, perhaps with overlapping boundaries (see, e.g., Bookheimer, 2002; Gandour, 2006; Hagoort, 2005; Price, 2000; Vigneau et al., 2006, for reviews). For example, phonological processing has been associated with neural activities in the left hemisphere in inferior frontal, superior temporal, and supramarginal areas (BA44/22/40; Brodmann’s areas), whereas processing of intonation and tones is often associated with activities in the right hemisphere in the anterior superior temporal gyrus (BA22) and the inferior frontal gyrus (operculus, BA44). By contrast, lexical semantic processing involves a stronger role of the temporal lobe, including the inferior temporal gyrus (BA20) and middle temporal gyrus (BA21) in the left hemisphere. The left inferior frontal gyrus (BA45/47) has also been implicated in semantic retrieval and word meaning selection. What has not become clear is how the different brain regions may interact and compete with each other during the same processing task.

In our study we manipulated four types of speech stimuli (see Zhao et al., 2008 for details): (a) synthesized speech with only rhythmic cues, in which all syllables were simple /s/ and /a/ combinations; (b) synthesized speech with both rhythmic and intonational (plus tonal) cues; (c) normal speech from Italian and Japanese, two unfamiliar languages (to Chinese listeners in the study) that have different patterns in prelexical cues; and (d) normal speech from Chinese (L1) and English (L2), familiar languages that contain both lexical and prelexical cues. Participants were exposed to these different types of stimuli and asked to perform an AX discrimination task (i.e., determining whether two successively presented

sentences A and X are from the same language). Functional images were acquired and grouped according to the four stimulus conditions and were compared between pairs of stimulus types. Fig. 7 presents the results from an ROI analysis to contrast patterns of activation elicited by familiar (Chinese and English) and unfamiliar languages (Japanese and Italian); this contrast involves patterns of neural response elicited by both lexical and prelexical cues.

Several observations can be made with regard to how prelexical and lexical cues compete and how such competitions are reflected as cortical activities. First, when processing unfamiliar languages such as Italian and Japanese, native Chinese listeners can only rely on prelexical cues, including phonological and prosodic regularities, but have no access to lexical semantic information. Thus, the left inferior frontal gyrus (along with superior temporal gyrus) associated with phonological and prosodic processing becomes highly activated. In contrast, when processing familiar languages in the L1 and L2, listeners can use both prelexical and lexical semantic knowledge, and thus, the left inferior temporal gyrus associated with semantic processing as well as the left inferior frontal gyrus become activated. These

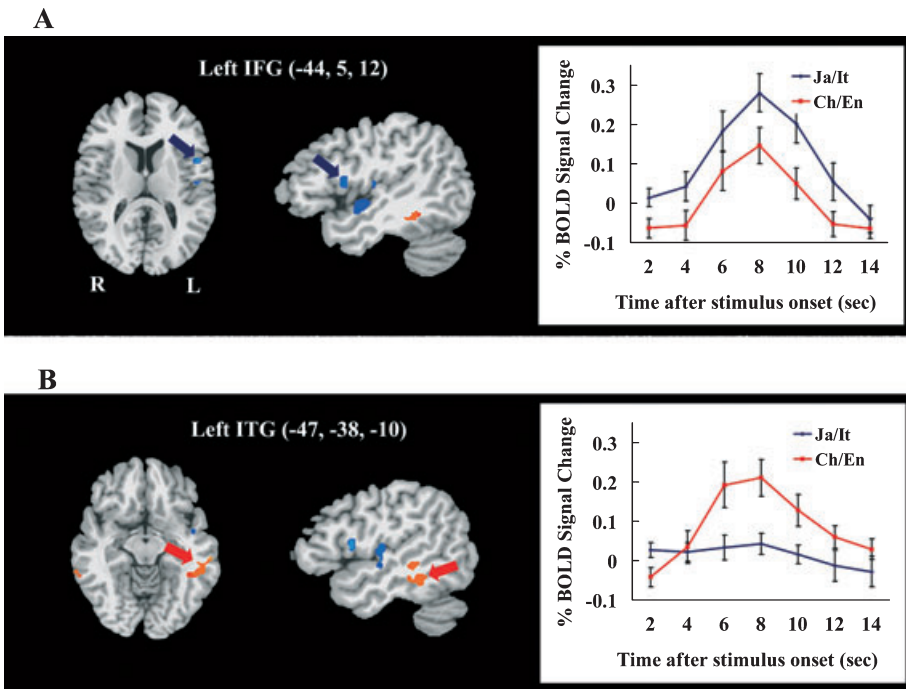


Fig. 7. Selected brain regions showing significant activation differences between familiar (Chinese/English) and unfamiliar (Italian/Japanese) stimulus conditions in the language discrimination task. Activation maps and time course results indicate that (A) unfamiliar languages elicited stronger activations than familiar languages in the left inferior frontal gyrus (IFG), while (B) familiar languages elicited stronger activations than unfamiliar languages in the left inferior temporal gyrus (ITG). Error bars indicate standard errors of the mean (based on Zhao et al., 2008; reproduced with permission from Elsevier).

patterns of activation show that complementary brain regions relevant to the processing task are recruited. Second, comparing Fig. 7A, B, we can see that when lexical semantic cues are available and semantic analysis is possible, as in the Chinese/English case, the brain regions associated with phonological or prosodic processing become less activated, as compared with the situation when only prelexical cues are available. This weaker activation may be due to cue competition, given that lexical semantics carry higher cue validities than prelexical cues and therefore lead more effectively and reliably to successful language identification and discrimination.

