Brain and Language 195 (2019) 104643

Contents lists available at ScienceDirect

Brain and Language

journal homepage: www.elsevier.com/locate/b&l

BiLex: A computational approach to the effects of age of acquisition and language exposure on bilingual lexical access

Claudia Peñaloza[a,*], Uli Grasemann[b], Maria Dekhtyar[a], Risto Miikkulainen[b], Swathi Kiran[a]

[a] Aphasia Research Laboratory, Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA 02215, USA
[b] Department of Computer Science, The University of Texas at Austin, TX 78712, USA

ARTICLE INFO

Keywords:
Bilingualism
Age of acquisition
Language exposure
Lexical access
Naming
Computational modeling
Neural network
Evolutionary algorithm

ABSTRACT

Lexical access in bilinguals can be modulated by multiple factors in their individual language learning history. We developed the BiLex computational model to examine the effects of L2 age of acquisition, language use and exposure on lexical retrieval in bilingual speakers. Twenty-eight Spanish-English bilinguals and five monolinguals recruited to test and validate the model were evaluated in their picture naming skills in each language and filled out a language use questionnaire. We examined whether BiLex can (i) simulate their naming performance in each language while taking into account their L2 age of acquisition, use and exposure to each language, and (ii) predict naming performance in other participants not used in model training. Our findings showed that BiLex could accurately simulate naming performance in bilinguals, suggesting that differences in L2 age of acquisition, language use and exposure can account for individual differences in bilingual lexical access.

1. Introduction

Lexical access is a fundamental aspect of language processing that provides an excellent window into the ability of bilingual speakers to handle two languages and select words from their mental lexicons. Lexical access in speech production can largely differ across bilinguals as mastering two languages depends on multiple factors that make the bilingual experience dynamic and almost unique to each speaker (Luk & Bialystok, 2013). There is considerable variation among bilinguals in terms of the degree of similarity between their two languages, their language learning contexts and their L2 age of acquisition (AoA), which leads to different patterns of learning, exposure, use, and attained proficiency for each language (Costa & Sebastián-Gallés, 2014). This natural variability and intrinsic complexity in bilingualism also poses methodological challenges for bilingual research. For instance, heterogeneity in language learning history can especially affect studies with small samples and variables that influence bilingual language acquisition and processing can often be confounded. Computational approaches may offer a potential solution to this problem by facilitating the principled and systematic investigation of important factors that influence bilingual lexical access in the context of individual variations (Fricke, Zirnstein, Navarro-Torres, & Kroll, 2018; Li, 2013). Current models of lexical access assume that word retrieval is achieved via spreading activation across at least two different representational stages in the lexical network from the conceptual system to phonology (Dell, 1986; Levelt, Roelofs, & Meyer, 1999). Because several lexical representations can be activated from the semantic system, the lexical unit with the highest degree of activation is ultimately chosen for production. In bilinguals, the semantic system activates the lexical nodes of the two languages (Colomé, 2001; Costa, 2005; Costa, Caramazza, & Sebastian-Galles, 2000; Gollan & Kroll, 2001) and cross-language activation and interaction in lexical processing are highly modulated by differences in L2 proficiency (van Hell Janet & Darren, 2012). Importantly, the Revised Hierarchical Model (RHM) (Kroll & Stewart, 1994) provides a developmental account of bilingual word production across varying degrees of L2 proficiency. Similar to most models of the bilingual mental lexicon (see French & Jacquet, 2004 for a review) the RHM assumes that concepts are stored in a shared semantic system with word forms stored in two separate lexical systems for each language. Active connections between representations in the conceptual and lexical systems vary in strength according to L2 fluency and relative dominance of L1 to L2 such that the mappings between L2 word forms and meanings are weak in early L2 learning, but become stronger with increased proficiency. This underlying asymmetry allows accounting for individual differences in language learning history in lexical processing (Kroll, Van Hell, Tokowicz, & Green, 2010).

Individual differences in lexical access can also be influenced by developmental and contextual factors including L2 AoA, and the degree

* Corresponding author at: Aphasia Research Laboratory, Department of Speech, Language and Hearing Sciences, Sargent College of Health & Rehabilitation Sciences, Boston University, 635 Commonwealth Ave, Boston, MA 02215, USA.
E-mail address: penaloza@bu.edu (C. Peñaloza).

https://doi.org/10.1016/j.bandl.2019.104643
Received 8 December 2018; Received in revised form 7 June 2019; Accepted 9 June 2019
Available online 24 June 2019

of lifetime exposure to each language and their relative frequency of use, which ultimately contribute to a bilingual's language proficiency (Kastenbaum et al., 2018). L2 AoA can facilitate or constrain the achievement of native-like L2 attainment (Birdsong, 2018) and late AoA has been associated with slower and less accurate lexical retrieval as compared to early AoA (Hirsh, Morrison, Gaset, & Carnicer, 2003; Kohnert, Hernandez, & Bates, 1998). Notably, beyond the maturational constraints related to the age of L2 learning onset, bilingual lexical access can also be influenced by the age at testing as the human lexicon tends naturally to increase over time from early infancy to adulthood, although normal aging can lead to a decay in word retrieval and overall lexical processing in both languages (Juncos-Rabadán, 1994). Also, language exposure and usage can influence L2 performance modulating lexical access (Kastenbaum et al., 2018) and tip-of-the-tongue states in bilinguals (Kreiner & Degani, 2015). Moreover, increased exposure and use may lead to both higher availability of the most exposed language for production and lower automaticity in word-finding in the less exposed language (Tu et al., 2015). Importantly, L1 and L2 also influence each other mutually and increased frequency of L2 use and immersion in an L2 environment can lead to decreased L1 proficiency in bilinguals (Baus, Costa, & Carreiras, 2013; Linck, Kroll, & Sunderman, 2009) affecting L1 lexical retrieval and often leading to L1 attrition (Schmid, 2010).

The examination of these sources of variation in language processing can help to better understand the architecture of the language system (Fricke et al., 2018) and to address unresolved questions regarding lexical acquisition and retrieval in bilinguals (Costa & Sebastián-Gallés, 2014). In this sense, computational modeling offers a promising framework to (i) systematically examine how lexical access is modulated by variations in factors that influence language learning and processing in bilinguals, (ii) disentangle the effects of influencing factors that often confound in behavioral research, (iii) provide an account on how such variations modulate bilingual language representation, and (iv) examine premorbid factors of the individual language learning history that influence language breakdown and recovery in bilinguals with language deficits. Computational models based on artificial neural networks and in particular Self-Organized Maps (SOMs, Kohonen, 2001) have made important contributions to language research (Li & Zhao, 2013; Miikkulainen, 1993). SOM models are trained using unsupervised learning algorithms with several properties that make them highly suitable to model the mental lexicon (Li & Farkas, 2002). SOM-based models of bilingualism have helped addressing multiple aspects of language acquisition and processing (Fang, Zinszer, Malt, & Li, 2016; Li & Farkas, 2002; Miikkulainen & Kiran, 2009; Shook & Marian, 2013; Zhao & Li, 2010, 2013). Among these, our previous bilingual DISLEX model was developed to represent lexical processing across varying combinations of L2 AoA and proficiency to fit a given combination of these factors to a particular individual's learning background and performance (Miikkulainen & Kiran, 2009). Importantly, its use was helpful to examine language breakdown and recovery in bilingual aphasia (Grasemann, Sandberg, Kiran, & Miikkulainen, 2011; Kiran, Grasemann, Sandberg, & Miikkulainen, 2013). However, no previous models including the bilingual DISLEX model, were designed to match behavioral lexical access performance in healthy bilinguals while examining the joint contribution of relevant variables that influence bilingual language representation and processing beyond AoA and proficiency.

Here, we sought to develop BiLex, a SOM-based computational connectionist model that can capture the natural variation in lexical access among adult bilinguals with varying degrees of L2 proficiency while accounting for developmental and contextual factors known to influence lexical retrieval and ultimate L2 competence. BiLex expands on our previously built DISLEX model (Miikkulainen & Kiran, 2009) that incorporated aspects of the RHM of the organization of the bilingual lexicon (Kroll & Stewart, 1994), to simulate lexical access in healthy bilinguals. The motivation for the development of BiLex was to extend our previous work based on computational simulations of language impairment and treatment response in bilinguals with aphasia (Grasemann et al., 2011; Kiran et al., 2013) to the prediction of treatment outcomes in this population.

The aim of the present study was twofold. We first aimed to examine whether BiLex can accurately simulate lexical access in Spanish-English adult bilinguals while taking into account their age at testing, L2 AoA, and a fine–grained characterization of their language exposure and usage, while incorporating aging and attrition effects in the overall model training. Our second aim was to determine whether the model is able to predict the naming performance of other participants whose data was not used in model training. In this way, these two specific aims can contribute to the validation of BiLex towards its ultimate goal: the prediction of treatment outcomes in bilinguals with aphasia by (i) using BiLex to simulate individual premorbid naming ability in each language, (ii) implementing a lesion component in each individual premorbid BiLex model to reflect the effects of brain damage on the bilingual language system and (iii) retraining such lesioned models to simulate and predict individual response to language treatment provided in one language versus the other.

2. Materials and methods

2.1. Participants

Participants were 33 healthy adults including 28 Spanish-English bilinguals (6 male, mean age = 42.89, SD = 16.2, range = 18–82; mean number of years of education = 17.13, SD = 4.46, range = 9–27), and 5 monolinguals (2 male, mean age = 56, SD = 5.15, range = 49–63; mean number of years of education = 13.14, SD = 4.56, range = 10–21) recruited to reflect maximum exposure to and use of Spanish (n = 2) and English (n = 3), thus allowing for a full range of inter-individual variability in terms of naming performance and language learning history across languages. The demographic information of all participants is provided in Table 1. All participants had normal or corrected-to-normal vision and hearing and no history of neurological or psychiatric illness. Following approved procedures by the Ethical Committee at Boston University, participants gave their written consent to undergo language testing in person or via videoconference using GoToMeeting (LogMeIn, Boston, MA, USA) and received a gift-card for their participation.

