2 A Parallel Distributed Processing Model of Language: Phonologic, Semantic, and Semantic–Phonologic (Lexical) Processing

The Wernicke–Lichtheim (W-L) information-processing model of language function has played a dominant role in understanding aphasic syndromes (Lichtheim 1885) and has stood the test of time in defining the topographical relationship between the modular domains (acoustic representations, articulatory motor representations, and concept representations) underlying spoken language function. Unfortunately, the W-L information-processing model does not specify the characteristics of the representations within these domains and how they might be stored in the brain. It also does not address the means by which these domains might interact. I have proposed a PDP model that uses the same general topography as the W-L model (Nadeau 2001; Roth et al. 2006) but also specifies how representations are generated in the modular domains and how knowledge is represented in the links between these domains (see figure 2.1). Though not tested in simulations, this model is neurally plausible and provides a cogent explanation for a broad range of psycholinguistic phenomena in normal subjects and subjects with aphasia.

The PDP modification of the W-L model posits that the acoustic domain (akin to Wernicke's area) contains large numbers of units located in auditory association cortices that represent acoustic features of phonemes.[1] The articulatory domain (analogous to Broca's area) contains units located predominantly in dominant frontal operculum that represent discrete articulatory features of speech, as opposed to continuously variable motor programs (e.g., phonemic distinctive features). The semantic or conceptual domain contains an array of units distributed throughout unimodal and polymodal association cortices that represent semantic features of concepts. For example, the representation of the concept of "house" might correspond to activation of units representing features of houses such as visual attributes, construction materials, contents (physical and human), and so on (each feature in turn a distributed representation over more primitive features). Each unit within a given domain is connected to many, if not most, of the other units in that same domain (symbolized by the small looping arrow appended to each domain in figure 2.1). Knowledge within each domain is represented as connection
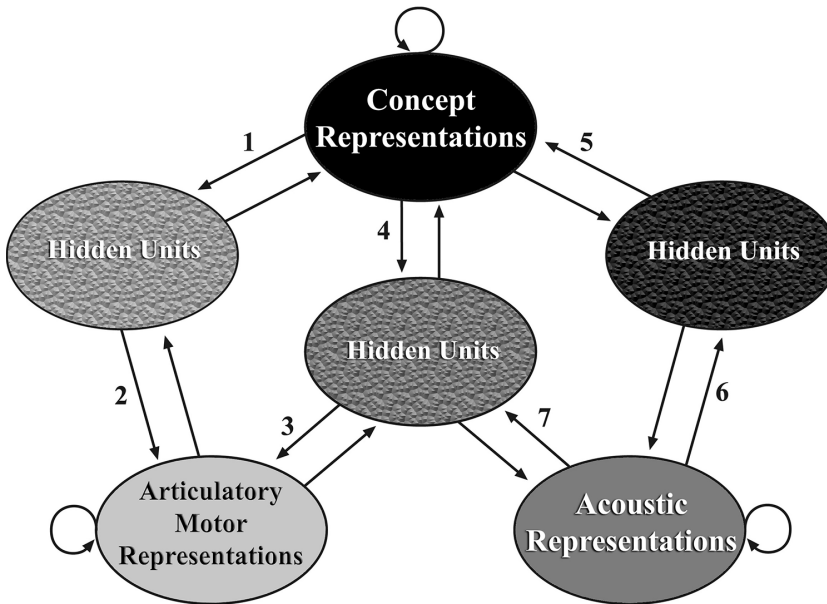
Figure 2.1
Proposed parallel distributed processing model of language. (Roth, H. L., S. E. Nadeau, A. L. Hollingsworth, A. M. Cimino-Knight, and K. M. Heilman. 2006. "Naming Concepts: Evidence of Two Routes." *Neurocase* 12:61–70.) Connectivity within the substrate for concept representations defines semantic knowledge. Connectivity within the acoustic–articulatory motor pattern associator defines phonologic sequence knowledge. Connectivity between the substrate for concept representations and the acoustic–articulatory motor pattern associator defines lexical knowledge (see text for details).

