LANGUAGE LEARNING
A Journal of Research in Language Studies
Check for updates
Language Learning ISSN 0023-8333

CONCEPTUAL REVIEW ARTICLE

Computational Modeling of Bilingual Language Learning: Current Models and Future Directions

Ping Li[a] and Qihui Xu[b]
[a]The Hong Kong Polytechnic University [b]City University of New York

Abstract: The last two decades have seen a significant amount of interest in bilingual language learning and processing. A number of computational models have also been developed to account for bilingualism, with varying degrees of success. In this article, we first briefly introduce the significance of computational approaches to bilingual language learning, along with a discussion of the major contributions of current models, their implications, and their limitations. We show that the current models have contributed to progress in understanding the bilingual mind, but significant gaps exist. We advocate a new research agenda integrating progress across different disciplines, such as computational neuroscience, natural language processing, and first language acquisition, to construct a pluralist computational account that combines high-level cognitive theories and neurobiological foundations for bilingual language learning. We outline the contributions and promises of this interdisciplinary approach in which we view bilingual language learning as a dynamic, interactive, and developmental process.

Keywords computational modeling; bilingualism; language learning

---

A one-page Accessible Summary of this article in non-technical language is freely available in the Supporting Information online and at https://oasis-database.org

Preparation of this article has been partially supported by a grant from the Hong Kong Research Grants Council (Project #PolyU15601520), a Research Startup Fund from the Hong Kong Polytechnic University, and a Sin Wai Kin Foundation grant provided to Ping Li. Qihui Xu has been supported by a Graduate Research Fellowship from the City University of New York. The authors wish to thank Jing Wang for her assistance in preparing the final version of the manuscript.

Correspondence concerning this article should be addressed to Ping Li, Department of Chinese and Bilingual Studies, Faculty of Humanities, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong SAR, China. Email: pi2li@polyu.edu.hk

The handling editor for this article was Scott Crossley.

DOI: 10.1111/lang.12529

Introduction

Computational methods and models have been widely used to study all natural phenomena and human behaviors, from modeling climate change to simulating brain interactions. Language learning, as a hallmark of human ability, has been considered as an important computational process, and therefore the understanding of this topic will also necessarily benefit from computational modeling (Meltzoff et al., 2009). A large amount of computational work has accumulated in the last decades, including the study of infant speech perception, first language (L1) learning, second language (L2) representation, and bilingual processing (e.g., Dijkstra et al., 2019; Li & Farkas, 2002; Saffran et al., 1996; Xu & Tenenbaum, 2007). The computational methodologies used have ranged from statistical learning to connectionist modeling to network analyses. The recent upsurge of interest in artificial intelligence, machine learning, and natural language processing (NLP) will only further accelerate the development and application of computational approaches to language learning research. In this article, we provide a synthesis of computational approaches with specific reference to bilingual language learning[1] and point to some exciting new directions that the field may pursue in future research.

The Importance of Computational Modeling

The cognitive revolution that started in the 1950s was based on both technological development in computing machines and theoretical thinking about how the human mind might work like a computer (for historic reviews, see H. Gardner, 1987; Leahey, 2004). The key to using the computer as a metaphor for the human mind lies in the ability of computing systems to process information (i.e., take input and use it to generate desired output). Since those beginnings, computational approaches and methodologies have flourished in all areas of investigation into human cognition, including the study of human language.

A number of important features make computational models particularly relevant in the context of language learning. First, “verbal models” that are based on box-and-arrow representations in classical cognitive psychology theories remain highly abstract and often do not lead to mechanistic accounts of the underlying processes or principles for a given behavior (Kriegeskorte & Douglas, 2018). Computational models force researchers to be explicit and specific about their research, including basic concepts, assumptions, and hypotheses that must be implementable in quantitative and algorithmic terms (e.g., “similarity” of two concepts and “association” between the concepts can be measured through vector spaces). Such specificity allows researchers to test theoretical hypotheses more rigorously and produce predictions more explicitly, satisfying the replicability criterion of the scientific process (Li, 2013).

Second, computational models allow researchers to explore cognitive processes through manipulating and testing specific variables while holding other potentially confounding variables constant, when the natural environment of concern may be too complex or confounding to allow for multiple variables of interest to be separated in a systematic way. In many cases it is difficult or impossible to directly manipulate these variables in empirical studies through parametric variations (e.g., to orthogonally cross all levels of one variable with all levels of another variable). In the case of language learning, for example, it is neither practical nor feasible to control the amount of input that a learner should receive at any given time when the researcher wants to examine the effect of input quantity only, whereas it is easy to vary and control the input to be provided to a computational model for training, in terms of the amount of words and sentences, as well as the characteristics of such input (e.g., length, type, and token frequency). As another example for bilingual language learning, age of acquisition (AoA) is often confounded with L2 proficiency; for example, in the well-known study by Kim et al. (1997), the early learners were also more proficient in their L2, thus causing doubt about the authors' claim that early learners were using different neural systems from late learners to handle their two languages. Computational models can systematically tease apart AoA and proficiency, for example, by introducing the L2 at different points in the model's learning of the L1, having the model learn different amount of L1 and L2, or having the model start with the L1 and L2 at different stages but training them with equal amount of data (Li, 2009; see later discussion). Thus, computational modeling allows researchers to manipulate variables of interest more flexibly and to study their interactions in a more systematic way, making it a particularly useful tool for researchers to study bilingual language learning with regard to manipulating the two languages in terms of L2 versus L1 onset time, amount of L2 versus L1 input, the order of learning in the two languages, and the frequency of items in each language.

Third, and perhaps more important for cognitive scientists, computational approaches allow us to explain the observed outcome through probing into the underlying processes that lead to the outcome. Researchers using verbal models of cognition often look at the relations between input (e.g., target material) and output (e.g., outcome of learning) and draw inferences about the internal processes involved; by contrast, computational models enable the internal representations to be examined directly, and across different stages (a "lifespan") of information processing and learning. For example, Elman (1990) used

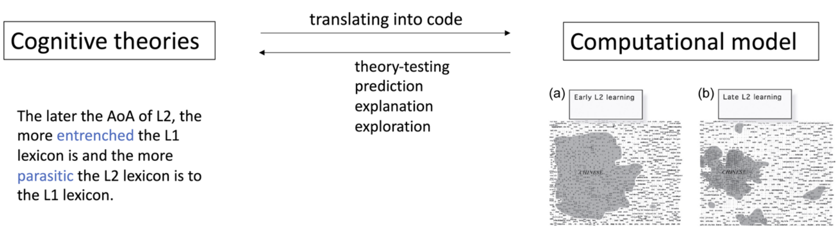


Figure 1 How computational models can shed light on cognitive theories, with an example of using DevLex-II (Li et al., 2007; see also Figure 2 and discussion in a later section) to model the competition model of bilingual language learning (Bates & MacWhinney, 1982).

hierarchical clustering analysis to probe into the underlying structure that an artificial neural network develops in the internal representation, showing that categories such as nouns, verbs, and adjectives emerge at certain stages of training of the network. In the case of learning an L2, understanding the internal processes is even more challenging because of the complex interplay between the two languages and the impossibility of systematically manipulating and controlling for all potentially confounding variables. Computational models and newer methods of data analysis make accessible and visible the internal representation and its gradual change and development (see further discussion in later sections). Figure 1 provides an overview of the value of computational models in shedding light on cognitive theories.

What Makes a Good Computational Model?

There are a number of criteria with which we can assess the success, utility, and power of computational models. The following discussion addresses several that are particularly relevant to language learning, although these are by no means exhaustive.

Validity

A computational model for language should first have validity. This means that the model itself should be tractable and psychologically plausible. The model should be configured such that the relevant parameters can be correspondingly adjusted, the size of the model is appropriate and relevant to the task, the training of the model is measurable in terms of success or failure, and the amount of training time is as practicable as possible, when measured against the parameter and size of the model, along with considerations of the researcher's own timeframe and resources. Further, the model should be configured as psychologically reasonable for human behavior and to perform tasks that can be matched to real learning tasks facing human learners. The validity is also related to the degree to which the model architecture and training can make direct contact with real-world situations in terms of the method, specific material, and manner in which people learn and use languages, which brings us to the next criterion.

Contact with Real Language

In many computational models, especially earlier ones, researchers use highly simplified or "synthetic" patterns to represent linguistic material. In the so-called "localist" representation, one unit or dimension in a vector can be flipped on, representing one sound or one word from a language, such that "01" can be used to represent one sound or one word and "10" another sound or word. This method contrasts with "distributed" representation in computational models, in which multiple units can be turned on (e.g., "001" can represent one word and "101" another word). For example, French (1998) used a localist representation to represent 24 words (12 for English and 12 for French), such that the vector contains 24 units and only one unit is turned on to indicate one of the 24 words (e.g., BOY = 100000000000000000000000, GIRL = 010000000000000000000000, MAN = 001000000000000000000000). Although such representations are easy to construct and can greatly simplify the modeling process and interpretation, it raises the issue of how accurately such input represents the complexity of language properties. We will provide examples of how computational modelers have attempted to solve this problem in the section on current models.

Interpretability

Simply simulating a behavior and reproducing the observed pattern from empirical studies is not the goal of modern computational modeling. Researchers need to understand the empirical patterns that have been identified and the hypotheses that have been formulated in the literature, and they should not be satisfied with a model's ability to simply reproduce the data or fit the data. As discussed previously, computational models aim at revealing the underlying cognitive processes and neural mechanisms that may lead to observable patterns. Thus, our modeling results must be interpretable within the framework of cognitive and linguistic theories, mechanisms, and principles. An important point to bear in mind in this regard is that computational models need both to be informed by empirical data and also to inform empirical studies. To achieve this goal, researchers should make predictions as to what the data would look like under one hypothesis versus another, thereby effectively evaluating competing hypotheses by systematically varying the parameters, size, configuration, and architecture of the model.

Predictive Power

In addition to the ability to describe the phenomenon in question, computational models should also be able to explain the underlying processes, principles, and mechanisms that lead to the observed input–output relations, and further to predict, based on the descriptions and explanations, what might occur in behavior and cognition downstream. For example, Frost et al. (2019) argued that statistical learning models should attempt to describe, explain, and predict empirical phenomena based on the interaction between learning/processing principles and statistical properties. Thus, a good model should be evaluated not only against empirical data (i.e., on how well it displays patterns observed in empirical studies), but also against its ability to generate testable predictions and new ideas, thereby inspiring future research. It has long been argued that good computational models should not merely simulate empirical data but should go beyond this to guide the design of experiments to collect empirical data (Li, 2013; McClelland, 2009; see further discussion of top-down vs. bottom-up approaches below). There will be some cases where empirical data have not been obtained or cannot be obtained; for example, one cannot go back to a patient's prelesion condition, and computational models may be especially helpful in creating and simulating such conditions, thereby informing both the design and conceptualization of an empirical study. If computational modeling of bilingualism can achieve this predictive power by testing different conditions and generating different hypotheses, it will help to provide a forum for inspiring new studies and novel ideas as well as formulating new theories. We return to the limitations of current bilingual learning models in the next section, assessing them against the above-mentioned features of what makes a good computational model.

Current Models and Their Implications

Because the goal of this article is not to give a comprehensive review, we refer readers to Li and Zhao (2018) for a review of computational models with respect to language research in general, and Shirai (2019) for a review of computational models in the field of L2 acquisition. Instead, in this section, we discuss only a few prominent computational models to illustrate key points about what makes them useful and pertinent for understanding bilingual language learning. These models demonstrate, on the one hand, the importance, utility, and predictive power of computational models in general, and on the other, the specific contribution such models offer to understanding the principles and mechanisms of bilingual language learning, beyond the contributions made by verbal models of bilingualism.

Connectionism and Bilingual Models of Representation

A connectionist model, whose architecture is inspired by neural networks in the human brain, is usually a network containing large numbers of interconnected *units* or "neurons." Connectionist models flourished in the 1980s, partly due to the recognition that the classical cognitive theories were based on the wrong computer metaphor, that of the digital serial processing of information (W. A. Gardner, 1984), whereas the human brain's information processing is massively *parallel* (with multiple neurons working simultaneously) and *distributed* (with multiple neurons forming a particular distributed pattern in response to given information; hence the term *parallel distributed processing* for connectionism). In the last 30 years, connectionism has become one of the most influential theoretical frameworks and analytic approaches for understanding human language learning as well as cognitive and linguistic behaviors in general. Table 1 provides brief explanations of basic concepts frequently used in connectionist models of language. Given that our article is not a technical overview of connectionist models, we refer readers to Goldberg's (2017) comprehensive introduction to neural networks in language processing.

