Bilingualism: Language and Cognition 16 (2), 2013, 288–303 © Cambridge University Press 2012 doi:10.1017/S1366728912000624

Simulating cross-language priming with a dynamic computational model of the lexicon*

XIAOWEI ZHAO
Emmanuel College
PING LI
Pennsylvania State University

(Received: December 30, 2011; final revision received: September 27, 2012; accepted: October 2, 2012; first published online 7 December 2012)

Cross-language priming is a widely used experimental paradigm in psycholinguistics to study how bilinguals' two languages are represented and organized. Researchers have observed a number of interesting patterns from the priming effects of both translation equivalents and semantically related word pairs across languages. In this study, we implement a self-organizing neural network model, DevLex–II, to simulate these two types of priming effects across Chinese and English. Specifically, our model incorporates a computational mechanism for simulating spreading activation based on the distance between bilingual words in the semantic space. The model also considers additional factors that modulate priming effects, such as the initial activation level of the prime words and the degree to which the target word can be recognized. Our model reveals differences with respect to the priming effects as a function of bilingual type (early versus late L2 learners), directions of priming (L1 to L2 versus L2 to L1), and types of priming (translation versus semantic priming). These simulated differences are compared with empirical findings from previous studies and discussed in the light of interactive and developmental theories of bilingual lexical representation.

Keywords: neural networks, cross-language priming, DevLex–II, spreading activation, AoA

Cross-language priming is a widely used experimental paradigm in psycholinguistics to study how bilinguals' two languages are represented and organized. In such a paradigm, cross-language word pairs (e.g., translation equivalents, or semantically related word pairs) are usually presented to participants sequentially in a reaction-time based task (such as lexical decision or word naming). The paradigm is designed to test if bilinguals show response time differences to pairs of prime–target words that differ in their relatedness. A faster reaction time to related pairs as compared to unrelated pairs across languages (e.g., prime from the first language and target from the second language) is usually interpreted as a facilitation effect due to the implicit spreading of activation from the prime word to the target word in the bilingual's mental lexicon, and a strong facilitation is often taken as an indicator of the shared or common conceptual memory representations of the two lexicons (cf. Pavlenko, 2009).

Many cross-language priming experiments have been conducted in the past decades (see a detailed review in Altarriba & Basnight-Brown, 2007). By and large, these experiments have shown effects of both translation priming[1] and semantic priming across languages, and have observed at least the following two interesting patterns: (i) facilitation for translation equivalents is usually larger than that for semantically related words (Basnight-Brown & Altarriba, 2007); and (ii) priming effects in the L1–L2 direction (from first language primes to second language targets) are often larger than those in the L2–L1 direction, and this pattern has been referred to as the PRIMING ASYMMETRY (Dimitropoulou, Duñabeitia & Carreiras, 2011; Jiang, 1999; Jiang & Forster, 2001).

Although it is widely accepted that cross-language priming effects are real, the exact nature of this phenomenon has not been studied systematically against important bilingual factors such as the participant's L2 learning history and language use habits, age of acquisition, and similarity distances between the bilingual's two languages, among other methodological issues discussed in Altarriba and Basnight-Brown (2007). As Grosjean (1998) has argued, in studying bilingual representation and the interaction between L1 and L2, researchers need to consider carefully factors such as the nature of the bilingual participant including bilingual proficiency, learning history, the nature of experimental

* Preparation of this article was supported by grants from the National Science Foundation (#0968369; #1057855) to PL and by a grant from the Faculty Development Committee at Emmanuel College to XZ. We would like to thank three anonymous reviewers for their valuable comments and suggestions on earlier versions of this article.

[1] In some studies translation priming has also been referred to as repetition priming (Forster & Davis, 1984).

Address for correspondence:
Xiaowei Zhao, Department of Psychology, Emmanuel College, 400 The Fenway Boston, MA 02115, USA
zhaox@emmanuel.edu

https://doi.org/10.1017/S1366728912000624 Published online by Cambridge University Press

tasks such as task characteristics (e.g., bilingual speech mode) and modality of testing (comprehension vs. production), and stimulus properties such as word length, frequency, and type (e.g. cognates vs. noncognates, abstract vs. concrete words; Van Hell & De Groot, 1998).

Computational models offer particular advantages in dealing with the complex interactions between variables by systematically bringing target variables under experimental control while holding other variables constant (McClelland, 2009). Although the vigor of experimental research lies in systematic control of variables, in natural language learning situations, especially in the bilingual case, it is often difficult to directly manipulate bilinguals' learning environment in parametric ways such as their L2 learning history. Given the flexibility of computational simulations in orthogonally manipulating variables of interest and relating to experimental hypotheses, in this study we test a computational model of cross-language bilingual priming.

A number of theoretical frameworks of the bilingual mental lexicon have been proposed in the literature, including the Bilingual Dual-Coding theory (Pavio & Desrochers, 1980), the Distributed Feature model (De Groot, 1992), the Revised Hierarchical model (Kroll & Stewart, 1994), and more recently, the Sense model (Finkbeiner, Foster, Nicol & Nakamura, 2004; see also Segalowitz & de Almeida, 2002). Most of these models have been designed to account for bilingual lexical processing at a conceptual level although they are based on specific experimental findings from a variety of paradigms including priming. In recent years, there has also been interest in building models that can be computationally implemented or verified (see Li & Farkas, 2002; Thomas & Van Heuven, 2005, for reviews). The Bilingual Interactive Activation (BIA) model (Dijkstra & Van Heuven, 1998, 2002) is one excellent example in computational modeling of bilingual language processing. However, the BIA model belongs to a class of "permanent" or "stationary" models because mechanisms of learning and adaptation for representation are missing in these models. Learning mechanisms are crucial, for example, in accounting for cross-language priming effects from bilinguals with different levels of L2 proficiency or different histories of learning, and such mechanisms have been incorporated into several models in the past (Jacquet & French, 2002; Li & Farkas, 2002; Zhao & Li, 2010; see Li, 2002, and Thomas & Van Heuven, 2005, for discussion).

In this study, we applied DevLex–II, a dynamic computational model that considers mechanisms of learning, to the study of bilingual lexical representation. DevLex–II is an unsupervised neural network model that learns lexical representations over time, and was originally used to simulate first language acquisition (Li, Zhao & MacWhinney, 2007). Here we apply it to simulate translation and semantic priming across two languages (Chinese and English) under two different learning situations, early versus late L2 learning. In addition, the model incorporated a computational mechanism for simulating spreading activation based on the distance of bilingual words in the semantic space. We examined the priming effects under the two learning situations with detailed statistical analyses of the simulation data. Our simulation data were largely consistent with the results from previous empirical studies of cross-language priming, including our own data in Zhao, Li, Liu, Fang and Shu (2011b). The simulations reported in this paper demonstrate the ability of computational methods to quantitatively capture the empirically observed patterns in cross-language priming and to motivate future empirical research in this domain.

Method

The model

DevLex–II is a multi-layer self-organizing neural network model, which includes three basic levels for the representation and organization of linguistic information as shown by the diagram in Figure 1. The core of the model is a connectionist network called a feature map that handles semantic/conceptual representations. A feature map is a self-organizing network that identifies input similarities in a high-dimensional space and projects these similarities on a two-dimensional space through a typography-preserving algorithm as discussed below (SOM; Kohonen 2001). This semantic feature map is connected to two other feature maps, one for input (auditory) phonology, and another for articulatory sequence of output phonology. Different from the BIA model, DevLex–II is a learning model with adaptable weights of connections among units in the network. Upon training, the linguistic information of a word is presented to the network, and on each map an area of nodes will become activated (the "activity bubbles") and the maximally active node (the Best Matching Unit or BMU) is taken to represent the input. As training progresses, representational patterns of activation become clearer and more focused on each layer. Meanwhile, the strength of certain links between layers becomes increasingly stronger since they connect co-activated units that represent linguistic contents (e.g., meaning and sound) of the same words (see Zhao & Li, 2007, 2010, for details of the DevLex–II model).