strengths between the units. Thus, semantic knowledge is represented as the pattern of connection strengths throughout the association cortices supporting this knowledge. Within any domain, a representation corresponds to a specific pattern of activity of all the units, hence the term distributed representation (a synonym for a population encoded representation). Each unit within each of these domains is connected via interposed hidden units to many, if not most, of the units in the other domains. During learning of a language, the strengths of the connections between the units are gradually adjusted so that a pattern of activity involving the units in one domain elicits the correct pattern of activity in the units of another domain. The entire set of connections between any two domains forms a pattern associator network. Hidden units are units whose activity cannot be directly interpreted in behavioral terms. The hidden-unit regions, in conjunction with nonlinear unit properties, enable the systematic association of representations in two connected domains that may be arbitrarily related to one another (e.g., word sound and word meaning). The model employs left–right position in acoustic and

articulatory motor representations as a surrogate for temporal order in precisely the same way as the reading model of Plaut et al. (1996). Thus, acoustic and articulatory motor representations would feature positions for each output phoneme or distinctive feature, ordered as they are in the phonologic word form. The use of left to right sequential order in lieu of temporal order is a device of convenience, but there is evidence of this temporal–geographic transform in the brain (Cheung et al. 2001). During any type of language processing, initiated by input to any domain of the network, there will be almost instantaneous engagement of all domains of the network. Thus, linguistic behavior is best viewed as the emergent product of the *entire* network.

I will now focus on particular components of the network in order to provide a more detailed understanding of how they work and the nature of the knowledge they support.

Concept Representations

As I have noted, the Wernicke–Lichtheim information-processing model provides no insight into the nature of the representations in the various domains. The nature of concept representations (depicted in figure 2.1) can be best illustrated by a particularly illuminating model developed by David Rumelhart and his colleagues (Rumelhart et al. 1986). This "rooms in a house" model was comprised of 40 "feature" units, each corresponding to an article typically found in particular rooms or an aspect of particular rooms. Each unit was connected with all the other units in the network—an attribute that defines the model as an *auto-associator network*. Auto-associator networks have the capacity for "settling" into a particular state that defines a representation. Connection strengths were defined by the likelihood that any two features might appear in conjunction in a typical house. When one or more units was clamped into the "on" state (as if the network had been shown these particular features or articles), activation spread throughout the model and the model eventually settled into a steady state that implicitly defined a particular room in a house. Thus, clamping "oven" ultimately resulted in activation of all the items one would expect to find in a kitchen and thereby *implicitly* defined, via a *distributed or population encoded representation*, the concept of a kitchen. No kitchen unit per se was turned on. Rather, kitchen was defined by the pattern of feature units that were activated. The network contained the knowledge, in the totality of its connections, that enabled this representation to be generated. The 40-unit model actually had the capability of generating distributed representations of a number of different rooms in a house (e.g., bathroom, bedroom, living room, study), subcomponents of rooms (e.g., easy chair and floor lamp, desk and desk chair, window and drapes) and blends of rooms that were not anticipated in the programming of the model

(e.g., clamping both bed and sofa led to a distributed representation of a large, fancy bedroom replete with a fireplace, television, and sofa).