A primary principle of connectionism is that there are multiple *units*—"neurons," or "nodes" that represent artificial neurons—interacting with one another to support information processing. For example, in the bilingual interactive activation (BIA) model (van Heuven et al., 1998), there are dedicated neurons at different levels to process visual features (e.g., vertical vs. horizontal bars), letters (e.g., *T*), and words (e.g., *Tom*) in two languages. The interactions are realized through the connections between a large assembly of neurons: When the right nodes are active due to the interactions, the relevant features, letters, or words become recognized, hence the interactive activation mechanism. In addition to the interactions within a language, there are also interactions across languages, simulating the effects of facilitation or interference from one language to the other. The BIA+ model extended the original BIA model by incorporating semantic and phonological representations in the word identification system, linguistic and nonlinguistic context effects, and a task-decision component, but similarly to BIA, it uses the basic interactive activation mechanism for modeling. A recent follow-up that further extends BIA and BIA+ is Multilink (Dijkstra et al., 2019). This model also provides a

Table 1 Concepts that frequently appear in connectionist models

Category	Term	Brief explanation
Elements	Units	Components representing artificial neurons in a connectionist model. Also called “neurons” or “nodes.”
	Weights	Values indicating how strongly multiple nodes in a neural network are connected with one another.
	Hidden units	Components comprising the layers of units used to represent intermediate states of input-to-output relations.
	Attention	A part of a neural architecture that can highlight relevant features of the input data, mimicking attention in human cognition.
Learning algorithms	Back-propagation	An algorithm used for training neural networks, which optimizes model prediction by computing the error between model output and the gold standard; this process iterates backward from the output layer to update weights.
	Supervised learning	A learning approach in which a model is trained for predicting human-labeled data where the output is predetermined or known.
	Unsupervised learning	A learning approach in which a model discovers hidden patterns in unlabeled datasets.
	Hebbian learning	A biologically inspired mechanism that adjust weights or neuronal efficiency according to the cooccurrence of neurons: neurons that fire together wire together (Hebb, 1949, p.70).

(Continued)

Table 1a Concepts that frequently appear in connectionist models

Category	Term	Brief explanation
Model types	Feedforward neural network	A class of neural networks where information moves in one direction from the input units to the output units.
	Self-organizing map	A class of neural networks where nodes are organized on a map like structure that captures the similarities in the input.
	Recurrent neural network (RNN)	A class of neural networks where information moves through consecutive hidden states; outputs from the previous state are used as inputs to the current state.
	Long short-term memory (LSTM) network	An advanced type of RNN that is good at storing past data in memory.
	Transformer	A class of attention-based neural networks where the significance of each part of the input data is differentially weighted by the attention mechanism.

much larger, scalable lexicon from the bilingual's two languages, including not only four-letter-word lexicons but also three-to-eight-letter words from English and Dutch. By considering the role of variables such as word frequency, word length, orthographic similarity, and phonological neighborhood, Multilink is poised to test and verify new empirical data from bilingual word recognition (e.g., simulating spreading activation from multiple lexicons).

Connectionist approaches to language and bilingualism have expanded far beyond the BIA, BIA+, and Multilink models pioneered by Dijkstra and colleagues since the 1990s, in part because of two directions of development: (a) the rapid development of new architectures and algorithms of connectionism, moving from interactive activation models to *feedforward neural networks* with multiple *hidden units* (Rumelhart et al., 1986) and to *recurrent neural networks* (RNNs) that can capture dynamic memory and information processing (Elman, 1990), and (b) new theoretical thinking about which variables are significant for successful bilingual language learning. Below we focus on how researchers can use the conceptual insights and learning mechanisms from connectionism to study important issues in bilingual language learning.

Bilingual Models Involving Dynamic Changes

Whereas models of bilingual representation such as BIA provided an early impetus to researchers in bilingualism, other models have focused on taking advantage of connectionist architectures to examine a number of important bilingual constructs, including L2 proficiency, AoA, and most importantly, the interaction between a bilingual's two linguistic systems. To understand the dynamic interaction between L1 and L2 in the process of learning has become a key goal of some computational models of bilingualism, as will become clear in this section.

Modeling the Emergence of Bilingual Lexicons

The BIA and BIA+ models were designed to account for proficient bilingual speakers' lexical knowledge and processing, but how different levels of proficiency might modulate lexical processes was an issue not examined in these models. Thomas (1997) developed the bilingual single network (BSN) to address this issue. The BSN uses a three-layer neural network with the *back-propagation* algorithm (Rumelhart et al., 1986) to transform a word's orthography (input nodes) to a word's semantic representation (output nodes) through the network's hidden units. The model simulated different levels of proficiency through different amount of training: either balanced, with equal amount of training, or unbalanced, with the L1 being trained three times as often as the

L2. An interesting finding is that the balanced training led to distinct internal representations for the L1 and L2 in the activation patterns of the hidden units, whereas the unbalanced training led to less clearly represented L2 words. The BSN model thus is a simple demonstration of how feedforward connectionist models can capture the role of the amount of training (and "proficiency") in modulating the quality of bilingual lexical representation within the same single network.

A critical notion in connectionist networks is *emergentism*, according to which higher level cognitive representations emerge naturally as a result of lower level simpler processes such as the interaction between a large number of processing neurons in a network (for a recent volume on emergentist approaches to language, see MacWhinney et al., 2022). The simple recurrent network (SRN; Elman, 1990) capitalizes on a task of predicting the next word in a string of words in a sentence to reflect structural linguistic properties; for example, "nouns" as a category can emerge in the recurrent hidden units through the network's learning of the statistical patterns of cooccurrence in corpus data (e.g., occurring in the same slots in a sentence). Relying on SRN mechanisms that were linguistically and psycholinguistically plausible, French and colleagues developed a bilingual SRN model (BSRN; French, 1998; see also French & Jacquet, 2004) to simulate the emergence of distinct patterns of representation as a result of learning sentences from two languages: Sentences from the L1 and L2 are mixed at different ratios such as bilingual learners would be exposed to in different learning contexts, based on which the BSRN model develops distinct linguistic categories from the two languages.

A major issue with previous models, including the BSN and BSRN, is that they have mostly used highly simplified patterns to represent linguistic input. As pointed out earlier, such patterns may not realistically reflect the actual input and linguistic material that L2 learners are acquiring. Several studies have considered how to realistically and faithfully represent the linguistic input or input features, in terms of deriving computational representations from corpora (e.g., Zhao et al., 2011) and from linguistic features in multiple languages (Li & MacWhinney, 2002; Zhao & Li, 2009). Most of these efforts, however, have focused on how to simulate L1 learning (see also Li & Shirai, 2000; Li et al., 2007).

In an early effort, Li and Farkas (2002) built the self-organizing model of bilingual processing (SOMBIP) to begin to tackle such issues in L2 learning, specifically by using input data derived from corpora of real language (see Goodman et al., 2008). Unlike the BSN and BSRN models, which were based on *supervised* connectionist learning, the SOMBIP was based on *unsupervised* connectionist learning using *self-organizing maps* (Kohonen, 2001, Chapter 6; see Table 1 for definitions of terms). Unlike the BSN and BSRN, the SOMBIP uses phonological representations based on articulatory features of phonemes and semantic representations based on cooccurrence statistics in child-directed parental speech, which gives the SOMBIP greater linguistic and developmental realism. These two types of representation are connected with each other through *Hebbian learning* (Hebb, 1949; see Table 1). The SOMBIP simultaneously learned mixed bilingual input (216 English words and 184 Chinese words), and the model produced patterns highly consistent with those of the BSN and BSRN, showing distinct lexical representations for Chinese and English after learning. The model also provides a different way to assess proficiency by having the network exposed to fewer sentences in the L2, simulating a novice learner having limited linguistic experience. This more natural way of modeling proficiency, interestingly, yielded comparable results to those from the BSN: The "novice" network's representation of the L2 was more compressed and less clearly delineated, as compared with the "proficient" network.

Modeling Bilingual Learning

Although the BSN, BSRN, and SOMBIP models clearly differ from BIA models in incorporating connectionist learning principles, they are what we would call "representation models," rather than "learning models" in the sense that the model simulates the learner's L2 development to different levels of proficiency over time. This was a significant gap, since learning implies developmental changes across time, progressing from less knowledge to more enriched knowledge representation at a higher level of proficiency. Simply dividing the models into two sets (one "proficient" and the other "less proficient"), as in the models discussed above, does not simulate developmental progression in the learner.

Recognizing this problem, researchers developed several connectionist models with the aim of simulating developmental changes, for example, the developmental lexicon (DevLex) model, especially the DevLex-II model for bilingual language learning (Li & Zhao, 2013; Zhao & Li, 2010). A key principle that supports connectionist learning is the adjustment of connection *weights* (see Table 1 for this term). It is the updating of these connection weights that accounts for changes as learning progresses, much as the brain undergoes functional changes in the form of increased or decreased connections between certain neurons, groups of neurons, or regions. For example, in the DevLex-II model, nodes that represent different modalities of linguistic

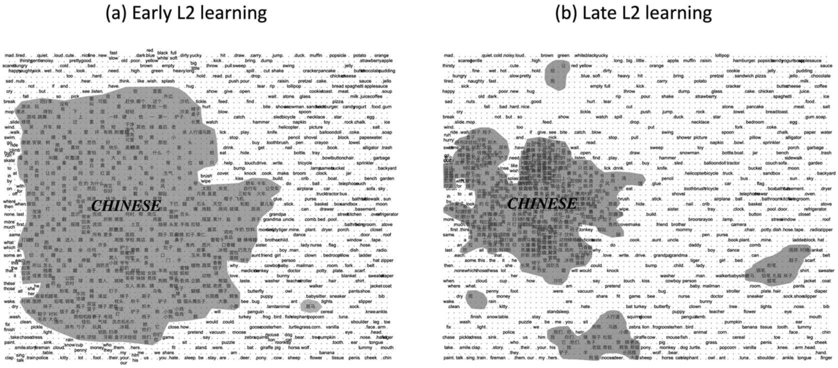


Figure 2 Different representation structures of the first language (L1) and second language (L2) lexicons in the developmental lexicon model DevLex-II as a function of (a) early L2 learning versus (b) late L2 learning. Shaded areas indicate Chinese (L2) representations (Zhao & Li, 2006) and nonshaded areas indicate English (L1) representations (adapted from Zhao & Li, 2006).

information such as orthography and phonology are connected through Hebbian learning (Hebb, 1949; see Table 1). As learning progresses, the use of the Hebbian learning rule in DevLex-II allows the model to adapt the weights based on cooccurrence between learning patterns (see Li & Zhao, 2013, for a discussion); for example, orthographic patterns ("CAT") and phonological patterns (/kæt/) that cooccur would become strongly connected, simulating the process of acquiring the mapping between the spelling and sound of a word. At the same time, the network would also undergo reorganization by adjusting connection weights between the input and output layers so that the output topographic structure (through neighborhood of ordered nodes) can capture the similarity in the input; this weight updating applies to orthography, phonology, and lexical semantics in the model (see Figure 2 of Li & Zhao, 2013, for illustration). It is through these connectionist learning principles that computational models can acquire enriched bilingual representations over time in learning. Figure 2 provides an example based on the DevLex-II model for L2 learning.

The DevLex-II model, in contrast to the BSN, BSRN, and SOMBIP models that received bilingual inputs simultaneously, learned a large L1 and L2 vocabulary sequentially (500 words each for the L1 and L2): (a) "Early bilingual language learning" involved an onset time of L2 input when only one fifth of the L1 vocabulary was trained in the model, and (b) "late bilingual language learning” involved an onset time of L2 input when four fifths of the L1 vocabulary was trained in the model. Figure 2 shows that AoA (early vs. late) plays an important role in modulating the overall representational structure. In late but not early bilingual language learning, the L2 structure is distinctly more fragmented as a whole than the L1 structure; the L2 structure is also more compressed within the L2 space, and has fuzzier representations. Detailed analyses also indicate that the L2 shows a pattern of representation “parasitic” on L1 patterns: Small L2 chunks are dispersed and interspersed within L1 regions, and the locations of the L2 words in the map are also dependent on the similarity structure and categories established by the L1 words in meaning (for the semantic map, see Figure 2 in this article) or in sound (for the phonological map, see Figure 4 of Li & Zhao, 2013). Such representational structures could also account for L2 learners’ difficulty in achieving efficient and effective lexical access and retrieval, given how different the representation of the L2 lexicon is from that of the L1 lexicon.