2.2. Assessment of language learning history

All participants filled out a Language Use Questionnaire (LUQ, Kastenbaum et al., 2018), where they reported their L2 AoA and provided a detailed profile of their language learning history in Spanish and English including current language use, lifetime exposure, lifetime confidence, family proficiency, educational history and language ability rating (see Kastenbaum et al., 2018 for a detailed description of all LUQ metrics). Of these measures, L2 AoA, lifetime exposure and current use of each language were selected as the most important inputs to train the BiLex model (see Section 2.7) because of their contribution to bilingual lexical access (Kastenbaum et al., 2018), and because they represent the most objective, fine-grained measures that could be effectively and reliably implemented in our computational simulations. L2 AoA reflected the age of L2 learning onset. The lifetime exposure section of the LUQ requested participants to indicate the percentage of time (expressed in 25% increments: 100% Spanish, 25% English-75% Spanish, 50% in each language, 75% English-25% Spanish, or 100% English) they spent hearing, speaking and reading in each language in each three-year interval from age 0 to 30 and a single final interval for age "30 and up". For current language use, participants needed to detail the languages they and their conversation partners used on weekdays and weekends on an hourly basis (Table 1).

For each participant, current age at testing, English and Spanish

Table 1
Demographics, LUQ metrics on L2 AoA, language exposure and use, and naming performance of all participants.

Participant	Sex	Age	Education (years)	Language profile	L1	L2 AoA	Lifetime exposure Spanish	Lifetime exposure English	Current usage Spanish	Current usage English	Comp. Naming Spanish	Comp. Naming English
P1	F	53	20	Bilingual	Spanish	6	0.63	0.37	0.13	0.87	0.88	0.82
P2	M	18	14	Bilingual	Spanish[1]	0	0.39	0.61	0.03	0.97	0.48	0.82
P3	M	36	12	Bilingual	Spanish	20	0.64	0.36	0.42	0.58	0.79	0.75
P4	F	18	15	Bilingual	Spanish	4	0.62	0.38	0.28	0.72	0.87	0.72
P5	F	36	21	Bilingual	Spanish	26	0.70	0.30	0.47	0.53	0.88	0.78
P6	M	45	27	Bilingual	Spanish	12	0.60	0.40	0	1	0.91	0.55
P7	F	30	23	Bilingual	Spanish	7	0.76	0.24	0.28	0.72	0.92	0.67
P8	F	48	15	Bilingual	Spanish	15	0.55	0.45	0.09	0.91	0.95	0.88
P9	F	39	21	Bilingual	Spanish	36	0.80	0.20	0.22	0.78	0.94	0.68
P10	M	30	26	Bilingual	Spanish	7	0.68	0.32	0.21	0.79	0.87	0.88
P11	F	27	18	Bilingual	Spanish	6	0.39	0.61	0.19	0.81	0.75	0.88
P12	F	25	22	Bilingual	English[1]	0	0.53	0.47	0.36	0.64	0.88	0.83
P13	F	37	16	Bilingual	Spanish[1]	0	0.37	0.63	0.22	0.78	0.81	0.97
P14	F	21	20	Bilingual	Spanish	7	0.81	0.19	0.82	0.18	0.73	0.48
P15	F	73	11	Bilingual	Spanish	23	0.89	0.11	0.87	0.13	0.73	0.45
P16	F	33	14	Bilingual	Spanish[1]	0	0.44	0.56	0.45	0.55	0.60	0.94
P17	F	63	14	Bilingual	Spanish	7	0.89	0.11	0.97	0.03	0.85	0.60
P18	F	33	18	Bilingual	English	19	0.19	0.81	0.5	0.5	0.82	0.99
P19	F	54	14	Bilingual	Spanish	5	0.45	0.55	0.37	0.63	0.61	0.88
P20	F	38	18	Bilingual	Spanish	21	0.64	0.36	0.07	0.93	0.79	0.83
P21	F	53	18	Bilingual	Spanish	18	0.51	0.49	0.17	0.83	0.77	0.75
P22	F	45	14	Bilingual	Spanish	26	0.65	0.35	0.73	0.27	0.73	0.62
P23	F	55	16	Bilingual	Spanish	12	0.26	0.74	0.07	0.93	0.80	0.94
P24	M	36	20	Bilingual	English	0	0.18	0.82	0.25	0.75	0.15	0.95
P25	F	52	12	Bilingual	Spanish	3	0.29	0.71	0.61	0.39	0.76	0.80
P26	F	82	12	Bilingual	Spanish[1]	40	0.89	0.10	1	0	0.7	0.35
P27	M	60	9	Bilingual	Spanish	27	0.74	0.26	0.86	0.14	0.68	0.68
P28	F	61	19	Bilingual	Spanish	3	0.69	0.31	0.96	0.04	0.85	0.69
P29	M	59	12	Monolingual	Spanish	–	1	0	1	0	0.92	0.06
P30	M	49	10	Monolingual	English	–	0	1	0.33	0.67	0	0.93
P31	F	63	10	Monolingual	Spanish	–	1	0	1	0	0.88	0.11
P32	M	56	21	Monolingual	English	–	0	1	0	1	0.05	0.98
P33	F	58	12	Monolingual	English	–	0	1	0	1	0	0.98

LUQ = Language Use Questionnaire; AoA = age of acquisition; L1 = native language, L2 = second language; Comp. Naming = Composite naming score; F = female; M = male.
Lifetime exposure and current use are shown as computed by the LUQ. Composite naming scores expressed as proportion of correct responses.
[1] Language the participant was most exposed to from birth although L2 was initially acquired at age 0.

lifetime exposure and English and Spanish current use were combined to estimate their percentage of English vs. Spanish overall language exposure on a yearly basis from birth to their current age. To compute this combined metric, lifetime exposure percentages were first calculated for each language as a weighted average of hearing, speaking, and reading for each three-year age interval, giving equal weight of 1 to speaking and listening, and 0.5 to reading.[1] The resulting percentage of lifetime exposure to English vs. Spanish was replicated for each year of life included in a given three-year age interval. Next, overall current use of English vs. Spanish was calculated as the average percentage of the overall time a participant and conversation partner spent each hour using each language during weekdays and weekends. The resulting percentage of current use was then combined with the lifetime exposure data over the five last years of life of the participant, shifting gradually from the original lifetime exposure percentage to the current use percentage for each language (e.g.: for a participant with age at testing = 50, English lifetime exposure at "30 and up" = 50%, and English current use = 60%, the final estimated overall language exposure for English for the last five years of life including ages 45–50 was 50, 52, 54, 56, 58, 60% respectively).

2.3. Assessment of lexical access in picture naming

Lexical access was examined in both languages in all participants using the Boston Naming Test (BNT- Kaplan, Goodglass, & Weintraub, 2001; Kohnert et al., 1998) and a 60-item naming screener involving 60 pictures of concrete words individually presented on a laptop computer. The lexical frequency (per million) of these words was estimated in English and Spanish using Clearpond (Marian, Bartolotti, Chabal, & Shook, 2012) and Espal (Duchon, Perea, Sebastián-Gallés, Martí, & Carreiras, 2013) respectively, and matched across languages (Spanish $M = 17.56$, $SD = 28.9$; English $M = 17.39$, $SD = 29.1$, $t(115) = -0.031$, $p = .976$). The order of language tested was counterbalanced across participants and both target words and acceptable dialectal or lexical variations were credited. Naming performance across tests in each language was averaged into a *composite naming score* for each participant and was used as an individual index of lexical access to be simulated using the BiLex model.

2.4. Word corpus for neural network training

The training corpus included 638 concrete nouns in English and their direct translations to Spanish. The word set was sufficiently representative as it included exemplars of thirteen semantic categories with different physical, categorical and functional features. For the purpose of model training, the semantic and phonetic representations of these words were developed as vectors of numbers that indicated the extent to which a representation had particular semantic or phonetic

[1] Reading skills had a smaller weight in the calculation of lifetime exposure because picture naming abilities do not directly rely on access to orthography and to account for bilinguals who do not acquire their L2 in formal education settings and thus, their lower reading abilities are less representative of their language exposure than their hearing and speaking skills.

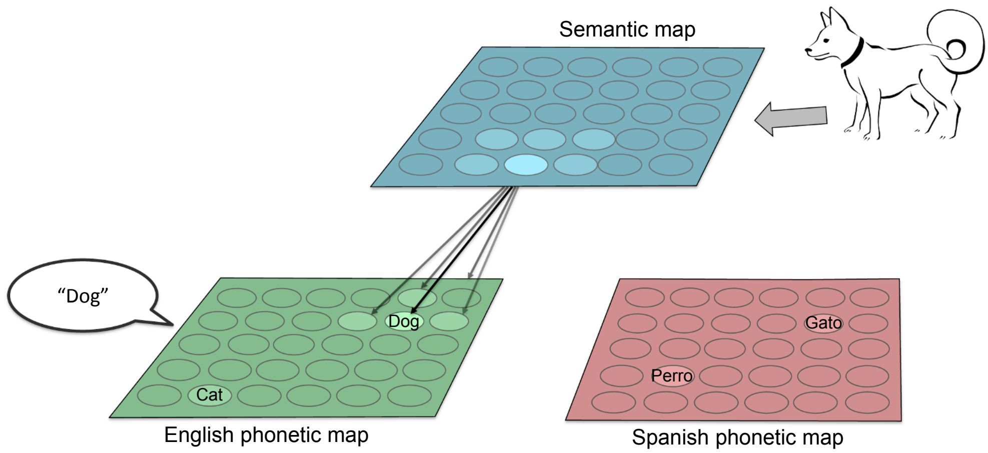

Fig. 1. Architecture of the BiLex model. The model consists of three SOMs one for the semantic representations of words shared across languages, and two for the phonetic representations of words in each language.

features.