This auto-associator model, simple though it is, has the essential attributes of a network that might instantiate semantic knowledge and be capable of generating the distributed representations corresponding to concepts. The brain's semantic auto-associator obviously is comprised of vastly more that 40 features, and enables an enormous repertoire of distributed representations corresponding to the vast number of concepts we are capable of representing. This particular model network is not compartmentalized, but there is nothing inherent in PDP models that precludes a semantic representation comprised of two or more subnetworks (see, e.g., Farah and McClelland 1991). There is good evidence that in the brain, the meaning of a given word is distributed over a host of networks, depending in part upon the semantic features that are most essential to that meaning, an idea proposed by Lissauer over a century ago (Lissauer 1988) and elaborated by Wernicke (as cited in Eggert 1977). For example, visual information makes a particularly large contribution to the meaning of living things, and consequently, subjects with damage to visual association cortex due to herpes simplex encephalitis exhibit category-specific naming and recognition deficits for living things (Forde and Humphreys 1999; Warrington and Shallice 1984). In this conceptualization, the distributed representation of the concept "dog" has a major component in visual association cortices made available by knowledge of the visual appearance of dogs in general, as well as particular dogs; a major component in auditory association cortices corresponding to the sounds that dogs characteristically make; a major component in the limbic system corresponding to one's feelings about dogs in general and specific dogs; a component in somatosensory cortex corresponding to the feel of dog fur, wet tongue, or cold nose; a predicative component involving frontal cortex that corresponds to our knowledge of what dogs do, hence a component of the semantic representation of a verb (to be considered in much detail below); a component in olfactory cortex, corresponding to the odors of dogs; and components in perisylvian language cortex that enable us to translate the semantic representation of dog into an articulatory motor representation (so we can say /*dawg*/) or an acoustic representation (so we can understand another person saying /*dawg*/; see figure 2.2a). Not all of these subnetworks need to be activated every time, or in exactly the same way by everyone. We can speak in terms of *working memory*—in this case a pattern of activity in particular subnetworks corresponding to dog, and we can speak in terms of *working associations*, meaning the elicitation of distributed concept representations in other connected subnetworks, either automatically or volitionally. Thus, nearly everyone would, when hearing /*dawg*/, develop visual and limbic distributed representations, each constituting a working memory, the two together comprising an automatic working association. However, the average person might need to
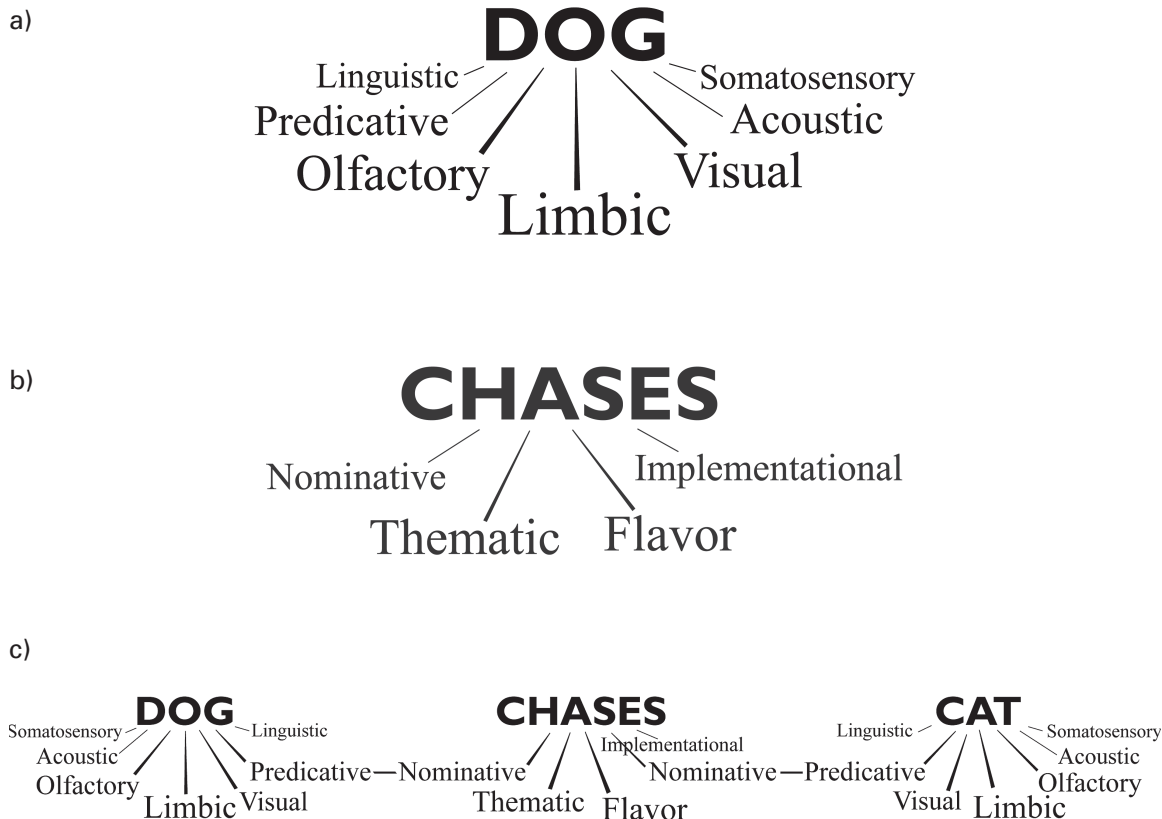
a)

DOG

Linguistic — Somatosensory
Predicative — Acoustic
Olfactory — Visual
Limbic

b)

