

Neurolinguistic Computational Models

BRIAN MACWHINNEY¹ and PING LI²

¹*Department of Psychology, Carnegie Mellon University Pittsburgh, PA, USA*

²*Department of Psychology, University of Richmond, Richmond, VA, USA*

ABSTRACT

It is tempting to think of the brain as functioning very much like a computer. Like the digital computer, the brain takes in data and outputs decisions and conclusions. However, unlike the computer, the brain does not store precise memories at specific locations. Instead, the brain reaches decisions through the dynamic interaction of diverse areas operating in functional neural circuits. The role of specific local areas in these functional neural circuits appears to be highly flexible and dynamic. Recent work has begun to provide detailed accounts of both the overall circuits supporting language and the detailed computations provided in smaller neural areas. These accounts take the shape of both structured and emergent models.

22.1. INTRODUCTION

Recent decades have seen enormous advances in linguistics, psycholinguistics, and neuroscience. Piecing these advances together, cognitive scientists have begun to formulate mechanistic accounts of how language is processed in the brain. Although these models are still very preliminary, they allow us to integrate information derived from a wide variety of studies and methodologies. They also yield predictions that can drive the search for specific neurolinguistic processing mechanisms.

22.2. THE COMPUTER AND THE BRAIN

Current neurocomputational models build on a core set of ideas deriving from the field of artificial intelligence, as it matured in the 1960s. At that time, researchers believed

that one could view the brain as a type of digital computer. The four crucial design features of the digital computer are: binary logic, seriality of computation, a fixed memory address space, and modularity of program design. Given this, we can ask whether the human brain also relies on binary logic, seriality, a fixed address space, and modularity.

Neuroscience has provided fairly clear answers to this question. First, we know that individual neurons do indeed fire in an on–off binary fashion. So this feature shows a close match. However, unlike the serial computer, the brain operates in a massively parallel fashion. Imaging studies have shown that, at a given moment in time, neurons are active throughout the brain. Because of this massive parallelism, the binary functioning of individual neurons takes on a very different role in the brain, serving to modulate decisions and activations, rather than making simple yes–no choices in a serial fashion. But it is not the case that the full parallel activations in the brain are equally in consciousness at a given moment. Although the brain has no central processing unit (CPU), there is a system of interrelated executive control processes localized in the frontal cortex that operates with additional support from posterior memory areas. We can think of this frontal system as a neural CPU. This system imposes its own form of seriality on thinking, allowing only a limited number of ideas or percepts to be active in working memory or focal attention at a given time. So, although the brain is massively parallel, it achieves a certain limited form of seriality for processes in focal attention.

Modularity is a crucial feature of program design in the digital computer. Modularity is not hard wired into the computer. Rather, it is enforced by the structure of computer languages and the methods used by particular programs. Some form of modularity is also clearly present in the brain. During neuroembryonic development, cells that

are initially undifferentiated migrate from the germinal matrix to specific cortical and sub-cortical areas. These migrating cells maintain their connections to other areas as they migrate but also begin to differentiate as they move to particular cortical areas. Thus, some cortical differentiation is present already at birth. The brain is initially highly plastic. Over time, as we will see in more detail later, areas become increasingly committed to particular computational functions. However, as Bates and colleagues (Elman *et al.*, 1996) have emphasized, “neural modules are made not born.” In this sense, modularity is an emergent fact in the brain, just as it is in computer programming. This insight can also help us understand the organization of processing modules in bilingualism (Hernandez *et al.*, 2005).

The biggest difference between the brain and the computer is the fact that the neural CPU cannot access memory via the systematic method used by the computer CPU. When a program is loaded onto a computer, it can reliably search in a particular memory location for each important piece of information. The neural CPU cannot rely on this scheme. In the late 1970s, there was an attempt to identify a system for memory addressing that might indeed parallel the system using by the digital computer (John, 1967). One idea was that individual neurons might be identified on the basis of codes embedded in the expressed portion of the DNA or RNA. If this were true, one might expect that memories could be encoded directly in the cells. To test this, biologists taught small planaria worms to navigate a series of turns in a maze to get some food. Once the worms had been trained, they ground them up and fed them to other untrained planaria. The hope was that the memories stored in the DNA of the trained planaria would be passed on to the untrained worms. At first, the results were promising. But later it appeared impossible to replicate the experiments outside of the original laboratories. The results of these experiments were chronicled in a series of papers called “The Worm Runner’s Digest.” Looking back, it seems remarkable that scientists could have thought that memories would be encoded this way. However, at that time, the strength of the analogy between the brain and the computer was so strong that the experimental hypothesis seemed perfectly reasonable.

