Expanding the Role of Connectionism in SLA Theory

Robert Nelson
University of Alabama

In this article, I explore how connectionism might expand its role in second language acquisition (SLA) theory by showing how some symbolic models of bilingual and second language lexical memory can be reduced to a biologically realistic (i.e., neurally plausible) connectionist model. This integration or hybridization of the two models follows the principles of what philosophers of science call intertheoretic reduction. Such a reduction serves two important purposes: It expands the explanatory scope of the symbolic models and it explains how some features of these models can actually emerge through learning in neural systems. To this end, I present a connectionist simulation of experimental data and show both the general feasibility of such a reduction and the specific manner in which the salient phenomenological distinction between form and meaning may be an emergent product of cortical memory processes. I argue this intertheoretic reduction of the symbolic to the neural serves an important goal of SLA, as these neural models can provide the theory of learning lacking in symbolic models of SLA.

Keywords SLA; connectionism; intertheoretic reduction; lexicon; lexical memory

Introduction

The emergence of connectionism in the mid 1980s (e.g., Rumelhart, Hinton, & Williams, 1986) resonated with many second language acquisition (SLA) researchers. Early discussions of connectionism in SLA focused on its potential as an alternative to the Universal Grammar (UG) orthodoxy (e.g., Sokolik, 1989). I thank Jan Hulstijn and Diane Larsen-Freeman for comments on an earlier draft of this article and Lourdes Ortega for expert and sorely needed editorial guidance. All errors and omissions remain my own.

Correspondence concerning this article should be sent to Robert Nelson, University of Alabama, Department of English, 103 Morgan Hall, Box 870244, Tuscaloosa, AL 35487-0244. Internet: rnnelson@bama.ua.edu

DOI: 10.1111/j.1467-9922.2012.00699.x
1990; also Broeder & Plunkett, 1994). Specifically, authors cited connectionism as providing an alternative to UG notions like the competence-performance distinction (e.g., Bialystok, 1990; Hatch, Shirai, & Fantuzzi, 1990) and the idea of a genetic pre-program for language that was resistant to instruction (e.g., Hatch et al., 1990; McLaughlin, 1990; Sokolik, 1990; Spolsky, 1988). Many early theorists also cited the neural plausibility of connectionist models (e.g., Ellis, 1999; Ney & Pearson, 1990; Sokolik, 1990) as an attractive feature, but stopped short of explaining how neural plausibility made connectionist models theoretically valuable.

In the later connectionist SLA (CSLA) literature, the implications of connectionism for a SLA theory of learning became more elaborate. In this work, the probabilistic nature of connectionist learning became a lens through which the relevant features of the input can be made clearer (e.g., Ellis, 2006a, 2006b; Harrington, 2002). Also, connectionist networks’ ability to integrate multiple, probabilistic cues (e.g., Ellis, 2002) and their ability to learn categories from exemplars (e.g., Larsen-Freeman & Cameron, 2008) made them ideal for use in support of a broader emergentist program informed by cognitive linguistics (e.g., Ellis, 2002, 2006a, 2006b; Larsen-Freeman & Cameron, 2008; O’Grady, 2008). The more recent CSLA models successfully implement this perspective, and networks have been used to support, for example, construction grammar, statistical learning, and cue integration interpretations of the learning of second language (L2) morphology, L2 case marking, L2 morphological gender, and verb-argument constructions (e.g., Ellis & Larsen-Freeman, 2009; Ellis & Schmidt, 1997; Kempe & MacWhinney, 1998; Taraban & Kempe, 1999). These later implementations attest to a mature CSLA—one that is thoroughly informed by linguistics and psycholinguistics and fully methodologically integrated, if still underrepresented by comparison to other more prolific research orientations in the wider SLA field.

Because connectionism views cognitive phenomena as interactions of many neuron-like units in a brain-like system, a touchstone of the discussions regarding the position of connectionism in SLA is its status as a theory vis-à-vis the more established information processing program that traditional psychology has made available for SLA (see Carroll, 1989). This latter approach views cognitive phenomena as the rule-governed operations of abstract processing modules on symbols in a computer-like system. That is, the predominant information processing view of cognitive SLA is based on symbolic models that are usually influenced by, and are sometimes attempts to instantiate, the UG perspective on SLA. From the beginning, authors have discussed the value of
hybridizing the connectionist and information processing approaches. Some have considered the benefits of such a union (e.g., Bialystok, 1990; Spolsky, 1988), while others have held it to be a priori impossible (e.g., Ney & Pearson, 1990; Sokolik, 1989). A smaller group has insisted that the role of connectionism in such a union would be limited to explaining peripheral phenomena (Truscott, 1998; Upsher, 1998). In contrast, Hulstijn (2002) predicts that an integration of connectionist and information-processing symbolic approaches, whatever form it takes, will lead to a unified account of L2 representation, learning, and processing.

My purpose in this article is to explore a fruitful line for the integration of neural and symbolic approaches to modeling L2 learning, a move which I believe has the potential to expand the role of connectionism in SLA theory. The article is organized as follows. In Part I, I discuss how the integration or hybridization of connectionist and symbolic models can be achieved through intertheoretic reduction (Churchland & Churchland, 1994; Rohrlich, 2001). I also introduce readers to form and meaning in lexical memory models and to issues of ontologies in intertheoretic reduction. In Part II, I demonstrate the general feasibility of a hypothetical intertheoretic reduction of symbolic and neural lexical memory models, discussing in some detail the necessary ontological correspondences that need to be made. In Part III, I present the design and actual implementation of a connectionist simulation study based on the bilingual and L2 lexical memory models discussed. This simulation study aims at modeling and replicating the emergence of Stroop-type form-meaning interference behavioral results obtained by Miller and Kroll (2002) with human participants. Throughout the article, I argue that intertheoretic reduction of the symbolic to the neural serves an important goal of SLA, as neural models can provide the theory of learning lacking in information-processing symbolic models of SLA, thereby expanding the role that connectionist research can play in SLA theory building.

**Part I: Connectionism in SLA and Intertheoretic Reduction**

The unification of connectionist and symbolic models is a promising direction of growth for CSLA. For this to occur, however, we have to recognize the fact that many (though not all) connectionist models are concerned with a different level of explanation than symbolic models. Loritz (1991) recognized this when he explicitly showed how a connectionist model might explain some central issues in SLA (e.g., fossilization and first language interference) as a product
of underlying cortical neurodynamics that do not need a higher-level symbolic interpretation. So did Fantuzzi (1992) when she asked if connectionist networks might only be low-level implementations of symbolic models.

Although she does not explicitly mention it, Fantuzzi (1992) assumes the framework first developed in Marr and Poggio (1976) for exploring visual cognition. Namely, she assumed, following these authors, that cognitive activity can be described at different levels (i.e., symbolic-computational, algorithmic, and implementational) and that the appropriate approach to explaining mental phenomena is top-down. This involves working out a so-called theory of overall computation at the symbolic level, and then worrying about how that could be implemented in a lower level of “transistors, neurons, diodes, and synapses” (Marr & Poggio, 1976, p. 2).