This example illustrates how computational modeling may provide us with insights into the dynamic interactions and competition between the L1 and L2. A model that examines such interactions can also efficiently account for the so-called “age effects” in bilingual language learning, because it takes into consideration the learning dynamics and neural plasticity of the learning system; see a recent formulation of this interaction in terms of emergentism and the ecosystem (Claussenius-Kalman et al., 2021). In particular, if L2 onset occurs at a time when L1 has been consolidated, as in late bilingual language learning, the learned structure in L1 will constrain what can be learned, and the plasticity of the network may also decrease because of the network’s commitment to L1. For late bilingual language learning, the more consolidated the representation of the L1, the more resistant to change (i.e., the more “entrenched”) the learning system will become in the face of new input from a new language. Furthermore, with a high degree of entrenchment (as in late bilingual language learning in the simulated network), the organization of the L2 will have to tap into existing L1 representational resources and its structure, thus exhibiting parasitic representations.

Limitations of Current Bilingual Computational Models

Despite progress made with computational models in bilingualism as discussed above, several major issues have limited their further development. Here we address those issues in relation to the four criteria discussed earlier, along with a consideration of open science practices to promote scientific communication and progress. When these limitations can be overcome, researchers will be able to develop better models that can provide insights into bilingual language representation and learning.

Validity

Although the current bilingual models are tractable, they often suffer from low psychological plausibility. For example, computational models of bilingual language processing have attempted to connect language with general cognition (e.g., language control and cognitive control mechanisms), but have so far failed to realistically model the interaction between language learning and general cognitive abilities. To do so requires models to simulate the interaction among multiple domain-specific and domain-general systems (see discussion in the later section on neural models). In addition, models should not only focus on the input and output of the two languages, but also make contact with cognitive computation in the context of learning and the specific features in the environment (see discussion in the later section on cognitive models).

In this connection, another important limitation of current computational models of bilingual language representation and learning is that most models have been either inspired by connectionist architectures (e.g., BIA and BIA+) or implemented in such architectures (e.g., BSN and SOMBIP), as discussed above. Although connectionist models are insightful and powerful, they remain limited in a number of ways. For example, Ellis (2005) and Shirai (2019) have pointed out the lack of ability of current connectionist models to simulate both implicit and explicit learning processes. Specifically, most connectionist models have focused on implicit learning processes (input based on implicit frequency) rather than explicit processes (input based on conscious knowledge). Unlike L1 acquisition, simply comprehending incoming input is generally not sufficient for L2 learning. Determining how future models can exploit explicit representations, especially prior world knowledge relevant to the context of learning, is crucial for pushing the boundaries in modeling bilingual learning. This point is related to the later section where we argue for a pluralist perspective on computational bilingual modeling.

Contact with Real Language

In earlier sections, we discussed the importance of representing the linguistic stimuli in an accurate and faithful manner so that models can make more direct contact with the input that the learner is exposed to in the real world. Many previous models have relied on localist representations, as illustrated above with the example from French (1998). Localist representations are simple and efficient to construct but may not accurately represent the input. Furthermore, models based on localist input (e.g., BIA+ or BSRN) are stationary models that are designed to capture bilingual speakers' lexicons in a developmental manner, and they are therefore not able to account for the learning mechanism.

Distributed representations may be more difficult to implement, but could be a better choice if the goal is to capture the similarities among sounds or concepts (e.g., for simulating effects such as similarity-based phonological or semantic priming). We previously mentioned the efforts that have been made to represent the input in greater detail than in earlier models through the use of distributed representations based on distributed statistics and faithful to language-specific properties. The faithful representation of input in computational models has become somewhat less of an issue in recent years, partly because of the rapid development of NLP models such as Word2vec, a vector space model of word representations (Mikolov et al., 2013), and other kinds of vector space modeling, which can automatically generate large-scale representations of linguistic items based on a large-scale corpus of speech or text.

Interpretability

Most computational models so far are designed to simulate a given bilingual phenomenon or pattern based on empirical data, but how these simulations provide theoretical insights beyond those offered by verbal models is often unclear. The advantages of modeling should be reflected not only in the flexibility of manipulation of variables of interest, but also in the ability of the models to adjudicate between competing theories and hypotheses. Currently, most computational models of bilingualism have been limited in this regard and have primarily served a confirmatory function of supporting given theoretical frameworks (which were based on empirical data in the first place). Exploring how computational models can generate new theoretical hypotheses or even inspire entirely new theoretical perspectives in the study of bilingual language learning is a significant direction for future research.

Predictive Power

Earlier computational models not only used simplified input, but also had simplified architecture with limited predictive power, and were often criticized as "toy models" by researchers opposed to the modeling enterprise. As computational modeling becomes more sophisticated and available computational power increases, the size of the model (e.g., the number of nodes or "processing units" and their connections) and the size of the input and output have both grown exponentially. This ability to scale up the power of our current computational models has implications for understanding bilingual language learning and processing. Li and Grant (2019) pointed out that the new Multilink model of Dijkstra et al. (2019) was able to account for a lexicon size of thousands of words, which is a significant improvement over previous models that handled dozens of words (e.g., French, 1998) or hundreds of words (e.g., Zhao & Li, 2010). Exploring how bilingual models can improve predictive power by taking advantage of the increasing computational power available and the increasing complexity of modeling (e.g., deep learning models; see the later section on neural models) will be a challenge in the coming years.

Open Science and Data Sharing

Modern research requires computational modeling to embrace the open science approach to further scientific progress in all domains of research (see discussion in the next section). Toward this end, many researchers share their computational models, programs, source codes, research protocols, original data, and metadata, and many open science platforms have been established (e.g., TalkBank, OpenNeuro, Github, Huggingface; see Appendix S1 in the Supporting Information online; see also the Open Science Framework at https://osf.io). So far, researchers developing computational models of bilingualism have not embraced this open science practice for data and code sharing. Modelers should heed the call by Addyman and French (2012) to make every effort to provide user-friendly interfaces and tools to nonmodelers, so that many more researchers of language science can use and test computational models, and can do so without fear of the technical hurdles posed by programming codes, simulating environments, and other modeling architectural concerns.

Toward Pluralist Bilingual Learning Models

Bilingual language learning is a complex process that involves multiple domain-specific and domain-general systems. The process not only deals with the input and output of the two languages, but also requires dynamic interactions among biological plasticity, cognitive computation, and the learning environment. As a result, there is an urgent need to broaden the scope of current bilingual models, employ pluralist learning mechanisms, and combine various computational algorithms and theories to advance the understanding of bilingual language learning.

In a critical review of cognitive computational neuroscience, Kriegeskorte and Douglas (2018) advocated an overarching goal of integrating both neural and cognitive models in our research efforts and cross-disciplinary collaborations. The neural models have been focused on using biologically plausible computational components to describe and explain cognitive processes, whereas the cognitive models incorporate cognitive mechanisms and principles that are abstracted away from the detailed biological structures and their computational processes. Although both approaches have made significant progress in the past, the field is in dire need of integrative models that can not only provide precise descriptions of important issues at hand (as implementable computational models have been designed to do), but also offer top-level theoretical guidance in how to develop and design task-performing models, so that we not only describe but also explain and predict human brain patterns and the relevant behavior.

The above perspective is directly relevant to our discussion of the modeling of bilingual language learning. Bilingual models need to extend their explorations of both neural and cognitive processes. Although the utility and applications of computational models for bilingual language learning are clear, our discussion above indicates that the major advances have been made only within the realm of connectionist modeling. Those developing connectionist models in bilingual research tend to focus on neural processes with the aim of implementing biologically plausible artificial neural networks. The limitations of those connectionist models in accounting for bilingual language learning and representation are also clear, as we pointed out earlier. Cognitive approaches such as Bayesian modeling and multimodal learning, although emerging as important for explaining monolingual L1 acquisition, have not been applied in bilingual L2 learning research. Despite our earlier criticism of the box-and-arrow models based on high-level cognitive theories (that they do not lead to mechanistic accounts and are out of touch with current neuroscience models), we recognize that theory-driven cognitive perspectives can nevertheless be important for developing theoretical insights (Kriegeskorte & Douglas, 2018).

In this section, we will first review a few relevant neurocognitive theories from empirical research into bilingual language learning. With those theories as an anchor, we will then indicate new directions in light of the framework laid out by Kriegeskorte and Douglas (2018) for neural and cognitive processes. Although these authors developed their framework for the new field called cognitive computational neuroscience, their overall argument applies to our perspective here equally well.

Neural and Cognitive Theories of Bilingual Language Learning

Age of Acquisition and Proficiency

AoA and L2 proficiency have been extensively examined in both empirical and computational work in bilingualism. It has been generally observed that the ability to learn a L2 declines when the AoA is late (for a review, see Hernandez & Li, 2007), although large individual variation exists in late AoA. In general, it is more challenging for a late learner to achieve high ultimate proficiency in a L2 (Flege et al., 1999). The emergentist accounts, such as the competition model (MacWhinney & Bates, 1989) and the sensorimotor integration hypothesis (Hernandez & Li, 2007; Hernandez et al., 2005), posit cascading effects of early learning on late learning and a competitive interplay between L1 and L2 (cf. our earlier discussion of L1 entrenchment and parasitic L2 representation). Such accounts are in stark contrast to biologically based explanations such as the critical period hypothesis (Lenneberg, 1967).

L2 proficiency is often confounded with AoA, and the roles played by these two variables remain unclear, as does their relative importance. Behavioral work suggests that proficiency, not AoA, determines naming latencies in lexical tasks when L2 acquisition occurs early in life (Hernandez & Reyes, 2002; Kohnert et al., 1999). Some neuroimaging work suggests that AoA effects on neural activity diminish or disappear when early and late learners are equated on L2 proficiency (Perani et al., 1998; Wartenburger et al., 2003). Hence, considerable evidence suggests that proficiency has a crucial role in L2 processing and may be at least partially independent from AoA (see review in Hernandez & Li, 2007). Recent research has recognized the possibility that AoA and L2 proficiency may play different roles in L2 learning and processing (Hakuta et al., 2003; Wartenburger et al., 2003; Weber-Fox & Neville, 1996). Some neuroimaging work has found that tasks involving syntactic processing showed larger AoA effects, whereas tasks involving semantic processing were largely constrained by proficiency (Wartenburger et al., 2003). Therefore, it is likely that AoA plays a role in syntactic processing whereas proficiency plays a role in semantic processing. The challenge is to find ways of using computational modeling to illuminate the computational mechanisms underlying such differences (see Hernandez & Li, 2007, for a review).

Bilingual Representations

Understanding how L2 learners represent and organize two languages has long been a fundamental area of research in bilingualism. The competition model proposes a theoretical framework in which distinct language modules emerge from the competitive interplay between two languages (Hernandez et al., 2005). Related to this issue is the question of whether L1 and L2 representations are distinct or distributed in neural substrates (see Li, 2009, for a discussion). Earlier evidence supported shared neural basis but with different computational demands between the L1 and L2 (Perani & Abutalebi, 2005).

However, more recent studies using more advanced neuroimaging analyses, such as multivoxel pattern analysis, have found distinct distributed patterns, with the two languages represented by interleaved (partially overlapping) but functionally independent neural populations (Xu et al., 2017); such evidence further suggests the importance of revisiting the issue from an emergentist developmental perspective (Claussenius-Kalman et al., 2021).

Cognitive Control

One common observation is that words in both languages become active in parallel when bilinguals use either of their two languages (Dijkstra & van Heuven, 1998; Kroll et al., 2013), suggesting a mechanism of cognitive control in place to help bilinguals avoid constant confusion (Green, 1998; see also the BIA+ model that has incorporated such a mechanism, Dijkstra & van Heuven, 2002). A large amount of literature has been devoted to studying cognitive control in bilingualism, either independently or in connection with the hypothesis that bilinguals have cognitive advantage over monolinguals (Bialystok, 2009; Bialystok et al., 2005). A new perspective on the bilingual cognitive advantage (e.g., DeLuca et al., 2019) treats bilingualism as a spectrum rather than a unitary concept or phenomenon, which is consistent with the dynamic emergentist perspective discussed above (see Li & Dong, 2020, Chapter 2, for a recent review). Further, the question of how domain-general cognitive control abilities impact bilingual language learning is also frequently raised in research (Woumans et al., 2019; see Wen et al., 2017, for a review of the role of working memory). Domain-general cognitive capacities include but are not limited to executive function, attention, working memory, and nonverbal IQ, and have both behavioral and neural correlates in individual differences in bilingual language learning (e.g., Yang et al., 2015; Yang & Li, 2019). For example, learners with high procedural but low declarative memory scores were able to learn simple but not complex rules (Ettlinger et al., 2014). Sheppard et al. (2012) also showed that the brain networks of more successful L2 learners exhibited greater global efficiency, an index considered to be positively associated with working memory capacity.