Semantic vector representations were created using the semantic feature data developed for another project (Sandberg, Gray, & Kiran, 2018). Briefly, feature validation using MTurk (https://www.mturk.com/), consisted in assigning 10–20 relevant semantic features per word (e.g.: feature "can fly" assigned to "vulture") and asking healthy adults whether or not a given word X had the feature Y. A total of 400 features were chosen on the basis that MTurk feature data were available across a minimum of five words while feature variability in the corpus would allow to distinguishing between otherwise similar words. Each word (e.g. "ant") was built a vector of numerical features that encoded the percentage of adults who thought a given feature was applicable to that particular word (e.g., "moves": 100%; "has legs": 94%; "swims": 16%). Because full input vectors containing all features were necessary for model training, the final dense semantic representations were compiled using zeros in the case of word-feature pairs that were not represented in the MTurk data set, resulting in vectors containing all 400 features for each word. These vectors were used as input for the semantic map during model training as initial experiments confirmed that the resulting representations were sufficiently detailed and accurate to enable well-organized SOM models of the semantic system.

Phonetic representations of English words and their Spanish translations were based on feature-based encodings of symbols in the International Phonetic Alphabet (IPA). All relevant phonemes (each consisting of a letter and possible diacritics) were encoded using four numeric values. IPA symbols for vowels were encoded according to four features: height, backness, length, and roundedness. For instance, /ə/, denoting a mid-central short unrounded vowel, was represented by (0.5, 0.5, 0, 0) whereas /œː/, an open front long rounded vowel, was encoded as (1, 1, 1, 1). IPA symbols for consonants were encoded using a different set of numeric features designed to approximate their phonetic properties: place and manner of articulation, phonation, and lateralization. Similar to roundedness for vowels, phonation and lateralization were binary features (e.g. 1 for a voiced consonant, 0 for a voiceless one). Place of articulation was encoded using a 0 for a bilabial sound, 1 for a glottal sound, and increasing values within this range for dental, palatal, and other intermediate sounds. Similarly, the manner of articulation was encoded as a single number ranging between 0 for "stop" consonants and 1 for nasal consonants (e.g., 0.3 for fricative consonants). Thus, feature-based encodings for vowels and consonants represent a tractable, simple encoding scheme that approximates the phonetics of words well enough to make meaningful comparisons between feature-based phonetic encodings.

Several constraints were implemented for the phonetic vector representations including: (a) numeric vectors of equal length for each word, (b) vowels and consonants were to be equated in the same way for all words within a language, and (c) word representations were aligned relative to their stress patterns instead of their absolute position within a word to enable differences in stress between similar sounding words (e.g. "insight" vs. "incite" in English). Considering these criteria, phonetic vector representations were created for all words in the corpus as follows. First, IPA transcriptions for each word were manually split into spoken syllables. Each syllable was then encoded as a CCVVCC structure. In cases where consonants were missing, those from the neighboring syllable were repeated. Single vowels or consonants were doubled if needed, and in rare cases where a syllable started or ended with three consonants, the features for the second and third were averaged. Thus, each syllable was represented by a list of 24 numeric attributes (6 phonemes with 4 features each) that could be meaningfully compared on an element-by-element basis. The representation of the entire word was obtained by concatenating the representations for each syllable and padding all words with neutral feature values (0.5) such that all had equal length, and the syllable carrying the primary stress was always in the same position.

The position of the primary stress, as well as the overall length of the phonetic representations, was determined by the maximum number of syllables before and after the primary stress that occurred in the corpus. In English, phonetic vector representations used 144 features (6 syllables with 24 features each) with the primary stress on the third syllable. In Spanish, phonetic vector representations used 192 features (8 syllables with 24 features each), with the primary stress on the fourth syllable. These vectors served as the training input for the phonetic maps of each language.

2.5. Model architecture and training

The architecture of the BiLex model is shown in Fig. 1 and is representative of all its individual instances simulating the naming performance of each participant. Following the assumptions of the organization of the bilingual mental lexicon proposed in the Revised Hierarchical Model (Kroll & Stewart, 1994), the architecture consists of three separated systems, a semantic system for word meanings and two phonetic systems for the phonetic representations of words in L1 and L2 respectively. All three systems are interconnected with one another by directional associative connections that vary in strength depending on relative language dominance and allow for network activation to propagate between systems. The semantic and phonetic systems were modeled as SOMs. Each SOM is a two-dimensional topographic grid of information processing units or neurons, trained to encode the semantic and phonetic vector representations developed for each word in the training corpus. The training of the entire BiLex model uses the standard SOM training algorithm, but trains the maps in parallel (i.e., the semantic map and one of the phonetic maps at a time) along with the associative connections that allow activation to flow between them.

In each training instance, a word from the entire corpus is randomly selected. Based on the current simulated age of the model (i.e. how far training has progressed), a language is selected probabilistically for training. For instance, if the current age of the model is 15 and the participant's *overall language exposure* at age 15 is 75% Spanish and 25% English, the training algorithm would train the selected word in Spanish with 75% probability. The semantic and phonetic vector representations of the selected word are then simultaneously presented to the corresponding maps leading to learning-induced changes in the architecture of the model. Within each map, every time an input vector is presented, the SOM algorithm computes for each neuron the Euclidean distance d between its weight w vector (i.e., connection strength) and the input vector or symbol representation.

The neuron with the smallest distance (i.e., winner unit) receives the highest degree of activation, and its weights and those of its neighbor neurons on the map grid are adjusted towards the input vector:

$$w'_{ij} = w_{ij} + \alpha(v - w_{ij})\theta_{ij,\sigma}$$

where α is the learning rate, w_{ij} is the weight vector of the map unit at grid position ij and v is the input vector. The Gaussian neighborhood function θ is centered on the winner unit; it determines how much each neighboring unit is adjusted. The width σ of the Gaussian determines how fast σ decreases with increasing map distance to the winner unit, and hence controls the size of the neighborhood that is adjusted meaningfully.

As a result of this process, the neuron's weight vector becomes a representation of the input vector, and the weights in the neighbor neurons become more similar to the input. Over time and with additional exposure, the neurons encoding input vector representations become increasingly organized in the semantic and phonetic space as a function of similarity, such that neurons located close in the map space tend to encode words with similar semantic features (semantic map) and phonetic properties (phonetic maps).

The effectiveness of SOM training depends on how many training patterns are presented to the map, and how the neighborhood size σ, and the learning rate α in Eq. (1) change over time. Initially, the size of the neighborhood is relatively large to establish the map global structure. To speed up training, learning rates are relatively high because the gross map structure is more important than precisely tuned weights. As training progresses, the size of the neighborhood is gradually reduced, leading the map to learn the similarity relations between input patterns at a progressively more fine-grained level. By the end of training, the neighborhood is usually reduced to a size where only the winner unit and, to a lesser degree, a few surrounding units adapt in each training cycle. Similar to neighborhood size, the learning rate is usually reduced over time, which allows the map to fine-tune its unit vectors and slowly settle into a locally more precise representation of the input space.

Presenting the semantic and phonetic vector representations of a given word to the maps also results in the activation of neurons in the corresponding semantic and phonetic maps. Activations are normalized to 1 for the unit closest to the input vector (i.e., winning neuron) and 0 for the one with the largest distance, and change linearly for distances in between (i.e., neighboring neurons). In this way, both the winning and neighboring neurons tend to react more strongly to the same or similar input vectors in future presentations. These activations are then used to adapt the associative connections between maps by strengthening the connections that link active neurons across maps using Hebbian learning. Thus, every time the neurons encoding the semantic and phonetic vector representations of a given word are co-activated across maps, their associative links become stronger and increase their likelihood of being co-activated in future instances. Cross-language connections between the two phonetic maps receive similar training. However, if the model is exposed to a word in one language, only the connections leading to the exposed map are trained (i.e., when training an L1 word, connections from L2 to L1 are trained and vice versa). As a result of this learning process, when a concept is presented to the semantic map, its associated phonetic representations in both phonetic maps are activated.

Additionally, because lexical access can decrease in humans due to aging or attrition, small amounts of noise were added to all associative connections during training (i.e., a random number between 0 and a threshold parameter γ was added to each associative connection). This approach to modeling aging and attrition was based on behavioral data showing that lexical retrieval (Connor, Spiro, Avron, Obler, & Albert, 2004; Kavé, Knafo, & Gilboa, 2010) and verbal memory performance (Salthouse, 2003) decline in older adulthood because the associations that link different concepts or attributes (e.g. words and meanings) become weaker in older adults (Naveh-Benjamin, 2000).

Finally, because Hebbian learning will always increase connections strengths, the overall sum of outgoing associative connections is normalized such that for each neuron, the L2 norm of outgoing connections to each target map is 1. This allows the overall output of a map to stay bounded, associative connections are able to adapt throughout model training, and the relative strength of outgoing connections remains the same. For efficiency, both normalization and connection noise are applied only once per training epoch (i.e. per year over the simulated lifetime of the model).

2.6. Simulations of human naming performance

Once training is completed, naming a specific word in either language can be simulated by presenting its semantic vector representation to the semantic map of the BiLex model. The resulting activation is then propagated to the phonetic map of the language simulated via the associative connections resulting in phonetic map activation. The weight vector of the most highly activated phonetic unit is then compared to all the phonetic representations in the corpus of that particular language, the word with the minimal distance is identified and produced as output. If the output word is the same as the original input, the word is counted as correctly named. The simulated naming performance is then calculated as the percentage of words that are correctly named in this way. In order to simulate human performance separately on each naming test, the corpus was split into two subsets. The first subset of words was used more frequently than the rest of the corpus for model training to reflect the high-frequency of words included in the 60-item naming screener. The second subset included 100 relatively rare words of the corpus that were used with less frequency for model training to reflect the inclusion of low-frequency words in the BNT. The higher and lower frequency subsets were then used for simulations of performance on the 60-item naming screener and the BNT respectively. As with the participants, a *simulated composite naming score* for each trained BiLex model was calculated as the average of both naming tests, and was used to evaluate how well the model's simulated scores were able to match actual naming performance (i.e., *composite naming score*).

2.7. Evolutionary algorithm, individual and global training parameters

A key feature of BiLex is its ability to capture the effects of an individual's language learning history on lexical access. Thus factors including *age at testing, overall language exposure*, and *L2 AoA* reflect the individual differences between participants (i.e., individual training parameters), whereas all other parameters governing model training remain constant across individuals (i.e., global training parameters). The individual training parameters were based on each individual's self-reported LUQ data. The *age at testing* indicated the number of epochs used for model training, one training epoch per simulated year of life for each simulated participant. The *overall language exposure* metric computed for each language determined the proportion of

English and Spanish words randomly selected for training during each simulated year (as described in Section 2.5). L2 *AoA* was accounted for as zero exposure in the L2 phonetic map up to the reported L2 *AoA* at which point training of the L2 phonetic map and its associative links with the semantic map commenced.