The problem of reliably accessing stored memories is part a more general problem in neural computation. Because individual neurons do not have addresses, they cannot be given unique names. We cannot imagine that a particular neuron represents “a yellow Volkswagen” or that another neuron represents “the past tense suffix.” It is clearly impossible to pass symbolic information down neuronal axons. Instead, neurons must acquire a functional significance that arises from their role as participants in connected neural networks. In part, this is because neurons are not as reliable as silicon. Neurons may die and, in some areas of the brain, new neurons may be born. Neural firing is subject to a variety of disruptions caused by conditions varying from fatigue to epilepsy. In

extreme cases, victims of stroke or other injuries may lose large portions of their brain, but maintain the ability to talk and think. However, if a computer has faults in even a few silicon gates in memory, it will be unable to function at all. Thus, the neural CPU must use a very different, more flexible, method for addressing memory. It was the realization of this fundamental fact that led in the 1980s to the rise of neural network models (Grossberg, 1987) as the major method for modeling the brain. All current work in neurocomputational models is illuminated by this basic insight.

22.3. STRUCTURED MODELS

There are two ways in which we can allow facts about the brain to constrain our neurolinguistic models. One method relies on structured modeling and the other on unstructured or emergent modeling. Within the framework of structured modeling, we can distinguish between module-level models and neuron-level models.

22.3.1. Module-Level Structured Models

In this section, we will examine work on module-level models. These models attempt to localize processing in particular neural modules. On an anatomical level, it is clear that the brain is rich in structure. For example, there are at least 54 separate processing areas in visual cortex (Van Essen *et al.*, 1990). But it is not clear whether these areas function as encapsulated modules or rather as interactive pieces of functional networks.

Evidence for neurolinguistic modules has come from three sources: aphasiology, brain imaging, and developmental disorders. The oldest of these sources is the evidence from differing patterns of language deficit in aphasia. One can study patients with lesions of different types in the hope of identifying double disassociations between information-processing skills and lesion types. For example, some patients will have damaged prosodic structure, but normal segmental phonology. Other patients will have damaged segmental structure, but normal prosody. This pattern of results would provide strong support for the notion that there is a localized cognitive module for the processing of prosody. In practice, however, evidence for such double dissociations is difficult to obtain without post hoc partitioning of subject groups. But this partitioning itself casts doubt on the underlying assumptions regarding modularity and dissociability.

A familiar example of this type of model for language is the Geschwind (1979) model of connected language modules. This model is designed to account for how we can listen to a sentence and then imitate it or reply to it. According to this model, language comprehension begins with the receipt of a linguistic signal by auditory cortex in the temporal lobe. This information is then passed on to Wernicke’s

area for lexical processing. From here, information is passed over the arcuate fasciculus to Broca's area, where the reply or imitation is planned. Finally, the output signal is sent to motor cortex for articulation. This model treats processing as the passing of information between modules. According to this model, damage in a given area will predict loss of the related ability. Thus, damage to Broca's area should lead to Broca's aphasia. Unfortunately, the neurological assumptions of this model have proven problematic. Originally, Wernicke's area was thought to be an association area at the juncture of the temporal and parietal lobes. However, there is little evidence that this area functions as association cortex with any specific linkage to lexical or linguistic processing. Some other components of the Geschwind model are less problematic. In particular, it is clear that sounds are controlled by temporal auditory cortex, and that the final stages of speech output are controlled by motor cortex.