In our simulations reported below, the network learned Chinese as the first language (L1) and English as the second language (L2). We used here as our basis the vocabulary from CDI, the MacArthur-Bates Communicative Development Inventories (Dale & Fenson 1996). Each lexicon included 500 words chosen from

https://doi.org/10.1017/S1366728912000624 Published online by Cambridge University Press

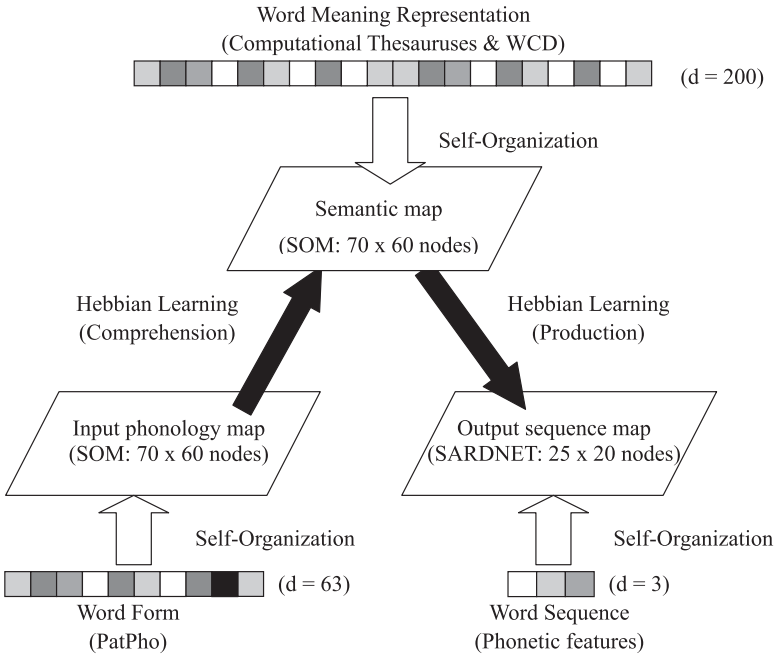


Figure 1. The architecture of the DevLex–II model (Figure from Zhao & Li, 2010; Reproduced with permission from Taylor and Francis). Each of the three self-organizing maps (SOM) takes input from the lexicon and organizes phonology, semantics, and phonemic sequence information of the vocabulary, respectively. The number of nodes in each map is indicated in parentheses. The dimension of the input vector for each map is indicated by "d = " in parentheses next to the input representation symbols. The maps are connected via associative links updated by Hebbian learning. SARDNET is a type of temporal or sequential SOM network (James & Miikkulainen, 1995; see details in Li et al., 2007, for its incorporation in DevLex–II). See text for further explanation of the model.

the Toddler List of the corresponding CDI. The English lexicon was identical to that of Li, Farkas and MacWhinney (2004), and the Chinese lexicon was derived from the Chinese version of the CDI (Tardif, Gelman & Xu 1999; Wu 1997). Use of the CDI has the advantage of deriving representations of frequently used words in the two languages, since CDIs reflect children's earliest vocabularies. We coded the linguistic information of the 500 words from each language as follows.

The sound pattern and phonemic makeup of a word was coded as the basic phonological input to the model according to PatPho, a generic phonological pattern generator for neural networks (Li & MacWhinney 2002; Zhao & Li, 2009; and see Zhao & Li, 2010, for how we made the phonological input from the two languages comparable). The semantic information of words was coded in the following two ways: (i) We used WCD, a special recurrent neural network that learns the lexical co-occurrence constraints of words, to read a stream of input sentences one word at a time, and learn the adjacent transitional probabilities between words which it represents as a matrix of weights. WCD computes two vectors that correspond to the left and the right context, respectively; it then transforms these probabilities into normalized vector representations for word meanings (Li et al., 2004, pp. 1348–1349). (ii) The second set of semantic representations was generated from word associations, synonym and hypernym relations as represented in computational thesauruses available for each of the two languages. For Chinese, it was derived from a Chinese computational database called HowNet (http://www.keenage.com). Through a program that calculates the similarity of Chinese words in the database (Liu & Li, 2002), we derived a matrix that represents the similarity of all the 500 Chinese words. For English, as in Li et al. (2004), we used a feature generation system developed by Harm (2002) to derive semantic features from the WordNet database (Miller, 1990), and the similarity of the 500 English words were further calculated according to these features. Finally, a Random Mapping (Kohonen, 2001) method was used to reduce the size of each set of the semantic representations to a lower dimension (from d = 500 to d = 100), and the two sets were then combined together to form each word's semantic vector. By combining the two methods described above, our model allows for a lexical representation with both syntactic and semantic information, which has the ability to introduce certain LANGUAGE-SPECIFIC INFORMATION into our representation (see Zhao, Li & Kohonen, 2011a, for a review of the advantages and disadvantages of semantic representation models).

https://doi.org/10.1017/S1366728912000624 Published online by Cambridge University Press

L2 learning (training in the model)

To simulate different L2 learning history and L2 proficiency, the 1000 words in the training lexicon were presented to the network according to two scenarios, early L2 learning versus late L2 learning. Specifically, we manipulated the onset time of lexical learning of L2: for early learning, the onset time of L2 was slightly delayed relative to that of L1, and for late learning, the onset time of L2 lagged significantly behind that of L1 (see details below). In the case of early L2 learning, the network was first trained on 100 L1 words (Chinese). Then the L2 words (English) were presented to the network stage by stage (each stage with 50 more new L2 words) along with the corresponding increment of L1 words. The training would end 10 stages later, when the entire set of 500 L2 words was seen by the network. Here, a training stage included 10 epochs, which means that each available word (including its meaning, sound, and articulatory sequence of output phonology) was presented to the network 10 times at each stage. In the case of late learning, L2 words began to join the training session 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 total exposure to L2 words in both the early and late scenario was 10 stages.

To better simulate the interactions between the two lexicons, we introduced a new type of associative connections within each layer in DevLex–II. Here, nodes on a map are fully connected with each other via lateral connections, and their weights are trained via Hebbian learning. Lateral connections within layers have been previously applied in the simulation of the primary visual cortex, particularly to simulate the long-range connections between areas that respond to similar visual features (e.g., neurons with the same line orientation preference, see Sirosh & Miikkulainen, 1994). In this study, we used lateral connections to specifically simulate the process of increased connections that develop between lexical items in the two languages during L2 learning. Through this mechanism, we wanted to simulate the effects of Long-Term Potentiation (LTP), a neural mechanism that instantiates the consolidation of long-term memory (Munakata & Pfaffly, 2004). In particular, we assumed that, after a new L2 (English) word is presented, its L1 translation equivalent (a Chinese word) is also activated in the system. Consequently, the map representations of the two words (including the BMUs and their neighbors) are activated and the lateral connections among them are strengthened via a Hebbian learning rule according to Equation (1):

$$\Delta w_{kl} = \beta.\alpha_k.\alpha_l \qquad (1)$$

Here, w_{kl} is the unidirectional associative weight going from node k to node l, α_k and α_l are the associated node activations corresponding to the input to the map, and β is a constant learning rate (which was set as 0.1 to be consistent with the learning rate between layers). To avoid uncontrolled weight growth, the associative weight vectors are then normalized to ensure that the largest possible lateral connection weight is no more than one.

In the simulations reported in this paper, the input phonology map and the semantic map each consisted of 70×60 nodes, and the output sequence map included 25×20 nodes. During training, the learning rate of SOMs and that of Hebbian learning (β) were kept constant (0.25 and 0.1, respectively). The radii of a winner's neighborhood on each map were changed automatically according to the self-adjustable neighborhood function introduced in Zhao and Li (2007). The initial radius on the SOM layer was set to be 20 and that on the SARDNET was 10. These numbers were chosen to be large enough to discriminate among the words and phonemes in the lexicon while keeping the computation process tractable.

Simulating spreading activation

The goal of our computational modeling is to capture mechanisms for cross-language priming. To make our model psychologically plausible, a spreading activation algorithm was incorporated into the model. Spreading activation has been the backbone theory underlying priming studies (Collins & Loftus, 1975; McNamara, 2005). The basic idea is that the activation of a concept (e.g., as represented in a prime word) implicitly spreads over to other related concepts in the mental representation and the residual activation of the concept of the prime word may facilitate its subsequent retrieval depending on how large the activation is. In cross-language priming, then, the residual activation of a concept in one language can spread over to the other language, causing similar or related concepts encoded in words of the other language to become active. Although priming effects can be obtained in several domains (phonological, orthographic, and semantic), in the current study we focused on simulating semantic priming using the spreading activation mechanism.[2] Specifically, the BMU of a target word could receive spreading activations from the BMU of a prime word via two paths, one

[2] Given the significant differences between Chinese phonology/orthography and English phonology/orthography, cross-language priming effects in these domains are less meaningful. Our model can be extended to the simulation of other language pairs where phonological and orthographic similarities are stronger. It has also been used previously in the study of orthographic processing, such as the simulation of children's learning of Chinese characters (see Xing, Shu & Li, 2004, 2007).

https://doi.org/10.1017/S1366728912000624 Published online by Cambridge University Press

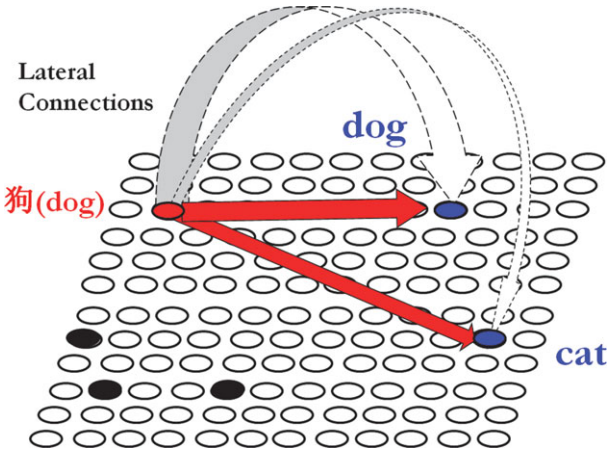


Figure 2. (Color online) An illustration of the two paths of activation spreading from the prime word to the target word. A shaded dot on the map represents the BMU of a word. The dashed arrows indicate the spreading activation via the lateral connections and the solid arrows the spreading activation within the semantic map. Both translation priming [狗 – *dog*] and semantic priming [狗 – *cat*] are depicted here (NB: Chinese 狗 = English "dog"). The lateral connection between semantically related cross-language word pairs is weaker (narrower) than that between translation equivalents, and such pattern was gradually developed as a function of learning/training in the model.

through their lateral connection (see earlier discussion) and one within the semantic map. An illustration of the two paths for both translational priming and cross-language semantic priming can be found in Figure 2.