The Marr and Poggio approach has had tremendous influence on cognitive science and has entered SLA as a part of the background knowledge of many researchers. In developing their taxonomy of levels, however, Marr and Poggio (1976) did not touch on a number of issues. Two of them are of concern for the argument developed in this article.

The first issue is that the formulation of the different levels ignores the manner in which lower-level (i.e., implementational) explanations serve to both constrain and select higher-level explanations. As Churchland and Sejnowski (1990) explain, the number of potential symbol processing systems that can account for any particular set of behaviors is vast, and so a top-down approach is, by itself, unlikely to “light on the correct theory” (p. 368). Lower-level accounts help by constraining the kinds of higher-level explanations that are put forward and by selecting, among all that are put forward, those most likely to be correct. This selection can be achieved through what is known in philosophy of science as intertheoretic reduction (see Churchland & Churchland, 1994; Rohrlick, 2001; and see more details in later sections). In the present article, the intertheoretic reduction I will present is of the higher- to the lower-level theory (Churchland, 1985): A symbolic model of lexical memory will be reduced to a connectionist model.

The second issue is that Marr and Poggio’s (1976) different levels employ entities from incommensurate categories in their explanations. In models of lexical memory, which will be our concern in this article, word form and meaning nodes in a model framed at a higher level have very different properties than the neural networks described at a lower level. Together, these entities (nodes vs. neural networks) and their properties (e.g., symbolic content vs. activation functions) invoke fundamentally different categories of explanation (i.e., different ontologies), and it seems clear that any final explanation will
have to establish exactly how the different ontologies of the lower and higher levels are causally related.

My purpose in this article is to illustrate a direction of growth for another branch of connectionist CSLA, one in which networks are “hybridized” with symbolic approaches but that also recognizes that they address a lower level of explanation than symbolic models. I will do this by showing how higher-level explanations of L2 lexical access can be reduced to lower-level explanations and how their different ontologies can be causally related. Interestingly, this reduction will allow me to show that an issue that is generally taken as unproblematic is actually in need of explanation. This issue is that, universally in models of lexical memory, the distinction between the form of a word (phonological and orthographic) and its meaning (i.e., associated concepts and conventions of use) is taken as preexisting in the structure of the memory system.

The absence of a developmental account of the emergence of dedicated form and meaning processing areas implies a general belief that the underlying neural systems are able to segregate incoming signals according to their information content. This assumption automatically assigns the job of creating the form–meaning distinction to the neural systems. So, one should expect that, when a symbolic model of lexical memory is reduced to a connectionist model, the reduction would show how the form–meaning distinction may be created. In response to this requirement, I develop a full theoretical articulation (in Part II) and implementation (in Part III) of a connectionist simulation in which a system of networks that does not embody or learn the form–meaning distinction is able to dynamically generate it in response to task demands.

**Form and Meaning**

Following Garrett (1975), all symbolic models of bilingual and L2 lexical learning and memory have held that the form and meaning of a word are represented and processed in different stages of a multistage lexical access system. This general framework places phonological/orthographic encoding in the first stage and conceptual/semantic processes in a structurally distinct later stage of the system. All processing-only (i.e., steady-state) symbolic models of bilingual and L2 lexical memory (e.g., Costa, Caramazza, & Sebastián-Gallés, 2000; de Groot, 1992; Dijkstra & Van Heuven, 2002; Green, 1998) maintain this position as do models of bilingual or L2 memory that take a developmental perspective (Jiang, 2002, 2004; Kroll & Stewart, 1994; Li & Farkas, 2002; Zhao & Li, 2007).

This position has received converging empirical support from studies that show a lack of cross-language priming effects for homographic stimuli
(e.g., Gerard & Scarborough, 1989), indicating that representations of formal features are not shared between the languages, and from studies that show cross-linguistic semantic priming (e.g., de Groot & Nas, 1991; de Zeeuw, Verhoeven, & Schreuder, 2012), indicating a shared semantic store. It is also consistent, in broad outline, with what recent Magnetic Resonance Imaging, Event Related Potential, and Magnetoencephalography studies have revealed about the time course of processing (e.g., Gold & Rastle, 2007; Hauk, Davis, Ford, Pulvermüller, & Marslen-Wilson, 2006), namely, that processing in perception proceeds from a form-controlled stage (e.g., sensitivity to n-gram frequency and word length) to a processing stage relying more on conceptual and abstract representations.

One may insist that a division between (at least) conceptual and phonological representation probably arises as the natural product of learning, created when the learner associates linguistic forms with the objects, events, and relations that she experiences as being subjectively distinct from the language used to describe them. Three characteristics of the neural pathways involved in language use prevent the automatic generation of the form–meaning distinction from passively available information. First, language is a series of behavioral events embedded in concurrent streams of nonlinguistic behavioral and environmental events, none of which come to our senses in ready-made categories. Even the broad distinction between signs and referents is absent in the world outside the perceiving person and must be generated by the learner-perceiver. Second, the cortex, which plays the major role in language and memory functions, shows a laminar (or layered) cytoarchitecture within which incoming signals are repeatedly converged, integrated, and reclassified without regard for their subjective status, that is, regardless of whether a set of stimulus patterns is a part of a word’s form or its meaning. This means that not only is it difficult for information about meaning to be segregated out for representation at a level separate from information about form, it also implies that these two kinds of information are repeatedly blended together as signals pass up through the hierarchy. Third and finally, further complicating these issues is the fact that we cannot rely on the content of the information being processed to justify any part of our explanation. Neural signaling is homogenous, consisting only of spiking frequencies, so that the function of a signal is determined only by its location in the system and not by its putative information content.

These structural and functional features of cortical networks directly prevent a straightforward neural reduction of symbolic models that assume discrete stages of form and meaning representation and processing. Nodes and links may make a model look like the brain, but these conventions alone do not
bestow the explanatory properties of cortical memory networks. Some bridging explanation is also needed to show how the information needed to create a form–meaning distinction is generated within cortical memory networks. It is the job of a proper reduction of the higher-level symbolic form–meaning architecture to a lower-level connectionist architecture to explain how attested neurodynamics could generate the conditions necessary to create this distinction during learning, as it is neither evident in the sensory stimulus nor an a priori of the structure of the brain. This issue is important to models of bilingual and L2 lexical memory because these models depend critically, even more so than their monolingual counterparts, on the status of form and meaning subprocessing in language comprehension.

The Role of Ontologies in Reductive Explanation

Let us now return to the concept of intertheoretic reduction introduced earlier in order to examine it in more detail. Many questions cannot be satisfactorily answered at the level at which they are framed. For these questions, we generally seek an answer at a lower level of explanation. To use Rorlich’s (2001) example, if we ask why ice floats, we may be told that it is because ice is lighter than water. This is a limited explanation, though, as we have explained the buoyancy of ice only by stating what is necessarily true of all things that float. A satisfactory explanation is one that adds empirical content (Lakatos, 1968), and this is often best achieved by formulating an explanation at a lower level. In our ice example, when water freezes, the average distance between molecules increases, leaving fewer molecules in a given volume of ice than in the same volume of water.