Learning Context

Whereas traditional bilingual language learning relies on associating L2 with L1 via translation or rote memory, social interactions and embodied experience can help to prevent L2 from becoming parasitic on L1. Recent neurocognitive studies have provided early evidence for the positive effects of social learning of L2. Learners in real or simulated social interactions have shown more embodied and nativelike neural representations (Jeong et al., 2010), greater learning efficiency (Hsiao et al., 2017), and less susceptibility to L1 interference (Linck et al., 2009). According to the social learning account (Li & Jeong, 2020), the parasitism of L2 on L1 can also be attributed to different learning contexts for the two languages. Studies in bilingual language learning and pedagogy have begun to examine how to leverage digital technologies to enhance social learning and social interaction, putting the context for L2 learning on a par with that of L1 learning in the natural environment (for reviews, see Li & Jeong, 2020; Li & Lan, 2021; Verga & Kotz, 2017). In this direction, researchers are also investigating how social learning, as compared with translation-based learning, might enable the L2 learner to develop stronger, more connected and integrated neural networks that can support better audio-visuo-spatial processing, multimodal integration, motor simulation, enriched semantic representation, and enhanced long-term memory retention (Jeong et al., 2010, 2021; Legault et al., 2019; Verga & Kotz, 2017).

Given the above discussion of core issues and theoretical hypotheses from neurocognitive studies and also the limitations of the current models (as discussed previously), it is clear that the current bilingual computational models have limited power in their explorations of both the neural and cognitive processes involved. For example, although theories posit important and perhaps distinct roles for AoA and proficiency in bilingual learning, few of the current computational models have considered both AoA and proficiency and how they affect learning differently. Understanding of bilingual representations may be hampered by the lack of faithful input representation in many bilingual models. Few studies have also incorporated nonlinguistic processing into their computational models, even though there is evidence of a tight interaction between cognitive control and bilingual processing (Green & Abutalebi, 2013).

Any computational account of bilingual language learning needs to be guided by theoretical considerations of how learning gives rise to the bilingual brain and to bilingual cognition and behavior. Thus, it will be crucial for computational models to incorporate both biologically plausible mechanisms and high-level cognitive principles of language and cognition, to which we now turn.

Neural Models

Computational models can be constructed by building biologically plausible computational components to implement high-level functions. A number of major initiatives (e.g., the European Brain Initiative) have adopted such an approach toward computational modeling of the brain for the future (see

Kriegeskorte & Douglas, 2018). Despite arguments about whether computational models should or can completely resemble human brains (Firestone, 2020), an increasing amount of literature has shown that some neural models can behave similarly to humans in some tasks, at both the biological level (Cadieu et al., 2014; Yamins et al., 2014) and the representational level (Tenney et al., 2019). For example, BERT, a powerful pretrained sentence encoder in NLP, has revealed layer-by-layer linguistic abstractions (Tenney et al., 2019) similar to those found in human language processing.

Alongside models' potential to be biologically plausible, learning theories can also benefit from pursuit of a neurobiological path. One central issue about learning is whether it can be independent of any symbolic system. Fodor and Pylyshyn (1988) argued that connectionist neural models cannot capture linguistic productivity and compositionality. On the other hand, many attempts have been made to enhance the productivity and compositionality of neural models (e.g., McCoy et al., 2020; see Linzen & Baroni, 2021, for a review); such attempts shed new light on the fundamental structure of human cognition and language.

Given current progress and the examples briefly introduced above, it is evident that bilingual modeling has not yet fulfilled its potential in taking the neurobiological research path. Both biological plausibility and the ability to simulate real cognitive functions need further improvement. A recent review by Pulvermüller et al. (2021) has summarized five types of biologically constrained neural models: localist, auto-associative, hetero-associative, deep neural, and whole-brain networks. Whereas most bilingual models have involved the first three types (Dijkstra et al., 2019; Dijkstra & van Heuven, 2002; Li & Farkas, 2002), deep neural networks (DNNs) and whole-brain networks are recent types that have greater power and biological plausibility not yet explored in bilingual computational models. Interestingly, a recent empirical study by Goldstein et al. (2022) showcased the shared computational principles for human brain's processing of language and for DNNs, in that both rely on contextual embeddings to represent words, which, in natural contexts, involves continuous predictive and surprisal evaluation processes. In the following sections, we discuss these two types of neural models, highlighting their connection to the biologically grounded L2 acquisition theories we discussed earlier.

Deep Neural Networks

A neural network model with three or more hidden layers will be considered deep, given its ability to perform complex and nonlinear computations (see LeCun et al., 2015, for a review). A DNN model is characterized by the following features: (a) distributed representations, in which knowledge is represented by activations spread over a large set of neurons, and different concepts are functionally represented by different but overlapping and interleaved neurons; (b) data-driven implicit learning mechanisms, allowing the model to automatically discover the representations and intricate structures in the raw data, with little help from human engineering; and (c) multiple levels, each transforming representations from one level to a higher and more abstract level, in many ways resembling information processing in the human brain (Christophel et al., 2017; LeCun et al., 2015).

These three characteristics of a DNN make it a realistic and biologically plausible framework for examining bilingual language learning, given empirical findings that (a) brain representations of L1 and L2 are distributed in neural substrates (Li, 2009; Xu et al., 2017); (b) bilingual language learning is sensitive to L2 input and can take place in the absence of explicit guidance or awareness (Rogers et al., 2015; Williams, 2005; see Andringa & Rebuschat, 2015, for a review); and (c) neural representations of both L1 and L2 have shown hierarchical patterns (Ding et al., 2016; Liberto et al., 2021), from smaller units to progressively more abstract ones (e.g., from words to phrases to sentences, and from phonemes to phonotactics to semantic meanings).

Astonishing progress has been made with recent DNNs, and some of them are even reaching near-human performance in certain language processing tasks (Devlin et al., 2019). *Long short-term memory networks* (LSTM networks; see Table 1 for a definition) enable the memory component to maintain information over time and thus facilitate language processing by integrating contextual information and information retrieved from memory. The *attention* mechanism provided in the *transformer* (Vaswani et al., 2017; see Table 1 for the terms) has further improved learning efficiency and outcomes by zooming in on the information most relevant to the language processing task.

In the L1 literature, modeling work using DNNs has yielded broad implications for L1 learning theories. Contemporary DNNs can acquire a surprising amount of linguistic knowledge at both surface and abstract levels, such as the representations of lexical semantics (Bojanowski et al., 2017) and abstract sentence structure (Gulordava et al., 2018). However, such linguistic abilities are susceptible to variability due to training and testing parameters, and still fall significantly short of human competence in linguistic productivity (see Lake et al., 2017, for a review). McCoy et al. (2020) found that a tree-based DNN, compared to a traditional sequence-based one, developed a stronger and more humanlike generalization ability. Such progress has ignited a broader and deeper discussion about the mechanisms that underlie human language processing and learning.

The architectures and algorithms employed in modern DNNs can also offer rich insights into bilingual language learning. Previous research has applied RNNs to the modeling of bilingual code-switching speech (Tsoukala et al., 2017, 2019, 2021), but more work needs to be done in the field of bilingual language learning to fulfill the potential of various types of DNN. For example, because the memory system in LSTM networks makes those models particularly good at maintaining relevant information while forgetting the irrelevant (Tato et al., 2018), one could use them to model bilingual language learning and the influence of memory. The attention mechanism in the transformer could be used to study the influence of attention on L2 learning and the attention mechanisms involved in processing. Although the implications of DNNs for bilingual language learning are still unclear, their use for neural machine translation is worth exploring given their capability of dealing with the interaction between two languages, which might mimic bilingual learning and processing. As of today, transformers are particularly adept at machine translation, suggesting that they may also hold great potential for understanding bilingual language learning (for current state-of-the-art translation models, see Edunov et al., 2018, and Liu et al., 2020; for a critical review of translation, see Saunders, 2022). Such integration would also overcome the limitations of connectionist models as discussed earlier, and enhance the power of computational models in simulating implicit and explicit processes and in connecting language, memory, and cognitive control.

Whole-Brain Networks

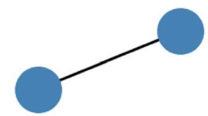
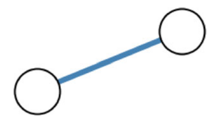
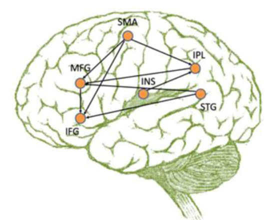
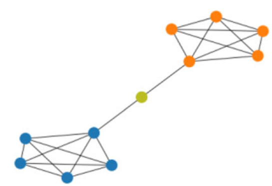
The human brain is a dynamic and complex system with numerous brain regions interconnected and interactively communicating with each other (Bullmore & Sporns, 2009; Karuza et al., 2016). The great importance of adopting a holistic view of learning has been brought to attention recently (Karuza et al., 2016; Mattar et al., 2016). Neuroscience researchers have highlighted the importance of several key networks in human learning and cognition, including the default mode network, the central executive network, and the salience network (Bressler & Menon, 2010; Raichle, 2010). The idea behind whole-brain networks is to view learning as a global system rather than a set of isolated processes, and to use brain network construction and network science to capture dynamic interaction and network topology. Construction of brain networks involves assembling relational data from neuroimaging measurements into a network. The network science approach quantitatively characterizes the topological structure of a network. Table 2 illustrates some concepts and measures used in whole-brain network studies (for a detailed review of brain network science, see Mattar et al., 2016, and Bassett & Sporns, 2017).

The whole-brain network approach has been applied in language studies to characterize neural organization for language processing. With *community* detection analysis (see Table 2 for an explanation of communities), Chai et al. (2016) found a modular structure of language representation in English L1 speakers, where the connections within hemispheres were stronger than those between hemispheres. One application of whole-brain network analysis in bilingual language learning is to study the relation between brain functional connectivity and L2 attainment. For example, Yang et al. (2015) showed that the *effective connectivity* (see Table 2) of the brain network at the beginning of learning predicted success in L2 Chinese tone learning, and that such relations were modulated by variables including L2 proficiency and auditory processing abilities (Yang & Li, 2019). However, the application of whole-brain networks in bilingualism can be further broadened. Bilingual models can harness the network science approach to study brain changes caused by learning a new language (see Li & Grant, 2016, for a review). Given evidence that bilingual representation changes with L2 proficiency (e.g., Wang et al., 2020, whole-brain network analysis might be well suited to capturing differences between L1 and L2 speakers and proficiency-based whole-brain changes, thus providing a global map of how bilingual learners acquire and represent multiple languages (e.g., see Zhang et al., 2020).

Cognitive Models

Cognitive models, from box-and-arrow cognitive hypotheses to specific computational implementations, can be abstracted away from brain data while directly contacting cognitive theories and human behavior. Although the neural models discussed above can provide constraints for computational theories, we also need cognitive models to enable progress on higher level cognitive processes such as attention, memory, and language, and most neuroscience models still fall short of accounting for features of the human mind such as abstract language representations. One example is the one-shot learning challenge. Lake et al. (2015) reported that a Bayesian model, but not a DNN, achieved human-level performance in one-shot learning of new visual concepts, showing an ability to generalize after learning just one sample (but see Brown et al., 2020, and Perez et al., 2021, for new insights and ongoing

Table 2 Some concepts and measures used in whole-brain network analyses with a network science approach

Term	Definition	Example
Node	Element of a network.	
Edge	The connection between two nodes.	
Effective connectivity	A whole-brain network analysis describing the causal influences that neural units exert over each other.	
Community	A group of nodes that are more densely connected to each other than to the other nodes of a network. In the example, there are three communities detected (marked with different colors).	

debates in DNN research). In Lake et al.'s study, the key to success in one-shot learning might be the Bayesian model's ability to capture the principles of causality, compositionality, and learning to learn, suggesting that the inductive bias might be a critical component determining learnability and learning efficiency. This example illustrates the importance of cognitive models; their contributions to understanding learning could be complementary to insights from neural network models.