The spreading activation from each path was defined as the product of the initial activation of the prime word and a Gaussian-like function:

$$Spread = Activation(prime) \times e^{-\frac{(a^* Dist)^2}{2b^2}} \quad (2)$$

Here, $a = 0.2$ and $b = 2$, which define the shape of the activation function.[3] *Dist* was a measure of the closeness between a prime's and a target's semantic representations. For the path via lateral associative connection, it was defined as the reciprocal of the weight; for the path within the semantic map, it was simply the Euclidean distance between the BMUs of the word pairs on the map. Similar, though not identical, Gaussian-like functions have been also used in previous computational studies to simulate spreading activations in semantic networks (Silberman, Bentin & Miikkulainen, 2007; Spitzer, 1999). Figure 3 presents the shape of this Gaussian function.

[3] Values of these free parameters in Equations (2) and (3) were set to control the range (a and b) and magnitude (c) of the activation functions so that (i) the range would not be too wide or too narrow compared with the size of the network, and (ii) the magnitude would not be too large.

Capturing time in lexical decision

Time is a critical variable that is controlled and measured in priming experiments. First and foremost, participants' reaction times to the target words need to be recorded so that the priming effect can be quantitatively measured. Second, a prime word needs to be displayed to the participants for a certain amount of time before the target word is shown (the Stimulus Onset Asynchrony, SOA) so that enough activation can be generated based on participants' semantic representation; at the same time the SOA should be brief enough to prevent participants from developing top–down strategies (e.g., expectancies from primes) during the experiment. SOAs were usually designed to be shorter than 200 milliseconds, and in some recent studies using the masked-priming paradigm, it can be as short as 50 milliseconds (see discussion in McNamara, 2005, p. 72; and Altarriba & Basnight-Brown, 2007). Our simulations are designed to both capture the reaction times and represent the SOAs from real lexical decision tasks.

It is crucial for our purposes to simulate the change of a target word's activation level in order to compare simulated priming effects to reaction times from real experiments. To achieve this, we defined a recognition threshold $T = 2$ for each node. We assumed that in a task like lexical decision, a target word can only be recognized when its representing node's activation level reaches the threshold. In addition, we defined how much the activation level of the target word's representing node increases as a function of the unit of elapsed time during the recognition phase of the target word:

$$\delta = c \times e^{-\frac{(a^* Density)^2}{2b^2}} \quad (3)$$

Here the free parameters are $c = 0.003$, $b = 2$ and $a = 1$. The increment of δ was also a Gaussian-like function, which changes with *density* – a variable representing how many neighboring words that the target word has on the semantic map. Specifically, we defined *density* as the number of words in its neighborhood (with radius of 1) divided by the total number of units of its neighborhood, which is usually nine, but could be six or four, depending on whether the tested word was on the border or at the corner of the map. The value of this density measure ranged from 1/9 when only the word itself is in the neighborhood, to 1.0 when all neighboring units of a word are occupied by other words. The larger the density is, the more interference there is among words, and the more difficult it is for the target word to be recognized due to the smaller increment of δ. Therefore, considering the residual/persisting activation spreading from the prime word, the total time units needed for the recognition of the target word (i.e. Reaction Time) is

$$RT = (2 - Spread)/\delta \quad (4)$$

https://doi.org/10.1017/S1366728912000624 Published online by Cambridge University Press

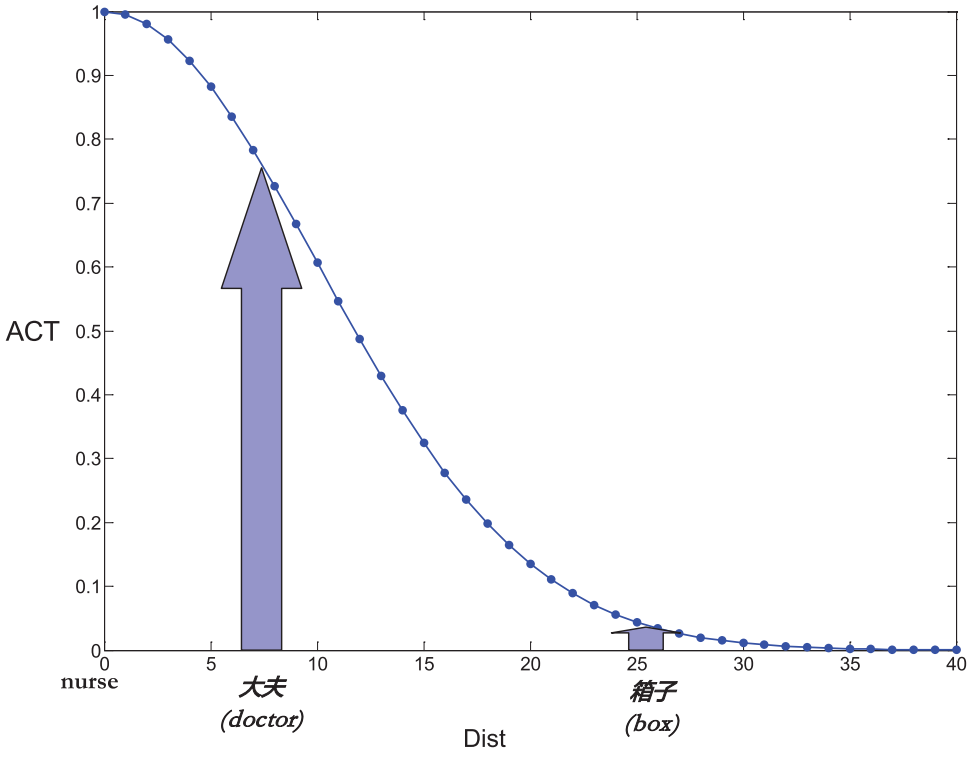


Figure 3. (Color online) The shape of spreading activation defined by Equation (2). The basic mechanism supporting priming effects is also depicted here: after the node *nurse* on the semantic map is activated, the node corresponding to 大夫 "doctor", which is closer to *nurse* than 箱子 "box" in the semantic representation, receives more activation and thus is more readily accessible from memory. The X axis indicates the Euclidean distance from different nodes to the node *nurse* on the semantic map. The Y axis indicates activation level on a scale from zero to one.

Obviously, the smaller the residual activation a target word receives from the prime word and the denser the target word's neighbourhood is, the longer the reaction time will be in the model.

Similarly, we assumed that the activation level of the prime word's representing node tends to accumulate during the SOA period. As a function of units of elapsed time, only a limited amount of activation is added, and the increment also follows the Gaussian-like function of density as shown in Equation (3) with a slight adjustment of the free parameter a to be 4. The basic rationale of such a mechanism is again to reflect the interference from neighboring words. The prime words from the less fluent language may be less activated than those from the dominant language in a brief display. In the current study, we ran our simulations under two SOA conditions: one with 150 elapsed time units and another with 50 time units.[4] The purpose of our design was to examine the potential impact of SOA differences on priming effects, and we used the two SOA conditions based on the SOAs used in previous lexical decision tasks (e.g. 150 ms in Zhao et al., 2011b, and 50 ms in other studies based on masked priming paradigm; see review in Altarriba & Basnight-Brown, 2007). In our simulation one time unit roughly represents one millisecond in a real experiment.

Test material and testing procedure

As described in section "L2 learning (training in the model)" above, we trained our model to simulate two scenarios: early L2 learning and late L2 learning. For each scenario, five networks were created and served as the basis for our further tests described below. Conceptually, each trained network could be likened to a proficient bilingual learner in realistic situations. The learner has acquired the two languages through a slow learning process over an extended period of time (corresponding to the training of our model; see the above section on L2 learning); the learner is then brought to the laboratory so that researchers can probe his or her mental representation in a fast testing procedure to identify cross-language priming effects (corresponding to the testing phase in our model).