What determines the level of the explanation is the ontology, or the system of meaningful categories, that characterizes the explanation. Explaining buoyancy by saying that ice is lighter than water uses the “ordinary or common level of human experience” (Rohrlich, 2001, p. 186) as the relevant ontology. That is, you can see ice float, and you can feel that it is lighter than the water, so you have explained its buoyancy at its own level. In contrast, the explanation at the lower level is characterized by an ontology that is not directly available to the senses: molecules and the nature of molecular bonds.

The position here is that SLA can add empirical content to its explanations by showing how they can be the products of the structure and dynamics of neural systems. To take a specific example, the interactions of first language and L2 lexical semantics in models of bilingual and L2 lexical memory depend on the partitioning of word memories into semantic and formal components (e.g., Jiang, 2002) or distributing them among different levels (e.g., Dijkstra & Van Heuven, 2002). These models explain interactions of form and meaning
by taking the phenomenological distinction between form and meaning (i.e., we can feel that form and meaning are different things and we can see it in other people’s responses to experimental manipulations) and assigning it to the underlying mechanism, such that the sensory ontology of the observations becomes the ontology of cognitive explanation. While these explanations may be correct, their assumptions need to be explained in lower-level terms. This is the purpose of a reductive explanation.

Reductive explanations become possible when the disparate ontologies of different theoretical levels are related by the successful intertheoretic reduction of the higher-level theory to the lower-level theory. These ontologies are linked through correspondence rules (Churchland, 1985) or ontological bridges (Rohrlich, 2001) that can take the form of laws, equations, or ordered pairs of features. These connections show that descriptions that are not logically equivalent to each other can in fact represent the same entities. Philosophers of science typically take their illustrations of these correspondences from physics (e.g., Churchland, 1985; Rohrlich, 2001). Equation (1),

\[ T = \frac{m v^2}{3 k B} \]  

is a correspondence rule saying that temperature \( T \) at the common level of experience is equivalent to the average mass and initial velocity of each molecule over Boltzmann’s constant \( k B \), at the molecular level). Because this particular correspondence is lawful, the relationship can be expressed as an equation. However, often the correspondence cannot be reduced with this regularity. For example, while the description of momentum in classical mechanics \( F = ma \) is accurate at the common level experience, it is also overfit to the ontology of the common level, and so it fails to be accurate at very high speeds. \( F = ma \) is therefore not equivalent to momentum in the more complete mechanics of relativity \( p = m_0 v \). We can formalize this correspondence rule as the ordered pair of equations in (2):

\[ \{ F = ma, \ p = \gamma m_0 v \} \]  

The paired expressions in (2) constitute a correspondence rule that relates classical to relativistic concepts of momentum, with the understanding that, at high speeds, one must use mathematics derived from the ontology of relativity to determine the momentum of a particle. Note that Equations (1) and (2) are
different from each other in another way: Equation (1) is a noneliminative reduction of temperature, while Equation (2) is an eliminative reduction. That is, the former explains temperature in terms of molecular activities without replacing the concept of temperature itself, whereas the latter shows that momentum in classical mechanics is actually incorrect and replaces it by explaining it as a special case of momentum in relativistic mechanics.

Employing much of the same logic as that just illustrated with the example of ice developed in this section, the structure and dynamics of symbolic models such as those used in most models of L2 lexical memory can also be reduced to the dynamics and states of a biologically derived artificial neural network. In the remainder of this article, I take up the challenge of demonstrating precisely just this process, first at the theoretical level (Part II) and later at the level of an actual simulation study (Part III).

**Part II: Theoretical Prerequisites to the Intertheoretic Reduction of SLA Models of the Mental Lexicon**

The goal in this part is to show that the structure and dynamics of one particular kind of symbolic model (used in most models of L2 lexical memory) can be reduced to the dynamics and states of a biologically derived artificial neural network. To lay the groundwork for this reduction, five things must be done: The relevant ontologies must be described, a suitable representation of the lower-level ontology must be implemented, the goals of the reduction must be explicitly formulated, correspondences between the relevant ontologies must be identified, and the correspondences must be operationalized as ordered pairs.

**Relevant Ontologies**

The relevant ontologies are (1) the symbolic ontology of information processing models and (2) the neurodynamic ontology of the brain, as it is represented in biologically faithful connectionist models. The symbolic ontology is semantic in nature. That is, it includes meaning as an a priori in the form of categories (embodied in preexisting symbols or nodes) and it views a linguistic stimulus as a meaning-bearing signal. The neurodynamic ontology is physical in nature. It includes only generic interactions (e.g., activation and inhibition) and a structure within which these interactions occur (i.e., a network). It also includes uniform principles specifying how these functions can interact with the structure (i.e., learning algorithms). In this ontology, linguistic stimuli are patterns of activity at a sensory surface or an input layer with no intrinsic semantic value (see Edelman, 1987).
The Lower-Level Ontology

The second step is the identification of a representation of the lower-level ontology. In a research project that began in the mid-1960s and that has continued on to the present day, Stephen Grossberg, Gail Carpenter, and their colleagues (Grossberg, 1967, 1973, 1976, 1986; also Carpenter & Grossberg, 1987; Grossberg, Boardman, & Cohen, 1997; Grossberg & Seidman, 2006) integrated the discoveries in neuroscience of Helmholtz (2005/1867), Hebb (1949), Hodgkin and Huxley (1952), Hartline and Ratliff (1957), and others into a single mathematical model of how behavior emerges from the learning and memory functions of the cerebral cortex. This theory, called Adaptive Resonance Theory (ART), has spawned a family of biologically derived connectionist networks. Because these models are derived from neuroscience, it is claimed here that they provide the basis for a strong neurodynamic interpretation of their behavior, making them suitable stand-ins for the neurodynamic ontology of biological tissue at the systems level.

Loritz (1991) describes ART learning in terms of its historical and biological foundations, while Nelson (2011) presents an algorithmic overview. Interested readers can find detailed C-style pseudocode for the ART2 networks used here in Freeman and Skapura (1992, pp. 329–335). The critical elements of ART learning are competition, vigilance, and resonance. Competition occurs when an ART network encounters a stimulus pattern, and it must decide which of its previously learned memory categories this stimulus belongs to. It does this by allowing all long-term memory (LTM) patterns to compete with each other so that the best matching pattern wins. Once a memory pattern has won, a process similar to the low-level orientation mechanism described in SLA by Tomlin and Villa (1994) determines whether the registered stimulus and the existing memory pattern are actually a good fit. Vigilance is the parameter that defines this goodness of fit (Carpenter & Grossberg, 1987). Vigilance “calibrates the network’s ‘attentive sensitivity’” (Carpenter & Grossberg, 1991, p. 314). Finally, resonance occurs if the similarity between the stimulus and stored memory exceeds that specified by the vigilance parameter, and the existing memory pattern is therefore adjusted to allow it to better respond to that pattern in the future. It is worth noting here that this process of resonance is strikingly similar to Schmidt’s (1995) conceptualization of awareness.