The role of Broca's area in the Geschwind model is also problematic. Although Broca's area is well defined anatomically, it has not been possible to locate specific perisylvian areas that are associated with specific aphasic symptoms or with specific patterns of disruption of naming in direct cortical stimulation (Ojemann *et al.*, 1989). However, recent work using functional magnetic resonance imaging (fMRI) methods has begun to clarify this issue. It has been shown that processing in inferior frontal gyrus (IFG) involves three clusters: (1) tasks emphasizing semantic processing with activation in anterior IFG in the *pars orbitalis*; (2) phonological tasks with activation in the posterior superior IFG; and (3) syntactic tasks with activation between the other two areas in middle IFG or *pars triangularis*, Brodman's area 44/45 (BA) (Bookheimer, 2002; Hagoort, 2005). These separations are not sharp and absolute, but they do seem to represent interesting differentiations in IFG that correspond with traditional linguistic distinctions. However, we must remember that the subtractive methodology used to analyze fMRI data tends to underestimate the contribution of other areas of the brain that are also involved in a particular task.

A final type of evidence for neurolinguistic modules comes from studies of children with developmental disorders. Here, researchers have applied the same logic of double dissociations used in the study of aphasia. In particular, it is often argued that children with Williams syndrome show an intense cognitive deficit with no serious disruption of language functioning. In contrast, children with Specific Language Impairment (SLI) are said to have intact cognitive functioning with marked impairment in language. However, this supposed double dissociation is not so clear in practice. Children with Williams syndrome do have effective control of language, but they achieve this control in ways that are far from normal (Karmiloff-Smith *et al.*, 1997; see Chapter 36, this volume). Moreover, many children with SLI also have problems with related areas of conceptual functioning (e.g., in temporal sequence processing) and there have not

yet been any successful attempts to link the module that is supposedly damaged in SLI to any particular brain region.

A somewhat different approach to localization views alternative cortical areas as participating in functional neural networks. In this framework, a particular cortical area may participate in a variety of functional networks. Within each of these various networks, the area would basically serve a similar computational role. However, because processing demands vary across networks, the specific products of this processing will vary depending on the network involved. Mason and Just (2006) argue that fMRI studies have provided evidence for five functional neural networks for discourse processing:

1. A right hemisphere network involving middle and superior temporal cortex that computes a coarse semantic representation.
2. A bilateral network in dorsolateral prefrontal cortex (DLPFC) that monitors conceptual coherence.
3. A left hemisphere network involving IFG and left anterior temporal for text integration.
4. A network involving medial frontal areas bilaterally and right temporal/parietal areas for perspective taking.
5. A bilateral, but left-dominant, network involving the intraparietal sulcus for spatial imagery processing.

In practice, it is difficult to understand the exact separation between these various processes. For example, there seems to be a conceptual overlap between the second and third of these networks. Also, it is not clear whether the additional activation recorded in right hemisphere sites indicates basic discourse processing or a spillover of processing from the left hemisphere. And it is not clear whether these areas would be involved in different ways for comprehending written versus spoken discourse. Furthermore, networks involved in broader conceptual tasks such as perspective taking probably involve more than just a few cortical areas. To the degree that perspective taking triggers empathy and body mapping (MacWhinney, 2005), it will also rely on additional frontal areas, basal ganglion, amygdala, and mirror neuron systems in both frontal and parietal cortex. Despite these various issues, it seems profitable to continue exploring the interpretation of fMRI results in terms of interlocking functional neural networks. These models allow researchers to study functional localization without forcing them to think of local areas as encapsulated neural modules.