From the 1000 words (500 words each in Chinese and English) in the training lexicon of our network, we chose for our test material a list of 32 translation equivalents (e.g., *sock* and 袜子 "sock") and a list of 32 semantically

[4] We also ran additional SOA conditions, as reported in Figure 5. These additional SOA conditions were entered into the ANOVA analyses but due to space limitation, are not reported here.

https://doi.org/10.1017/S1366728912000624 Published online by Cambridge University Press

Table 1. *Mean reaction times from the late L2 learning experiment with SOA of 150 units.*

Target language	TR	TU	Priming effects	SR	SU	Priming effects
English (L1–L2)	544.84 (30.00)	712.41 (22.77)	+167.56	680.08 (27.20)	720.14 (39.01)	+40.06
Chinese (L2–L1)	595.36 (18.60)	671.89 (7.00)	+76.54	647.37 (9.74)	667.45 (6.24)	+20.07
Priming asymmetry			+91.02			+19.99

TR = Translation equivalents, TU = Translation unrelated, SR = Semantically related, SU = Semantically unrelated. Same for Tables 2–4.
Note: Average results based on 20 simulation runs. Numbers in parentheses represent standard deviations. Same for Tables 2–4.

related word pairs (e.g., 大 夫 "doctor" and *nurse*). A complete list of these word pairs is given in the Appendix. In addition, we created two more lists of unrelated word pairs that are matched with related words by shuffling the words in the two related lists, so that unrelated words are put into a pair. From these words, four versions of the experimental material were further constructed, each including 16 translation equivalents (TR), 16 unrelated translation pairs (TU), 16 semantically related cross-language pairs (SR), and 16 unrelated semantic pairs (SU). A Latin Square method was applied to ensure that no two versions of the experimental material shared the same word pairs and no words were repeated twice in a single version of the material. In each category, half of the pairs had English words (L2) as the target words, and the other half had Chinese words (L1) as the target words. This setup ensured that we could study the priming effects for both the L1-to-L2 direction and the L2-to-L1 direction.

For each SOA condition (150 or 50 elapsed units, respectively), each trained network was tested in four simulations, resulting in 20 simulations for each learning scenario. During testing, a version of the experimental material (with 64 prime–target word pairs) was presented to the trained network. Here, the presentation of a prime–target word pair is roughly comparable to a trial in a real lexical decision experiment. For each pair, the prime word was first presented to the trained network for a certain amount of time (SOA), and the target word was then presented to the network for recognition. Based on our spreading activation mechanism discussed above, the residual activation that the target word received was calculated, and the reaction time for the target word to be recognized was calculated based on Equation (4) above. Considering all the 64 target words in a single simulation, the priming effects were calculated by subtracting the mean RT of related word pairs from that of the matched unrelated word pairs.

Results

In this section we report simulation results based on the modified DevLex–II model as discussed in the Methods section. Under each SOA condition, our focus is on comparing the priming effects based on the two L2 learning scenarios and on comparing our modeling results with empirical findings from previous studies.

SOA of 150 elapsed time units

Late L2 learning

The mean reaction times and the priming data for the late L2 learning scenario averaged across 20 individual simulations are presented in Table 1. A participant-based 2 (Direction: L1–L2 vs. L2–L1) by 2 (Type: semantic pairs vs. translation equivalents) by 2 (Relatedness: related vs. unrelated) factorial ANOVA was conducted. Significant main effects were found for all three factors. The main effect of Direction [$F(1,19) = 41.99$, $MSE = 338$, $p < .001$, partial $\eta^2 = .69$] suggested that over all our network responded significantly faster to Chinese targets (L1: 645.52 time units) than to English targets (L2: 664.37 time units). The main effect of Type [$F(1,19) = 145.99$, $MSE = 621$, $p < .001$, partial $\eta^2 = .89$] showed that our network recognized targets in the translation group (631.13) faster than those in the semantic group (678.76). Finally, the main effect of Relatedness [$F(1,19) = 494.23$, $MSE = 468$, $p < .001$, partial $\eta^2 = .96$] indicated that our network was significantly faster in responding to the related word pairs than to unrelated pairs (616.91 vs. 692.97; i.e., a priming effect).

Significant interactions were also observed in our data. The interaction between Type and Relatedness [$F(1,19) = 96.76$, $p < .001$, $MSE = 874$, partial $\eta^2 = .84$] clearly showed that the magnitudes of priming effects were not equal for the translation equivalents and the semantic related word pairs, reflected in that the translation priming effects (+122.05, $p < .001$) were larger than the semantic priming effects (30.07, $p = .001$). The interaction between Direction and Relatedness [$F(1,19) = 55.62$, $p < .001$, $MSE = 553$, partial $\eta^2 = .75$] showed that the magnitudes of priming effects were not equal for the L1–L2 and L2–L1 directions, in that priming from L1 to L2 was larger (103.82) than priming from L2 to L1 (48.31), though both priming effects were significant ($p < .001$). Finally,

https://doi.org/10.1017/S1366728912000624 Published online by Cambridge University Press

Table 2. *Mean reaction times for the early L2 learning experiment with SOA of 150 units.*

Target language	TR	TU	Priming effects	SR	SU	Priming effects
English (L1–L2)	542.87 (16.18)	689.61 (10.58)	+146.77	677.92 (11.81)	686.87 (16.55)	+8.96
Chinese (L2–L1)	569.53 (18.23)	668.54 (4.47)	+99.01	664.23 (8.71)	667.79 (10.36)	+3.36
Priming asymmetry			+47.76			+5.6

the significant interaction between Direction and Type [$F(1,19) = 51.45$, $p < .001$, $MSE = 442$, partial $\eta^2 = .73$] showed that, when the target words were from L2, the difference between semantic priming and translation priming (71.48, $p < .001$) was larger as compared with that when the target words were from L1 (23.78, $p < .001$).

The three-way interaction among Direction, Type and Relatedness was also significant [$F(1,19) = 29.38$, $MSE = 429$, $p < .001$, partial $\eta^2 = .61$]. To better understand this interaction, we conducted a series of pair-wise comparisons to study individual priming effects, using Bonferroni adjustments to control for the overall Type I error. As can be seen on the columns of "Priming effects" in Table 1, the results showed significant translation-priming effects and semantic priming effects for both Chinese (L1) and English (L2) targets. There were significant translation-priming effects of +167.56 for English (L2) targets [$t(19) = 40.48$, $p < .001$] and +76.54 for Chinese (L1) targets [$t(19) = 16.61$, $p < .001$]. With regard to semantic priming, there were significant effects of +40.06 for English targets [$t(19) = 2.93$, $p = .009$] and +20.07 for Chinese targets [$t(19) = 7.30$, $p < .001$]. Critically, for both types, the priming obtained in L1 to L2 direction was larger than the one obtained in L2 to L1 direction, revealing a clear "priming asymmetry". In addition, the asymmetry for translation priming was much larger than that for semantic priming (91.02 vs. 19.99; see last row of Table 1).

Early L2 learning

While our results from the late L2 learning scenario confirmed a number of classic priming effects in the literature, including the priming asymmetry patterns, one could hypothesize that a different pattern of priming effects may be obtained with the early L2 learning scenario, since early L2 learning generally leads to high L2 proficiency. Our simulation data from the early L2 learning scenario are presented in Table 2, showing the mean reaction times and priming data under SOA of 150 averaged across 20 individual simulations. As with the late L2 learning situation, we performed a participant-based $2 \times 2 \times 2$ factorial ANOVA on the data. Significant main effects were also found for Direction [$F(1,19) = 10.20$, $MSE = 183$, $p = .005$, partial $\eta^2 = .35$], for Type [$F(1,19) = 808.89$, $MSE = 158$, $p < .001$, partial $\eta^2 = .98$], and for Relatedness [$F(1,19) = 739.27$, $MSE = 225$, $p < .001$, partial $\eta^2 = .98$].

Again, significant interactions were observed under the early L2 learning scenario. The interaction between Type and Relatedness [$F(1,19) = 478.05$, $p < .001$, $MSE = 285$ and partial $\eta^2 = .96$] showed larger translation priming effect (122.89, $p < .001$) than semantic priming effects (6.16, $p = .097$). The interaction between Direction and Relatedness [$F(1,19) = 67.19$, $p < .001$, $MSE = 106$, partial $\eta^2 = .78$] revealed larger priming from L1 to L2 (77.87, $p < .001$) than from L2 to L1 (51.18, $p < .001$), again a "priming asymmetry". Finally, the significant interaction between Direction and Type [$F(1,19) = 26.92$, $p < .001$, $MSE = 138$, partial $\eta^2 = .59$] showed that, when the target words were from L2, the difference between semantic priming and translation priming (66.17, $p < .001$) was larger as compared with that when the target words were from L1 (46.87, $p < .001$).

For the early L2 learning scenario, there was also a significant three-way interaction among Direction, Type and Relatedness [$F(1,19) = 27.83$, $MSE = 160$, $p < .001$, partial $\eta^2 = .60$]. As with the late L2 learning scenario, we performed a series of subsequent pair-wise comparisons to study individual priming effects, using Bonferroni adjustments. As can be seen on the columns of "Priming effects" in Table 2, significant translation-priming effects of +146.77 were obtained for English targets [$t(19) = 29.20$, $p < .001$] and +99.01 for Chinese targets [$t(19) = 22.71$, $p < .001$]. Although this is also a "priming asymmetry", the magnitude of this asymmetry (47.76) was much smaller than that in the late L2 learning situation (91.02). With regard to the semantic priming effects, a priming of +8.96 was found for English targets [$t(19) = 1.98$, $p = .063$] and +3.36 for Chinese targets [$t(19) = 0.94$, $p = .361$]. Thus, although there was a "priming asymmetry" trend, unlike with the late learning scenario, neither the L1 nor L2 targets produced statistically significant semantic priming.