Finally, because our experimental design involved a hierarchically organized structure in which progressively more cues were included at higher levels (i.e., from prosody to phonology to semantics), we were able to tease apart patterns of cortical activities through identifying the competing cues at each level. Thus, we could see that depending on the amount and kind of cues available, different brain regions independently or jointly contribute to the processing. For example, comparing the conditions of rhythm versus rhythm plus intonation (see Zhao et al., 2008), we could see that while only the superior temporal gyrus (STG) is activated in the former, both the STG and the inferior frontal gyrus (IFG) are activated in the latter. Interestingly, this rather “additive” pattern of brain activities applies more to the prelexical than to the lexical cues. When both prelexical and lexical cues are present, cortical activities do not simply multiply but rather depend on the validity of the relevant cues for the task at hand (e.g., the lexical semantic cue wins out because it is more reliable). This progression from “additive” to “competitive” patterns of neural response could have significant implications for understanding the child-adult differences in language acquisition, given that infants initially focus on prelexical cues (prosody, rhythm, tones, intonations) and gradually acquire lexical semantic information, while adults have to attend to both prelexical and lexical cues at the same time in learning a second language. I return to this point in the General Discussion.

5. General discussion

The rapid acquisition of a large lexicon in early childhood is a remarkable human linguistic capacity. For decades, scholars have attempted to understand this capacity from multiple research angles. The general picture of how children acquire words has become increasingly clear now, thanks to the significant progresses made in the understanding of speech perception in preverbal infants (e.g., Jusczyk, 1999; Kuhl, 2000, 2004), infant statistical learning (e.g., GIoizzi et al., 2009; Saffran et al., 1996), fast mapping (e.g., Markson & Bloom, 1997; Werker, Cohen, Lloyd, Casacola, & Stager, 1998), and more recently, multimodal, real-time child–parent interactions (e.g., Yu, Ballard, & Aslin, 2005; Yu & Smith, 2007). But it was only until recently we have started to look at the acquisition, representation, and organization of the lexicon from a dynamical systems perspective, due in large part to Jeff Elman, Liz Bates, and other colleagues’ work (e.g., Bates, 1999; Bates & Elman, 1993; Bates et al., 1994; Elman, 1990, 1995, 2004; Elman et al., 1996; Smith & Thelen, 1994; Van Geert, 1991).

In this article, I outlined a proposal based on our research that attempts to examine the lexicon as a dynamical system, and I discussed how this proposal can be applied to account for lexical organization in development and lexical competition between and within languages. In particular, I provided computational evidence based on our DevLex model, and neuroimaging data based on the fMRI method, to illustrate how children and adults learn and represent the lexicon in their first and second languages. In the computational research, our goal has been to identify, through linguistically and developmentally realistic models, detailed cognitive mechanisms underlying the dynamic self-organizing processes in children's word learning and representation; in the neuroimaging research, our goal has been to identify the neural substrates that subserve lexical organization and competition in the monolingual and the bilingual brain. Our computational and neuroimaging approaches are complementary in that the former allows for the investigation of the representational structure of a large-scale mental lexicon, whereas the latter the study of neural correlates of organization and competition in vivo. While caution is needed in comparative interpretations given the different levels of granularity associated with the two kinds of data, our modeling and neuroimaging findings are often consistent in that they both point to how lexical representational structures can arise and change as a result of the competitive dynamic processes in learning.

One particular focus that has emerged from both our computational and neural studies, along with recent trends in cognitive and developmental sciences, is our attempt to understand the interactive dynamics involved in the acquisition and representation of one or multiple languages. In particular, our model provides insights into how early learning impacts later development, more specifically, how early learning leads to dedicated cognitive and neural structures that affect or shape the process and outcome of later development. Such effects may be cascade-like, nonlinear, and could be either positive or negative. For example, in the case of lexical acquisition in first language, early learned words establish a basic semantic framework or structure upon which later word learning can be built, leading to vocabulary spurt; in the case of bilingual lexical representation, previously established lexical structure in a first language often acts to impede optimal learning and representation of a second language lexicon. The latter is particularly true when the structural consolidation in L1 has reached a point where further reorganization becomes difficult or impossible, and if the learning of L2 occurs at this point, a fundamentally different structure will result for the representation of the L2 lexicon. Our analyses in Fig. 5 and Table 1 indicate clearly how the reduced plasticity of learning can cause compressed space, fragmented structure, and a parasitic lexicon overall. These results inform us of the intricate relationships between competition, entrenchment, and plasticity in the learning of first and second languages.

Understanding of such developmental dynamics is important, and it is possible once the dynamics are fleshed out mechanistically in a model such as the DevLex, in which learning outcomes and developmental trajectories are determined by the joint forces of learner-and-learning variables (e.g., timing of learning, characteristics of the input, and resources and capacity available to learning). The acquisition of structured representation has been looked at previously in other contexts as well, but it has not been systematically investigated with regard to the structural changes that occur in the lexicon as a dynamic interactive whole, as

examined in our research. For example, Bowerman (1982, 1988) was the first to point out that the representation of something like Whorf's (1956) covert semantic category—a cryptotype—could explain why children overgeneralize the English reversive prefix “un-” and produce errors like *unbury*, *unhang*, *unpress*, and *unsqueeze*. However, neither Whorf nor Bowerman provided a method to define a cryptotype or to capture how cryptotypes might emerge in children's mental representation.¹⁰ Li (1993, 2003) and Li and MacWhinney (1996) showed that cryptotypes can be acquired as distributed patterns through a connectionist discovery process, in that the system learns a weighted matrix of semantics-to-morphology relationships, and verbs that fall within a cryptotype tend to be the ones that are closest in this similarity space. Our connectionist model provides a mechanistic account of the semantic structure underlying cryptotypes and the developmental errors based on the structure (see empirical reports of errors in Bowerman, 1982; Clark, Carpenter, & Deutsch, 1995). The same type of account also successfully captures child-adult differences in the acquisition of tense-aspect markers in both first and second languages (see Li & Shirai, 2000; Li & Zhao, 2009; Zhao & Li, 2009). These examples demonstrate that structural changes in the organization of the lexicon are important, not only for the acquisition of the lexicon itself but also for related systems such as grammatical morphology (Li, 2003, 2006, 2008).