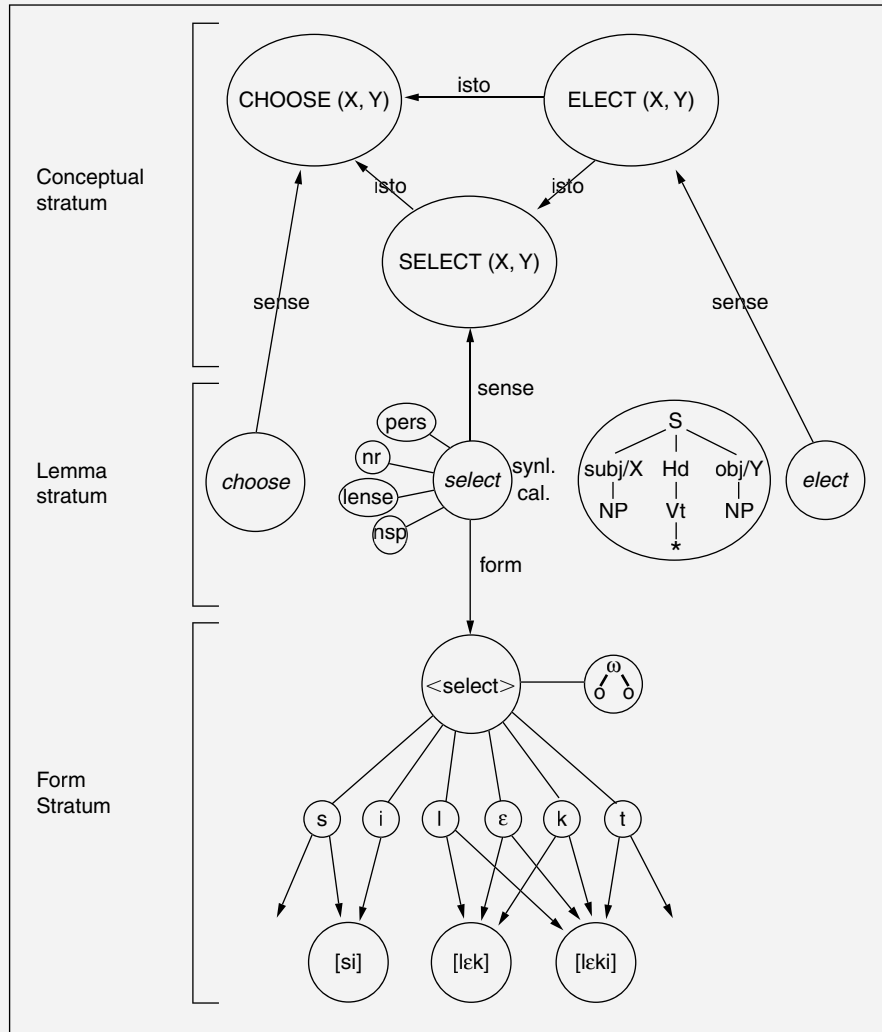
22.3.2. Neuron-Level Structured Models

Structured modeling can also be conducted on a level that avoids direct reference to modules. Morton's (1970) logogen model was a particularly successful model of this type. At the center of each logogen network was a master neuron devoted to the particular word. Activation of this central unit could then trigger further activation of units

encoding its phonology, orthography, meaning, and syntax. Pushing the notion of spreading neural activation still further, McClelland and Rumelhart (1981) developed an interactive activation (IA) account of context effects in letter perception. IA models have succeeded in providing clear accounts for a wide variety of phenomena in speech production, reading, auditory perception, lexical semantics,

and second language learning. Because each unit in these models can be clearly mapped to a particular linguistic construct, such as a word, phoneme, or semantic feature, the generation of predictions from IA models is very straightforward (see Box 22.1). Because these models do well at representing the processes of activation and competition, they usually provide good models of experimental

Box 22.1 Interactive-activation account of speech production



The IA account from Levelt (2004) illustrates the activation of the word “select” and its surrounding syntactic context. The diagram shows how, on the level of form, the word “select” is composed of phonemes that activate specific syllables. Here [lek] competes for activation with [lekɪ]. Also, the word has a weak–strong stress pattern. On the level of the lemma, the word has the syntactic feature of present tense and specifies a

subject and an object. At this level, it competes with forms like “choose” and “elect.” Finally on the conceptual level, notions of *selecting*, *choosing*, and *electing* are all activated, but only *selecting* is chosen in this case.

Levelt, W.J.M. (2004). Models of word production. *Trends in Cognitive Sciences*, 4, 223–232.

data with both healthy and aphasic individuals. Moreover, the core idea of IA maps up well with what we know about how neurons interact in terms of excitatory and inhibitory synaptic connections.