SOA of 50 elapsed time units

As discussed above, in the current empirical literature of priming research, many researchers have used a very

https://doi.org/10.1017/S1366728912000624 Published online by Cambridge University Press

Table 3. *Mean reaction times from the late L2 learning experiment with a shorter SOA of 50 units.*

Target language	TR	TU	Priming effects	SR	SU	Priming effects
English (L1–L2)	663.39 (22.16)	719.25 (20.56)	+55.85	718.14 (31.73)	731.49 (36.44)	+13.35
Chinese (L2–L1)	649.58 (8.74)	675.09 (7.02)	+25.51	665.37 (5.02)	672.06 (4.48)	+6.69
Priming asymmetry			+30.34			+6.66

Table 4. *Mean reaction times for the early L2 learning experiment with a shorter SOA of 50 units.*

Target language	TR	TU	Priming effects	SR	SU	Priming effects
English (L1–L2)	643.11 (10.93)	692.04 (9.83)	+48.93	688.38 (10.59)	691.37 (12.54)	+2.99
Chinese (L2–L1)	637.34 (7.48)	670.34 (3.58)	+33.00	670.01 (7.02)	671.13 (7.39)	+1.12
Priming asymmetry			+15.93			+1.87

short SOA (e.g., as short as 50 milliseconds), along with the masked priming paradigm, to maximally reduce the influence of potential top–down processing strategies such as effects due to expectancy of upcoming targets. It is worth noting again in this context that the advantage of computational modeling is the tight control/elimination of potential confounds that are commonly found in empirical studies. Our model does not include a component for top–down processing and therefore even when the SOA is 150 units the model will not have performance due to expectancy effects. Nevertheless, we wanted to investigate if varying SOAs can indeed influence cross-language priming, given that our model does simulate the time needed in making lexical decisions. As with simulations presented before, for each learning scenario, the reported results are based on 20 simulations.

Late L2 learning

The mean reaction times and the priming data for this scenario are presented in Table 3. A participant-based 2 × 2 × 2 factorial ANOVA was conducted, as with the SOA of 150 units condition. Significant main effects were found for Direction [$F(1,19) = 348.35$, $MSE = 207$, $p < .001$, partial $\eta^2 = .95$], for Type [$F(1,19) = 36.85$, $MSE = 431$, $p < .001$, partial $\eta^2 = .66$], and for Relatedness [$F(1,19) = 64.87$, $MSE = 396$, $p < .001$, partial $\eta^2 = .77$].

Again, significant interactions were observed under the late L2 learning scenario when SOA was 50 units. The interaction between Type and Relatedness [$F(1,19) = 12.93$, $p = .002$, $MSE = 727.43$, partial $\eta^2 = .41$] showed larger overall translation priming effects (40.68, $p < .001$) than semantic priming effects (10.02, $p = .183$). The interaction between Direction and Relatedness [$F(1,19) = 6.27$, $p = .022$, $MSE = 546$, partial $\eta^2 = .25$] revealed larger overall priming from L1 to L2 (34.6, $p < .001$) than from L2 to L1 (16.10, $p < .001$), again a "priming asymmetry" though in a smaller amount compared with that the previous situation with a SOA of 150 units. Finally, the interaction between Direction and Type was also significant [$F(1,19) = 18.99$, $p < .001$, $MSE = 387$, partial $\eta^2 = .5$] and showed that, when the target words were from L2, the difference between semantic priming and translation priming (33.49, $p < .001$) was larger as compared with that when the target words were from L1 (6.25, $p < .001$). Unlike the SOA with 150 units condition, the three-way interaction among Direction, Type and Relatedness was not significant [$F(1,19) = 2.94$, $MSE = 477$, $p = .103$, partial $\eta^2 = .13$]. The magnitude of the individual priming effects can be found in Table 3. For both translation priming and semantic priming, the priming obtained in the L1-to-L2 direction was larger than the one obtained in the L2-to-L1 direction, revealing a "priming asymmetry" again. In addition, the asymmetry for translation priming was larger than that for semantic priming (30.34 vs. 6.66; see last row of Table 3).

Early L2 learning

With 50 units as the SOA, the mean reaction times and priming data for the early L2 learning scenario are presented in Table 4 and we similarly performed a participant-based 2 × 2 × 2 factorial ANOVA on the data. Significant main effects were found for Direction [$F(1,19) = 96.04$, $MSE = 113$, $p < .001$, partial $\eta^2 = .84$], for Type [$F(1,19) = 242.36$, $MSE = 62.88$, $p < .001$,

https://doi.org/10.1017/S1366728912000624 Published online by Cambridge University Press

partial $\eta^2 = .93$], and for Relatedness [$F(1,19) = 235.93$, $MSE = 78$, $p < .001$, partial $\eta^2 = .93$].

Significant interactions were also observed under the early L2 learning scenario. The interaction between Type and Relatedness [$F(1,19) = 84.96$, $p < .001$, $MSE = 178$, partial $\eta^2 = .82$] showed larger translation priming effect (40.96, $p < .001$) than semantic priming effects (2.05, $p = .466$). The interaction between Direction and Relatedness [$F(1,19) = 19.92$, $p < .001$, $MSE = 40$, partial $\eta^2 = .51$] revealed larger priming from L1 to L2 (25.96, $p < .001$) than from L2 to L1 (17.06, $p < .001$), again a "priming asymmetry". Finally, the significant interaction between Direction and Type [$F(1,19) = 5.36$, $p = .032$, $MSE = 58$, partial $\eta^2 = .22$] showed that, when the target words were from the L2, the difference between semantic priming and translation priming (22.31, $p < .001$) was larger as compared with that when the target words were from L1 (16.73, $p < .001$).

There was a significant three-way interaction among Direction, Type and Relatedness, [$F(1,19) = 8.38$, $MSE = 58.91$, $p = .009$, partial $\eta^2 = .31$]. We again conducted a series of pair-wise comparisons to study individual priming effects, using Bonferroni adjustments. As can be seen in the columns of "Priming effects" in Table 4, significant translation-priming effects of +48.93 were obtained for English targets [$t(19) = 13.63$, $p < .001$] and +33.00 for Chinese targets [$t(19) = 17.16$, $p < .001$]. This is a clear "priming asymmetry", with a magnitude of 15.93 in difference. With regard to the semantic priming effects, no significant effect was found for either English targets (+2.99) or Chinese targets (+1.11) at the $\alpha = .05$ level. Given such small magnitudes of semantic priming, the "priming asymmetry" is not salient for the early L2 learning scenario under the SOA of 50 units condition.

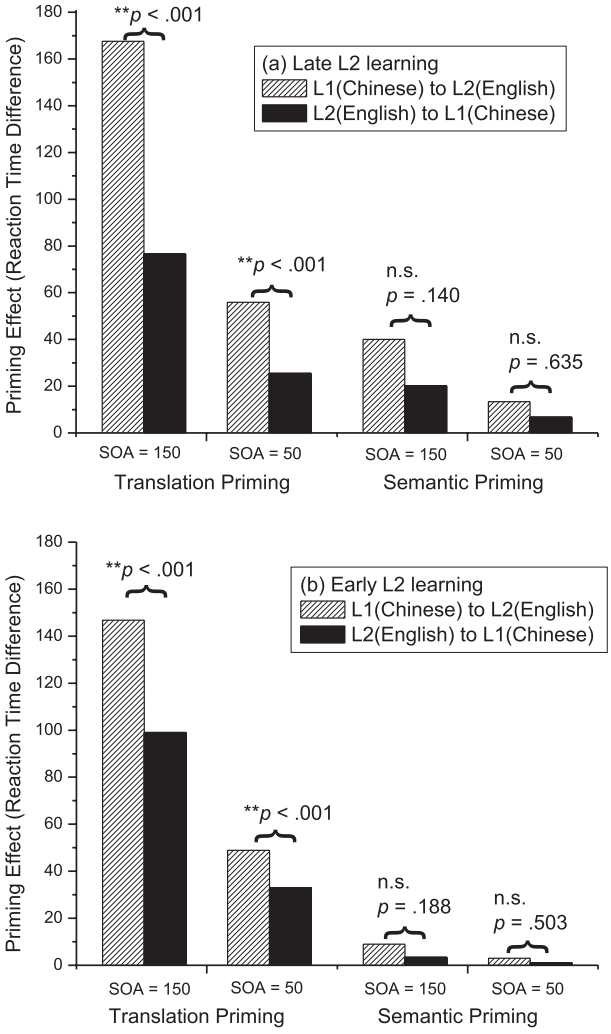


Figure 4. Priming effects from our simulations. (a) Late L2 learning. (b) Early L2 learning. Regardless of SOA (150 or 50 elapsed time units), translation priming is always stronger than semantic priming. Priming effects are calculated by subtracting the RTs of related word pairs from the RTs of unrelated word pairs. The priming effects from L1 (Chinese) to L2 (English) are always larger than those from L2 to L1. This priming asymmetry is also larger in the late L2 learning situation than in the early L2 learning situation. As SOA decreases, the priming effects become smaller and the priming asymmetry also reduces (see also Figure 5). The *p*-values indicate the significance level of the priming asymmetry under the different conditions (paired-samples *t*-test of the 20 simulations under each condition: ** = significant priming asymmetry; n.s. = not significant).