The global training parameters included the learning rate α, and the neighborhood size θ at different simulated ages, the scale γ of the random noise added to associative connections each epoch to simulate aging and language attrition effects, the size N of grid neurons for each of the three SOMs and word frequency. We used an evolutionary algorithm (EA, Bäck, 1996) to find the best-fit set of global parameters that when combined with the individual parameters of each participant would result in a set of individually-tailored BiLex models that were able to reproduce each participant's naming performance in each language. EAs are population-based optimization algorithms that use biologically inspired mechanisms to evolve a population of candidate solutions to a given problem (i.e., candidate sets of global training parameters). We used a steady-state EA which started out with a random population of candidate solutions that were continuously optimized and assessed using an evaluation function (see Section 2.8 for details). This optimization took place by repeatedly selecting, recombining, and mutating high-fitness candidate solutions, discarding old ones and replacing low-fitness ones with the offspring of high-fitness ones according to the evaluation of the quality of each candidate solution (i.e., goodness-of-fit) while keeping each population at an approximately constant size. A candidate solution in the EA consisted of 20 numeric values encoding all global training parameters (see Appendix A for details on the optimization of training parameters). An initial population of 100 candidate solutions was generated by choosing random values within reasonable intervals chosen empirically for each parameter. For both learning rates α and neighborhood sizes θ, a parameter value was required for every year of training. An important goal was to speed up the production of reasonable candidate solutions and maintain population diversity while limiting the number of optimized parameters to avoid overfitting. Thus, both parameters were optimized only for certain fixed ages and parameter values for the ages in between were determined using linear interpolation. Also, to ensure that parameters stayed reasonable over time both parameters were constrained to non-increasing values after age 4 during model training.

New candidate solutions were added to the population whenever the population size fell below a threshold of 30. For efficiency, instead of adding new candidates one at a time, the population size was increased to a predetermined maximum of 70. The cycle of evaluation, selection, recombination, and mutation was repeated as long as a new best-fit parameter set was found at least once in every 500 candidates added to the population. However, if none of the most recent 500 solutions was able to improve on the previous best solution, new candidates were instead added by just selecting and mutating (i.e., skipping the recombination step, but increasing the mutation rate to 0.5). This kind of mutation burst allowed maintaining population diversity while preventing the EA from converging on a sub-optimal solution prematurely. The EA optimization was run until no new best-fit candidate solution was found in the most recent 1000 candidates added to the population. All parameter settings governing the EA, such as population sizes and mutation rates, were set empirically and are explained elsewhere (Grasemann, Peñaloza, Kiran, & Miikkulainen, in press).

2.8. Fitness evaluation of optimized candidate solutions

To evaluate how well a particular candidate solution in the population was able to match the naming performance of a given participant, a BiLex model was trained using the combined global and individual training parameters. Naming scores in each language were then simulated separately for the 60-item naming screener and the BNT using the trained BiLex model. The goodness-of-fit (GOF) for candidate solution $\boldsymbol{c}$ in the population on individual participant $\boldsymbol{i}(\boldsymbol{GOF_{ci}})$was then calculated as the sum of squared residual fitting errors over the four tests scores (both naming tests in English and Spanish). Additionally, to evaluate how well candidate solution $\boldsymbol{c}$ was able to match naming performance in general, its GOF was averaged over all individual training parameters sets available for model fitting by the EA. The fitness of candidate $\boldsymbol{c}$ was then calculated as

$$fitness_c = \sqrt{\frac{N}{\sum_{i=1}^{N} GOF_{ci}}}$$

where $\boldsymbol{N}$ is the number of individual parameter sets on which $\boldsymbol{c}$ was evaluated.

Model training and testing was implemented on GPU hardware using TensorFlow (Abadi, Agarwal, Barham, Brevdo, Chn, Citro, & Ghemawat, 2015). To increase efficient use of available computing resources, a modified evaluation function based on the age layering technique (Shahrzad, Hodjat, & Miikkulainen, 2016) was used to limit full evaluations to just the most promising candidate solutions in the population (for details see Grasemann et al., in press).

2.9. Cross-validation

We conducted a five-fold cross-validation through random data splitting to evaluate the simulation performance of BiLex and the EA parameter fitting procedure. This analysis allows to ensure the best-fit candidate solution is reliably comparable to other high-fitness solutions (i.e., and therefore is not an outlier in the population) while minimizing the possibility of data over-fitting. Each of the 33 participants were randomly assigned to five test sets (n = 6 or 7), with the exception that each of the five test sets included one monolingual participant to ensure that each EA was exposed to the full range of both language exposure and naming performance. For each test set, the remaining participants were assigned to the corresponding training set (n = 27 or 26), and their individual training parameters were used for GOF evaluation by the EA during evolution. The five-best fit candidate solutions from each EA were used to train individual BiLex models for each participant in its corresponding test set using their individual training parameters together with the global training parameters evolved to match other participants' naming scores.

3. Results

3.1. LUQ metrics and naming performance in Spanish and English

Bilingual participants presented varying profiles of language learning history in terms of their L2 AoA (M = 12.50, SD = 11.38; range = 0–40) their lifetime exposure to Spanish (M = 0.58, SD = 0.21; range = 0.18–0.89) and English (M = 0.42, SD = 0.21; range = 0.10–0.82) and their current use of Spanish (M = 0.41, SD = 0.32; range = 0–1) and English (M = 0.59, SD = 0.32; range = 0–1). They also showed a wide range of naming performance in both languages. This variability was evident on the Spanish BNT (M = 0.72, SD = 0.17; range = 0.27–0.95), the 60-item naming screener (M = 0.82, SD = 0.15; range = 0.23–0.95), and on the resulting composite naming score in Spanish (M = 0.77, SD = 0.15; range = 0.25–0.95). Likewise, a large individual variation in naming performance was also observed on the English BNT (M = 0.70, SD = 0.17; range = 0.30–0.98), the 60-item naming screener (M = 0.82, SD = 0.15; range = 0.46–1) and consequently in the *composite naming score* in English (*M* = 0.76, *SD* = 0.16; *range* = 0.40–0.99) (Table 1).

Thus, the behavioral data available for simulations are representative of a large range of bilingual speakers across different ages at testing, *L2 AoA*, *lifetime exposure* and *current language use* (i.e., reflected in the individual *overall language exposure),* and levels of naming ability in both languages. The inter-individual variability in bilingual

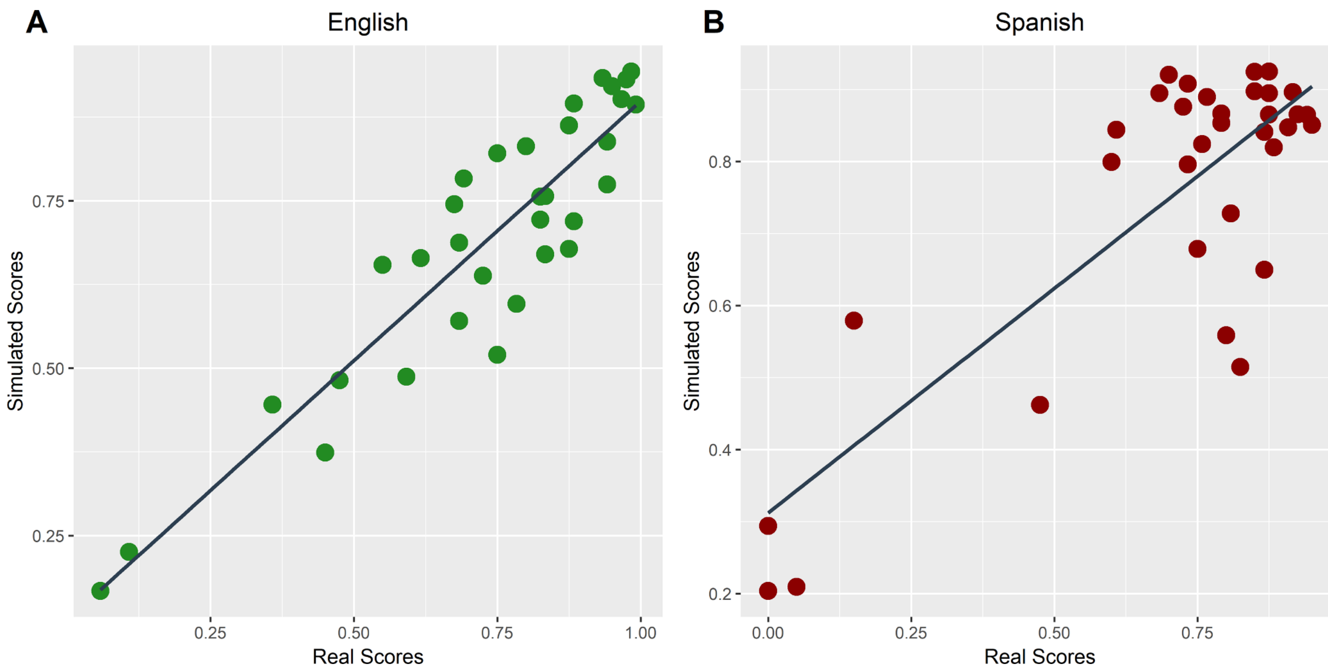

Fig. 2. Results of five-fold cross validation. Scatterplots show how the simulation composite scores (y axis) predicted actual composite naming scores (x axis) in all participants in both English (left) and Spanish (right) across all five best-fit candidate solutions.

language learning history is important for the validation of the BiLex model, as it allows to test the assumption that when considering *L2 AoA*, *lifetime exposure* and *current language use* as individual parameters for model training, BiLex can account for individual differences in naming performance in each language across different profiles of bilingualism. As expected, the monolingual participants showed extremely high *lifetime exposure* and *current use* of their native language accompanied by an equally high naming performance in their native language and only minimal ability in the other language (Table 1). This confirms that the test sets could benefit of extreme profiles of high naming performance in the context of high exposure and use of just one language.