Goals for the Proposed Intertheoretic Reduction

The third step is to explicitly formulate the goals of the reduction. The goal of our particular reduction of symbolic models of SLA lexical memory to their neural systems counterparts is to account for what must be an emergent
property of the higher-level model, namely, the structurally encoded form–meaning distinction. Any property is emergent when it occurs at some level of description, and has definite properties at that level, but does not occur at the lower level of description (see Ellis, 1998). Many of the entities assumed by symbolic models do not have apparent neural correlates. However, neural system-level models exhibit dynamic states that can be viewed as the neural correlate of a durable entity in the higher-level symbolic model. These states would emerge as the product of competition or cooperation between LTM dispositions that are reliably enacted when specific classes of stimuli are present, and as such they can be examined only through dynamic neural models of the relevant structures and situations.

**Correspondences Between Ontologies**

The fourth step in intertheoretic reduction is the identification of correspondences between the relevant ontologies. As it is the goal here to reduce the symbolic models of lexical access to a neural model with regard to the emergence of the form–meaning distinction, we have to look at both models in general outline. Especially, we are interested in what each account holds to be (a) the primitive unit of processing and (b) the principle or principles by which these primitive units interact.

Most symbolic models of bilingual and L2 lexical memory assume the general form of de Bot’s (1992) adaptation of Levelt’s (1989) monolingual model (e.g., Costa et al., 2000; de Groot, 1992; Jiang, 2002) or follow the Interactive Activation framework of McClelland and Rumelhart (1981), as do, for example, Dijkstra and Van Heuven (2002). In these models, there are levels corresponding to concepts, lemmas, and phonological codes, and in each level there are nodes that represent individual concepts, lemmas, or phones that are joined between levels by links. Word recognition occurs when the appropriate nodes are activated through links. For example, if one wishes to say *chair*, the concept node CHAIR is activated at the concept level, this activation then travels through links to the lemma nodes, where the information needed to associate the word with an appropriate construction is activated. Then the activation flows to the nodes at the phonological level, where the correct string of phonemes is activated. In feature-based (de Groot, 1992) and sense models (Finkbeiner, Forster, Nicol, & Nakamura, 2004), conceptual nodes are further broken down into nodes for semantic features. In all of these models, the nodes are the primitive units of processing and the links function as their principle of interaction.
By contrast, in a lower-level cortical memory network, lexical access is achieved through pattern completion. In this way, word recognition is not different from any other type of memory recall: It is the retroactivation (Damasio, 1989) or reactivation (Fuster, 2003) of a complete memory pattern after the activation of a subpart of that pattern by an external or internal event that was associated with that pattern by prior learning. This causes a number of larger memory networks to become partially active. This short-term memory (STM) activation transfers instantly and automatically from the smaller to the larger networks because multiple memories are partially superposed over the same connections. These larger memory networks may be associated with other networks encoding fragmental (i.e., noncomprehensive) records of phonological, orthographic, and syntactic information, and they will compete by inhibiting each other’s activation. These inhibitory signals will be proportional to the amount of activation each larger network receives from the active smaller networks and the larger network that receives the most activation from all of the active smaller networks will win, activating associated networks (see Damasio, 1989; Grossberg, 1980; Fuster, 1997). In this account, the STM pattern is the primitive unit of processing, and the incidence of memory superpositioning is the principle of interaction.

**Operationalization of Correspondences as Ordered Pairs**

Given the ontological differences just identified, the fifth step can proceed, and the following operationalization of identified correspondences as ordered pairs is proposed.

**Primitive Units of Processing: Nodes and STM Patterns**

The first bridge between ontologies can be achieved in the pairing of the nodes of symbolic models and the STM patterns of biologically derived neural networks, which are the minimal units of processing in their respective theoretical environments. In spite of their common functionality, the different ontologies in which they are realized give them different natures.

Nodes have two properties that allow them to function as the primitive units of processing in a symbolic ontology: They are durable and they have content. They are durable because they are discrete entities whose existence does not depend upon the states of the overall system. That is, they function independently of any context (see Hulstijn, 2002) and so must acquire their function from their semantic content. The primitive units of processing in a biologically derived connectionist model are STM patterns. These are patterns of activation caused by an encounter with an external stimulus or the activation
of an associated memory pattern. STM patterns are enacted and content free: They are enacted in that they exist (i.e., only have causal efficacy) when a LTM network of which they are a part is activated by a stimulus. They are content free because they acquire their function solely from their position in the processing stream (making them context dependent).

Principles of Interaction: Links and Superpositions
In both ontologies, the primitive units of processing must interact in some principled way for any work to be done by the system. In most models of L2 and bilingual lexical memory, these principles are links that allow nodes to transfer activation from one to another. The links of symbolic models may be interpreted as mapping onto neural pathways that allow the activity of a STM pattern in network A to activate a STM pattern in network B in the way that, for example, a concept node might activate a set of phonological nodes. This is only a partial interpretation of the complete process, however. It is more accurate to say that the STM patterns are both superposed throughout connections of a superordinate cortical memory pattern X and that the activity of STM pattern A causes the activation of X and all relevant patterns that are superposed in the connectivity structure of X, including B. One advantage of describing the transfer of activation as occurring between superpositioned representations is that the empirically attested co-activations of similar conceptual memories (e.g., translation equivalents, cognates) are inherent in the structure of the system.

Emergent Phenomena
There is still a discontinuity between the symbolic and neural interpretations of lexical access that prevents the easy reduction of the former to the latter. That is, it is still unclear how multiple, discrete, content-specific stages of representation map onto an integrated and distributed network memory. It is the goal here to show that this division can be dynamically constructed by competing STM patterns in integrated lexical memory networks so that the form–meaning distinction at least begins as an emergent property of the interactions between cortical memory networks.

Target for Simulation: Miller and Kroll (2002)
What follows in the remainder of this part, before the actual simulation study is described in Part III, is a hypothetical connectionist simulation of the emergence of concept and form levels in a self-organizing network, in which a network that integrates the formal and conceptual dimensions of word representation into a single LTM pattern is able to respond to manipulations
of form and meaning as if these subjective distinctions had been designed into it.