While the use of computational models and tools has significantly advanced the field of cognitive science, neuroimaging techniques, particularly fMRI, have swiftly revolutionized our understanding of language, brain, and culture. The frontiers of cognitive neuroscience, however, will no longer be the identification of component cognitive functions for individual cortical regions, but the understanding of the brain as a dynamical system of coordinated neural networks that are involved in information processing or problem solving. In particular, it is important to understand the interactive dynamics that lead to the recruitment of complementary and competing cortical regions, depending on the task demand and the nature of the problem. In section 4 I illustrated this aspect with an fMRI study, in which the lexicon is considered at different levels of complexity with regard to the validity of the relevant informational cue. The study shows that the perception and discrimination of language in a single task is the result of joint contributions from a network of multiple brain regions, each being weighted by the validity of the corresponding cues in processing. The competitive and interactive patterns provide a neural instantiation of the dynamical systems perspective on language, but they are at odds with theories that postulate “encapsulated” processes of language processing at different levels (c.f. Fodor, 1983).

The neuroimaging methodology has also quickly found its way in recent years into the study of bilingual language representation and acquisition. An underlying hypothesis of many of these studies is that an individual's experience with a second language carries not only cognitive consequences but also brings neuroanatomical changes to the brain: In the former case, for example, Bialystok (2001) and Bialystok, Craik, Klein, and Viswanathan (2004) have argued for a bilingual advantage in executive control, and in the latter case, for example, Mechelli et al. (2004) have proposed that bilingual experiences may lead to structural reorganization in the brain (e.g., increased gray-matter density in the inferior parietal

cortex). Such perspectives have been discussed in the literature as the “experience-mediated changes” (Trainor, 2005; Werker & Tees, 2005) or “experience-dependent synaptic change” (Bates, 1999). Recent neuroimaging evidence shows clearly that the brain is highly responsive to characteristics of the learning environment, and in the case of bilingual processing, responsive to the properties of the specific language (see, e.g., Perfetti et al., 2007; Zhang & Wang, 2007 for reviews). Our fMRI study of early Chinese-English bilinguals indicates that the neural patterns of response to nouns and verbs in the bilingual’s two languages are modulated by the bilingual learner’s experience with (and sensitivity to) the specific structural characteristics of the corresponding language. Our computational model also demonstrates distinct bilingual representations if L2 is learned sufficiently early to compete effectively with the representation of L1.

While we are still far from understanding the computational and neural mechanisms responsible for the learning differences between children and adults and between L1 and L2 learning, evidence has been accumulating and pointing to promising directions. Hernandez and Li (2007), reviewing the literature on age of acquisition, propose that differences in sensorimotor integration during early versus late stages of learning could be a key factor in accounting for age-related effects in first and second language acquisition. Specifically, they point out that early versus late learned words are fundamentally different in the way they are represented and retrieved in the monolingual and the bilingual brain. Early learned words tend to rely more strongly on auditory processing, activating brain regions near Heschl’s gyrus, while late learned words tend to rely more strongly on semantic processing and hence require more effortful articulatory and motor processes during word retrieval or picture naming. This disparity in neural patterns suggests that early on, language learning taps more directly into cortices that deal with acoustic and phonological processing, skills that place high demands on sensorimotor integration; by contrast, later acquisition taps more directly into cortical regions responsible for semantic integration rather than automatic phonological processes.

When pitted against our cortical competition data discussed in section 4, these analyses suggest that children might be particularly privileged early on by focusing on sensorimotor integration and working on auditory and acoustic processing, and by taking advantage of the gradual, “additive” neural effects of prelexical cues in building up the early preverbal linguistic representation. The acquisition of a language in infancy through early childhood tends to progress from low-level perceptual, suprasegmental cues (e.g., rhythm and prosody) to segmental (phonology and phonemic sequences), and finally to lexical semantic cues. By contrast, acquisition of a new language in adulthood does not give the learner the luxury of this type of progressive development; instead, the learner has to face the “competitive” processes of prosodic, phonological, and semantic integration all at once. At this point, we can only speculate about what brain correlates this kind of learning difference might have; for example, it is possible that crucial neural systems for sensorimotor integration and coordination (e.g., the frontal-basal ganglia neural circuitry) undergo progressive though rapid organization during early perceptual and auditory learning, whereas neural systems for semantic integration and world knowledge (e.g., the inferior and middle temporal cortex) develop at a somewhat later stage in learning. If this is the case, then adults would be highly

disadvantaged by having possessed a fully developed competitive neural system when learning a second language (one could consider this a neural example of “less is more”; c.f. Johnson & Newport, 1989). This disadvantage may be particularly biased against the learning of low-level perceptual prelexical cues such as rhythm, prosody, and phonotactics, given that (a) lexical semantic cues, as suggested by our study, are more favorably selected by the competition process because of their higher validity for language processing; and (b) adult learners have already in place an established lexical semantic system and a large knowledge base due to their native language.