Discussion

In this study we have implemented a self-organizing connectionist network to study bilingual priming. Our study is a first systematic attempt to use computational models to specifically simulate the developmental patterns in cross-language semantic priming and translation priming. Models of this type allow us to parametrically control a number of variables that are thought to affect bilingual lexical representation, including age of L2 acquisition and directions of interaction among L1 and L2. Our simulation results across the varying conditions are summarized and presented in Figure 4. From this figure, along with the statistical analyses reported above, we see that our model has successfully captured several important patterns found in previous empirical studies:

1. Bilinguals on average respond faster to target words in their first language than in their second language, as reflected in the significant main effect of Direction in our simulations.
2. The main effect of Relatedness shows clearly significant cross-language priming effects.

https://doi.org/10.1017/S1366728912000624 Published online by Cambridge University Press

3. The main effect of Type and the Type × Relatedness interaction indicate that translation priming effects are always stronger than semantic priming effects (as observed in the empirical results of Basnight-Brown & Altarriba, 2007, and Zhao et al., 2011b).
4. There was a clear "priming asymmetry" in our model as revealed by the Direction × Relatedness interaction, the three-way interaction and the post-hoc pair-wise comparisons. The average priming effect from Chinese primes (L1) to English targets (L2) was always larger than that of the opposite direction, as presented in Figure 4. This asymmetry pattern is consistent with existing evidence in the literature on cross-language priming (see Table 1 in Dimitropoulou et al., 2011).

Effects of SOA

An interesting pattern from our data is that cross-language priming effects decrease as the SOA of the testing procedure reduces from 150 to 50 time units. A tentative explanation of such a pattern is that a brief exposure of the prime word (shorter SOA) may only trigger a small amount of initial activation, which in turn leads to smaller amount of spreading activation to the target word, thus a less salient or non-existing priming effect. To further explore how the priming effects may change as a function of SOA, we conducted additional simulations based on two other SOA situations (SOA = 10 and 100 elapsed time units, respectively).[5] Results from these simulations are reported in Figure 5, in combination with the results from the two SOA conditions as discussed in the Results section. The line graphs show a clear decreasing tendency of priming effects as SOA decreases from 150 to 10. To extrapolate from this figure, we can consider an extreme case in which the SOA is zero where there is no exposure to the prime word, and in this case we will find no priming effects in either our simulations or in real experiments.

In addition, the extent of the priming asymmetry seems also to decrease as SOA decreases. For example, comparing the significance levels of asymmetry of semantic priming under two SOA conditions shown on Figure 4, we found that although both are non-significant, the *t*-values for the shorter SOA are smaller (less significant) than those for the longer SOA. The cause of such decrement of priming asymmetry might be a floor effect related to the relatively smaller priming effects at both priming directions with a shorter SOA. We cannot yet fully evaluate these findings against empirical data as few previous empirical cross-language priming studies have been done conducted to systematically investigate the effect of varying SOAs. One exception was Schoonbaert, Duyck, Brysbaert and Hartsuiker (2009), who conducted a series of masked-priming experiments with Dutch–English bilinguals, and their data are largely consistent with our simulation results showing that cross-language priming effects are smaller when SOA is shorter (100 ms vs. 250 ms in the longer SOA).

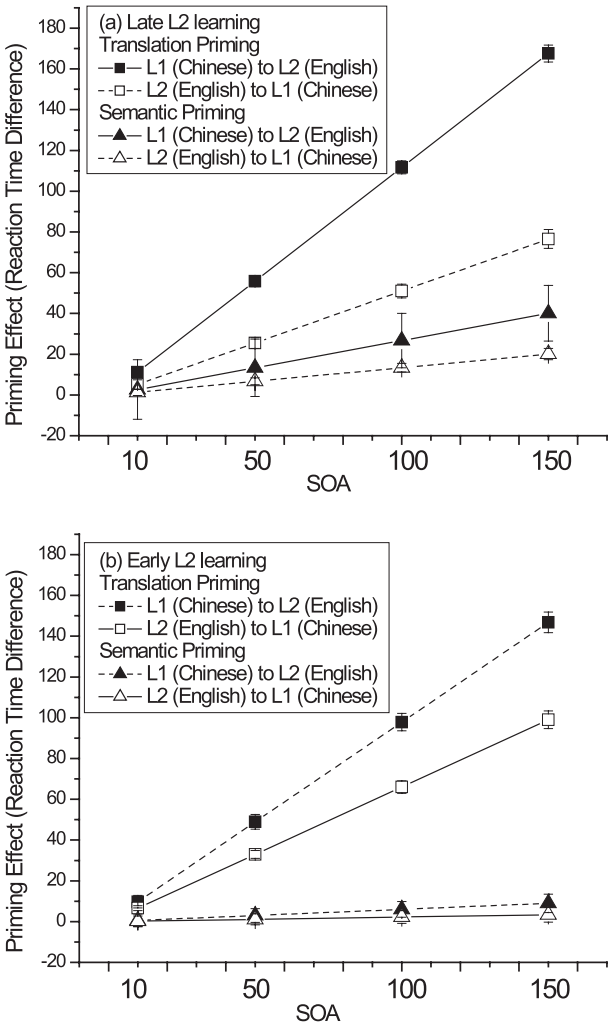


Figure 5. Priming effects as a function of SOA (from 10 to 150 elapsed time units). (a) Late L2 learning. (b) Early L2 learning. Error bars indicate the standard errors based on 20 simulations under each condition. The figure shows that, as SOA decreases, the priming effects and the priming asymmetry both reduce in our simulations.

Role of L2 age of acquisition

Other than capturing main empirical data with regard to SOA, our model provides insights into how priming effects

[5] We also conducted ANOVA analyses on the simulation results of these two SOA situations, but due to the length limitation of the paper, we do not report them here.

https://doi.org/10.1017/S1366728912000624 Published online by Cambridge University Press

may differ as a function of different learning history of the bilingual individual and therefore inspire future empirical work. For example, the most salient priming asymmetry in our simulations was found in translation priming for the late L2 group, but the size of this priming asymmetry in the early L2 group was much smaller (for translation priming) or non-existent (for semantic priming). Such early versus late learning differences suggest that the direction of cross-language priming may become less salient for bilinguals who acquire their L2 early in life. In other words, early learners may have reached a level of proficiency for both languages, such that priming from one language to the other is more or less equal, regardless of direction, due to equal amount of spreading activation across languages.

Given the significant debate on the role of age of acquisition (AoA) versus that of proficiency in L2 language representation and processing (see Hernandez & Li, 2007, for review), our study cannot yet distinguish whether the priming asymmetry differences are truly due to AoA or due to proficiency as these two variables are often correlated and as we have not independently manipulated proficiency in this study. There is already preliminary evidence that proficiency in L1 versus L2 might contribute to priming asymmetry. For example, in a study based on Spanish–English bilinguals, Kiran and Lebel (2007) found that the bilinguals less fluent in L2 had overall larger priming effects and larger priming asymmetry than those who were more balanced in L1 and L2, although the differences were not statistically significant partly due to the relatively small sample size of the unbalanced bilinguals in their study ($n = 4$). In another work, Dimitropoulou et al. (2011) showed clear asymmetry patterns of translation priming in Greek–English bilinguals, but the magnitudes of the priming and of the asymmetry did not decrease as the participants' L2 proficiency level increased. The authors indicated that this puzzling result might be due to the fact that their participants were all late L2 learners, and suggested that the asymmetry could change when the AoA of the bilingual speakers had been taken into consideration.

An important feature of the DevLex–II model in accounting for bilingual lexical memory is its dynamic semantic representation of the two lexicons. We believe that this feature plays a significant role in allowing for cross-language priming effects to occur in our model. In Zhao and Li (2007, 2010) we showed that the lexical representation structure differs significantly as a function of early versus late L2 learning. Analyses of the network's semantic representation revealed that words from the two learning situations are not evenly distributed in the semantic map. Figure 6 shows an example based on Zhao and Li's (2010) simulation data, indicating that for early L2 learning (Figure 6a), clearly separated representations emerge in the semantic map, while for late L2 learning, a disjointed and more compressed L2 representation pattern (Figure 6b) has occurred. When the L2 words are densely represented, the competition between them is stronger, resulting in higher confusion rates and retrieval errors. In the late learning situation, the L2 words, compared to L1 words, occupied only small and fragmented regions, and were interspersed with L1 regions. This is a situation to which we refer as parasitic L2 lexicon (see also Hernandez, Li & MacWhinney, 2005), because the representation of L2 words is dependent on the established L1 lexical structure. These modeling results point to the effects that the structural consolidation of the L1 lexicon can have on the representation and processing of L2 words, depending on whether L2 learning occurs early (leading to representation less dependent on L1) or late (leading to a parasitic L2 representation).

Role of semantic representation in priming

Our model further highlights the development of the richness of semantic representation in the L1 versus L2 contexts, which in turn contributes to the development to priming effects. Our interpretation of cross-language priming patterns is conceptually consistent with previous theoretical frameworks such as the Distributed Feature model (De Groot, 1992) and the Sense model (Finkbeiner et al., 2004), and we provide a computational implementation of the effects. The previous theoretical frameworks propose that the existence of cross-language priming effects depends on the amount of shared semantic features (or senses) between the prime word and target word. The translation priming effect is often larger than semantic priming effect because the translation equivalents share more features than cross-languages semantically related pairs, which in turn have more common features than unrelated word pairs.