A division of labor between the two stages of representation and processing emerges most clearly in an experimental paradigm developed by La Heij, de Bruyn, Elens, Hartsuiker, & Helaha (1990) and used in Miller and Kroll (2002). In this paradigm, participants are shown a word and asked to translate it. However, shortly after the word to be translated is presented, a different word is briefly flashed on the screen as a Stroop-type distractor. In Experiment 1 in Miller and Kroll, Spanish–English bilinguals saw a word in either Spanish or English that they were to translate into their other language. However, shortly after the presentation of the word to be translated, either a form-relative of the translation or a meaning-relative of the translation was presented. Two stimulus-onset asynchronies (SOAs) were used, 200 milliseconds and 500 milliseconds. Results from both SOA conditions showed that semantically related distractor words produced translation interference (i.e., increased the time it took to translate the word) while form related distractor words facilitated translation (i.e., decreased the time it took to translate the word). These results are consistent with a multistage processing scheme in which the intended language plays a role in the activation of the concept nodes stored at the concept level (e.g., Broersma & de Bot, 2006; Costa et al., 2000; de Bot, 1992; Green, 1998).

Both symbolic and connectionist accounts of the dynamics of the lexical memory system agree that, in order for exposure to a priming word to facilitate the recognition of a target word, it must somehow raise the readiness (or activation) level of the mental representation of the target. Likewise, for an exposure to one word to interfere with the recognition of a target, it must somehow inhibit the activation of the mental representation of the target. Symbolic and connectionist accounts disagree, however, over the mechanisms that enact these influences. In Miller and Kroll (2002), the interference effect is interpreted as occurring when a semantically related distractor word activates a language node at the concept level, which functions as a bias unit and inhibits activation at the phonological level (p. 262). The facilitation effect occurs because the language cue in the conceptual level is not activated by the distractor word, while (presumably) the lexical form cues are activated at the form level.

The present connectionist account is substantially different. Figure 1 presents two different (but overlapping) three-unit anatomies from an idealized (i.e., not actually modeled) neural system. This hypothetical neural system consists of an input layer (F₁) and a classification layer (F₂). The units in the F₂ layer are in exploded view, while only the relative STM activations of the
Figure 1 Interactions in a network with lateral inhibition receiving activation. Frame (a) shows that recognition of the stimulus niño (N) also co-activates the translation boy (B), while laterally inhibiting it. In frame (b), the distractor word girl is presented too briefly for recognition to happen, but allows (g) to continue the inhibition of (B) by virtue of its formal properties only. Frame (i) shows how the recognition of niño affects the units for toy (T) and (B) differently. Frame (ii) shows how the formal properties of the word toy can facilitate the activation of the boy (B) unit, continuing the co-activation begun by the recognition of niño.

units at the F₁ level are indicated. Patterns of STM activation are shown as gray polygons above the levels. The first anatomy (frames [a] and [b] in the top row of the Figure) consists of three F₂ units that have learned to respond to the stimuli niño (“N,” in Figure 1), girl (“G”), and boy (the translation of niño; “B”). Frame (a) shows how this anatomy responds to the complete activation of the STM pattern for niño, as it would occur when the word is recognized (i.e., after the orthographic stimulus is associated with the complete memory pattern). Frame (a) also shows that all units in this anatomy receive activation from this STM response, with the F₂ “N” unit receiving the most activation and the “G” and “B” units receiving partial activation. This happens because
the semantic/conceptual components for all of these memory categories are partially superposed at the F\textsubscript{1} layer. Patterns of inhibition within the F\textsubscript{2} level are also shown, filled boxes denote strong inhibition, and open boxes denote moderate inhibition. The inhibitory signal sent out by a unit is proportional to the activation it receives.

Frame (b) shows the same anatomy immediately after the orthographic form of the meaning distractor, \textit{girl}, is briefly presented to the F\textsubscript{1} layer (as in Miller & Kroll, 2002). Here, only the “G” unit receives significant activation—as there is little overlap in form between the words \textit{girl, niño, and toy}. Frame (b) also shows that there is only partial F\textsubscript{1} activity generated by the orthographic stimulus, as resonance cannot occur during this brief presentation. Here, relative to their previous activation states, activation in the “G” unit increases and activations in the “N” and “B” units decrease, as they receive moderate inhibition and little or no activation. This activity pattern would delay, or inhibit, any behavior that depended upon the activation of the memory pattern corresponding to the stimulus pattern \textit{boy}.

Frames (i) and (ii) in the bottom row of Figure 1 show the “N”, “T”, and “B” unit anatomy corresponding to the memory patterns for the words \textit{niño, toy, and boy}. Frame (i) shows that, when the complete STM pattern for \textit{niño} is active, the “N” and “B” units are active, with the “N” unit most active, because of resonance with the stimulus pattern \textit{niño}. Frame (ii) shows the same anatomy when the orthographic stimulus pattern \textit{toy} is presented. Here, the “B” unit, which was previously partially active, receives activation from the STM pattern for the form-similar stimulus, \textit{toy}. This activity pattern would advantage, or facilitate, any behavior that depended upon the activation of the memory pattern corresponding to the stimulus pattern \textit{boy}.

The foregoing shows that it is at least conceptually possible for patterns of behavior consistent with separate stages of form and concept representation to emerge from a system that does not instantiate these stages, or even their content. In the case of Stroop-like interference, these patterns may thus emerge because of dynamic interactions between different STM patterns that are superposed over the same memory connections. What remains is to see if these patterns actually will emerge in an appropriately trained network.

**Part III: The Simulation Study**

What follows in this final part is an actual connectionist simulation study of the emergence of Stroop-type interference (La Heij et al., 1990; Miller & Kroll, 2002) similar to the results obtained in Miller and Kroll’s (2002) Experiment 1.
This is not an attempt to simulate the complete results of that study, however, only the phenomenon of different effects of form and meaning in a laboratory word translation task.

Data
The data were taken from the Spanish–English bilingual corpora of the CHILDES database (MacWhinney, 2005). Within these data, it was possible to identify twelve translation pairs along with words that were related to one of the translations either formally (having at least a 50% orthographic overlap) or semantically (being close in meaning) for a total of 48 words, presented in the Appendix. An additional 246 words were also chosen at random from the database sample to serve as controls and to extend the training sample toward a better representation of real learning situations. Of the total 294 words, 110 were assigned orthographic representations in the final training and simulation. The input data (i.e., words) were coded according to grammatical category, semantic category, phonological feature, and orthography.

The lemmatization of the data involved coding according to 11 grammatical features, which included the basic parts of speech, plurality, question word, and social function word (e.g., “thanks”). They were then coded according to 21 semantic categories. For nouns, these categories included basic thematic roles, derived from Van Valin (2005), and animacy. For verbs, the categories used were taken from the verb matrix in Cook (1979). For adjectives, the words were coded according to the four basic adjective types (dimension, age, value, and color) given in Dixon (2004). Because adverbs are a highly variable category, they were coded using the criteria for verbs and adjectives together, while determiners were coded only as grammatical function words and prepositions were coded as function words plus the nominal role assignment typically associated with them, so that, for example, “at” was coded as a function word assigning the role of location.