3.2. Cross-validation of the EA parameter fitting method

The five EAs were run on the five training sets and each produced a number of highly fit candidate solutions with comparable best-fitness values across runs $M = 12.83$; $SD = 0.72$. The final five best-fit candidate solutions (Appendix B) were found by the EAs after evaluating approximately 2750 candidate solutions ($M = 2749$, $SD = 1023$, *range*: 1024–3727). Although training sets differed for all runs, the EAs converged on similar parameters in many cases, e.g. low but finite minimum exposure ε ($M = 0.04$, $SD = 0.0137$, *range* = 0.019–0.08), and large initial neighborhood size ($M = 08.06$, $SD = 1.17$, *range* = 6.29–9.46) that dropped to a much lower size ($M = 0.59$, $SD = 0.049$, *range*: 0.5–0.63) by age 25.

To evaluate the overall simulation performance of BiLex, we conducted multiple regression analyses for each language separately. For English, the statistical model included the English simulation composite score (i.e., the average of all five *simulated composite naming scores* for this language) as predictor, the actual *composite naming scores* in English as the dependent variable, and language (whether English is L1 or L2) as a covariate. The same analysis was conducted for Spanish. The regression analysis for English showed that the English simulation composite scores significantly predicted actual *composite naming scores* in English F (2, 30) = 73.68, $p < .0001$, $R^2 = 0.83$ after controlling for language ($b = 1.08$, $SE = 0.10$, $t = 10.72$, $p < .0001$). Likewise, the regression analysis for Spanish revealed that the Spanish simulation composite scores significantly predicted actual *composite naming scores* in Spanish F (2, 30) = 31.9, $p < .0001$, $R^2 = 0.68$ after controlling for language ($b = 0.86$, $SE = 0.19$, $t = 4.39$, $p < .0001$). These results demonstrate that by using EA-optimized parameter sets, BiLex is also able to predict the naming performance of unknown individual participants in each language regardless of whether the language is the native or the second language (Fig. 2).

3.3. Effects of L2 AoA and language exposure on map organization

The best-fit candidate solution identified by the EA allowed the Bilex model to acquire highly organized SOMs in both the semantic and phonetic space reflecting the organization of an adult bilingual lexicon. Appropriate map organization is reflected in its local structure (i.e., units encoding word input representations cluster in semantic and phonetic categories and differentiate well from each other) and global structure (i.e., similar word categories tend to be neighbors and all the map space is used appropriately for representation) (Fig. 3). Importantly, BiLex was able to capture how the representation of words in the L2 phonetic system can be modulated by differences in L2 AoA and language exposure leading to different degrees of L2 naming performance (Fig. 4). More specifically, early AoA and high exposure lead to well-organized L2 phonetic maps and highest naming performance (> 90% accuracy). As long as AoA was early, the L2 phonetic map organized well globally and locally (i.e., words cluster in categories with no overlapping representations) even for relatively low exposure, with still good naming performance (~80% accuracy). Late AoA led to a decreased global organization of the L2 phonetic map and a slightly decreased naming performance even at high exposure (~70% accuracy). Finally, late AoA and low exposure led to a deficient global and local organization of the L2 phonetic map and low naming performance (~40% accuracy).

4. Discussion

The present study sought to examine (i) whether our BiLex computational model can accurately simulate lexical access in picture naming in Spanish-English bilinguals while taking into account their L2

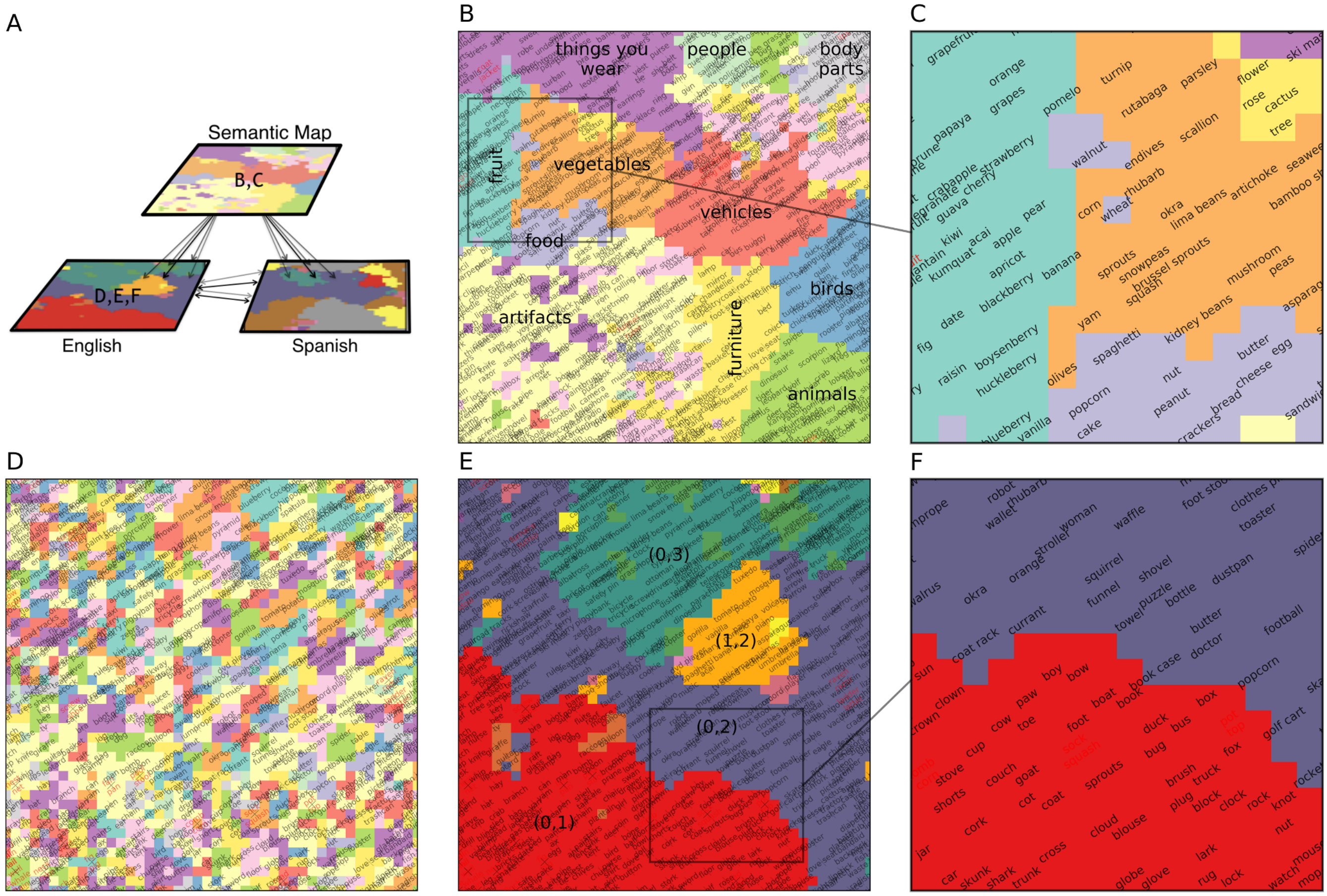

Fig. 3. Organization of the semantic and English phonetic maps after training a BiLex model that simulates naming performance of a highly proficient bilingual. Representative BiLex model that achieved close to 100% naming accuracy in both languages with high exposure and L2 AoA = 0 for both languages. (A) Full architecture of BiLex depicted for reference, with one semantic map and one phonetic map for each language. (B and C) Semantic map: global (B) and local (C) depictions of an entire semantic map after training. Each map unit is colored according to the semantic category associated with the closest word (e.g. the category of "apple" is "fruit"). Words tend to cluster under similar categories and categories tend to be neighbors (e.g. "fruit" is next to "vegetables"). (D–F) English phonetic map after training: (D) full phonetic map colored according to semantic categories for contrast, showing that English words do not form color-coded semantic clusters as in figure (B) because they are organized in space according to their phonetic properties. This panel also shows global (E) and local (F) depictions of the phonetic map this time colored according to rough phonetic categories reflecting the number of syllables before and after the stress of each word (e.g. map units colored purple under the cluster 0,2 encode words with no syllables before and two syllables after the main stress). The Spanish phonetic map organization (not shown in detail) resembles that shown for the English phonetic map (D–F).

AoA, degree of lifetime exposure and current use of each language and (ii) whether the model can predict naming accuracy in participants whose data was not used for model training. Our findings indicate that the BiLex model was able to simulate a wide range of naming performance in all the participants, explaining 83% of the variance in their naming scores in English and 68% in Spanish irrespective of whether the language was L1 or L2. Moreover, the cross-validation approach showed that when using the five best-fit candidate solutions determined by the EAs, BiLex was also able to accurately match the naming performance of other participants whose data were not part of model training, which validates its predictive capacity on naming performance across different individual profiles of bilingualism. These findings align with previous behavioral research suggesting that age of L2 onset and degree of exposure and use of each language modulate lexical access in bilinguals (Hirsh et al., 2003; Kastenbaum et al., 2018; Kohnert et al., 1998) and provide additional evidence that SOM-based computational approaches can contribute to our understanding of how such factors affect bilingual language representation and processing (see Li, 2013; Li & Zhao, 2013 for a review).