Within the L2 literature, there are as yet few attempts to use cognitive computational approaches (but see the earlier discussion of BIA models). We argue that computational modeling should bring high-level cognitive models into the research agenda and should widen its scope by integrating cognitive theories from bilingual language learning with neurobiological models. In this section, we present three cognitive computational frameworks: Bayesian statistical learning, multimodal learning, and network science modeling. As acquiring an L2 requires both learning itself and the learner's interaction with the learning context, we will highlight the potential of using Bayesian statistical learning for characterizing the learning mechanism, and multimodal learning for the interaction with the L2 learning context. Finally, we will return to the network science approach, but this time considering it from a perspective independent of neural mechanisms, aimed at scaling up the power of L2 computational models with both local patterns and global architectures.

Bayesian Statistical Learning

The Bayesian framework is grounded on the assumption that learners use background knowledge to make statistically optimal inferences from incomplete data (Perfors et al., 2011; Tenenbaum et al., 2011). According to Tenenbaum et al. (2011), Bayesian inference is for answering the question of how abstract knowledge guides inference from incomplete data. Bayesian inference allows for efficient learning and reasoning based on prior knowledge, which explains why a Bayesian model could sometimes outperform a neural network model when learning from a small sample (Lake et al., 2015).

The general Bayesian rule assigns a probability to a hypothesis based on prior knowledge of conditions that might be related to the hypothesis (see Perfors et al., 2001, for details). Suppose that a child heard a novel word *fep* while seeing three objects: a table, a plate, and an apple. The probability of a hypothesis that *fep* means the object "apple" could be determined by the child's prior knowledge that the apple, unlike the table and the plate, is edible, and the situation that the child's parent was calling the child when there was food on the table. Inductive computations resembling Bayesian rules have been observed in many cognitive domains (e.g., Cheng & Almor, 2017, 2019; Feldman et al., 2009; Frank et al., 2009; Steyvers et al., 2006; Xu & Tenenbaum, 2007). For example, Xu and Tenenbaum (2007) found that the way in which adults and children learn new words could be well explained and predicted using a Bayesian framework.

Whereas the Bayesian approach has been frequently used to understand L1 acquisition (Feldman et al., 2009; Frank et al., 2009; Xu & Tenenbaum, 2007), few in the L2 field have made similar attempts (but see Zinszer et al., 2018, for an example). However, understanding bilingual language learning from the Bayesian perspective is both theoretically and empirically plausible. First, it is commonly believed that L2 learning also involves statistical learning (Ellis, 2002, 2005; Hamrick, 2014), where rational statistical inference could be made from the learning input. Zinszer et al. (2018) showed that modification of a monolingual Bayesian model with a different prior probability to weaken the mutual exclusivity bias (i.e., a belief that one object can only have one label; Markman & Wachtel, 1988) could be used to simulate word learning in bilingual children, suggesting a significant role for the Bayesian framework in bilingual learning research. Second, two recent studies (Cheng & Almor, 2017, 2019) reported L2 biases during the processing of within-sentence coreference, with beliefs about pronoun use differing from those of native speakers; these biases were successfully captured by the Bayesian model, further attesting to the important role of Bayesian inference in bilingual processing.

Given the above, a promising direction for future research is to understand the statistical learning mechanisms of L2 learning from a cognitive Bayesian perspective. Bayesian learning theories assume that the human mind builds mental models of the world, and that these models can then be used to learn new concepts in a generative and productive way (Tenenbaum et al., 2011). The mental models of an L2 learner would clearly include knowledge of the L1. Given previous work examining how consolidation of L1 might influence L2 learning (as discussed earlier), future research could use the Bayesian approach to establish the dynamic interplay between L1 and L2 involving key variables such as AoA and L2 proficiency. For example, the entrenchment of L1 knowledge, usually strengthened with increasing AoA, might serve as prior knowledge for facilitating or interfering with L2 learning (Hernandez & Li, 2007). Hence, a Bayesian interpretation of L1 transfer as a function of AoA can be established. Early bilinguals would not experience interference from L1, because the knowledge of the two languages accumulates simultaneously; however, L2 learners with a later AoA might have stronger prior biases arising from L1, causing misbeliefs when interpreting L2 input. On the other hand, it is likely that some existing knowledge of L1 allows learners to make correct predictions about L2 input and hence enhances learning efficiency, reflecting a positive L1 transfer.

Unlike AoA, which affects L2 knowledge as a function of how entrenched L1 knowledge is, L2 proficiency primarily reflects the status of L2 knowledge construction. To achieve similarly high proficiency, early learners might proceed by constructing L1 and L2 knowledge simultaneously, whereas late learners might need to accumulate L2 knowledge from scratch when L1 knowledge is already in place. Therefore, L2 learners with a late AoA will require more effort to overcome the strong prior biases from the L1 in order to attain high proficiency. A Bayesian model can offer an explicit way to instantiate the distinct mechanisms of AoA and L2 proficiency through simulating the specific procedures of constructing L1 and L2 knowledge.

Multimodal Learning

As mentioned earlier, there has recently been an increasing appreciation of the role of the learning environment in the L2 literature. A significant part of language learning relies on the social environment with which learners actively interact (Kuhl, 2004; Li & Jeong, 2020). Social learning theories received earlier attention in L1 acquisition, both empirically and computationally (e.g., Ross et al., 2018; Smith et al., 2016; Yu & Ballard, 2007; Yurovsky et al., 2013), and therefore computational modeling of L2 social learning may benefit from advances in L1 work.

Yu and Ballard (2004) developed a statistical model to simulate, in an integrated fashion, multimodal word learning with speech input, embodiment experience, and the learning environment. During training, they asked a real person to perform everyday tasks while verbally describing their actions. Meanwhile, various sensors were attached to the person to track their sensorimotor experiences. Therefore, the model received input from speech modalities, such as phonology and meaning, and nonspeech modalities, such as visual perception and body movement. In an evaluation, the model showed considerably high accuracy in word–meaning association tasks, suggesting that a multimodal learning system can derive perceptually grounded meanings of words from observing users' everyday activities. In another study (Yu & Ballard, 2007), a statistical model learned word–meaning associations from adult–child communication data presented in audio and video recordings. The results showed that the multimodal model that incorporated social communication information (e.g., joint attention) outperformed the baseline model that did not contain such information. Thus, language input and nonlinguistic social cues work in tandem to contribute to language acquisition.

The development of social learning theories in the L2 learning field can benefit from using multimodal learning models such as those discussed. For example, the model trained with body movements (Yu & Ballard, 2004) can be used to test the role of embodied experience in bilingual language learning (Jeong et al., 2010); the cues provided by joint attention (Yu & Ballard, 2007) can be adapted for studies of L2 learning through social interactions (Hsiao et al., 2017; Jeong et al., 2010; Linck et al., 2009; Verga & Kotz, 2017). As being immersed in L2 can inhibit interference from L1 (Linck et al., 2009), a multimodal learning model can be used to examine what social cues in an immersion context inhibit L1 access and why. Work on L2 social learning is still at an early stage of progress, and joint efforts from empirical and computational studies in cognitive science, neuroscience, and education will be essential to the understanding of its mechanisms (see Li & Lan, 2021, for a recent discussion).

Network Science Approach

Besides being used in the biologically driven whole-brain network (discussed in a previous section), the network science approach is also under the spotlight of theory-driven studies in language and cognition (Chan & Vitevitch, 2009; Hills et al., 2009; Karuza et al., 2016; Sizemore et al., 2018; Vitevitch, 2019; Xu et al., 2021). Unlike whole-brain network models that link different brain regions, theory-driven network science models characterize local patterns and global structures of the high-level cognitive system (see Karuza et al., 2016, for a review).

The network science approach has offered new insight into language learning. Hills et al. (2009) considered networks with different growth principles to simulate the construction of the child learner's semantic network and tested which network could better predict the child's AoA of words. One network tested was the preferential attachment model, which assumes that the more connected are the known words to which a new word is related, the more likely it is that the new word will enter the lexicon. The other network tested was the preferential acquisition model, which assumes that the more connected a new word is to other words in the learning environment, either known or unknown, the more likely it is that the new word will enter the lexicon (see Figure 3 for a simplified illustration). They found that the preferential acquisition network outperformed the preferential attachment network, suggesting that a word is

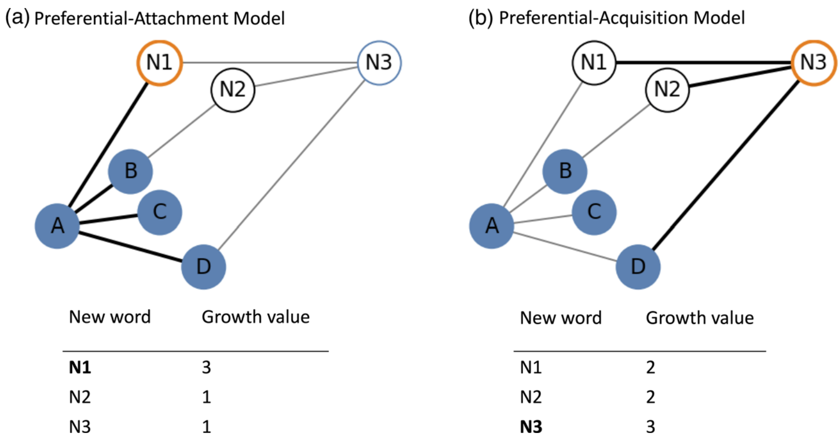


Figure 3 An illustration of the preferential attachment model (a) and the preferential acquisition model (b) (adapted from Hills et al., 2009). A, B, C, and D represent known words in the lexicon, whereas nodes N1, N2, and N3 represent words to be learned. The greater the growth value, the more likely that the word will be learned. The preferential attachment model (a) predicts that N1 would have a higher chance of being learned, because it is connected to node A, which has the most connections (i.e., to B, C, and D) in the lexicon. The preferential acquisition model (b) predicts that N3 would have a higher chance of being learned, because it has the highest number of connections with other nodes (i.e., with N1, N2, and D).

more likely to enter a child's lexicon if it is well connected to other words in the learning environment. With topological analysis of the child's semantic network, Sizemore et al. (2018) found word learning to be a process of gap forming and filling. In the child's semantic network, earlier acquired words left sparse space where later acquired words could enter and fill gaps (Figure 4), much like a process of filling the holes in Swiss cheese (see Li et al., 2007, for computational modeling of this process). Sizemore et al.'s study instantiates how higher order connectivity patterns between words can constrain the development of semantic feature representations during language learning.

The network science approach has been recently explored (Tiv et al., 2020; Xu et al., 2021) in bilingual processing but not in bilingual language learning. Tiv et al. (2020) established networks separately for the L1 and L2 to map conversational topics that bilinguals use in each language. A comparison of the structures of L1 and L2 networks showed that the L1 network is greater in size (i.e., number of *nodes*; see Table 2 for terms)[2] and density (i.e., number of *edges* in a network out of the total number of possible edges), suggest-

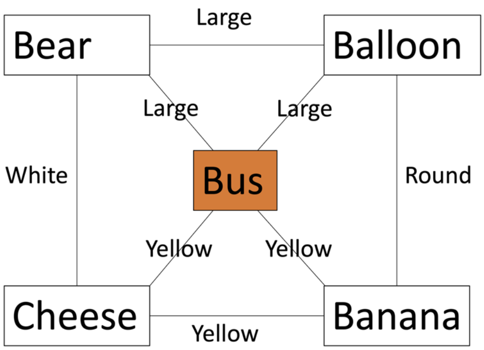


Figure 4 An illustration of a gap-filling network. Earlier acquired words left sparse space where later acquired words could enter and fill the gap (adapted from Sizemore et al., 2018). In this example, the feature-based connections between *balloon*, *bear*, *cheese*, and *banana* left a gap within the semantic network, which was later filled in with the new word *bus*.

ing that bilinguals' communication topics are broader and more diversified in L1 compared with L2. Xu et al. (2021) employed one unified semantic network to examine the bilingual lexicon and semantic representation in naturalistic speech. The authors trained a Word2vec model (Mikolov et al., 2013) on a spontaneous code-switching speech corpus and then obtained the semantic similarity between words to determine network edges. Although bilinguals often mix the two languages in naturalistic conversations, the authors still found, by using community detections, separate modular representations of words in different languages in the semantic network. These findings are consistent with earlier connectionist modeling results (cf. our earlier discussion of the patterns in French, 1998; Li & Farkas, 2002; Thomas, 1998). The study offers a holistic view of the organization and dynamic competition (MacWhinney, 2012; MacWhinney & Bates, 1989) of the bilingual lexicon.