IA models suffer from a fundamental weakness. It is difficult to imagine how real neurons could achieve the local conceptual labeling required by these models. How could the learner manage to tag one neuron as “fork” and another neuron as “spoon?” How could the learner manage to connect up exactly these specific cells to the correct phonemic components? Moreover, should we really imagine that each word or concept is represented by one and only one central neuron? Given the fact that neurons are subject to death and replacement, would we not expect to find even normal speakers continually losing words or phonemes when their controlling cells die? If there were some learning method associated with this architecture, these various problems with IA models could be solved. However, it is not clear how one could formulate a learning algorithm for this type of localist model.

22.4. EMERGENT MODELS

To address this problem, Rumelhart *et al.* (1986) developed a neural network learning algorithm called back propagation. This algorithm makes no assumptions regarding “labels” on neurons. Instead, the functioning of individual neurons emerges through a competitive training process in which connections between neurons are adjusted in proportion to the mismatch (error) in the input-to-output mapping. Because there are no labels on units, there is a tendency for information controlling a given pattern to become distributed across the network. This distribution of information can be characterized as parallel distributed processing (PDP). Because PDP models based on back propagation algorithm are easy to develop and train, this framework has generated a proliferation of models of language learning. PDP models have been successfully applied to a wide variety of language learning areas including the English past tense, Dutch word stress, universal metrical features, German participle acquisition, German plurals, Italian articles, Spanish articles, English derivation for reversives, lexical learning from perceptual input, deictic reference, personal pronouns, polysemic patterns in word meaning, vowel harmony, historical change, early auditory processing, the phonological loop, early phonological output processes, ambiguity resolution, relative clause processing, word class learning, speech errors, bilingualism, and the vocabulary spurt.

22.4.1. Self-Organizing Maps

Although PDP models succeed in capturing the distributed nature of memory in the brain, they do this by relying on reciprocal connections between neurons. In fact, there is

little evidence that individual neurons are connected in this way. Also, PDP relies on back propagation of an explicit error correction signal. Again, evidence that this is present is fairly weak. MacWhinney (2000a) argued that distributed PDP models fail to provide an emergent localist representation of the word, making further morphological and syntactic processing difficult. As we noted earlier, one of the great strengths of IA models is their ability to provide localist representations for words and other linguistic constructs. However, because these forms cannot be learned from the input, researchers have turned to back propagation and PDP to provide learning mechanisms. A more ideal solution to the problem would be to rely on a learning algorithm that managed to induce flexible, localist representations that correspond in a statistical sense to concepts and words.

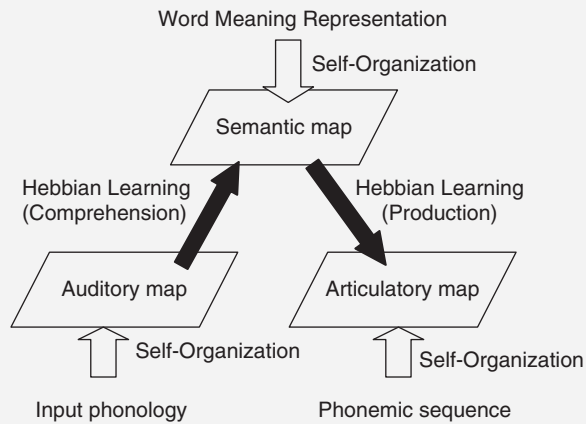
The self-organizing feature map (SOFM) algorithm of Kohonen (2001) provides one way of tackling these problems. This framework has been used to account in great detail for the development and organization of the visual system (Miikkulainen *et al.*, 2005). Applications of the model to language are more difficult, but are also showing continual progress. To illustrate the operation of SOFM for language, let us consider the DevLex-II word learning model of Li *et al.* (2007). This model uses three local maps to account for word learning in children: an auditory map, a concept map, and an articulatory map (see Box 22.2). The actual word representations are computed from the contextual features derived from realistic child–parent interactions in the CHILDES database (MacWhinney, 2000b). In effect, this meant that words that occurred in similar sentence contexts are represented similarly, and consequently organized by SOFM in nearby neighborhoods in the map. For example, nouns were organized together because they appeared consistently in slots such as “my X” or “the X.” These representations can also be coupled with perceptual features to capture the child’s early perceptual experiences (Li *et al.*, 2004).