The EA approach was useful in determining efficient candidate solutions that by gradually reducing both neighborhood size and learning rate allowed to establish a well-organized model architecture representing a fully developed adult bilingual mental lexicon. It is generally assumed that the semantic and phonetic systems are structured as topographic maps, where concepts are organized spatially according to some degree of similarity (Caramazza, Hillis, & Leek, 1994; Farah & Wallace, 1992). In BiLex, this organization is reflected in the proximity of words with similar features in the semantic and phonetic space, and the observation that similar categories neighbor each other even when information regarding category membership was not part of the training input. It is worth noting that because the semantic representations of words were always activated regardless of the language being trained in each learning instance, the resulting semantic map was well-organized in global and local structure and conceptual knowledge was largely distributed and well-differentiated in the semantic space for all participants. However, the organization of the L2 phonetic map was crucially affected by L2 AoA and its interaction with L2 exposure. For instance, very early L2 AoA led to highly functional and very well-organized maps in terms of their global and local structure resulting in high L2 naming performance. On the other hand, the later the L2 AoA the more L2 exposure was needed for the model to achieve a still functional but less-efficient organization and distribution of representations in the phonetic space, which in turn resulted in already lower degrees of L2 naming performance. At the end of this

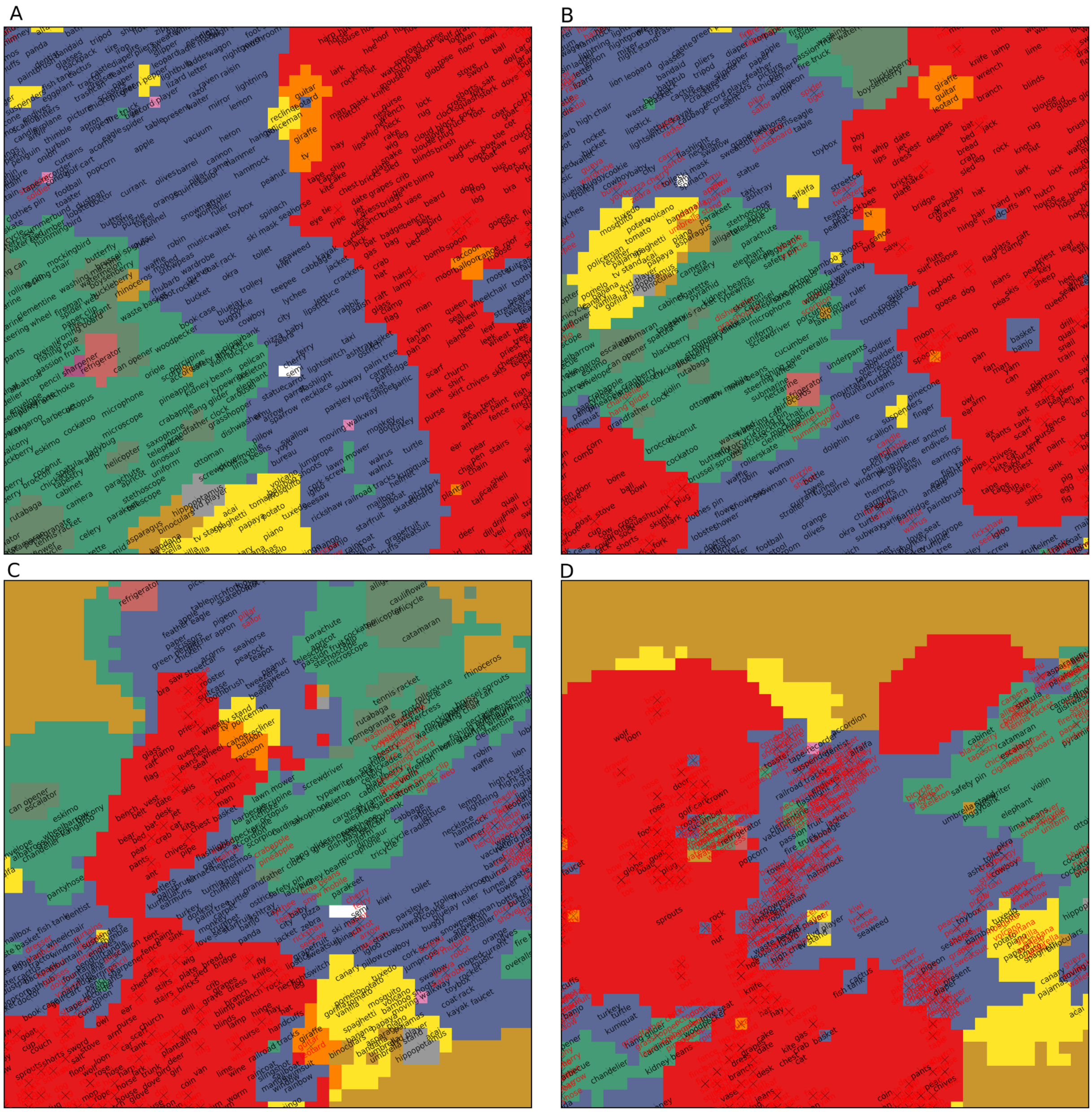

Fig. 4. Effects of L2 AoA and exposure on the global structure of an L2 phonetic map. Depiction of differences in the organization of the L2 English phonetic map across four individual BiLex models simulating naming performance in four different bilinguals across a range of L2 AoA (early–late) and exposure (high-low). (A) Early AoA and high exposure leads to well-organized L2 phonetic maps and high naming performance. (B) Early AoA leads to well-organized maps and high naming performance even for relatively low exposure. (C) Late L2 AoA impacts both global organization of the phonetic map and naming performance even at high exposure. (D) Late AoA and low exposure, lead to deficient global and local map organization (i.e., less map area used with high overlap between word representations) and low naming performance.

spectrum, very late L2 AoA heavily impacted the phonetic map with a poor global and local organization and no amount of exposure being able to aid native-like functionality resulting in decreased L2 naming performance. Models with late L2 AoA had neighborhood sizes and learning rates that started out too small leading to phonetic features of words to be less efficiently distributed in the global structure of the phonetic space and less well differentiated from other representations to the extent that some overlapped in high word density areas of the map. These findings are in line with previous computational accounts of bilingual processing showing that representational structure is highly dependent on L2 AoA (Grasemann et al., 2011; Miikkulainen & Kiran, 2009; Zhao & Li, 2010) with better L2 lexical organization for early as opposed to late L2 learning (Zhao & Li, 2007) and an overall less well-organized lexical network for L2 as compared to L1 even in proficient bilinguals (Borodkin, Kenett, Faust, & Mashal, 2016). Moreover, it has been proposed that this type of overlapping organization of word representations in late L2 AoA can lead to increased difficulty, higher confusion rates and errors in word retrieval due to competition (Zhao & Li, 2007, 2010).

Importantly, a central working assumption underlying the BiLex model is that, similar to an efficient set of training parameters (i.e. candidate solution) necessary to achieve well-organized SOMs, language acquisition during human development requires an equivalent progression of factors governing learning. In other words, the cortical structures that underlie the human lexicon start out highly flexible and adaptive, later in life adapt only to a smaller degree, both in terms of learning intensity and overall flexibility while plasticity and functionality decrease with late L2 learning. Thus, such a process can also provide a mechanistic explanation for age-related limitations on L2 learning and the modulatory effect of language exposure in bilinguals as revealed by previous behavioral research showing that: (i) in early L2 AoA, language exposure and use contribute to receptive and expressive vocabulary growth in the first years of life (Legacy, Zesiger, Friend, & Poulin-Dubois, 2018; Ribot, Hoff, & Burridge, 2017) while proficiency (which is ultimately dependent on language exposure) determines naming performance in adulthood (Hernandez & Kohnert, 1999; Kohnert et al., 1998), (ii) in late L2 AoA naming patterns can approximate those of monolingual speakers with increased L2 use and exposure (Malt, Li, Pavlenko, Zhu, & Ameel, 2015) and (iii) in late L2 AoA but limited L2 exposure, a lower performance in picture naming and verbal fluency can be expected (Bethlehem, de Picciotto, & Watt, 2003; Hernandez & Li, 2007).

It is worth noting that BiLex was able to simulate naming performance reasonably well for individuals above the age of 30, considering that the LUQ collected fine-grained information about language exposure in 3-year intervals up to this age, but included a single interval for the age range "30 years and up". While language exposure can differ later in adulthood, the model accounted for current language use during the last five years of life of each participant to capture more recent language exposure and usage, which may have helped improving the naming simulations for older individuals, especially for those with late L2 AoA. It is also possible that fine-grained exposure metrics from birth to early adulthood are more crucial to simulate naming performance in older individuals who acquired their L2 before the age of 30 as these metrics can better reflect changes in vocabulary due to development, and environmental influences from formal education and working experience. However, the model may have achieved better simulations for older bilinguals if equally fine-grained exposure data for later years in life would have been available as input for model training. Future research will need to extend the LUQ fine-grained assessments of language exposure to the entire adulthood lifespan to test this possibility.

Finally, our model also incorporated small amounts of noise to the associative connections in order to reflect aging and attrition effects on lexical access in older bilinguals immersed in an L2 speaking environment. As in previous work (Grasemann et al., 2011) noise led to decreased performance in the bilingual models thus capturing naming ability also in older bilinguals. Examining the differential effects of aging and attrition in bilingual representation and lexical processing was beyond the scope of this study. However, because most of the attrition literature typically involves older adults and therefore their impact in lexical access could be at least partially confounded (Rossi & Diaz, 2016), adaptations of the BiLex model could potentially help to disentangle the effects of aging from those of attrition in bilingual lexical access across different established L1 and L2 attrition profiles (van Els, 1986). Similarly, computational experiments with BiLex could allow examining the effects of decreased language exposure and use and the influence of the other language as main factors influencing language attrition (Goral, 2004).

The architecture of Bilex used the RHM (Kroll & Stewart, 1994; Kroll et al., 2010) as a reference framework for bilingual word production with special focus on language proficiency. However, our findings do not prove against alternative bilingual models. In fact, BiLex may conform to other relevant models that make similar assumptions about the organization of the bilingual mental lexicon (see French & Jacquet, 2004 for a review). Although beyond the goal of the study, it is also worth considering the architecture of BiLex as a plausibility model of the cortical organization of language in the bilingual brain. The representation of a single semantic system shared across languages in our model is in line with neuroimaging findings of a substantial neural overlap of semantic representations for equivalent words in L1 and L2 in the bilateral occipito-temporal cortex and the anterior temporal lobes (van de Putte, De Baene, Brass, & Duyck, 2017). This last region may operate as a crucial hub for semantic processing as evidenced by lesion studies of semantic dementia (see Lambon Ralph, Jefferies, Patterson, & Rogers, 2016 for a review). However, the separation of L1 and L2 phonetic representations would be expected at the local neuronal level of the cortical maps involved in specific aspects of processing these representations (i.e., audition, articulation, sequential processing, etc.) instead of a strict separation at the higher neuroanatomical level of cortical organization (Hernandez, Li, & MacWhinney, 2005). At this gross neuroanatomical level, neuroimaging studies suggest a large neural overlap for L1 and L2 processing that only extends to additional regions presumably to compensate for the additional cognitive demands related to low L2 proficiency (Sebastian, Laird, & Kiran, 2011), and both languages are typically affected in bilinguals with aphasia presenting lesions in perisylvian language regions (Peñaloza & Kiran, 2019).