With regard to our model, on the semantic map the closeness of the BMUs of different words reflect their overlap in terms of semantic features since our SOM-based model has the ability to capture, on a two-dimensional map, semantic similarity in a distributed high dimensional space (e.g., each dimension representing a semantic feature) (Figure 5). Specifically, close overlap of two lexicons in semantic representation allows spreading activation to occur more easily from words in one language to their semantically related words in the other language, in turn leading to overall large cross-language priming effects. Since there is more overlap in meaning between translation equivalents than between semantically related words across two languages, translation equivalents are often located close to each other in the map's representation, and therefore translation priming is often

https://doi.org/10.1017/S1366728912000624 Published online by Cambridge University Press

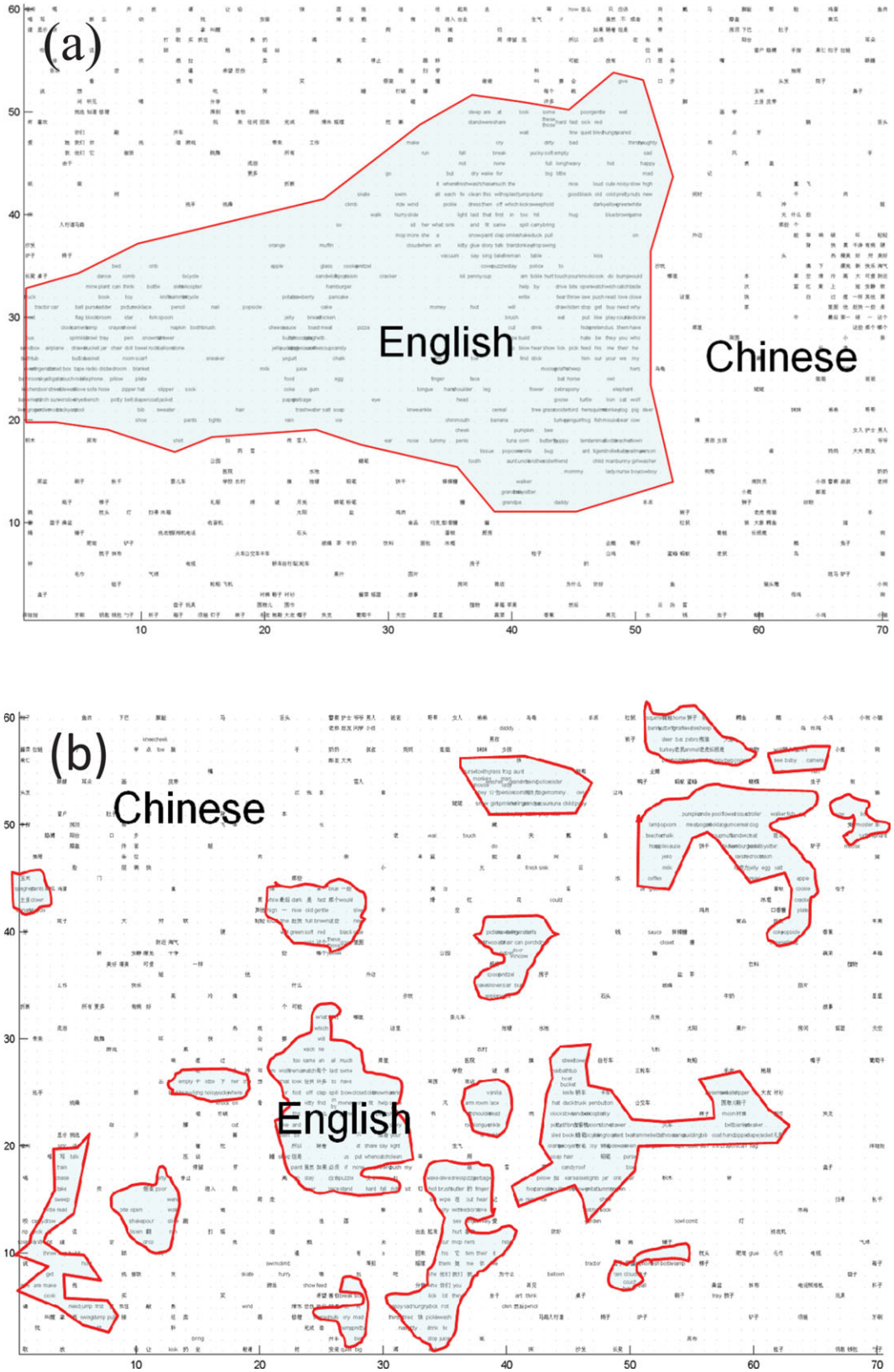


Figure 6. (Color online) An example of bilingual lexical representation on the semantic map, as a function of (a) early versus (b) late L2 learning. Shaded areas correspond to L2 words (English). Similar results have been observed with Chinese as the L2. See Zhao and Li (2007, 2010) for further discussion.

https://doi.org/10.1017/S1366728912000624 Published online by Cambridge University Press

larger than semantic priming. Moreover, the lateral connections that gradually develop among nodes within the map also contribute to the cross-language priming patterns seen in our simulations. As learning progresses, the connections between translation equivalents become increasingly stronger. When a prime word is activated on the map, the activation can quickly spread to the node corresponding to its translation equivalent via the strong lateral connection, causing the node to be more readily accessible in the semantic space. The effect of these "short paths" among translation equivalents could also contribute to the larger magnitude of translation priming than semantic priming shown in our model, which could be useful in accounting for patterns in empirical data.

In terms of priming asymmetry patterns, previous theoretical frameworks often assume a less rich semantic representation of L2 than L1 (i.e., fewer semantic features/senses of L2 words that are understood by bilinguals). Therefore the same amount of activated features/senses will cause a larger priming of L2 targets than L1 targets since relatively larger proportion of the senses of a target is activated when it belongs to L2. For example, consider an L1 word and its L2 translation that share three common features: it is possible that the three features are all that a bilingual understands about the L2 word but the bilingual speaker may know more than three features (e.g., six) about the corresponding L1 word. Under this situation, priming from L1 to L2 will cause 100% features of the L2 target word to become activated but only 50% features of the L1 target word activated when L2 to L1 priming occurs. This is how priming asymmetry occurs according to some accounts (e.g., see Figure 2 in Schoonbaert et al., 2009, and discussions in Wang & Foster, 2010). As discussed elsewhere, our model highlights the role of lexical competition and lexical confusion among bilinguals. In our view, the richness of semantic representation and the potential lexical competition are inversely related: the richer or better understanding a bilingual has for a word, the less confusion or competition he or she may experience between the word and other lexical items. Consequently, depending on a bilingual's L2 level, more lexical competitions or confusions may exist among L2 words than among L1 words (see Zhao & Li, 2010). Such a difference may contribute greatly to the "priming asymmetry" (see further discussion below).

The above discussion suggests two important points: (i) L2 items are represented in more densely populated neighborhoods and hence have increased lexical competition from their nearby lexical items. When they serve as primes, a very brief exposure (SOA) to them may not trigger initial activations strong enough to spread to the target L1 items not directly adjacent in the representation. In contrast, activations of L1 items could be much stronger given that they are more sparsely represented (thus having less competition). Indeed, some recent Event-Related Potential (ERP) studies of bilingual translation priming show that late L2 learners often "were slower and less efficient in processing L2 primes" (see a recent review by Van Hell & Kroll, in press). In other words, an early asymmetry exists even when the initial activations are generated by the prime words, which could in turn cause the priming asymmetry to occur. (ii) When L2 words serve as the targets, their dense distribution and the strong competitions among them result in the difficulty of lexical retrieval and subsequent word naming, leading to slower reaction times for L2 than for L1 words in lexical decision tasks (which is the main effect of Direction in our simulations). In a recent ERP study of noncognate translation priming, Schoonbaert, Holcomb, Grainger and Hartsuiker (2011) found a 100 ms processing delay for L2 targets on ERP compared with L1 targets, and this delay has been associated with the asymmetric priming pattern shown in behavioral data. In short, the above two points could explain why priming asymmetry was more salient for late L2 learning than for early L2 learning.

To conclude, in this study we have attempted to provide a computational account for cross-language priming effects by extending the DevLex–II model to simulate the Chinese–English bilingual priming. The aim was to investigate how two lexicons are organized in semantic representation and how they interact with each other from a developmental perspective. The consistency between our simulation results and previous empirical findings suggests that the nature of bilingual conceptual representation is the result of a highly dynamic process shaped by the interactions between the learning of L1 and L2. Our model differs from previous bilingual computational models by using learning algorithms based on developmental and neurally plausible mechanisms such as Hebbian learning, unsupervised learning, and spreading activation to account for cross-language bilingual priming. Future computational and empirical studies should be conducted to verify the role of AoA and proficiency in cross-language priming, and to understand more generally the nature of cross-language interaction and its impact on bilingual representation and processing.