Despite a lack of network science research in bilingual language learning, previous relevant work in L1 acquisition (Hills et al., 2009; Sizemore et al., 2018) and bilingual processing (Tiv et al., 2020; Xu et al., 2021) points to its potential to benefit bilingual language learning theories. For instance, in lexical development, since bilingual language learning involves a dynamic interaction between words in the two languages (as discussed earlier), the network science approach provides a basis for addressing issues such as (a) how L1 words may constrain the acquisition and organization of L2 words, (b) how acquiring L2 lexical items may in turn affect the organization of L1 words, (c) whether there are individual differences in the patterns of learning L2 words, and (d) whether the different learning patterns can predict ultimate attainment in L2. Answers to some of these questions would also enable connections with other empirical and computational modeling approaches (e.g., the work discussed in the section on current models). For example, one could apply the three growth mechanisms in Hills et al. (2009) or the gap-filling mechanism introduced by Sizemore et al. (2018) to track changes in L1 versus L2 lexicons.

In regard to issue (a) above, considering the parasitism of L2 on L1 (Hernandez et al., 2005), it is likely that the preferential attachment mechanism could better account for L2 word learning than the preferential acquisition mechanism (Hills et al., 2009); an L2 word might be more likely to enter the lexicon if the word is more connected to L1 words in the lexicon (see Li, 2009, for argumentation). In regard to issue (b), some L2 words may in turn fill the semantic gap left by L1 words to facilitate semantic representations and word retrieval, since studies have reported the benefit of using two languages together in production compared to staying within the L1 (e.g., Kleinman & Gollan, 2016). In regard to issue (c), individual learners may differ in learning patterns such that gifted L2 learners may adopt the preferential acquisition mechanism, whereas less successful learners may rely more on the preferential attachment mechanism due to stronger parasitism of the L2 on the L1 in these learners. In regard to issue (d), preferential acquisition may result in greater L2 attainment, as it helps learners construct a more integrated and independent knowledge representation of the L2. One can even further ask whether the enhanced access to the L2 when learners were immersed in the L2 environment (Linck et al., 2009) occurs because features in the environment can better support the preferential acquisition of words. Examples of such approaches would enable us to extend the scope of network science application, especially from static and cross-sectional data to longitudinal data that incorporate dynamic changes in bilingual language learning (Hills et al., 2009; Sizemore et al., 2018). Computational modeling of bilingual language learning should thus combine multiple methods to make use of the network science approach with the greatest possible vigor (Li & Grant, 2019).

Integrating Neural and Cognitive Approaches

Just as it is important to investigate the bilingual brain and bilingual cognition empirically, it is imperative to consider how best to bridge the gap between neuroscientific and high-level cognitive paths of computational modeling (Griffiths et al., 2010; Kriegeskorte & Douglas, 2018; McClelland et al., 2010). Without an understanding of neural implementations, the formulation of cognitive theories would become a purely intellectual exercise. On the other hand, without an understanding of how high-level cognition is organized, neural implementations can become completely reductionist. To characterize the interplay of the bilingual mind and brain, we should aim at constructing an integrated computational account by combining cognitive theories and biological foundations for bilingual language learning.

Here we advocate that this integration can be achieved through a pluralist investigation into the shared and distinct features of diverse learning models in terms of their competition and compatibility (see Mitchell, 2002, for a philosophical discussion on pluralism). Competitive pluralism highlights exclusivity, according to which different theories should be pitted against each other so that some can be accepted and others rejected, whereas compatible pluralism allows diverse and mutually compatible models to account for different aspects or dimensions of a phenomenon in light of the complexity of nature.

From a competitive pluralism perspective, investigating the competition between different neural or cognitive models may advance the understanding of bilingual language learning. For example, a connectionist and a Bayesian account may have incompatible premises on bilingual learning. The former assumes an emergence of knowledge structure via learning, whereas the latter assumes the availability of structured knowledge for efficient learning with inductive bias. In a long-standing debate on general learning theories between advocates of the two approaches (e.g., Griffiths et al., 2010; McClelland et al., 2010), Bayesians challenged the sampling efficiency of connectionist learning models, whereas connectionists questioned Bayesians' prior specification of hypothesis space and its psychological plausibility. As the debate continues, the field makes progress in understanding learning through the comparison and competition between the two types of models.

In contrast to the above, it is also useful to embrace the perspective that different models can be compatible and complementary such that neither approach can handle the full complexity of learning alone. For example, research on the "Bayesian brain" attempts to bridge cognitive and neural levels through implementing Bayesian inference in a biologically plausible way or empowering neural nets with top-down predictions (Ali et al., 2021; Deneve, 2008; Ma et al., 2008). For lexical-semantic representation, there have been promising attempts to link whole biological brain networks with word-embedding models, that is, DNN models for the semantic representation of words (e.g., fastText, Bojanowski et al., 2017; BERT, Devlin et al., 2019; Ramakrishnan & Deniz, 2021; Ruan et al., 2016). Cognitive theories of social multimodal learning have also been drawn on in research on DNNs or deep learning in general. Many studies have explored the theoretical and empirical benefits of combining information from multiple modalities in improving the performance of deep learning (e.g., Ngiam et al., 2011; Reed et al., 2016; Wang et al., 2019).

It is clear from the above discussion that both competitive and compatible pluralism perspectives are important. Although it is debatable whether the competitive or the compatible approach is desirable for computational modeling, underlying both perspectives is the requirement that researchers pursue and understand diverse models with the greatest vigor. However, current bilingual models are often limited in diversity and depth. As we discussed earlier, most of the existing bilingual models have adopted a connectionist framework, and many modeling architectures date to the 1980s and 1990s. They are in dire need of enhancement in terms of their validity, contact with real language, interpretability, and predictive power, especially in light of the rapid developments in computational modeling methods and applications. At the same time, computational approaches inspired by cognitive theory have been largely ignored in the current models, which hampers the theoretical contribution that computational modeling aims for. If the pluralist approach that we advocate here is adopted, this should lead to the emergence of novel computational models that contribute significantly to the understanding of the mechanisms and processes of bilingual learning and representation.

Concluding Remarks

Computational approaches have much to offer to the understanding of bilingual language learning, because they force explicit hypothesis specification and generate testable internal representations, reliability, and experimental manipulability. We have shown that current models have contributed to progress in understanding the bilingual mind, but that a significant gap exists between the promise offered by theoretical constructs and the explanatory and predictive power achieved so far. Current bilingual models are often limited in terms of oversimplified knowledge representation and architecture, and a lack of contact with the complexity of learner experience and the learning environment. For connectionist neural models, an enhancement of biological plausibility and predictive power is sorely needed to overcome the challenges ahead. At the same time, theory-driven cognitive approaches can provide important insights in theories of bilingual language learning, yet they have been largely ignored in the current models.

In this article, we have advocated a new research agenda for computational modeling in bilingual learning, through linking progress across different disciplines, such as computational neuroscience, NLP, and L1 acquisition, to construct an integrated computational account combining cognitive theories and biological foundations for bilingual learning. Those constructing bilingual language learning models should widen the scope of these models and take advantage of the progress and rapid advances in other fields. As highlighted by Kriegeskorte and Douglas (2018), the neural and cognitive paths of computational modeling are the extremes of a continuum rather than a dichotomy. The two paths have a common goal of explaining "how our brains give rise to our minds" (Kriegeskorte & Douglas, 2018, p. 1152; this is also an exciting avenue of future computational research in bilingual learning. Our approach outlined here resonates strongly with a recent call for placing language in an integrated system of understanding and communicating (McClelland et al., 2020); that is, exploring the full range of multisensory (auditory, visual, tactile) contexts in which language is used and represented, rather than treating language as an isolated linguistic system. Future challenges for researchers will include building bridges to enable computational models to integrate behavioral, cognitive, neuropsychological, and neuroimaging findings to arrive at a converging picture of language learning. Cross-disciplinary joint work in bilingual language learning, as in other modern scientific domains, is not a luxury but a necessity for success.

Final revised version accepted 17 May 2022

Notes

1 In this article, we will use the term "bilingual language learning" rather than "second language learning" or "second language acquisition," to highlight the fact that learning an L2 inevitably involves interaction with the L1 in the bilingual mind.

2 As the whole-brain network and the network science approach at the cognitive level share the same graphic analysis, readers can refer to Table 2 for basic terms mentioned here (node, edge, community).

References

Addyman, C., & French, R. M. (2012). Computational modeling in cognitive science: A manifesto for change. *Topics in Cognitive Science*, *4*(3), 332–341. https://doi.org/10.1111/j.1756-8765.2012.01206.x

Ali, A., Ahmad, N., de Groot, E., van Gerven, M. A., & Kietzmann, T. C. (2021). *Predictive coding is a consequence of energy efficiency in recurrent neural networks* (2021.02.16.430904). bioRxiv. https://doi.org/10.1101/2021.02.16.430904

Andringa, S., & Rebuschat, P. (2015). New directions in the study of implicit and explicit learning: An introduction. *Studies in Second Language Acquisition*, *37*(2), 185–196. https://doi.org/10.1017/S027226311500008X

Bassett, D. S., & Sporns, O. (2017). Network neuroscience. *Nature Neuroscience*, *20*(3), 353–364. https://doi.org/10.1038/nn.4502

Bates, E., & MacWhinney, B. (1982). Functionalist approaches to grammar. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 174–218). Cambridge University Press.

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition*, *12*(1), 3–11. https://doi.org/10.1017/S1366728908003477

Bialystok, E., Martin, M. M., & Viswanathan, M. (2005). Bilingualism across the lifespan: The rise and fall of inhibitory control. *International Journal of Bilingualism*, *9*(1), 103–119. https://doi.org/10.1177/13670069050090010701

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, *5*, 135–146. https://doi.org/10.1162/tacl_a_00051

Bressler, S. L., & Menon, V. (2010). Large-scale brain networks in cognition: Emerging methods and principles. *Trends in Cognitive Sciences*, *14*(6), 277–290. https://doi.org/10.1016/j.tics.2010.04.004

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). *Language models are few-shot learners* (arXiv:2005.14165). arXiv. https://doi.org/10.48550/arXiv.2005.14165

Bullmore, E., & Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, *10*(3), 186–198. https://doi.org/10.1038/nrn2575

Cadieu, C. F., Hong, H., Yamins, D. L. K., Pinto, N., Ardila, D., Solomon, E. A., Majaj, N. J., & DiCarlo, J. J. (2014). Deep neural networks rival the representation of primate IT cortex for core visual object recognition. *PLoS Computational Biology*, *10*(12), e1003963. https://doi.org/10.1371/journal.pcbi.1003963

Chai, L. R., Mattar, M. G., Blank, I. A., Fedorenko, E., & Bassett, D. S. (2016). Functional network dynamics of the language system. *Cerebral Cortex*, *26*(11), 4148–4159. https://doi.org/10.1093/cercor/bhw238

Chan, K. Y., & Vitevitch, M. S. (2009). The influence of the phonological neighborhood clustering coefficient on spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(6), 1934–1949. https://doi.org/10.1037/a0016902

Cheng, W., & Almor, A. (2017). Referent predictability is affected by syntactic structure: Evidence from Chinese. *Journal of Psycholinguistic Research*, *46*(1), 157–174. https://doi.org/10.1007/s10936-016-9430-6

Cheng, W., & Almor, A. (2019). A Bayesian approach to establishing coreference in second language discourse: Evidence from implicit causality and consequentiality verbs. *Bilingualism: Language and Cognition*, *22*(3), 1–20. https://doi.org/10.1017/S136672891800055X

Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., & Haynes, J. D. (2017). The distributed nature of working memory. *Trends in Cognitive Sciences*, *21*(2), 111–124. https://doi.org/10.1016/j.tics.2016.12.007

Claussenius-Kalman, H., Hernandez, A., & Li, P. (2021). Expertise, ecosystem, and emergentism: Dynamic developmental bilingualism. *Brain and Language*, *222*, Article 105013. https://doi.org/10.1016/j.bandl.2021.105013

DeLuca, V., Rothman, J., Bialystok, E., & Pliatsikas, C. (2019). Redefining bilingualism as a spectrum of experiences that differentially affects brain structure and function. *Proceedings of the National Academy of Sciences*, *116*(15), 7565–7574. https://doi.org/10.1073/pnas.1811513116

Deneve, S. (2008). Bayesian spiking neurons I: Inference. *Neural Computation*, *20*(1), 91–117.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, *1*, 4171–4186. https://doi.org/10.18653/v1/N19-1423

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

Dijkstra, T., & van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*(3), 175–197. https://doi.org/10.1017/S1366728902003012