Importantly, our findings also suggest potential avenues for future research. For instance, while BiLex was originally designed to simulate lexical access in word production, it can also allow for simulations of word comprehension as the bidirectional connections between SOMS allow to propagate activation from the two phonetic maps to the semantic map. Also, the architecture of BiLex can facilitate the examination of how individual differences in L2 AoA and language exposure modulate cross-language co-activation during lexical access. Moreover, BiLex could further improve its simulation capacity by incorporating cognitive factors that influence lexical access in bilinguals including cognitive control (Green & Abutalebi, 2013) and working memory (Linck, Osthus, Koeth, & Bunting, 2014) and biological factors such as dopamine-related genes known to modulate the effects of AoA and exposure in L2 proficiency (Vaughn & Hernandez, 2018). Finally, BiLex could contribute to both a better understanding of language breakdown and the prediction of treatment-induced recovery in bilingual adults with language dysfunction. Specifically, our findings suggest that BiLex can be used to simulate a healthy bilingual language system that reflects the premorbid language processing abilities of bilinguals with language deficits using their L2 AoA, language exposure and usage as individual training parameters. Such individual BiLex models can be then used in further simulations of language impairment and rehabilitation outcomes that can ultimately help predicting individual response to treatment provided in one language versus the other.

5. Conclusions

The present study provides evidence of the important contribution of computational modeling to the examination of relevant aspects in the language learning history of bilingual speakers that can modulate their lexical processing abilities in each language. Using simulation and cross-validation experiments, we demonstrate that BiLex can offer a computational account of the influence of L2 AoA and language exposure and use on (i) individual differences in lexical access in speakers of different ages and with varying bilingual backgrounds, and (ii) individual variation in the representational structure of the bilingual mental lexicon. Future research based on BiLex can help examining the effects of other relevant factors such as aging and attrition, and further our understanding of lexical access deficits and recovery.

Statement of significance

The present study aimed to develop and validate BiLex, a computational connectionist model of the bilingual lexicon. Here we demonstrate that BiLex is able to account for individual differences in bilingual lexical access by taking into account the L2 age of acquisition and the degree of relative use and exposure to each language in healthy bilingual speakers. Our findings demonstrate that these factors of the language learning history of bilingual speakers can influence the organization of the bilingual mental lexicon and its functionality in word retrieval. Importantly, they set the stage for continued work on the impact of lesions and rehabilitation on bilingual word retrieval.

Acknowledgments

This work was supported by the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health [grant U01DC014922] awarded to Swathi Kiran. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors would like to thank all the participants involved in this study. We also thank Marina Calleja and Bianelkis Ramos for their assistance with participant recruitment and Leela Rao for help with data collection.

Declaration of Competing Interest

Swathi Kiran serves as a consultant for The Learning Corporation with no scientific overlap with the present study. Risto Miikkulainen serves as vice-president for Sentient Inc. with no scientific overlap with the present study.

Appendix A

A.1. Optimization of global training parameters

Minimum exposure parameter ε. Because monolingual participants tended to show above zero naming performance in their non-native language, a minimum exposure parameter ε was added to allow the model to adapt to these data. During model training, the exposure percentage for each language was then clipped to values between ε and (1- ε).

Learning rate and neighborhood size. To minimize potential data overfitting (i.e., by allowing the EA to evolve too fine-grained encodings in limited participant data across training samples), both learning rate α and neighborhood size σ were encoded as piecewise-linear functions of time, i.e. their specific values were evolved at a small number of simulated ages (1, 4, 7, 10, 13, 19, 25, and 50) and interpolated linearly for intermediate values.

Associative connections. Similar to the SOMs, the associative connections also required a learning rate at each time during training. A single factor k was added to limit the overall number of evolved global training parameters, such that at each time the learning rate used for associative connections was $\alpha' = k \times \alpha$. In this way, the scale of the learning rate for associative connections was independent of that for SOMs, but changed in the same way over time.

Minimum word frequency. The word frequency training parameter consisting in the minimum word frequency used for the rarest word in the corpus was included in order for the model to simulate differences in word frequencies between the BNT and the naming screener.

Number of words trained per simulated year. Although initially an evolved parameter, the number of words trained per simulated year was fixed to 1.5 as early experiments showed that any value above 1.5 × corpus size allowed well-organized maps and accurate naming simulations.

Appendix B

See Table A.

Table A
Global training parameters of the five best-fit candidate solutions optimized by the EA.

Parameter	Candidate solution 1	Candidate solution 2	Candidate solution 3	Candidate solution 4	Candidate solution 5
α1	0.2714	0.4364	0.0933	0.3302	0.3200
α4	0.1179	0.1297	0.2497	0.3286	0.2415
α7	0.1179	0.1297	0.2337	0.3286	0.2415
α10	0.1179	0.1297	0.2337	0.1956	0.2415
α13	0.1179	0.1297	0.2245	0.1956	0.2371
α19	0.1179	0.1297	0.2245	0.1956	0.2371
α25	0.1179	0.1249	0.1071	0.1956	0.1511
α50	0.0775	0.0451	0.0670	0.0574	0.1189
σ1	7.8035	9.4620	6.2886	8.1986	8.5671
σ4	5.8057	4.1175	1.7832	6.6013	4.8228
σ7	2.3171	3.4691	1.7832	1.9787	4.8228
σ10	1.0063	1.8927	1.7832	1.9787	0.6296
σ13	0.6166	0.5001	0.6264	1.2336	0.6296
σ19	0.6166	0.5001	0.6264	0.5714	0.6296
σ25	0.6166	0.5001	0.6264	0.5714	0.6296
σ50	0.6166	0.5001	0.6264	0.3589	0.6296
Associated α factor	1.5545	1.3324	0.8664	1.5571	1.9174
Minimal exposure	0.0258	0.0587	0.0365	0.0215	0.0317
Noise *SD*	0.0096	0.0082	0.0059	0.0004	0.0033
Min. word frequency	0.2671	0.2558	0.4932	0.0254	0.5071

EA = evolutionary algorithm; α = learning rate; σ = neighborhood size; *SD* = standard deviation; Min. word frequency = minimal word frequency.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chn, Z., Citro, C., … & Ghemawat, S. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. URL http://tensorflow.org/. Software available from tensorflow.org.

Bäck, T. (1996). *Evolutionary algorithms in theory and practice.* New York: Oxford University Press.

Baus, C., Costa, A., & Carreiras, M. (2013). On the effects of second language immersion on first language production. *Acta Psychologica, 142*(3), 402–409. https://doi.org/10.1016/j.actpsy.2013.01.010.

Bethlehem, D., de Picciotto, J., & Watt, N. (2003). assessment of verbal fluency in bilingual zulu-english speakers. *South African Journal of Psychology, 33*(4), 236–240. https://doi.org/10.1177/008124630303300406.

Birdsong, D. (2018). Plasticity, variability and age in second language acquisition and bilingualism. *Frontiers in Psychology, 9*, 81. https://doi.org/10.3389/fpsyg.2018.00081.

Borodkin, K., Kenett, Y. N., Faust, M., & Mashal, N. (2016). When pumpkin is closer to onion than to squash: The structure of the second language lexicon. *Cognition, 156*, 60–70. https://doi.org/10.1016/j.cognition.2016.07.014.

Caramazza, A., Hillis, A., & Leek, E. C. (1994). The organization of lexical knowledge in the brain: Evidence from category- and modality-specific deficits. In L. A. Hirschfeld, & S. A. Gelman (Eds.). *Mapping the mind: Domain specificity in cognition and culture* (pp. 68–84). New York, NY, US: Cambridge University Press. https://doi.org/10.1017/CBO9780511752902.004.

Colomé, À. (2001). Lexical activation in bilinguals' speech production: Language-specific or language-independent? *Journal of Memory and Language, 45*(4), 721–736. https://doi.org/10.1006/jmla.2001.2793.

Connor, L. T., Spiro, A., Obler, L. K., & Albert, M. L. (2004). Change in object naming ability during adulthood. *The Journals of Gerontology: Series B, 59*(5), P203–P209. https://doi.org/10.1093/geronb/59.5.P203.

Costa, A. (2005). Lexical access in bilingual production. In J. F. Kroll, & A. M. B. De Groot (Eds.). *Handbook of bilingualism: Psycholinguistic approaches* (pp. 308–325). New York: Oxford University Press.

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(5), 1283–1296. https://doi.org/10.1037/0278-7393.26.5.1283.

Costa, A., & Sebastián-Gallés, N. (2014). How does the bilingual experience sculpt the brain? *Nature Reviews Neuroscience, 15*, 336.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*(3), 283–321. https://doi.org/10.1037/0033-295X.93.3.283.

Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., & Carreiras, M. (2013). EsPal: One-stop shopping for Spanish word properties. *Behavior Research Methods, 45*(4), 1246–1258. https://doi.org/10.3758/s13428-013-0326-1.

Fang, S. Y., Zinszer, B. D., Malt, B. C., & Li, P. (2016). Bilingual object naming: A connectionist model. *Frontiers in Psychology, 7*, 644. https://doi.org/10.3389/fpsyg.2016.00644.

Farah, M. J., & Wallace, M. A. (1992). Semantically-bounded anomia: Implications for the neural implementation of naming. *Neuropsychologia, 30*(7), 609–621. https://doi.org/10.1016/0028-3932(92)90066-U.

French, R. M., & Jacquet, M. (2004). Understanding bilingual memory: Models and data. *Trends in Cognitive Sciences, 8*(2), 87–93. https://doi.org/10.1016/j.tics.2003.12.011.

Fricke, M., Zirnstein, M., Navarro-Torres, C., & Kroll, J. F. (2018). Bilingualism reveals fundamental variation in language processing. *Bilingualism: Language and Cognition,* 1–8. https://doi.org/10.1017/S1366728918000482.