Appendix. Test material

(A) The 32 pairs of cross-language translation equivalents used in our study.
(B) The 32 cross-language semantically related word pairs used in our study.

https://doi.org/10.1017/S1366728912000624 Published online by Cambridge University Press

(A)		(B)		
Sock	袜子	Bear	狮子	(lion)
Ant	蚂蚁	Juice	牛奶	(milk)
Pillow	枕头	Sidewalk	马路	(road)
Bug	虫子	Apple	草莓	(strawberry)
Star	星星	Toe	手指	(finger)
Cup	杯子	Moon	太阳	(sun)
Ear	耳朵	Kitchen	房间	(room)
Hair	头发	Bench	桌子	(table)
Spoon	勺子	Mouth	牙齿	(tooth)
Monkey	猴子	Kitty	小狗	(puppy)
Bird	小鸟	Egg	食品	(food)
Button	扣子	Pen	铅笔	(pencil)
Train	火车	Bus	轿车	(car)
Bottle	瓶子	Towel	肥皂	(soap)
Coat	大衣	Banana	桔子	(orange)
Nail	钉子	Brother	姐姐	(sister)
Duck	鸭子	Shovel	锤子	(hammer)
Rock	石头	Lady	男人	(man)
Hat	帽子	Basket	箱子	(box)
Tongue	舌头	Uncle	阿姨	(aunt)
Mouse	老鼠	Pants	皮带	(belt)
Toy	玩具	Goose	企鹅	(penguin)
Basket	篮子	Bread	蛋糕	(cake)
Flower	花朵	Boat	飞机	(plane)
Sky	天空	Jacket	衬衫	(shirt)
Sweater	毛衣	Teacher	小孩	(child)
Shoe	鞋子	Roof	窗户	(window)
Animal	动物	Leg	胳膊	(arm)
Drawer	抽屉	Nurse	大夫	(doctor)
Tiger	老虎	Bowl	盘子	(plate)
Corn	玉米	Hen	公鸡	(cock)
Nose	鼻子	Sofa	椅子	(chair)

References

Altarriba, J., & Basnight-Brown, D. M. (2007). Methodological considerations in performing semantic- and translation-priming experiments across languages. *Behavior Research Methods, 39 (1),* 1–18.

Basnight-Brown, D., & Altarriba, J. (2007). Differences in semantic and translation priming across languages: The role of language direction and language dominance. *Memory & Cognition, 35 (5),* 953–965.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82,* 407–128.

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

De Groot, A. M. B. (1992). Bilingual lexical representation: A closer look at conceptual representations. In R. Frost & L. Katz (eds.), *Orthography, phonology, morphology, and meaning*, pp. 389–412. Amsterdam: Elsevier.

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

Dijkstra, T., & Van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition, 5 (3),* 175–197.

Dimitropoulou, M., Duñabeitia, J. A., & Carreiras, M. (2011). Two words, one meaning: Evidence of automatic co-activation of translation equivalents. *Frontiers in Psychology, 2,* 188.

Finkbeiner, M., Forster, K., Nicol, J., & Nakamura, K. (2004). The role of polysemy in masked semantic and translation priming. *Journal of Memory and Language, 51 (1),* 1–22.

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 680–698.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition, 1,* 131–149.

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

Hernandez, A., & Li, P. (2007). Age of acquisition: Its neural and computational mechanisms. *Psychological Bulletin, 133,* 638–650.

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

Jacquet, M., & French, R. (2002). The BIA++: Extending the BIA+ to a dynamical distributed connectionist framework. Comment. *Bilingualism: Language and Cognition, 5 (3),* 202–205.

James, D., & Miikkulainen, R. (1995). SARDNET: A self-organizing feature map for sequences. In G. Tesauro, D. S. Touretzky & T. K. Leen (eds.), *Advances in neural information processing systems 7*, pp. 577–584. Cambridge, MA: MIT Press.

Jiang, N. (1999). Testing processing explanations for the asymmetry in masked cross-language priming. *Bilingualism: Lang and Cognition, 2,* 59–75

Jiang, N., & Forster, K. (2001). Cross-language priming asymmetries in lexical decision and episodic recognition. *Journal of Memory and Language, 44,* 32–51.

Kiran, S., & Lebel, K. R. (2007). Crosslinguistic semantic and translation priming in normal bilingual individuals and bilingual aphasia. *Clinical Linguistics & Phonetics, 21,* 277–303.

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

Kroll, J., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connection between bilingual memory representations. *Journal of Memory and Language, 33 (2),* 149–174.

Li, P. (2002). Bilingualism is in dire need of formal models. *Bilingualism: Language and Cognition, 5,* 213.

https://doi.org/10.1017/S1366728912000624 Published online by Cambridge University Press

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

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

Li, P., & MacWhinney, B. (2002). PatPho: A phonological pattern generator for neural networks. *Behavior Research Methods, Instruments & Computers, 34 (3),* 408–415.

Li, P., Zhao, X., & MacWhinney, B. (2007). Dynamic Self-Organization and children's word learning. *Cognitive Science, 31,* 581–612.

Liu, Q., & Li, S. (2002). Word similarity computing based on How-net. *Computational Linguistics and Chinese Language Processing, 7,* 59–76.

McClelland, J. L. (2009). The place of modeling in cognitive science. *Topics in Cognitive Science, 1,* 11–28.

McNamara, T. (2005). *Semantic priming: Perspectives from memory and word recognition*. New York: Psychology Press.

Miller, G. A. (1990). WordNet: An on-line lexical database. *International Journal of Lexicography, 3,* 235–312.

Munakata, Y., & Pfaffly, J. (2004). Hebbian learning and development. *Developmental Science, 7,* 141–148.

Paivio, A., & Desrochers, A. (1980). A dual-coding approach to bilingual memory. *Canadian Journal of Psychology, 34,* 388–399.

Pavlenko, A. (2009). Conceptual representation in the bilingual lexicon and second language vocabulary learning. In A. Pavlenko (ed.), *The bilingual mental lexicon: Interdisciplinary approaches*, pp. 125–160. Tonawanda, NY: Multilingual Matters.

Schoonbaert, S., Duyck, W., Brysbaert, M., & Hartsuiker, R. J. (2009). Semantic and translation priming from a first language to a second and back: Making sense of the findings. *Memory & Cognition, 37 (5),* 569–586.

Schoonbaert, S., Holcomb, P. J., Grainger, J., & Hartsuiker, R. J. (2011). Testing asymmetries in noncognate translation priming: Evidence from RTs and ERPs. *Psychophysiology, 48 (1),* 74–81.

Segalowitz, N., & de Almeida, R. (2002). Conceptual representation of verbs in bilinguals: Semantic field effects and a second-language performance paradox. *Brain and Language, 81 (1),* 517–531.

Silberman, Y., Bentin, S., & Miikkulainen, R. (2007). Semantic boost on episodic associations: An empirically-based computational model. *Cognitive Science: A Multidisciplinary Journal, 31 (4),* 645–671.

Sirosh, J., & Miikkulainen, R. (1994). Cooperative self-organization of afferent and lateral connections in cortical maps. *Biological Cybernetics, 71,* 66–78.

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

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

Thomas, M. S. C., & Van Heuven, W. J. B. (2005). Computational models of bilingual comprehension. In J. F. Kroll & A. M. B. De Groot (eds.), *Handbook of bilingualism: Psycholinguistic approaches*, pp. 202–225. New York: Oxford University Press.

Van Hell, J. G., & De Groot, A. M. B. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition, 1 (3),* 193–211.

Van Hell, J. G., & Kroll, J. F. (in press). Using electrophysiological measures to track the mapping of words to concepts in the bilingual brain: A focus on translation. In J. Altarriba & L. Isurin (eds.), *Memory, language, and bilingualism: Theoretical and applied approaches*. New York: Cambridge University Press.

Wang, X., & Forster, K. I. (2010). Masked translation priming with semantic categorization: Testing the Sense Model. *Bilingualism: Language and Cognition, 13 (3),* 327–340.

Xing, H., Shu, H., & Li, P. (2004). The acquisition of Chinese characters: Corpus analyses and connectionist simulations. *Journal of Cognitive Science, 5,* 1–49.

Xing, H., Shu, H., & Li, P. (2007). A self-organizing model of vocabulary acquisition by elementary school children. *Contemporary Linguistics, 9,* 193–207. [In Chinese.]

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

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 Cognitive Science Society*, pp. 755–760. Nashville, TN.

Zhao, X., & Li, P. (2009). An online database of phnological representation for Mandarin Chinese monosyllables. *Behavior Research Methods, 41,* 575–583.

Zhao, X., & Li, P. (2010). Bilingual lexical interactions in an unsupervised neural network model. *International Journal of Bilingual Education and Bilingualism, 13,* 505–524.

Zhao, X., Li, P., & Kohonen, T. (2011a). Contextual self-organizing map: Software for constructing semantic representation. *Behavior Research Methods, 43,* 77–88.

Zhao, X., Li, P., Liu, Y., Fang, X., & Shu, H. (2011b). Cross-language priming in Chinese–English bilinguals with different second language proficiency levels. In L. Carlson, C. Hölscher & T. Shipley (eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, pp. 801–806. Austin, TX: Cognitive Science Society.

https://doi.org/10.1017/S1366728912000624 Published online by Cambridge University Press