If properties of the lexicon provide a foundation to the learner for cracking the puzzle of language, the understanding of the interactive dynamics involved in the acquisition, representation, and organization of lexical systems will provide an important vantage point for understanding the computational and neural bases of how single or multiple languages are acquired and represented by the brain. Future challenges will be placed on the understanding of how various acoustic, visual, and semantic cues in the learning environment are integrated by the monolingual and the bilingual brain, how that integration occurs in real time, how previously acquired representations guide and select the integration process, and finally, what structural and functional changes in the dynamic neural system accompany the process and affect the timing and the outcome of the integration.

Notes

1. According to lexically motivated theories of language acquisition (e.g., item-based learning; MacWhinney, 1987, 2004; Tomasello, 2000, 2003), the acquisition of grammar is accompanied, and at the same time constrained, by the acquisition of the lexicon.
2. A mechanism for inserting or growing neurons in the semantic map in response to learning demands was described in Li, Farkas, and MacWhinney (2004a).
3. It is also called MCDI, since the work was sponsored by the MacArthur Foundation. It was later named MacArthur-Bates Communicative Development Inventory, in memory of Elizabeth Bates’s contribution to the study of language development.
4. In a more recent formulation, Tardif (2006) does not argue for a “verb bias” but instead suggests that nouns and verbs show parallel growth patterns in Chinese (with perhaps only a marginal verb advantage at the very earliest stages). This revised view is more consistent with our simulation patterns presented here.
5. This is especially true for high-frequency words; according to the *Corpus for Modern Chinese Research* (Sun, Sun, Huang, Li, & Xing, 1996), among the 1,000 most frequent words 46% are monosyllables.
6. Our semantic representations are language-specific rather than language-independent conceptual representations, given that they are generated from word co-occurrences in the linguistic input and word-specific semantic features of each language. Most bilingual lexical memory research does not make this semantic-conceptual distinction, which might explain why prevalent bilingual lexical models assume a single

conceptual level for both languages (represented by one box in diagrammatical sketches) but separate L1 and L2 lexical forms (represented by separate boxes, one for each language).

7. In separate simulations we obtained similar results with Chinese being the L1 and English the L2. Note also that the choice of an initial 100 words was based on results from earlier simulations (Li et al., 2007) that indicate that the basic lexical structure may be established at this level (see also discussion in section 2.2).
8. Our results from simultaneous learning are consistent with the Dual Language System Hypothesis (Genesee, 1989; Genesee, Paradis, & Crago, 2004), according to which children exposed to two languages from the beginning can quickly establish two separate linguistic systems at early stages of acquisition.
9. It should be noted that in making this comparison between fMRI-derived neural patterns and computational modeling results we are not making the claim that our model has the necessary neural structures or mechanisms that lend itself to modeling neural patterns of language representation. Cautions are needed in interpreting neuroimaging patterns along with computational representations, as the two may be structurally different (we are grateful to an anonymous reviewer for this point). Still, some parallels may be drawn in functional terms as to what impact specific variables may have in determining patterns of activation and patterns of representation as a result of specific learning experience.
10. According to Whorf (1956), cryptotypes are elusive and subtle, difficult to pin down with a single label, and hence “intangible.” Connectionist networks provide a mechanism for cryptotypes, as discussed by Li (2003), Li (2008), and Li and MacWhinney (1996).

Acknowledgments

Preparation of this article was supported by a grant from the National Science Foundation (No. BCS-0642586). I would like to thank Elizabeth Bates, Jeff Elman, Igor Farkas, Brian MacWhinney, Hua Shu, Li Hai Tan, and Xiaowei Zhao for their comments, discussions, and collaborations.

This article was completed while the author was working for the National Science Foundation. The opinions expressed in this article are those of the author and do not necessarily reflect the views of the National Science Foundation.

References

- Abutalebi, J., Cappa, S. F., & Perani, D. (2001). The bilingual brain as revealed by functional neuroimaging. *Bilingualism: Language and Cognition*, 4, 179–190.
- Abutalebi, J., Cappa, S. F., & Perani, D. (2005). What can functional neuroimaging tell us about the bilingual brain? In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism* (pp. 497–515). Oxford, England: Oxford University Press.