The phonological form of each word was presented to the network as a series of overlapping pairs of phones (e.g., [trΛk] = tr, rΛ, Λk) and each phone was coded according to feature (manner, place, and voicing) for consonants, and according to actual first and second formant values for vowels. These values were taken from a native speaker of U.S. English and a native speaker of Peruvian Spanish. Each phone was coded as a string of 25 comma-separated values, so each input pair consisted of 50 values. For the vowels, formant values were divided by 100 so that they would fit into the 25-value array.

Orthographic inputs were presented to the network as a series of overlapping character triplets (e.g., TRUCK = TRU, RUC, UCK) with each character in
the form of a $5 \times 5$ array (Figure 2 shows actual input patterns). While this input scheme is simple, it operationalizes the same type of information made available at the fovea during the fixation of the eye while learning to read, and it does so without using the more abstract featural descriptions of other models, thus providing for better contact with actual learning data. Because each letter was a $5 \times 5$ (25-unit) array, and letters were presented three at a time, these input patterns contained 75 values.

**Training**

It is desirable to remove coding bias from the training data, otherwise a network model can be criticized for including “a priori solutions to the problem” (Ellis, 2005, p. 321). It is also desirable for the model to recognize and represent the layered architecture of the cortex. To achieve both of these goals, the raw inputs were learned by six ART2 networks before final training, as shown in Figure 2. Only the semantic input to network 1 (in Figure 2) is open to coding bias, as all other inputs are determined by physical characteristics of the actual data or by established categories. Any bias that was not eliminated from the semantic input would have been obscured by being filtered through at least two networks before reaching the highest network in the hierarchy, the Converge-Integrate network (see Figure 2). That is, only networks 1, 2, and 5 were exposed to coded data. All other networks learned only the classifications of lower level networks. The learning of the Converge-Integrate network, which is the final network, was the subject of analysis and testing.

Traditional training parameters for ART2 networks are vigilance (or $\rho$), number of units in the input ($F_1$) layer, number of units in the classification
(F₂) layer, A, B, C, D, and θ. A, B, C, D, and θ are constants for the F₁ level unit activation functions.² For all networks shown in Figure 2, parameters A, B, C, and D were set to 10, 10, 0.1, and 0.9, respectively, and θ was set to 0.1. Except for θ (see note 2), these are the parameter settings described for the original ART2 model in Carpenter and Grossberg (1987).³ Testing the robustness of this simulation under different parameter settings would be an interesting and informative project that is unfortunately outside of the scope of this article (however, the results reported below are robust for all A and B greater than 0). When the above parameters are held to these values and vigilance is left as the only free parameter, the criterion level of 110 unique word categories in the highest Converge-Integrate network is attained in each condition. Vigilance levels for each network are given below.

The combined grammatical and semantic features were fed into an ART2 network (1 in Figure 2), which derived 162 grammatical-semantic categories (ρ = 0.99999). The phonological inputs were fed into a second ART2 network, which generated 419 categories (one for each unique phone pair, ρ = 0.9999). Phonological representations of entire words were developed by buffering these 419 patterns in a third ART2 network, which learned them as 294 categories (one per word, ρ = 0.9999). The output of this network and the grammatical-semantic network was fed into a fourth ART2 network, which derived 294 unique representations (one per word) that integrated phonological, semantic, and grammatical information. Orthographic trigrams were learned as 302 unique patterns in a fifth ART2 network (ρ = 0.9999). These patterns were buffered into whole-word orthographic representations and then presented to a sixth ART2 network that categorized them as 110 whole-word orthographic patterns. The final Converge-Integrate network, the subject of analysis, took as its input patterns the outputs of the fourth and sixth networks, as shown in Figure 2.⁴ The Converge-Integrate network was trained only on the subset of data for which orthographic representations had been specified, and developed 110 unique categories (ρ = 0.99995). Each one of these categories integrated the entire grammatical, semantic, phonological, and orthographic representation of each word into a single, nondecomposable memory pattern.⁵

Testing

In the Stroop-like task of Miller and Kroll (2002), participants are exposed to a word and asked to translate it. However, following the presentation of the word to be translated, a distractor word is briefly shown on the screen. Measurements are then taken of the length of time it takes the participants to produce the translation of the target word. In Experiment 1 of Miller and Kroll,
the distractor words were semantically related to the translation, formally related to the translation, or they were unrelated controls. Miller and Kroll presented the target words for 200 milliseconds, and the distractors at SOAs of 200 or 500 milliseconds. The distractors were presented for 100 milliseconds. This scheme is replicated here: The target words are presented to the network for 200 processing cycles, and distractors are presented at SOAs of 200 and 500 processing cycles and the distractors are presented for 100 cycles (the intent is not to suggest identity between cycles and seconds, but to ensure that the between-condition differences in SOA are faithfully represented). In Miller and Kroll, all stimuli are presented visually. This scheme is replicated here, as all stimuli are presented through networks 5 and 6 (see Figure 1).

The relative activation levels of memory patterns encoding the target, the translation of the target, and the distractor were recorded at the final cycle of distractor presentation, that is, at 300 (short SOA) and 600 (long SOA) cycles of processing. These results were compared with the following measures: a $t$ test (recommended in Flexer, 1994), correlations (Pearson’s $r$ and Spearman’s rho, $r_s$), and effect size (Hedge’s $g$) with accompanying confidence intervals at 95% level (CI.95).

**Research Question and Hypotheses**

Miller and Kroll (2002) found that words followed by meaning-related distractors took longer to translate than words followed by an unrelated control, while words followed by a form-related distractor were translated more quickly than words followed by a control. Both the symbolic account of this Stroop-like effect and the connectionist account outlined above agree that facilitation occurs because the distractor somehow raises the readiness (or activation) level of the mental representation of the translation and that interference occurs because the distractor inhibits the activation of the mental representation of the translation. The question here is whether a system that is not structurally biased to produce that result can show it. If the network is to accurately match the behavioral results obtained by human participants by Miller and Kroll, we should expect to see, at the end of the presentation of the distractor, evidence that the activation of the memory pattern for the translation of the target word is: (1) higher in the form distractor condition than in the control condition and (2) lower in the meaning distractor condition than in the control condition. It is noted here that, during testing, the networks do not produce the translations of the critical words, as did the subjects in Miller and Kroll. What the networks are expected to show are *biases for* and *biases against* the production of words that bear particular relationships to other words that are used as priming stimuli (i.e.,
form- or meaning-related distractors), while not showing these biases when unrelated controls are used as priming stimuli. These biases can be measured directly from the activation levels of $F_2$ units and should be proportional to the latencies described in the results of Miller and Kroll, where “semantically related distractors produced interference and form-related distractors produced facilitation, relative to unrelated controls” (p. 623).

Results and Discussion

Figure 3 shows an example of how the activation level of the memory pattern boy changed over time during network processing in the relevant conditions.

In all frames in Figure 3, the level of activation of the units is shown by the $y$ axis and the processing cycle is shown on the $x$ axis. Words at an angle above each frame show the stimulus presented to the network during those processing cycles. For all frames, from cycle 0 to 100, all STM activity is from only the orthographic pattern for niño at the input. At cycle 101, the winning memory pattern was allowed to respond and activate its entire STM pattern until the stimulus word (niño) was removed at cycle 200. In the short SOA condition, the distractor was presented during cycles 200 to 300, and in the long SOA condition, the distractor was presented during cycles 500 to 600.