Self-organizing maps offer a promising method for achieving emergent localist encodings. There are also several ways in which the capacity of maps can be expanded through the addition of neurons or overlays with new coding features. Which of these methods is actually used in the brain remains unclear. We know that the brain has an enormous capacity for the storage of memories. However, efficient use of this storage space may rely on hippocampal storage mechanisms to organize memories for efficient retrieval (Miikkulainen *et al.*, 2005). We will discuss this issue further in the final section.

22.4.2. Syntactic Emergence

Although these emergentist models have succeeded in modeling a variety of features in word learning and phonology, it has been more difficult to apply them to the task of modeling syntactic processing. Elman’s (1990) simple recurrent network (SRN) model uses recurrent back-propagation

Box 22.2 DevLex-II: A SOFM-based model of lexical development



The DevLex-II model is a neural network model of early lexical development based on SOFM (Li *et al.*, 2007). The model has three sub-networks (maps) to process and organize linguistic information: Auditory map that takes the phonological information of words as input, semantic map that organizes the meaning representations based on lexical co-occurrence statistics, and articulatory map that outputs phonemic sequences of words. The semantic map is connected to both the auditory map and the articulatory map through associative links that are trained by Hebbian learning. Such connections allow for the effective modeling of comprehension (auditory to semantic) and production (semantic to articulatory) processes in language acquisition.

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

connections to update the network's memory after it reads each word. The network's task is to predict the next word. This framework views language comprehension as a highly constructive process in which the major goal is trying to predict what will come next. An alternative to the predictive framework relies on the mechanisms of spreading activation and competition. For example, MacDonald *et al.* (1994) have presented a model of ambiguity resolution in sentence processing that is grounded on competition between lexical items. Models of this type do an excellent job of modeling the temporal properties of sentence processing. Such models assume that the problem of lexical learning in neural networks has been solved. They then proceed to use localist representations to control IA during sentence processing.

Another lexicalist approach uses a linguistic framework known as Construction Grammar. This framework emphasizes the role of individual lexical items in early grammatical learning (MacWhinney, 1987). Early on, children learn

to use simple frames such as *my + X* or *his + X* to indicate possession. As development progresses, these frames are merged into general constructions, such as the possessive construction. In effect, each construction emerges from a lexical gang. Sentence processing then relies on the child's ability to combine constructions online. When two alternative constructions compete, errors appear. An example would be **say me that story*, instead of *tell me that story*. In this error, the child has treated "say" as a member of the group of verbs that forms the dative construction. In the classic theory of generative grammar, recovery from this error is supposed to trigger a learnability problem, since such errors are seldom overtly corrected and, when they are, children tend to ignore the feedback. Neural network implementations of Construction Grammar address this problem by emphasizing the direct competition between *say* and *tell* during production. The child can rely on positive data to strengthen the verb *tell* and its link to the dative construction, thereby eliminating this error without corrective feedback. In this way, models that implement competition provide solutions to the logical problem of language acquisition.

These various approaches to syntactic learning must eventually find a way of dealing with the compositional nature of syntax. A noun phrase such as "my big dog and his ball" can be further decomposed into two segments conjoined by the "and." Each of the segments is further composed of a head noun and its modifiers. Our ability to recursively combine words into larger phrases stands as a major challenge to connectionist modeling. One likely solution would use predicate constructions to activate arguments that are then combined in a short-term memory buffer during sentence planning and interpretation. To build a model of this type, we need to develop a clearer mechanistic link between constructions as lexical items and constructions as controllers of the on-the-fly process of syntactic combination.

22.4.3. Lesioning Emergent Models

Once one has constructed networks that are capable of learning basic syntactic processes, it is an easy matter to test the ability of these models to simulate aphasic symptoms by subjecting the model to lesions, just as the brain of the real aphasic person was subjected to real lesions. Lesioning can be done by removing hidden units, removing input units, removing connections, rescaling weights, or simply adding noise to the system. It has been shown that these various methods of lesioning networks all produce similar effects (Bullinaria & Chater, 1995). However, work on lesioning of models can also capture patterns of dissociation. For example, in lesioned models, high-frequency items are typically preserved better than low-frequency items. Similarly, patterns with many members will be preserved better than

patterns with fewer members, and items with associations to many neighbors better than items with fewer neighbors.