- Abutalebi, J., & Green, D. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics*, 20, 242–275.
- Altmann, G. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, 33.
- Bates, E. (1999). Plasticity, localization and language development. In S. Broman & J. M. Fletcher (Eds.), *The changing nervous system: Neurobehavioral consequences of early brain disorders* (pp. 214–253). New York: Oxford University Press.
- Bates, E., & Carnevale, G. (1993). New directions in research on language development. *Developmental Review*, 13, 436–470.
- Bates, E., Chen, S., Tzeng, O., Li, P., & Opie, M. (1991). The noun-verb problem in Chinese aphasia. *Brain and Language*, 41, 203–233.
- Bates, E., & Elman, J. (1993). Connectionism and the study of change. In M. Johnson (Ed.), *Brain development and cognition: A reader* (pp. 623–642). Oxford, England: Blackwell Publishers.
- Bates, E., & MacWhinney, B. (1982). Functionalist approaches to grammar. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 173–218). New York: Cambridge University Press.
- Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 157–194). Hillsdale, NJ: Lawrence Erlbaum.
- Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 3–76). New York: Cambridge University Press.
- Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P. S., Reznick, J. S., Reilly, J., & Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language*, 21, 85–123.
- Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition*. Cambridge, England: Cambridge University Press.
- Bialystok, E., Craik, F., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging*, 19, 290–303.
- Birdsong, D. (Ed.) (1999). *Second language acquisition and the critical period hypothesis*. Mahwah, NJ: Lawrence Erlbaum.
- Bookheimer, S. (2002). Functional MRI of language: New approaches to understanding cortical organization of semantic processing. *Annual Review of Neuroscience*, 25, 151–188.
- Bowerman, M. (1982). Reorganizational processes in lexical and syntactic development. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 319–346). Cambridge, England: Cambridge University Press.
- Bowerman, M. (1988). The “no negative evidence” problem: How do children avoid constructing an overly general grammar? In J. Hawkins (Ed.), *Explaining language universals* (pp. 73–101). Oxford, England: Blackwell.
- Burgess, C., & Lund, K. (1997). Modelling parsing constraints with high-dimensional context space. *Language and Cognitive Processes*, 12, 177–210.
- Burgess, C., & Lund, K. (1999). The dynamics of meaning in memory. In E. Dietrich & A. Markman (Eds.), *Cognitive dynamics: Conceptual change in humans and machines* (pp. 17–56). Hillsdale, NJ: Lawrence Erlbaum.
- Caramazza, A., & Hillis, A. E. (1991). Lexical organization of nouns and verbs in the brain. *Nature*, 349, 788–790.
- Chan, A., Luke, K., Li, P., Yip, V., Li, G., Weekes, B., & Tan, L. (2008). Neural correlates of nouns and verbs in early bilinguals. *Annals of the New York Academy of Sciences*, 1145, 30–40.
- Chee, M. W. L. (2006). Language processing in bilingual as revealed by functional imaging: A contemporary synthesis. In P. Li, L. H. Tan, E. Bates & O. Tzeng, *Handbook of East Asian psycholinguistics* (Vol. 1: Chinese, pp. 287–297). Cambridge, England: Cambridge University Press.
- Chee, M. W. L., Tan, E. W. L., & Thiel, T. (1999). Mandarin and English single word processing studied with fMRI. *Journal of Neuroscience*, 19, 3050–3056.
- Choi, S. (2000). Caregiver input in English and Korean: Use of nouns and verbs in book-reading and toy-play contexts. *Journal of Child Language*, 27, 69–96.

- Clark, E., Carpenter, K., & Deutsch, W. (1995). Reference states and reversals: Undoing actions with verbs. *Journal of Child Language*, 22, 633–662.
- Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers*, 28, 125–127.
- Damasio, A., & Tranel, D. (1993). Nouns and verbs are retrieved with differently distributed neural systems. *Proceedings of the National Academy of Sciences of the United States of America*, 90, 4957–4960.
- Dong, Y., Gui, S., & MacWhinney, B. (2005). Shared and separate meanings in the bilingual lexical memory. *Bilingualism: Language and Cognition*, 8 (3), 221–238.
- Dromi, E. (1987). *Early lexical development*. Cambridge, England: Cambridge University Press.
- Ellis, A. W., & Lambon Ralph, M. A. (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, 1103–1123.
- Elman, J. (1990). Finding structure in time. *Cognitive Science*, 14, 179–211.
- Elman, J. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, 48, 71–99.
- Elman, J. (1995). Language as a dynamical system. In R. Port & T. van Gelder (Eds.), *Mind as motion: Dynamical perspectives on behavior and cognition* (pp. 195–223). Cambridge, MA: MIT Press.
- Elman, J. (2004). An alternative view of the mental lexicon. *Trends in Cognitive Sciences*, 8, 301–306.
- Elman, J. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science*, 33.
- Elman, J., Bates, A., Johnson, A., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.
- Farkas, I., & Li, P. (2001). A self-organizing neural network model of the acquisition of word meaning. In E. M. Altmann, A. Cleeremans, C. D. Schunn & W. D. Gray (Eds.), *Proceedings of the Fourth International Conference on Cognitive Modeling* (pp. 67–72). Mahwah, NJ: Lawrence Erlbaum.
- Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, 41, 78–104.
- Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: Bradford.
- French, R. M., & Jacquet, M. (2004). Understanding bilingual memory. *Trends in Cognitive Science*, 8, 87–93.
- Gandour, J. (2006). Brain mapping of Chinese speech prosody. In P. Li, L. H. Tan, E. Bates & O. J. L. Tzeng (Eds.), *Handbook of East Asian psycholinguistics* (Vol. 1: Chinese, pp. 308–319). Cambridge, England: Cambridge University Press.
- Genesee, F. (1989). Early bilingual development: One language or two? *Journal of Child Language*, 6, 161–179.
- Genesee, F., Paradis, J., & Crago, M. (2004). *Dual language development and disorders*. Baltimore, MD: Paul H. Brookes Publishing Co.
- Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj (Ed.), *Language development: Vol. 2. Language, thought and culture* (pp. 301–334). Hillsdale, NJ: Lawrence Erlbaum.
- Gliozzi, V., Mayor, J., Hu, J.-F., & Plunkett, K. (2009). Labels as features (not names) for infant categorisation: A neuro-computational approach. *Cognitive Science*, 33.
- Goldfield, B. A., & Reznick, J. S. (1990). Early lexical acquisition: Rate, content, and the vocabulary spurt. *Journal of Child Language*, 17, 171–183.
- Gollan, T., Montoya, R., Fennema-Notestine, C., & Morris, S. (2005). Bilingualism affects picture naming but not picture classification. *Memory and Cognition*, 33, 1220–1234.
- Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, 35, 515–531.
- de Groot, A. M. B. (1993). Word-type effects in bilingual processing tasks: Support for a mixed representational system. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 27–51). Amsterdam: John Benjamins.