Figure 3 shows that, at the end of the presentation of the distractor in the short SOA condition, the activation level of the memory pattern for boy (the translation of niño) is higher than the control in the form distractor condition (5.5 vs. 5.11) and lower than the control in the meaning distractor condition (4.35 vs. 5.11). In the long SOA condition, the same pattern holds, with boy active at a higher level in the form distractor condition than in the control condition (2.64 vs. 1.84) and less active than the control in the meaning distractor condition (1.55 vs. 1.84).

Tests of all 12 sets reveal that this pattern occurs every time: All reported measurements are activation levels of the translation of the target at the end of the presentation of the distractor—300th cycle for short SOA and 600th cycle for long SOA. In the short SOA condition, the mean of the translation activations in the meaning distractor condition ($M = 3.63, SD = 0.97$) was reliably lower than that in the control condition ($M = 4.44, SD = 1.16$), $t(11) = -5.6, p < .001$, with a considerable effect size, $r(10) = 0.91, p < .001, r_s(10) = 0.87, p < .001, g = 0.73$ (CI.95 = −0.09, 1.55). Also in the short SOA condition, the mean of the translation activations in the form distractor condition ($M = 6.46, SD = 1.23$) was reliably higher than in the control condition ($M = 4.44, SD = 1.16$), $t(11) = 8.82, p < 0.001$, with a considerable effect size, $r(10) = 0.75, p = 0.005, r_s(10) = 0.71, p = 0.005, g = 1.63$ (CI.95 = −0.71, 2.56).
Figure 3  Activation profiles of units classifying niño, girl, toy, and boy as they evolve over time. These patterns fit the Miller and Kroll (2002) data closely, indicating that the network dynamically constructs the form-meaning distinction in response to task demands, through the interactions illustrated in Figure 1.

In the long SOA condition, the mean of the translation activations in the meaning distractor condition (\(M = 1.46, SD = 0.35\)) was reliably lower than that of the controls (\(M = 2.26, SD = 0.57\)), \(t(11) = -7.5, p < .001\), with a substantial effect size, \(r(10) = 0.73, p = .007\), \(r_s(10) = 0.64, p = .013\), \(g = 1.63\) (CI.95 = 0.71, 2.56). The mean of translation activations in the form distractor condition (\(M = 4.34, SD = 1.16\)) was higher than that of controls (\(M = 2.26, SD = 0.57\)), \(t(11) = 7.47, p < .001\), with a strong effect size,
Figure 4  Averages for all translations in all conditions as they evolve over time.

$g = 2.24$ (CI.95 = 1.22, 3.27), but not with a significant correlation, $r(10) = 0.52, p = .085, r_s(10) = 0.43, p = .084$. Figure 4 presents the average activations for only the translations for all twelve sets. Here, the effects of the various distractor conditions are clear.

The simulations revealed that a neural system that is structurally insensitive to the subjective distinction between the form and the meaning of a word is able to dynamically construct that difference in response to task demands. This difference, once constructed, is available as data for higher-level learning, where it can then be durably instantiated in a cortical memory network. Something like these interactions may be responsible for our ability to disassociate the form and meaning of a word, both consciously and preconsciously, and “enact” them as objects of attention.

As a final note to the discussion of the simulation study results, the touchstone criteria of (primary) data contact, task veridicality, and input representativeness, set out by Christiansen and Chater (2001), are met in the tested model: The network learned in an environment composed of veridical representations of actual linguistic data and then showed human-like performance on a close simulation of a psycholinguistic task. An additional criterion of neurodynamic interpretability is also met by employing network algorithms and structures that are an attested subset of actual cortical dynamics and architectures.

**General Discussion and Conclusion**

In this article, I have elaborated on intertheoretic reduction as a way in which connectionism might expand the role it can play in SLA theory building in
the future, specifically by demonstrating how a symbolic model of bilingual and second language lexical memory can be reduced to a biologically realistic (i.e., neurally plausible) connectionist model. A question that may be asked is this: What value is gained from a reduction? I have argued that the gain is an increase in explanatory scope—specifically, the ability to link representation to learning, perception, and attention. It may be objected that this initial attempt at an intertheoretic reduction achieved little, as it only moved from a model constructed from symbolic nodes and synapse-like links to a more biologically informed version of connectionism. After all, these two formalisms are often grouped together under the connectionist rubric (see Hulstijn, 2002). However, these two approaches are radically different in terms of their underlying ontology, and the conflation of these two kinds of model is a category error based on the similarities they seem to show when graphically presented. Node-and-link type symbolic models are no less symbolic than the information processing models of UG-SLA, and they are no more capable of embodying the dynamics of change and growth than these models.

Connectionist models can answer questions that symbolic models cannot even express. In the models of second language lexical memory that feature symbolic nodes joined by implicational connections, specific groups of nodes and specific patterns of links are designed into the model because it is those arrangements that lead the model to success, typically with particular nodes linked because they show either a named semantic relationship or a named implicational relationship (e.g., X is cognate to Y or X is a subset of Y). For example, a researcher who wants to study the effects of cross-language cognates on lexical recall in bilinguals may begin by constructing a symbolic model in a node-and-link form for Spanish–English bilingual lexical memory. He or she may choose to link the nodes for “globe” and “globo” to overlapping sets of meaning nodes because they are translations and then link them to many of the same phonological and orthographic nodes because they share subsets of forms. While these linkages may be motivated by experimental data, linguistic facts, and classroom observation, they also automatically build the cognate status of the two words into the model. This creates a circular explanation that is actually seen in models of second language and bilingual lexical memory, and that is in fact unavoidable if we build models from categories that are derived from a high-level (if not folk-level) analysis of the problem. The effect of this circularity is to curtail any lower-level explanation involving learning, perception, or attention, forcing us to explain cognate facilitation by referring only to what is necessarily true of all cognates. The question to address is not how, for example, form and
meaning nodes might be linked to word nodes in order to bring a model to
success, but rather how perception, experience, learning, and attention cause
phenomenological categories, like form and meaning, to arise in the system in
the first place.

The question I have attempted to take seriously in this article is how we
transition from explanation framed in the folk-psychological terminology of the
lexicon (i.e., concepts, lemma, phonological encoding) and toward explanations
framed in the lower-level physical dynamics of STM and LTM interactions in
distributed cortical memory networks. I believe this simulation model, albeit
limited, has demonstrated a nontrivial reduction of the ontology of symbolic
models to the physical ontology of cortical neurodynamics.

Revised version accepted 4 January 2012

Notes

1 The relationship between models and theories is the subject of much debate (see
Odenbaugh, 2011). For this article, a model is a mental image of a complex
phenomenon that aids understanding by framing the phenomenon in terms of more
familiar elements and relations. A model may be a pretheoretic hypothesis or it may
be a projection of a formal theory into a more familiar and testable form. The
models of lexical memory presented in this article are pretheoretic hypotheses.