A particularly interesting application of the network lesioning technique is the model of deep dyslexia developed by Plaut and Shallice (1993). The model had connections from orthography to semantics and from semantics to phonology, as well as hidden unit layers and layers of clean-up units for semantics and phonology. Lesions to the orthographic layer led to loss of the ability to read abstract words. However, lesions to the semantic clean-up units led to problem in reading concrete words. This work illustrates how one could use the analysis of lesions to emergent models to elaborate ideas about area-wide patterns of connectivity in the brain.

Like models based on back propagation, SOFM-based models can also be used to study lesion and recovery from lesion. Li *et al.* (2007) showed how DevLex-II, when lesioned with noise in the semantic and phonological representations and their associative links, can display early plasticity in the recovery from lesion induced at different time points of learning. Their lesion data matched up with empirical findings from children who suffer from early focal brain injury with respect to lexical learning and U-shaped behavior (MacWhinney *et al.*, 2000).

22.5. CHALLENGES AND FUTURE DIRECTIONS

Despite the various successes of both structured and emergent models, neurolinguistic computational models continue to suffer from a core limitation. This is the failure of these models to decode the basic addressing system of the brain. However, recent work in the area of embodied cognition may be pointing toward a resolution of this problem. This work emphasizes the ways in which the shape of human thought emerges from the fact that the brain is situated inside the body and the body is embedded in concrete physical and social interactions. In neural terms, embodied cognition relies heavily on a perception-action cycle that works to interpret new perceptions in terms of the actions needed to produce them. For example, when watching an experimenter grab a nut, a monkey will activate neurons in motor cortex that correspond to the areas that the monkey itself uses when grabbing a nut. These so-called mirror neurons (see Chapter 23) form the backbone of a rich system for mirroring the actions of others by interpreting them in terms of our own actions.

A variety of recent computational models have tried to capture aspects of these perceptual-motor linkages. One approach emphasizes the ways in which distal learning processes can train action patterns such as speech production on the basis of their perceptual products (Westermann & Miranda, 2004). A very different approach, developed in MacWhinney (2005), works out the consequences of the online construction of an embodied mental model of a

sentence for linguistic structures. This account views the frontal lobes as working to encode a virtual homunculus that can be used to enact the various actions, stances, and perspective shifts involved in linguistic discourse.

By itself the notion of an embodied perceptual-motor cycle will not solve the core problem of memory addressing in the brain. However, it seems to point us in a very interesting direction. Starting with the core observation that the brain is located in the body and fully connected to all the senses and motor effectors, we could imagine that the primary code used in the brain might be the code of the body itself. When we look at the tonotopic and retinotopic organization of sensory areas or the multiple representations of effectors in motor areas, this view of the brain as an encoder of the body seems transparently true. However, it is less clear how this code might extend beyond these primary areas for use throughout the brain. One radical possibility is that the brain may rely on embodied codes throughout and that the body itself could provide the fundamental language of neural computation.