Dijkstra, T., Wahl, A., Buytenhuijs, F., Van Halem, N., Al-Jibouri, Z., De Korte, M., & Rekké, S. (2019). Multilink: A computational model for bilingual word recognition and word translation. *Bilingualism: Language and Cognition*, *22*(4), 657–679. https://doi.org/10.1017/S1366728918000287

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*(1), 158–164. https://doi.org/10.1038/nn.4186

Edunov, S., Ott, M., Auli, M., & Grangier, D. (2018). *Understanding back-translation at scale* (1808.09381). arXiv. http://arxiv.org/abs/1808.09381

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*(2), 143–188. https://doi.org/10.1017/S0272263102002024

Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, *27*(2), 305–352. https://doi.org/10.1017/S027226310505014X

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211. https://doi.org/10.1207/s15516709cog1402_1

Ettlinger, M., Bradlow, A. R., & Wong, P. C. (2014). Variability in the learning of complex morphophonology. *Applied Psycholinguistics*, *35*(4), 807–831. https://doi.org/10.1017/S0142716412000586

Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review*, *116*(4), 752–782. https://doi.org/10.1037/a0017196

Firestone, C. (2020). Performance vs. competence in human–machine comparisons. *Proceedings of the National Academy of Sciences*, *117*(43), 26562–26571. https://doi.org/10.1073/pnas.1905334117

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, *41*(1), 78–104. https://doi.org/10.1006/jmla.1999.2638

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*(1–2), 3–71. https://doi.org/10.1016/0010-0277(88)90031-5

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science*, *20*(5), 578–585. https://doi.org/10.1111/j.1467-9280.2009.02335.x

French, R. (1998). A simple recurrent network model of bilingual memory. In M. A. Gernsbacher & S. A. Derry (Eds.), *Proceedings of the 20th Annual Conference of the Cognitive Science Society* (pp. 368–373). Routledge. https://doi.org/10.4324/9781315782416-73.

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

Frost, R., Armstrong, B. C., & Christiansen, M. H. (2019). Statistical learning research: A critical review and possible new directions. *Psychological Bulletin*, *145*(12), 1128–1153. https://doi.org/10.1037/bul0000210

Gardner, H. (1987). *The mind's new science: A history of the cognitive revolution*. Basic Books.

Gardner, W. A. (1984). Learning characteristics of stochastic-gradient-descent algorithms: A general study, analysis, and critique. *Signal Processing*, *6*(2), 113–133. https://doi.org/10.1016/0165-1684(84)90013-6

Goldberg, Y. (2017). Neural network methods for natural language processing. *Synthesis Lectures on Human Language Technologies*, *10*(1), 1–309. https://doi.org/10.2200/S00762ED1V01Y201703HLT037

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, *35*(3), 515–531. https://doi.org/10.1017/S0305000907008641

Goldstein, A., Zada, Z., Buchnik, E., Schain, M., Price, A., Aubrey, B., Nastase, S. A., Feder, A., Emanuel, D., Cohen, A., Jansen, A., Gazula, H., Choe, G., Rao, A., Kim, C., Casto, C., Fanda, L., Doyle, W., Friedman, D., … Hasson, U. (2022). Shared computational principles for language processing in humans and deep language models. *Nature Neuroscience*, *25*(3), 369–380. https://doi.org/10.1038/s41593-022-01026-4

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*(2), 67–81. https://doi.org/10.1017/S1366728998000133

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

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences*, *14*(8), 357–364. https://doi.org/10.1016/j.tics.2010.05.004

Gulordava, K., Bojanowski, P., Grave, E., Linzen, T., & Baroni, M. (2018). *Colorless green recurrent networks dream hierarchically* (1803.11138). arXiv. https://doi.org/10.48550/arXiv.1803.11138

Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical evidence: A test of the critical-period hypothesis for second-language acquisition. *Psychological Science*, *14*(1), 31–38. https://doi.org/10.1111/1467-9280.01415

Hamrick, P. (2014). A role for chunk formation in statistical learning of second language syntax. *Language Learning*, *64*(2), 247–278. https://doi.org/10.1111/lang.12049

Hebb, D. O. (1949). *The organisation of behaviour: A neuropsychological theory*. Science Editions.

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

Hernandez, A. E., & Reyes, I. (2002). Within- and between-language priming differ: Evidence from repetition of pictures in Spanish-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 726–734. https://doi.org/10.1037//0278-7393.28.4.726

Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009). Categorical structure among shared features in networks of early-learned nouns. *Cognition*, *112*(3), 381–396. https://doi.org/10.1016/j.cognition.2009.06.002

Hsiao, I. Y. T., Lan, Y.-J., Kao, C.-L., & Li, P. (2017). Visualization analytics for second language vocabulary learning in virtual worlds. *Journal of Educational Technology & Society*, *20*(2), 161–175.

Jeong, H., Li, P., Suzuki, W., Kawashima, R., & Sugiura, M. (2021). Neural mechanisms of language learning from social contexts. *Brain and Language*, *212*, Article 104874. https://doi.org/10.1016/j.bandl.2020.104874

Jeong, H., Sugiura, M., Sassa, Y., Wakusawa, K., Horie, K., Sato, S., & Kawashima, R. (2010). Learning second language vocabulary: Neural dissociation of situation-based learning and text-based learning. *NeuroImage*, *50*(2), 802–809. https://doi.org/10.1016/j.neuroimage.2009.12.038

Karuza, E. A., Thompson-Schill, S. L., & Bassett, D. S. (2016). Local patterns to global architectures: Influences of network topology on human learning. *Trends in Cognitive Sciences*, *20*(8), 629–640. https://doi.org/10.1016/j.tics.2016.06.003

Kim, K. H., Relkin, N. R., Lee, K. M., & Hirsch, J. (1997). Distinct cortical areas associated with native and second languages. *Nature*, *388*, 171–174. https://doi.org/10.1038/40623

Kleinman, D., & Gollan, T. H. (2016). Speaking two languages for the price of one: Bypassing language control mechanisms via accessibility-driven switches. *Psychological Science*, *27*(5), 700–714. https://doi.org/10.1177/0956797616634633

Kohnert, K. J., Bates, E., & Hernandez, A. E. (1999). Balancing bilinguals: Lexical-semantic production and cognitive processing in children learning Spanish and English. *Journal of Speech, Language, and Hearing Research*, *42*(6), 1400–1413. https://doi.org/10.1044/jslhr.4206.1400

Kohonen, T. (2001). *Self-organizing maps* (3rd ed.). Springer. https://doi.org/10.1007/978-3-642-97610-0_6

Kriegeskorte, N., & Douglas, P. K. (2018). Cognitive computational neuroscience. *Nature Neuroscience*, *21*(9), 1148–1160. https://doi.org/10.1038/s41593-018-0210-5

Kroll, J. F., Gullifer, J. W., & Rossi, E. (2013). The multilingual lexicon: The cognitive and neural basis of lexical comprehension and production in two or more languages. *Annual Review of Applied Linguistics*, *33*, 102–127. https://doi.org/10.1017/S0267190513000111

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–843. https://doi.org/10.1038/nrn1533

Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, *350*(6266), 1332–1338. https://doi.org/10.1126/science.aab3050

Lake, B., Ullman, T., Tenenbaum, J., & Gershman, S. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, 40, Article e253. https://doi.org/10.1017/S0140525/16001837

Leahey, T. (2004). *A history of psychology: Main currents in psychological thought* (6th ed.). Pearson Prentice Hall.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Legault, J., Fang, S. Y., Lan, Y. J., & Li, P. (2019). Structural brain changes as a function of second language vocabulary training: Effects of learning context. *Brain and Cognition*, *134*, 90–102. https://doi.org/10.1016/j.bandc.2018.09.004

Lenneberg, E. H. (1967). The biological foundations of language. *Hospital Practice*, *2*(12), 59–67. https://doi.org/10.1080/21548331.1967.11707799

Li, P. (2009). Lexical organization and competition in first and second languages: Computational and neural mechanisms. *Cognitive Science*, *33*(4), 629–664. https://doi.org/10.1111/j.1551-6709.2009.01028.x

Li, P. (2013). Computational modeling of bilingualism. *Bilingualism: Language and Cognition*, *16*(2), 241–366. https://doi.org/10.1017/S1366728913000059

Li, P., & Dong, Y. (2020). Language experience and cognitive control: A dynamic perspective. *Psychology of Learning and Motivation*, *72*, 27–52. https://doi.org/10.1016/bs.plm.2020.02.002

Li, P., & Farkas, I. (2002). A self-organizing connectionist model of bilingual processing. In R. R. Heredia & J. Altarriba (Eds.), *Bilingual sentence processing* (pp. 59–85). North-Holland. https://doi.org/10.1016/S0166-4115(02)80006-1

Li, P., & Grant, A. (2016). Second language learning success revealed by brain networks. *Bilingualism: Language and Cognition*, *19*, 657–664. https://doi.org/10.1017/S1366728915000280

Li, P., & Grant, A. (2019). Scaling up: How computational models can propel bilingualism research forward. *Bilingualism: Language and Cognition*, *22*, 1–3. https://doi.org/10.1017/S1366728918000755

Li, P., & Jeong, H. (2020). The social brain of language: Grounding second language learning in social interaction. *npj Science of Learning*, *5*, Article 8. https://doi.org/10.1038/s41539-020-0068-7

Li, P., & Lan, Y.J. (2021). Digital Language Learning (DLL): Insights from behavior, cognition, and the brain. *Bilingualism: Language and Cognition*, *25*(3), 361–378. https://doi.org/10.1017/S1366728921000353

Li, P., & MacWhinney, B. (2002). PatPho: A phonological pattern generator for neural networks. *Behavior Research Methods, Instruments, & Computers*, *34*(3), 408–415. https://doi.org/10.3758/bf03195469

Li, P., & Shirai, Y. (2000). *The acquisition of lexical and grammatical aspect*. Walter de Gruyter. https://doi.org/10.1515/9783110800715

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

Li, P., & Zhao, X. (2018). Computational modeling. In A. M. B. de Groot & P. Hagoort (Eds.), *Research methods in psycholinguistics and the neurobiology of language: A practical guide* (pp. 208–229). Wiley.

Li, P., Zhao, X., & MacWhinney, B. (2007). Dynamic self-organization and early lexical development in children. *Cognitive Science*, *31*, 581–612. https://doi.org/10.1080/15326900701399905

Liberto, G. M., Nie, J., Yeaton, J., Khalighinejad, B., Shamma, S. A., & Mesgarani, N. (2021). Neural representation of linguistic feature hierarchy reflects second-language proficiency. *NeuroImage*, *227*, Article 117586. https://doi.org/10.1016/j.neuroimage.2020.117586

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

Linzen, T., & Baroni, M. (2021). Syntactic structure from deep learning. *Annual Review of Linguistics*, *7*(1), 195–212. https://doi.org/10.1146/annurev-linguistics-032020-051035

Liu, X., Duh, K., Liu, L., & Gao, J. (2020). *Very deep transformers for neural machine translation*. arXiv. http://arxiv.org/abs/2008.07772

Ma, W. J., Beck, J. M., & Pouget, A. (2008). Spiking networks for Bayesian inference and choice. *Current Opinion in Neurobiology*, *18*(2), 217–222. https://doi.org/10.1016/j.conb.2008.07.004

MacWhinney, B. (2012). The logic of the unified model. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 229–245). Routledge.

MacWhinney, B., & Bates, E. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 3–76). Cambridge University Press.

MacWhinney, B., Kempe, V., Li, P., & Brooks, P. J. (Eds.). (2022). *Emergentist approaches to language*. Frontiers Media SA.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121–157. https://doi.org/10.1016/0010-0285(88)90017-5

Mattar, M. G., Betzel, R. F., & Bassett, D. S. (2016). The flexible brain. *Brain*, *139*(8), 2110–2112. https://doi.org/10.1093/brain/aww151

McClelland, J. L. (2009). The place of modeling in cognitive science. *Topics in Cognitive Science*, *1*(1), 11–38. https://doi.org/10.1111/j.1756-8765.2008.01003.x

McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., & Smith, L. B. (2010). Letting structure emerge: Connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Sciences*, *14*(8), 348–356. https://doi.org/10.1016/j.tics.2010.06.002

McClelland, J. L., Hill, F., Rudolph, M., Baldridge, J., & Schütze, H. (2020). Placing language in an integrated understanding system: Next steps toward human-level performance in neural language models. *Proceedings of the National Academy of Sciences*, *117*(42), 25966–25974. https://doi.org/10.1073/pnas.1910416117

McCoy, R. T., Frank, R., & Linzen, T. (2020). Does syntax need to grow on trees? Sources of hierarchical inductive bias in sequence-to-sequence networks. *Transactions of the Association for Computational Linguistics*, *8*, 125–140. https://doi.org/10.48550/arXiv.2001.03632

Meltzoff, A. N., Kuhl, P. K., Movellan, J., & Sejnowski, T. J. (2009). Foundations for a new science of learning. *Science*, *325*(5938), 284–288. https://doi.org/10.1126/science.1175626

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient estimation of word representations in vector space* (1301.3781). arXiv. https://doi.org/10.48550/arXiv.1301.3781

Mitchell, S. D. (2002). Integrative pluralism. *Biology and Philosophy*, *17*(1), 55–70. https://doi.org/10.1023/A:1012990030867

Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., & Ng, A. Y. (2011). Multimodal deep learning. *Proceedings of the 28th International Conference on Machine Learning* (pp. 689–696).