2 The parameters listed in the text operate in the F₁ and orienting subsystem layers of
the ART2 network. The ART2 F₁ layer consists of six sublayers, which together
normalize the input pattern, and three stand-alone inhibitory units that provide gain
control on sublayer activation levels. Carpenter and Grossberg (1987) indexes these
layers as w, x, u, v, p, and q. The F₁ sublayers are interconnected on a unit-to-unit
basis. Units in all sublayers obey Equation (1). In Equation (1), \( x_k \) is the activation
level of a particular unit, \( J_k^+ \) is the excitatory input from lower F₁ layers, and \( J_k^- \)
is inhibitory input from competing units in the same layer.

\[
\frac{dx_k}{dt} = -Ax_k + (1 - Bx_k)J_k^+ - (C + Dx_k)J_k^- \tag{1}
\]

\( \theta \) is the F₁ level noise threshold employed in sublayer v. In this model, units in v
obey linear threshold function (2) as in Carpenter and Grossberg (1987), but
sigmoid functions are also sometimes used. \( \theta \) is set lower in this model than in
Carpenter and Grossberg (1987) because the lower activation values in this
implementation were more critical to category determination than were those in
\[ f(x) = \begin{cases} 0 & 0 \leq x \leq \theta \\ x & x > \theta. \end{cases} \] (2)

The vigilance parameter sets the reset condition for the network in the orienting subsystem, which is a layer of units that computes a vector, \( r \), from the activity of sublayers \( u, c, \) and \( p \). The activity of a unit on \( r, r_i \), is given in (3):

\[ r_i = \frac{u_i + cp_i}{|u| + |cp|} \] (3)

and the vigilance parameter, \( \rho \), sets the condition for reset as (4):

\[ \frac{\rho}{|r|} > 1 \] (4)

3 ART2 \( F_2 \) layer units also obey Equation (1), with \( J_k^+ \) representing input from \( F_1 \) and \( J_k^- \) representing lateral inhibition (i.e., from other \( F_2 \) units). In this model, \( J_k^- \) is (5):

\[ J_k^- = \frac{1}{n} \sum_{j \neq k}^{n} x_j \] (5)

ART recognizes that making inhibitory connections subject to change through learning leads to unstable representations (Grossberg, 1987), and so inhibition is always as in (5). Inhibition drives competition between \( F_2 \) units. Resonance, or learning, occurs in the intralevel connections (\( F_1 \rightarrow F_2, F_2 \rightarrow F_1 \)) when a \( F_2 \) unit, \( T_j \), at the \( F_2 \) level, \( T \), wins the competition, according to (6), where \( D \) is the constant from Equation (1).

\[ T_j = \begin{cases} D & \text{if } T_j = \text{Max}\{T\} \\ 0 & \text{otherwise} \end{cases} \] (6)

After winning, the intralevel connections are updated according to (7), where \( T_{ij} \) denotes a top-down connection between \( F_1 \) layer unit \( i \) and \( F_2 \) layer unit \( j \). Both top-down and bottom-up connections obey the same equation in ART2.

\[ \frac{dT_{ij}}{dx} = F_{2j}(F_{ij} - T_j) \] (7)
4 This scenario implies the operation of entities that function something like the conjunctive neurons that Simmons and Barsalou (2003) propose in their interpretation of Damasio’s (1989) Convergence Zones. Damasio holds that “(i)n recollecting a perceived object, conjunctive neurons in CZs (Convergence Zones) re-enact the sensory-motor states active while encoding it. Similarly, when representing a concept, conjunctive neurons in CZs reactivate the sensory-motor states characteristic of its instances” (1989, p. 455), while Simmons and Barsalou additionally posit that these neurons also have representational status.

5 Figure 2 shows that the auditory/phonological stream is joined to the conceptual stream in network 4 before union with the visual/orthographic stream in the Converge-Integrate network. This design choice was made for two reasons. First, joining the phonological to orthographic stream would design the solution into the system. Second, during development lexical memory is established first as the union between phonological form and other memory systems. I thank an anonymous reviewer for pointing out that this choice may also be consistent with the experimental support for the existence of a phonological loop in STM (Baddeley & Della Salla, 1966). This scheme is also coherent with the observation that the phonetic pathway (into lexical memory) is shallower than the conceptual pathway. That is, linguistic forms are first processed and related to other linguistic forms before being processed for conceptual content (for a review, see Barsalou, Santos, Simmons, & Wilson, 2008).

6 Hedge’s g is used as the effect size measure because it corrects for biases due to small sample size.

References


Grossberg, S. (1986). The adaptive self-organization of serial order in behavior:


Appendix

Stimulus Sets Used in the Testing of the Network Model

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Meaning Distractor</th>
<th>Form Distractor</th>
<th>Control Distractor</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIÑO</td>
<td>GIRL</td>
<td>TOY</td>
<td>SHOP</td>
<td>BOY</td>
</tr>
<tr>
<td>ESTRELLA</td>
<td>MOON</td>
<td>STAIRS</td>
<td>GRANDMA</td>
<td>STAR</td>
</tr>
<tr>
<td>MILK</td>
<td>COMIDA</td>
<td>COCHE</td>
<td>GALLO</td>
<td>LECHE</td>
</tr>
<tr>
<td>CABALLO</td>
<td>COW</td>
<td>HOUSE</td>
<td>TOYS</td>
<td>HORSE</td>
</tr>
<tr>
<td>CHICKEN</td>
<td>GANSO</td>
<td>GALLETA</td>
<td>SEMANA</td>
<td>GALLINA</td>
</tr>
<tr>
<td>DOG</td>
<td>GATO</td>
<td>HIERRO</td>
<td>DOLOR</td>
<td>PERRO</td>
</tr>
<tr>
<td>TIENDA</td>
<td>SHOP</td>
<td>SORE</td>
<td>HIERRO</td>
<td>STORE</td>
</tr>
<tr>
<td>COMIDA</td>
<td>MILK</td>
<td>FOOT</td>
<td>PEOPLE</td>
<td>FOOD</td>
</tr>
<tr>
<td>ÁRBOL</td>
<td>LEAVES</td>
<td>THREE</td>
<td>GIRL</td>
<td>TREE</td>
</tr>
<tr>
<td>TRUCK</td>
<td>COCHE</td>
<td>CAMISA</td>
<td>TIENDA</td>
<td>CAMIÓN</td>
</tr>
<tr>
<td>SORE</td>
<td>DAÑO</td>
<td>DUERME</td>
<td>COCHE</td>
<td>DUELE</td>
</tr>
<tr>
<td>STAR</td>
<td>LUNA</td>
<td>TORTILLA</td>
<td>MAMITA</td>
<td>ESTRELLA</td>
</tr>
</tbody>
</table>