References

- Bookheimer, S. (2002). Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience*, 25, 151–188.
- Bullinaria, J.A., & Chater, N. (1995). Connectionist modelling: Implications for cognitive neuropsychology. *Language and Cognitive Processes*, 10, 227–264.
- Elman, J.L. (1990). Finding structure in time. *Cognitive Science*, 14, 179–212.
- Elman, J.L., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness*. Cambridge, MA: MIT Press.
- Geschwind, N. (1979). Specializations of the human brain. *Scientific American*, 241, 180–199.
- Grossberg, S. (1987). Competitive learning: From interactive activation to adaptive resonance. *Cognitive Science*, 11, 23–63.
- Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences*, 9, 416–422.
- Hernandez, A., Li, P., & MacWhinney, B. (2005). The emergence of competing modules in bilingualism. *Trends in Cognitive Sciences*, 9, 220–225.
- John, E.R. (1967). *Memory*. New York: Academic Press.
- Karmiloff-Smith, A., Grant, J., Berthoud, I., Davies, M., Howlin, P., & Udwin, O. (1997). *Language and Williams syndrome: How intact is "intact?"* *Child Development*, 68, 246–262.
- Kohonen, T. (2001). *Self-organizing maps* (3rd edn). Berlin: Springer.
- Li, P., Farkas, I., & MacWhinney, B. (2004). Early lexical development in a self-organizing neural network. *Neural Networks*, 17, 1345–1362.
- Li, P., Zhao, X., & MacWhinney, B. (2007). Dynamic self-organization and early lexical development in children. *Cognitive Science*, 31, 581–612.
- MacDonald, M.C., Pearlmutter, N.J., & Seidenberg, M.S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101(4), 676–703.
- MacWhinney, B. (1987). The competition model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 249–308). Hillsdale, NJ: Lawrence Erlbaum.
- MacWhinney, B. (2000a). Lexicalist connectionism. In P. Broeder & J. Murre (Eds.), *Models of language acquisition: Inductive and deductive approaches* (pp. 9–32). Cambridge, MA: MIT Press.

- MacWhinney, B. (2000b). *The CHILDES project: Tools for analyzing talk*. Mahwah, NJ: Lawrence Erlbaum Associates.
- MacWhinney, B. (2005). The emergence of grammar from perspective. In D. Pecher & R.A. Zwaan (Eds.), *The grounding of cognition: The role of perception and action in memory, language, and thinking* (pp. 198–223). Mahwah, NJ: Lawrence Erlbaum Associates.
- MacWhinney, B., Feldman, H.M., Sacco, K., & Valdes-Perez, R. (2000). Online measures of basic language skills in children with early focal brain lesions. *Brain and Language, 71*, 400–431.
- Mason, R., & Just, M. (2006). Neuroimaging contributions to the understanding of discourse processes. In M. Traxler & M. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 765–799). Mahwah, NJ: Lawrence Erlbaum Associates.
- McClelland, J.L., & Rumelhart, D.E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of the basic findings. *Psychological Review, 88*, 375–402.
- Miikkulainen, R., Bednar, J.A., Choe, Y., & Sirosh, J. (2005). *Computational maps in the visual cortex*. New York: Springer.
- Moll, M., & Miikkulainen, R. (1997). Convergence-zone episodic memory: Analysis and simulations. *Neural Networks, 10*, 1017–1036.
- Morton, J. (1970). A functional model for memory. In D.A. Norman (Ed.), *Models of human memory* (pp. 203–248). New York: Academic Press.
- Ojemann, G., Ojemann, J., Lettich, E., & Berger, M. (1989). Cortical language localization in left, dominant hemisphere: An electrical stimulation mapping investigation in 117 patients. *Journal of Neurosurgery, 71*, 316–326.
- Plaut, D., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology, 10*, 377–500.
- Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning internal representations by back propagation. In D. Rumelhart & J. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (pp. 318–362). Cambridge, MA: MIT Press.
- Van Essen, D.C., Felleman, D.F., DeYoe, E.A., Olavarria, J.F., & Knierim, J.J. (1990). Modular and hierarchical organization of extrastriate visual cortex in the macaque monkey. *Cold Spring Harbor Symposium on Quantitative Biology, 55*, 679–696.
- Westermann, G., & Miranda, E.R. (2004). A new model of sensorimotor coupling in the development of speech. *Brain and Language, 89*, 393–400.

Further Readings

Just, M.A., & Varma, S. (2007). The organization of thinking: what functional brain imaging reveals about the neuroarchitecture of complex cognition. *Cognitive, Affective, and Behavioral Neuroscience, 7*, 153–191.

This article compares three recent neurocomputational approaches that provide overall maps of high-level functional neural circuits for much of cognition.

Miikkulainen, R., Bednar, J.A., Choe, Y., & Sirosh, J. (2005). *Computational maps in the visual cortex*. New York: Springer.

This book shows how self-organizing feature maps can be used to describe many of the detailed results from the neuroscience of vision.

Ullman, M. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition, 92*, 231–270.

An interesting, albeit speculative, application of a neurocomputational model to issues in second language learning.