CHASES

Nominative — Implementational
Thematic — Flavor

c)

Somatosensory — DOG — Linguistic
Acoustic
Olfactory — Predicative — Nominative
Limbic — Visual

CHASES
Implementational
Nominative — Predicative
Thematic — Flavor

Linguistic — CAT — Somatosensory
Acoustic
Visual — Olfactory
Limbic

Figure 2.2
(a) The multifocal distributed representation of nouns. The predicative component represents our knowledge of what a noun concept can or likely will do. (b) The multifocal distributed representation of verbs. The nominative component, by analogy to the predicative component of nouns, represents all the nouns that a verb is likely to modify. (c) The linking of multifocal distributed representations of nouns and verbs through their mechanisms for mutual engagement.

volitionally develop the working association that brings in the olfactory component of the meaning of "dog." Glancing at a calculator may not be sufficient to evoke a poke component or pick-up gesture component to its distributed concept representation, but looking at a calculator in the context of an action paradigm is (Bub, Masson, and Cree 2008).

This very old idea of Lissauer and Wernicke, because it is grounded in neuroanatomy, has always retained currency among neurologists, cognitive neuropsychologists, and others strongly guided by principles of strict functional–anatomic correlation, even as it has been repeatedly questioned by investigators, most notably cognitive psychologists, who have worked from models that do not take anatomy

into account. In these models, concepts have been posited to be symbolic, amodal entities that somehow are mapped to or grounded in the real world. However, with the advancement of cognitive science, we are now beginning to see a universal recognition that neuroanatomic considerations are essential, and the concept of amodal processing in the brain is steadily being undermined (see, e.g., Barsalou 2008). The old theory of Lissauer and Wernicke has been resurrected as the embodied cognition framework (Buxbaum and Kalénine 2010; Fernandino and Iacoboni 2010; Spivey 2007). A very influential factor in this renaissance has been the discovery of mirror neurons by Rizzolati and colleagues (Rizzolatti and Craighero 2004), neurons in frontal cortex that receive projections from parietal cortex such that they respond to the perception of goal-directed movement—for example, firing in response to the perception of an investigator picking up a morsel of food. Such responses were entirely predictable from the landmark neuroanatomic paper of Chavis and Pandya (1976), which demonstrated that postcentral cortices hold no secrets from the frontal lobes because of a vast and intricately organized system of reciprocal frontal–postcentral connections (see, e.g., Evangeliou et al. 2009 and Raos, Evangeliou, and Savaki 2007). Indeed, it is this connectivity that provides the basis for the vast repertoire of daily activities we pursue more or less automatically (Nadeau and Heilman 2007).

The Acoustic–Articulatory Motor Pattern Associator Network

The knowledge that allows a person to translate heard sound sequences into articulatory motor sequences, and thereby mediates repetition of both real words and nonwords, is contained in the network that connects the acoustic domain to the articulatory motor domain (the acoustic–articulatory motor pattern associator, figure 2.1, pathway 7–3). Because this network has acquired, through experience, knowledge of the systematic relationships between acoustic sequences and articulatory sequences, it has learned the sound *sequence* regularities of the language: the phonemic sequences of joint phonemes, rhymes, syllables, affixes, morphemes, and words characteristic of the language (Nadeau 2001). Because the role of sequence knowledge will be a major theme of this book, it is important at this stage to have a clear understanding of how sequence might be encoded in a distributed neural network architecture. The reading model developed by Plaut and colleagues (Plaut et al. 1996; Seidenberg and McClelland 1989) provides a particularly clear illustration. I will therefore describe it in some detail even though reading per se represents somewhat of a digression.

This reading model fundamentally recapitulates the acoustic–articulatory motor pathway of figure 2.1, the major difference (inconsequential to this discussion) being that in place of acoustic representations, it incorporated orthographic