Gollan, T. H., & Kroll, J. F. (2001). Bilingual lexical access. In B. Rapp (Ed.). *The handbook of cognitive neuropsychology: What deficits reveal about the human mind* (pp. 321–345). New York: Psychology Press.

Goral, M. (2004). First-language decline in healthy aging: Implications for attrition in bilingualism. *Attrition, 17*(1), 31–52. https://doi.org/10.1016/S0911-6044(03)00052-6.

Grasemann, U., Peñaloza, C., Kiran, S., & Miikkulainen, R. (2019). Evolutionary model-fitting for a neural network model of lexical access in bilinguals. In *Proceedings of the 17th International Conference on Cognitive Modeling (ICCM 2019)* (in press).

Grasemann, U., Sandberg, C., Kiran, S., & Miikkulainen, R. (2011). Impairment and rehabilitation in bilingual aphasia: A SOM-based model. In J. Laaksonen, & T. Honkela (Eds.). *Advances in self-organizing maps* (pp. 207–217). Berlin Heidelberg: Springer.

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology, 25*(5), 515–530. https://doi.org/10.1080/20445911.2013.796377.

Hernandez, A. E., & Kohnert, K. J. (1999). Aging and language switching in bilinguals. *Aging, Neuropsychology, and Cognition, 6*(2), 69–83. https://doi.org/10.1076/anec.6.2.69.783.

Hernandez, A. E., & Li, P. (2007). Age of acquisition: Its neural and computational mechanisms. *Psychological Bulletin, 133*(4), 638–650. https://doi.org/10.1037/0033-2909.133.4.638.

Hernandez, A., Li, P., & MacWhinney, B. (2005). The emergence of competing modules in bilingualism. *Trends in Cognitive Sciences, 9*(5), 220–225. https://doi.org/10.1016/j.tics.2005.03.003.

Hirsh, K. W., Morrison, C. M., Gaset, S., & Carnicer, E. (2003). Age of acquisition and speech production in L2. *Bilingualism: Language and Cognition, 6*(2), 117–128. https://doi.org/10.1017/S136672890300107X.

Juncos-Rabadán, O. (1994). The assessment of bilingualism in normal aging with the bilingual aphasia test. *Journal of Neurolinguistics, 8*(1), 67–73. https://doi.org/10.1016/0911-6044(94)90008-6.

Kaplan, E., Goodglass, H., & Weintraub, S. (2001). *The Boston naming test.* Philadelphia: Lippincott, Williams & Wilkins.

Kastenbaum, J. G., Bedore, L. M., Peña, E. D., Sheng, L., Mavis, I., Sebastian-Vaytadden, R., ... Kiran, S. (2018). The influence of proficiency and language combination on bilingual lexical access. *Bilingualism: Language and Cognition, 1–31*. https://doi.org/10.1017/S1366728918000366.

Kavé, G., Knafo, A., & Gilboa, A. (2010). The rise and fall of word retrieval across the lifespan. *Psychology and Aging, 25*(3), 719–724. https://doi.org/10.1037/a0018927.

Kiran, S., Grasemann, U., Sandberg, C., & Miikkulainen, R. (2013). A computational account of bilingual aphasia rehabilitation. *Bilingualism: Language and Cognition, 16*(2), 325–342. https://doi.org/10.1017/S1366728912000533.

Kohnert, K. J., Hernandez, A. E., & Bates, E. (1998). Bilingual performance on the Boston Naming Test: Preliminary norms in Spanish and English. *Brain and Language, 65*(3), 422–440. https://doi.org/10.1006/brln.1998.2001.

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

Kreiner, H., & Degani, T. (2015). Tip-of-the-tongue in a second language: The effects of brief first-language exposure and long-term use. *Cognition, 137*, 106–114. https://doi.org/10.1016/j.cognition.2014.12.011.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 33*(2), 149–174. https://doi.org/10.1006/jmla.1994.1008.

Kroll, J. F., Van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The revised hierarchical model: A critical review and assessment. *Bilingualism: Language and Cognition, 13*(3), 373–381. https://doi.org/10.1017/S136672891000009X.

Lambon Ralph, M. A., Jefferies, E., Patterson, K., & Rogers, T. T. (2016). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience, 18*(1), 42–55. https://doi.org/10.1038/nrn.2016.150.

Legacy, J., Zesiger, P., Friend, M., & Poulin-Dubois, D. (2018). Vocabulary size and speed of word recognition in very young French-English bilinguals: A longitudinal study. *Bilingualism: Language and Cognition, 21*(1), 137–149. https://doi.org/10.1017/S1366728916000833.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(1), 1–38.

Li, P. (2013). Computational modeling of bilingualism: How can models tell us more about the bilingual mind? *Bilingualism: Language and Cognition, 16*(2), 241–245. https://doi.org/10.1017/S1366728913000059.

Li, P., & Farkas, I. (2002). A self-organizing connectionist model of bilingual processing. In Bilingual sentence processing. (pp. 59–85). Amsterdam, Netherlands: North-Holland/Elsevier Science Publishers. https://doi.org/10.1016/S0166-4115(02)80006-1.

Li, P., & Zhao, X. (2013). Self-organizing map models of language acquisition. *Frontiers in Psychology, 4*, 828. https://doi.org/10.3389/fpsyg.2013.00828.

Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning. *Psychological Science, 20*(12), 1507–1515. https://doi.org/10.1111/j.1467-9280.2009.02480.x.

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review, 21*(4), 861–883. https://doi.org/10.3758/s13423-013-0565-2.

Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology, 25*(5), 605–621. https://doi.org/10.1080/20445911.2013.795574.

Malt, B. C., Li, P., Pavlenko, A., Zhu, H., & Ameel, E. (2015). Bidirectional lexical interaction in late immersed Mandarin-English bilinguals. *Journal of Memory and Language, 82*, 86–104. https://doi.org/10.1016/j.jml.2015.03.001.

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLOS ONE, 7*(8), e43230. https://doi.org/10.1371/journal.pone.0043230.

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

Miikkulainen, R., & Kiran, S. (2009). Modeling the bilingual lexicon of an individual subject. In J. C. Príncipe, & R. Miikkulainen (Eds.). *Advances in self-organizing maps* (pp. 191–199). Berlin Heidelberg: Springer.

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(5), 1170–1187. https://doi.org/10.1037/0278-7393.26.5.1170.

Peñaloza, C., & Kiran, S. (2019). Recovery patterns in multilingual aphasia. In J. W. Schwieter (Ed.). *The handbook of the neuroscience of multilingualism* (pp. 553–571). New Jersey: Wiley-Blackwell Publishing.

Ribot, K. M., Hoff, E., & Burridge, A. (2017). Language use contributes to expressive language growth: Evidence from bilingual children. *Child Development, 89*(3), 929–940. https://doi.org/10.1111/cdev.12770.

Rossi, E., & Diaz, M. (2016). How aging and bilingualism influence language processing: Theoretical and neural models. *Linguistic Approaches to Bilingualism, 6*(1), 9–42. https://doi.org/10.1075/lab.14029.ros.

Salthouse, T. A. (2003). Memory aging from 18 to 80. *Alzheimer Disease & Associated Disorders, 17*(3), 162–167.

Sandberg, C. W., Gray, T., & Kiran, S. (2018). *Development of a free online interactive naming therapy for bilingual aphasia.* Boston, MA: American Speech Language Hearing Association Convention.

Sebastian, R., Laird, A. R., & Kiran, S. (2011). Meta-analysis of the neural representation of first language and second language. *Applied Psycholinguistics, 32*(4), 799–819. https://doi.org/10.1017/S0142716411000075.

Schmid, M. S. (2010). Languages at play: The relevance of L1 attrition to the study of bilingualism. *Bilingualism: Language and Cognition, 13*(1), 1–7. https://doi.org/10.1017/S1366728909990368.

Shahrzad, H. Hodjat, B., & Miikkulainen, R. (2016). Estimating the advantage of age-layering in evolutionary algorithms. In T. Friedrich (Ed.), Proceedings of the genetic and evolutionary computation conference 2016 (GECCO 2016) (pp. 693–699). New York: ACM. https://doi.org/10.1145/2908812.2908911.

Shook, A., & Marian, V. (2013). The bilingual language interaction network for comprehension of speech. *Bilingualism: Language and Cognition, 16*(2), 304–324. https://doi.org/10.1017/S1366728912000466.

Tu, L., Wang, J., Abutalebi, J., Jiang, B., Pan, X., Li, M., ... Huang, R. (2015). Language exposure induced neuroplasticity in the bilingual brain: A follow-up fMRI study. *Cortex, 64*, 8–19. https://doi.org/10.1016/j.cortex.2014.09.019.

van de Putte, E., De Baene, W., Brass, M., & Duyck, W. (2017). Neural overlap of L1 and L2 semantic representations in speech: A decoding approach. *NeuroImage, 162*, 106–116. https://doi.org/10.1016/j.neuroimage.2017.08.082.

van Els, T. (1986). An overview of European research on language attrition. In B. Weltens, K. de Bot, & T. Van Els (Eds.). *Language attrition in progress* (pp. 3–18). Dordrecht: Foris.

van Hell Janet, G., & Darren, T. (2012). Second language proficiency and cross-language lexical activation. *Language Learning, 62*(s2), 148–171. https://doi.org/10.1111/j.1467-9922.2012.00710.x.

Vaughn, K. A., & Hernandez, A. E. (2018). Becoming a balanced, proficient bilingual: Predictions from age of acquisition & genetic background. *Journal of Neurolinguistics, 46*, 69–77. https://doi.org/10.1016/j.jneuroling.2017.12.012.

Zhao, X., & Li, P. (2007). *Bilingual lexical representation in a self-organizing neural network model. Proceedings of the annual meeting of the cognitive science society*.

Zhao, X., & Li, P. (2010). Bilingual lexical interactions in an unsupervised neural network model. *International Journal of Bilingual Education and Bilingualism, 13*(5), 505–524. https://doi.org/10.1080/13670050.2010.488284.

Zhao, X., & Li, P. (2013). Simulating cross-language priming with a dynamic computational model of the lexicon. *Bilingualism: Language and Cognition, 16*(2), 288–303. https://doi.org/10.1017/S1366728912000624.