Perani, D., & Abutalebi, J. (2005). The neural basis of first and second language processing. *Current Opinion in Neurobiology*, *15*(2), 202–206. https://doi.org/10.1016/j.conb.2005.03.007

Perani, D., Paulesu, E., Galles, N. S., Dupoux, E., Dehaene, S., Bettinardi, V., Cappa, S. F., Fazio, F., & Mehler, J. (1998). The bilingual brain: Proficiency and age of acquisition of the second language. *Brain*, *121*(10), 1841–1852. https://doi.org/10.1093/brain/121.10.1841

Perez, E., Kiela, D., & Cho, K. (2021). *True few-shot learning with language models*. arXiv. https://doi.org/10.48550/arXiv.2105.11447

Perfors, A., Tenenbaum, J. B., Griffiths, T. L., & Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition*, *120*(3), 302–321. https://doi.org/10.1016/j.cognition.2010.11.015

Pulvermüller, F., Tomasello, R., Henningsen-Schomers, M. R., & Wennekers, T. (2021). Biological constraints on neural network models of cognitive function. *Nature Reviews Neuroscience*, *22*, 488–502. https://doi.org/10.1038/s41583-021-00473-5

Raichle, M. E. (2010). Two views of brain function. *Trends in Cognitive Sciences*, *14*(4), 180–190. https://doi.org/10.1016/j.tics.2010.01.008

Ramakrishnan, K., & Deniz, F. (2021). Non-complementarity of information in word-embedding and brain representations in distinguishing between concrete and abstract words. In E. Chersoni, N. Hollenstein, C. Jacobs, Y. Oseki, L. Prévot, & E. Santus (Eds.), *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics* (pp. 1–11). https://doi.org/10.18653/v1/2021.cmcl-1.1

Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., & Lee, H. (2016). Generative adversarial text to image synthesis. In M. F. Balcan & K. Q. Weinberger (Eds.), *International Conference on Machine Learning* (pp. 1060–1069). PMLR. https://doi.org/10.48550/arXiv.1605.05396

Rogers, J., Révész, A., & Rebuschat, P. (2015). Challenges in implicit learning research. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 275–300). John Benjamins. https://doi.org/10.1075/sibil.48.12rog

Ross, C., Barbu, A., Berzak, Y., Myanganbayar, B., & Katz, B. (2018). Grounding language acquisition by training semantic parsers using captioned videos. In E. Riloff, D. Chiang, J. Hockenmaier, & J. I. Tsujii (Eds.), *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 2647–2656). https://doi.org/10.18653/v1/D18-1285

Ruan, Y. P., Ling, Z. H., & Hu, Y. (2016). Exploring semantic representation in brain activity using word embeddings. In J. Su, K. Duh, & X. Carreras (Eds.), *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 669–679). https://doi.org/10.18653/v1/D16-1064

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Saunders, D. (2022). *Domain adaptation and multi-domain adaptation for neural machine translation: A survey*. arXiv. http://arxiv.org/abs/2104.06951

Sheppard, J. P., Wang, J. P., & Wong, P. C. (2012). Large-scale cortical network properties predict future sound-to-word learning success. *Journal of Cognitive Neuroscience*, *24*(5), 1087–1103. https://doi.org/10.1162/jocn_a_00210

Shirai, Y. (2019). *Connectionism and second language acquisition*. Routledge. https://doi.org/10.4324/9780203118085

Sizemore, A. E., Karuza, E. A., Giusti, C., & Bassett, D. S. (2018). Knowledge gaps in the early growth of semantic feature networks. *Nature Human Behaviour*, *2*(9), 682–692. https://doi.org/10.1038/s41562-018-0422-4

Smith, A. C., Monaghan, P., & Huettig, F. (2016). Complex word recognition behaviour emerges from the richness of the word learning environment. In K. Twomey, A. C. Smith, G. Westermann, & P. Monaghan (Eds.), *Neurocomputational Models of Cognitive Development and Processing: Proceedings of the 14th Neural Computation and Psychology Workshop* (pp. 99–114). https://doi.org/10.1142/9789814699341_0007

Steyvers, M., Griffiths, T. L., & Dennis, S. (2006). Probabilistic inference in human semantic memory. *Trends in Cognitive Sciences*, *10*(7), 327–334. https://doi.org/10.1016/j.tics.2006.05.005

Tato, A., Nkambou, R., & Frasson, C. (2018). Predicting emotions from multimodal users' data using a deep long term memory. In G. Houben, G. McCalla, F. Pianesi,

& M. Zancanaro (Eds.), *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization* (pp. 369–370). Association for Computing Machinery. https://doi.org/10.1145/3209219.3209264

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*(6022), 1279–1285. https://doi.org/10.1126/science.1192788

Tenney, I., Das, D., & Pavlick, E. (2019). *BERT rediscovers the classical NLP pipeline*. arXiv. https://doi.org/10.48550/arXiv.1905.05950

Thomas, M. S. C. (1997). Distributed representations and the bilingual lexicon: One store or two? In J. Bullinaria, D. Glasspool, & G. Houghton (Eds.), *Proceedings of the 4th Annual Neural Computation and Psychology Workshop* (pp. 9–11). Springer. https://doi.org/10.1007/978-1-4471-1546-5_19

Tiv, M., Gullifer, J. W., Feng, R. Y., & Titone, D. (2020). Using network science to map what Montréal bilinguals talk about across languages and communicative contexts. *Journal of Neurolinguistics*, *56*, Article 100913. https://doi.org/10.1016/j.jneuroling.2020.100913

Tsoukala, C., Broersma, M., Van den Bosch, A., & Frank, S. L. (2021). Simulating code-switching using a neural network model of bilingual sentence production. *Computational Brain & Behavior*, *4*(1), 87–100. https://doi.org/10.1007/s42113-020-00088-6

Tsoukala, C., Frank, S. L., & Broersma, M. (2017). "He's pregnant": Simulating the confusing case of gender pronoun errors in L2 English. In *The 39th Annual Meeting of the Cognitive Science Society* (pp. 3392–3397). Cognitive Science Society.

Tsoukala, C., Frank, S. L., van den Bosch, A. P. J., Valdes Kroff, J., & Broersma, M. (2019). Simulating Spanish-English code-switching: El modelo está generating code-switches. In E. Chersoni, N. Hollenstein, C. Jacobs, Y. Oseki, L. Prévot, & E. Santus (Eds.), *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics* (pp. 20–29). https://doi.org/10.18653/v1/W19-2903

van Heuven, W. J., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language*, *39*(3), 458–483. https://doi.org/10.1006/jmla.1998.2584

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention is all you need* (1706.03762). arXiv. https://doi.org/10.48550/arXiv.1706.03762

Verga, L., & Kotz, S. A. (2017). Help me if I can't: Social interaction effects in adult contextual word learning. *Cognition*, *168*, 76–90. https://doi.org/10.1016/j.cognition.2017.06.018

Vitevitch, M. S. (2019). *Network science in cognitive psychology*. Routledge.

Wang, R., Ke, S., Zhang, Q., Zhou, K., Li, P., & Yang, J. (2020). Functional and structural neuroplasticity associated with second language proficiency: An MRI study of Chinese-English bilinguals. *Journal of Neurolinguistics*, *56*, Article 100940. https://doi.org/10.1016/j.jneuroling.2020.100940

Wang, Y., Shen, Y., Liu, Z., Liang, P. P., Zadeh, A., & Morency, L. P. (2019). Words can shift: Dynamically adjusting word representations using nonverbal behaviors. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 33, No. 01, pp. 7216–7223). http://arxiv.org/abs/1811.09362

Wartenburger, I., Heekeren, H. R., Abutalebi, J., Cappa, S. F., Villringer, A., & Perani, D. (2003). Early setting of grammatical processing in the bilingual brain. *Neuron*, *37*(1), 159–170. https://doi.org/10.1016/S0896-6273(02)01150-9

Weber-Fox, C. M., & Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience*, *8*(3), 231–256. https://doi.org/10.1162/jocn.1996.8.3.231

Wen, Z, Biedroń, A, & Skehan, P. (2017). Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching*, *50*(1), 1–31. https://doi.org/10.1017/S0261444816000276

Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition*, *27*(2), 269–304. https://doi.org/10.1017/S0272263105050138

Woumans, E., Ameloot, S., Keuleers, E., & Van Assche, E. (2019). The relationship between second language acquisition and nonverbal cognitive abilities. *Journal of Experimental Psychology: General*, *148*(7), 1169–1177. https://doi.org/10.1037/xge0000536

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*(2), 245–272. https://doi.org/10.1037/0033-295X.114.2.245

Xu, M., Baldauf, D., Chang, C. Q., Desimone, R., & Tan, L. H. (2017). Distinct distributed patterns of neural activity are associated with two languages in the bilingual brain. *Science Advances*, *3*(7), Article e1603309. https://doi.org/10.1126/sciadv.1603309

Xu, Q., Markowska, M., Chodorow, M., & Li, P. (2021). Modeling bilingual lexical processing through code-switching speech: A network science approach. *Frontiers in Psychology*, *12*, Article 662409. https://doi.org/10.3389/fpsyg.2021.816653

Yamins, D. L., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, *111*(23), 8619–8624. https://doi.org/10.1073/pnas.1403112111

Yang, J., Gates, K., Molenaar, P., & Li, P. (2015). Neural changes underlying successful second language word learning: An fMRI study. *Journal of Neurolinguistics*, *33*, 29–49. https://doi.org/10.1016/j.jneuroling.2014.09.004

Yang, J. & Li, P. (2019). Mechanisms for auditory perception: A neurocognitive study of second language learning of Mandarin Chinese. *Brain Sciences*, *9*(6), Article 139. https://doi.org/10.3390/brainsci9060139

Yu, C., & Ballard, D. H. (2004). A multimodal learning interface for grounding spoken language in sensory perceptions. *ACM Transactions on Applied Perception (TAP)*, *1*(1), 57–80. https://doi.org/10.1145/1008722.1008727

Yu, C., & Ballard, D. H. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing*, *70*(13–15), 2149–2165. https://doi.org/10.1016/j.neucom.2006.01.034

Yurovsky, D., Smith, L. B., & Yu, C. (2013). Statistical word learning at scale: The baby's view is better. *Developmental Science*, *16*(6), 959–966. https://doi.org/10.1111/desc.12036

Zhang, X., Yang, J., Wang, R., & Li, P. (2020). A neuroimaging study of semantic representation in first and second languages. *Language, Cognition and Neuroscience*, *35*(10), 1223–1238. https://doi.org/10.1080/23273798.2020.1738509

Zhao, X., & Li, P. (2006). A self-organizing connectionist model of bilingual lexical development. In *Proceedings of the Twenty-Eighth Annual Conference of the Cognitive Science Society* (p. 2639). Lawrence Erlbaum.

Zhao, X., & Li, P. (2009). The acquisition of aspect in self-organizing connectionist models. *Linguistics*, *47*(5), 1075–1112. https://doi.org/10.1515/LING.2009.038

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., & Kohonen, T. (2011). Contextual self-organizing map: Software for constructing semantic representations. *Behavior Research Methods*, *43*(1), 77–88. https://doi.org/10.3758/s13428-010-0042-z

Zinszer, B. D., Rolotti, S. V., Li, F., & Li, P. (2018). Bayesian word learning in multiple language environments. *Cognitive Science*, *42*, 439–462. https://doi.org/10.1111/cogs.12567

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Accessible Summary

Appendix S1. Some Practical Considerations for Computational Modeling.