- Grosjean, F. (1989). Neurolinguist, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, 36, 3–15.
- Grosjean, F., Li, P., Münte, T., & Rodriguez-Fornells, A. (2003). Imaging bilinguals: When the neurosciences meet the language sciences. *Bilingualism: Language and Cognition*, 6, 159–165.
- Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences*, 9 (9), 416–423.
- Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical evidence: A test of the critical-period hypothesis for second-language acquisition. *Psychological Science*, 14, 31–38.
- Hare, M., McRae, K., & Elman, J. (2003). Sense and structure: Meaning as a determinant of verb subcategorization preferences. *Journal of Memory and Language*, 48, 281–303.
- Hare, M., McRae, K., & Elman, J. (2009). This wind chilled the spectators, but the wind just chilled: Sense, structure, and sentence comprehension. *Cognitive Science*, 33.
- Harley, B., & Wang, W. (1997). The critical period hypothesis: Where are we now? In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 1–50). Hillsdale, NJ: Lawrence Erlbaum.
- Hebb, D. (1949). *The organization of behavior: A neuropsychological theory*. New York: Wiley.
- Hernandez, A., & Li, P. (2007). Age of acquisition: Its neural and computational mechanisms. *Psychological Bulletin*, 133 (4), 638–650.
- Hernandez, A., Li, P., & MacWhinney, B. (2005). The emergence of competing modules in bilingualism. *Trends Cognitive Science*, 9 (5), 220–225.
- Jackendoff, R. (2002). *Foundations of language*. New York: Oxford University Press.
- James, D. L., & Miikkulainen, R. (1995). SARDNET: A self-organizing feature map for sequences. In G. Tesauro, D. Touretzky & T. Leen (Eds.), *Advances in neural information processing systems 7* (pp. 577–584). Cambridge, MA: MIT Press.
- Johnson, J. S., & Newport, E. K. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21, 60–99.
- Jusczyk, P. (1999). How infants begin to extract words from speech. *Trends in Cognitive Sciences*, 3, 323–328.
- Jusczyk, P., Houston, D., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, 39, 159–207.
- Kim, K., Relkin, N., Lee, K., & Hirsh, J. (1997). Distinct cortical areas associated with native and second languages. *Nature*, 388, 171–174.
- Kirsner, K., Smith, M. C., Lockhart, R. S., King, M. L., & Jain, M. (1984). The bilingual lexicon: Language-specific units in an integrated network. *Journal of Verbal Learning and Verbal Behavior*, 23, 519–539.
- Klein, D., Milner, B., Zatorre, R. J., Meyer, E., & Evans, A. C. (1995). The neural substrates underlying word generation: A bilingual functional-imaging study. *Proceedings of the National Academy of Sciences of the United States of America*, 92, 2899–2903.
- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43, 59–69.
- Kohonen, T. (2001). *Self-organizing maps* (3rd ed.). Berlin: Springer.
- Kroll, J. F., & de Groot, A. M. . B. (1997). Lexical and conceptual memory in the bilingual: Mapping form to meaning in two languages. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 169–199). Mahwah, NJ: Erlbaum.
- Kroll, J. F., & Stewart, E. (1994). Category interferences in translation and picture naming: Evidence for asymmetric connection between bilingual memory representation. *Journal of Memory and Language*, 33, 149–174.
- Kroll, J. F., & Tokowicz, N. (2005). Models of bilingual representation and processing. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 531–553). New York: Oxford University Press.
- Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 11850–11857.

- Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, 5, 831–843.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211–240.
- Lenneberg, E. (1967). *Biological foundations of language*. New York: Wiley.
- Li, P. (1993). Cryptotypes, form-meaning mappings, and overgeneralizations. In E. V. Clark (Ed.), *Proceedings of the Twenty-fourth Child Language Research Forum* (pp. 162–178). Stanford, CA: Center for the Study of Language and Information.
- Li, P. (2003). Language acquisition in a self-organising neural network model. In P. Quinlan (Ed.), *Connectionist models of development: Developmental processes in real and artificial neural networks* (pp. 115–149). Hove, England: Psychology Press.
- Li, P. (2006). In search of meaning: The acquisition of semantic structure and morphological systems. In J. Luchjenbroers (Ed.), *Cognitive linguistics investigations across languages, fields, and philosophical boundaries* (pp. 109–137). Amsterdam: John Benjamins, Inc.
- Li, P. (2008). Meaning in acquisition: Semantic structure, lexical organization, and crosslinguistic variation. In V. Gathercole (Ed.), *Routes to language: Studies in honor of Melissa Bowerman* (pp. 257–284). London: Taylor & Francis.
- Li, P., Burgess, C., & Lund, K. (2000). The acquisition of word meaning through global lexical co-occurrences. In E. V. Clark (Ed.), *Proceedings of the Thirtieth Annual Child Language Research Forum* (pp. 167–178). Stanford, CA: Center for the Study of Language and Information.
- 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, B. (2004a). Early lexical development in a self-organizing neural networks. *Neural Networks*, 17, 1345–1362.
- Li, P., & Green D. (Eds.) (2007). Neurocognitive approaches to bilingualism: Asian languages. A special issue of *Bilingualism: Language and Cognition*, 10, 117–210.
- Li, P., Jin, Z., & Tan, L. (2004b). Neural representations of nouns and verbs in Chinese: An fMRI study. *Neuro-Image*, 21, 1533–1541.
- Li, P., & MacWhinney, B. (1996). Cryptotype, overgeneralization, and competition: A connectionist model of the learning of English reversive prefixes. *Connection Science*, 8, 3–30.
- Li, P., & Shirai, Y. (2000). *The acquisition of lexical and grammatical aspect*. Berlin: Mouton de Gruyter.
- Li, X., Shu, H., Liu, Y. Y., & Li, P. (2006). Mental representation of verb meanings: Behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience*, 18, 1774–1787.
- Li, P., & Zhao, X. (2009). Computational modeling of the expression of time. In W. Klein & P. Li (Eds.), *The expression of time in language*. Berlin: Mouton.
- Li, P., Zhao, X., & MacWhinney, B. (2007). Dynamic self-organization and early lexical development in children. *Cognitive Science*, 31, 581–612.
- Liu, H., Bates, E., & Li, P. (1992). Sentence interpretation in bilingual speakers of English and Chinese. *Applied Psycholinguistics*, 13, 451–484.
- Liu, S., Zhao, X., & Li, P. (2008). Early lexical development: A corpus-based study of three languages. In B. C. Love, K. McRae & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 1343–1348). Austin, TX: Cognitive Science Society.
- Long, M. (1990). Maturation constraints on language development. *Studies in Second Language Acquisition*, 12, 251–285.
- MacWhinney, B. (1987). The competition model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 249–308). Hillsdale, NJ: Erlbaum.
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Hillsdale, NJ: Lawrence Erlbaum.
- MacWhinney, B. (2002). Lexical connectionism. In P. Broeder & J. M. Murre (Eds.), *Models of acquisition: Inductive and deductive approaches* (pp. 9–31). Oxford, England: Oxford University Press.