representations. The model was composed of three layers: (1) an input layer of 105 grapheme units grouped into clusters, the first cluster including all possibilities for the one or more consonants of the onset, the second cluster including all the possible vowels in the nucleus, and the third cluster including all possibilities for the one or more consonants in the coda; (2) a hidden-unit layer of 100 units; and (3) an output layer of 61 phoneme units grouped into clusters including all the possibilities for onset, nucleus, and coda, respectively (as for the graphemes). Local representations were used for the graphemes and phonemes. There were one-way connections from each of the grapheme input units to each of the hidden units, and there were two-way connections between each of the hidden units and each of the phoneme output units. Every output unit was connected to every other output unit, providing the network the auto-associator capability for "settling into" the best solution (as opposed to its own approximate solution). The model was trained using a mathematical algorithm (backpropagation) that incrementally alters the strengths of connections in proportion to their contribution to the error, which was computed as the difference between the actual product of the network and the desired product of the network.[2] The orthographic representations of 3,000 English single-syllable words and their corresponding phonologic forms were presented, one pair at a time, cycling repeatedly through the entire corpus. In this way, the model ultimately learned to produce the correct pronunciation of all the words it had read. One of the most striking things about the trained model is that it also was able to produce correct pronunciations of plausible English nonwords (i.e., orthographic sequences it had never encountered before). How was this possible?

One might have inferred that the model was simply learning the pronunciation of all the words by rote. If this had been the case, however, the model would have been incapable of applying what it had learned to novel words. In fact, what the model learned was the statistical relationships between *sequences* of graphemes and *sequences* of phonemes that are characteristic of the English language. To the extent that there is a limited repertoire of sequence types, the model was able to learn it and then apply that knowledge to novel forms that incorporated some of the sequential relationships in this repertoire. Certain sequences, those most commonly found in English single-syllable words, were more thoroughly etched in network connectivity. Thus, it was very fast with high-frequency words. It was also very fast with words with an absolutely consistent orthographic–phonologic sequence relationship—for example, words ending in "ust," which are always pronounced /ʌst/ (must, bust, trust, lust, crust, etc.). The model encountered difficulty (reflected in prolonged reading latency) only with low-frequency words, and only to the extent that it had learned different, competing pronunciations of the same orthographic sequence. Thus, it was slow to read "pint" because in every case but

"pint," the sequence "int" is pronounced /*Int*/ (e.g., mint, tint, flint, lint). It was also slow, though not quite so slow, to read words like "shown" because there are two equally frequent alternatives to the pronunciation of "own" (gown, down, town vs. shown, blown, flown). It was very slow with words that are unique in their ortho-graphic–phonologic sequence relationship (e.g., aisle, guide, and fugue). These behaviors precisely recapitulate the behavior of normal human subjects given reading tasks.

The knowledge the model acquires reflects competing effects of type frequency and token frequency. If a single word is sufficiently common (high token frequency), the model acquires enough experience with it that competing orthographic–phonologic sequential relationships have a negligible impact on naming latency. However, if a word is relatively uncommon (e.g., pint), its naming latency will be significantly affected by the knowledge of other words that, though equally uncom-mon, together belong to a competing type—an "enemy" (e.g., mint, flint, tint, sprint).

The implicit knowledge of various competing regularities captured by the model (and the brain) through experience defines *quasi-regular domains*. For example, in the case of words ending in "own," orthographic–phonologic regularity exists but it is only quasi-regular because there is not one but two alternatives (shown vs. gown), a particular alternative being determined by the onset cluster. Quasi-regular domains may be composed of more or less equally competing subdomains, each correspond-ing to a regularity, as in the case of "own" words, or a domain that is regular but for a single member (e.g., mint, tint, splint, etc. vs. pint). In some cases they may be fully regular (e.g., the "_ust" words). The higher the frequency of a word, the more deeply its orthographic–phonologic connectivity becomes etched in neural connectivity, the less its production is influenced by similarity to neighbors, and the more it approaches a regular domain that has only one member. Whether or not linguistic forms belong to particular quasi-regular domains depends upon the particular regularities that the network is endowed to capture through experience. The term quasi-regular domains will be referred to repeatedly in this book because it succinctly character-izes patterns of knowledge that are directly related to neural instantiation and because aphasia can, to a substantial degree, be viewed as lesion-induced per-turbation of the competition between regularities in quasi-regular domains in which the competitive strength of lower frequency or less typical representations is reduced. Grammatic function is driven by competition between regularities within quasi-regular domains of knowledge no less than the orthographic–phonologic translation discussed in this section. Grammatic function differs only in that a considerably larger number and variety of quasi-regular domains are in play at any given time. When a neural network supporting a quasi-regular domain settles into the wrong state, the result is a paraphasic error: phonemic, verbal, or semantic, but also mor-phologic (paragrammatic) or syntactic.