- MacWhinney, B. (2004). A multiple process solution to the logical problem of language acquisition. *Journal of Child Language*, 31, 883–914.
- Mahendra, N., Plante, E., Magloire, J., Milman, L., & Trouard, T. P. (2003). fMRI variability and the localization of languages in the bilingual brain. *Neuroreport*, 14, 1225–1228.
- Markson, L., & Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature*, 385, 813–815.
- Martin, A., Haxby, J. V., Lalonde, F. M., Wiggs, C. L., & Ungerleider, L. G. (1995). Discrete cortical regions associated with knowledge of color and knowledge of action. *Science*, 270, 102–105.
- McCarthy, D. (1946). Language development in children. In L. Carmichael (Ed.), *Manual of child psychology* (pp. 476–581). New York: John Wiley & Sons.
- McRae, K. (2009). Conceptual hierarchies in a flat attractor network: Dynamics of learning and computations. *Cognitive Science*, 33.
- McShane, J. (1979). The development of naming. *Linguistics*, 17, 879–905.
- Mechelli, A., Grinioni, J., Noppeney, U., O’Doherty, J., Ashburner, J., Frackowiak, R., Price, C. (2004). Structural plasticity in the bilingual brain. *Nature*, 432, 757.
- Miceli, G., Silveri, M. C., Noccioni, U., & Caramazza, A. (1988). Patterns of dissociation in comprehension and production of nouns and verbs. *Aphasiology*, 2, 351–358.
- Miikkulainen, R. (1993). *Subsymbolic natural language processing: An integrated model of scripts, lexicon, and memory*. Cambridge MA: MIT Press.
- Miikkulainen, R. (1997). Dyslexic and category-specific aphasic impairments in a self organizing feature map model of the lexicon. *Brain and Language*, 59, 334–366.
- Miikkulainen, R., Bednar, J., Choe, Y., & Sirosh, J. (2005). *Computational maps in the visual cortex*. New York: Springer.
- Morgan, J. L., & Demuth, K. (1996). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Mahwah, NJ: Erlbaum.
- Nelson, K. (1973). *Structure and strategy in learning to talk*. Chicago: University of Chicago Press.
- Perani, D., Paulesu, E., Sebastian-Galles, N., Dupoux, E., Dehaene, S., Bettinardi, V. et al. (1998). The bilingual brain: Proficiency and age of acquisition of the second language. *Brain*, 121, 1841–1852.
- Perfetti, C. A., Liu, Y., Fiez, J., Nelson, J., Bolger, D. J., & Tan, L. H. (2007). Reading in two writing systems: Accommodation and assimilation of the brain’s reading network. *Bilingualism: Language and Cognition*, 10, 131–146.
- Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perceptron: Implications for child language acquisition. *Cognition*, 38, 43–102.
- Plunkett, K., Sinha, C. G., Moller, M. F., & Strandsby, O. (1992). Symbol grounding or the emergence of symbols? Vocabulary growth in children and a connectionist net. *Connection Science*, 4, 293–312.
- Potter, M. C., So, K. F., Von Eckhart, B., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior*, 23, 23–38.
- Price, C. J. (2000). The anatomy of language: Contributions from functional neuroimaging. *Journal of Anatomy*, 197, 335–359.
- Pulvermuller, F. (1999). Words in the brain’s language. *Behavioral and Brain Sciences*, 22, 253–336.
- Ramus, F., & Mehler, J. (1999). Language identification with suprasegmental cues: A study based on speech resynthesis. *Journal of the Acoustical Society of America*, 105 (1), 512–521.
- Regier, T. (2005). The emergence of words: Attentional learning in form and meaning. *Cognitive Science*, 29, 819–865.
- Ritter, H., & Kohonen, T. (1989). Self-organizing semantic maps. *Biological Cybernetics*, 61, 241–254.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928.
- Siskind, J. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, 61, 39–91.

- Smith, M. A., Cottrell, G. W., & Anderson, K. L. (2001). The early word catches the weights. In T. K. Leen, T. G. Dietterich & V. Tresp (Eds.), *Advances in neural information processing systems* (pp. 52–58). Cambridge, MA: MIT Press.
- Smith, L. B., & Thelen, E. (1994). *A dynamical systems approach to the development of cognition and action*. Bradford Books: MIT Press.
- Snow, C., & Hoefnagel-Hohle, M. (1978). The critical period for language acquisition: Evidence from second language learning. *Child Development*, 49, 1114–1128.
- Spitzer, M. (1999). *The mind within the net: Models of learning, thinking, and acting*. Cambridge, MA: MIT Press.
- Sun, H., Sun, D., Huang, J., Li, D., & Xing, H. (1996). Survey on *Xiandai hanyu yanjiu yuliaoku* (Corpus for Modern Chinese Research). In Z. Luo & Y. Yuan (Eds.), *Studies of Chinese language and characters in the computer era* (pp. 283–294). Beijing, China: Tsinghua University Press.
- Tan, L. H., Spinks, J. A., Feng, C. M., Siok, W. T., Perfetti, C. A., Xiong, J., Fox, P. T., & Gao, J. H. (2003). Neural systems of second language reading are shaped by native language. *Human Brain Mapping*, 18, 158–166.
- Tardif, T. (1996). Nouns are not always learned before verbs: Evidence from Mandarin speaker's early vocabularies. *Developmental Psychology*, 32, 492–504.
- Tardif, T. (2006). The importance of verbs in Chinese. In P. Li, L. H. Tan, E. Bates & O. Tzeng (Eds.), *Handbook of East Asian Psycholinguistics* (Vol. 1: Chinese, pp. 124–135). Cambridge, England: Cambridge University Press.
- Tardif, T., Fletch, P., Liang, W. L., & Zhang, Z. X. (2002). Nouns and verbs in children's early vocabularies: A cross-linguistic study of the MacArthur Communicative Developmental Inventory in English, Mandarin, and Cantonese. Poster presented at the Joint Conference of International Association for the Study of Child Language and Society for Research in Communication Disorders. Madison, WI, July 2002.
- 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: Psycholinguistic approach*. New York: Oxford University Press.
- Tomasello, M. (1999). *The cultural origin of human cognition*. Cambridge, MA: Harvard University Press.
- Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, 74, 209–253.
- Tomasello, M. (2003). *Constructing a language. A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Trainor, L. J. (2005). Are there critical periods for music development? *Developmental Psychobiology*, 46, 262–278.
- Tyler, L. K., Russell, R., Fadili, J., & Moss, H. E. (2001). The neural representation of nouns and verbs: PET studies. *Brain*, 124, 1619–1634.
- Vaid, J., & Hull, R. (2002). Re-envisioning the bilingual brain using functional neuroimaging: Methodological and interpretive issues. In F. Fabbro (Ed.), *Advances in the neurolinguistic study of bilingualism* (pp. 315–355). Udine, Italy: Forum.
- Van Geert, P. (1991). A dynamic systems model of cognitive and language growth. *Psychological Review*, 98, 3–53.
- Vigneau, M., Beaucousin, V., Herve, P. Y., Duffau, H., Crivello, F., Houde, O., Mazoyer, B., & Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *NeuroImage*, 30, 1414–1432.
- Wartenburger, I., Heekeren, H. R., Abutalebi, J., Cappa, S. F., Villringer, R. A., & Perani, D. (2003). Early setting of grammatical processing in the bilingual brain. *Neuron*, 37, 159–170.
- Werker, J. F., Cohen, L. B., Lloyd, V. L., Casacola, M., & Stager, C. L. (1998). Acquisition of word-object associations by 14-month-old infants. *Developmental Psychology*, 34, 1289–1309.
- Werker, J. F., & Tees, R. C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology*, 46, 233–251.
- Whorf, B. (1956). A linguistic consideration of thinking in primitive communities. In J. B. Carroll (Ed.), *Language, thought, and reality* (pp. 65–86). Cambridge, MA: The MIT Press. (Original manuscript from 1936)

- Yip, V., & Matthews, S. (2000). Syntactic transfer in a Cantonese-English bilingual child. *Bilingualism: Language and Cognition*, 3, 193–208.
- Yu, C., Ballard, D. H., & Aslin, R. N. (2005). The role of embodied intention in early lexical acquisition. *Cognitive Science*, 29 (6), 961–1005.
- Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, 18, 414–420.
- Zhang, Y., & Wang, Y. (2007). Neural plasticity in speech acquisition and learning. *Bilingualism: Language and Cognition*, 10, 147–160.
- Zhao, X., & Li, P. (2006). Bilingual lexical representation in a self-organizing neural network. In R. Sun & N. Miyaki (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 26–39). Vancouver, Canada: Cognitive Science Society.
- Zhao, X., & Li, P. (2007). Bilingual lexical representation in a self-organizing neural network. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (pp. 755–760). Austin, TX: Cognitive Science Society.
- Zhao, X., & Li, P. (2008). Vocabulary development in English and Chinese: A comparative study with self-organizing neural networks. In B. C. Love, K. McRae & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 1900–1905). Austin, TX: Cognitive Science Society.
- Zhao, X., & Li, P. (in press). The acquisition of lexical and grammatical aspect in a developmental lexicon model. *Linguistics: An Interdisciplinary Journal of the Language Sciences*.
- Zhao, J., Shu, H., Zhang, L., Wang, X., Gong, Q., & Li, P. (2008). Cortical competition during language discrimination. *NeuroImage*, 43, 624–633.