The capacity of the model of Plaut and colleagues to read nonwords reflects its ability to capture patterns in the sequential relationships between orthographic and articulatory word forms and to apply this knowledge to novel word forms. Plaut et al. (1996), as well as Seidenberg and McClelland (1989), in their earlier work on this reading model, focused on differences in rhyme components of single-syllable words because these are the major determinants of whether a word is orthographically regular (e.g., mint) or irregular (e.g., pint). However, as Seidenberg and McClelland pointed out, the network architecture in these models is capable of capturing any kind of regularity in the orthographic and phonologic sequences it is exposed to, limited only by the extent of exposure. Such regularities would include joint phonemes other than rhymes (e.g., "str" of stream, street, stray, and strum) and, in a multisyllabic version, syllables and morphemes (affixes and the root forms of nouns and verbs, as well as free grammatic morphology, e.g., articles, auxiliary verbs, conjunctions, and prepositions). The simple architecture of the model of Plaut et al. provides adequate support for the orthographic phonologic sequence regularities discussed here to illustrate the basic idea of network instantiation of sequence knowledge. However, as we shall see, somewhat more complex networks are necessary to adequately support the sequence knowledge required of phonologic and grammatic processes (see chapter 3, "Sequence: The Basis of Syntax").

The acoustic–articulatory motor pathway in the model of figure 2.1 would capture analogous patterns in the sequential relationships between acoustic and articulatory word forms. These sequential relationship patterns (captured in part by measures of biphone frequencies) potentially involve sequences of varying length, from phoneme pairs (joint phonemes) and syllables up to and including whole words and possibly, multiple word compounds. These patterns represent the repository of knowledge about subword (sublexical) entities in general, as well as our knowledge of phonotactic constraints (the rules that determine whether or not a given phonologic sequence is permissible in a particular language). This repository of sequence knowledge is additionally influenced by "neighborhood" effects (Vitevitch 1997). Any given phoneme sequence represented in the acoustic–articulatory motor pattern associator comprises part of one or more words. These words define the phonologic neighborhood of that sequence. They are engaged through bottom-up/top-down interactions between the acoustic–articulatory motor pattern associator and the domains of concept representations (semantics). Thus, the final production of a phoneme sequence is a reflection of the combined effects of input (e.g., acoustic), phonotactic effects, neighborhood effects, and noise in the system. Phonotactic frequency effects are seen in tendencies to produce near miss errors that are off by a small number of phonemic distinctive features (usually one or two)—for example, /pat/ in lieu of /bat/. Phonologic neighborhood effects are seen in larger errors that

substantially reflect regularities in patterns of semantic–phonologic knowledge, that is, lexical effects (e.g., /rat/ in lieu of /cat/).

The effect of competing regularities in the acoustic–articulatory motor pattern associator network is dramatically illustrated in the production of past tense forms of verbs. The rich literature on this subject will be discussed later in this book after I have had the opportunity to lay down some additional essential groundwork.

Lexicons

Understanding the meaning of a word that is heard is achieved through the connections between the neural domain that contains the sound features of language and the neural domain that contains concept features (the acoustic–concepts representations pattern associator, figure 2.1, pathway 6–5). This pattern associator network corresponds to the cognitive neuropsychological concept of a phonologic input lexicon (Ellis and Young 1988). It contains neither knowledge of acoustics nor knowledge of semantics—it serves only to translate a representation in the acoustic domain into a representation in the concepts–semantics domain (where meaning is instantiated). This conceptualization of a lexicon as a vast number of connections between two network domains, though well accepted in the connectionist literature, is not intuitive and is strongly at odds with traditional conceptualizations of lexicons as repositories of abstract local representations of single words. However, all representations in the central nervous system are distributed (like the "kitchen" representation), not local, and the knowledge that underlies the capacity to generate a representation lies in connection strengths and is not a piece of data at a memory location (as in a digital computer).[3]

The knowledge that enables a person to translate a concept into a spoken word (the phonologic output lexicon; Ellis and Young 1988) is contained in two different pattern associator networks that connect the concept representations domain to the articulatory motor domain (see figure 2.1, pathways 1–2 and 4–3). These two interacting pattern associator networks support different forms of knowledge. The indirect concept representations–articulatory motor pathway (pathway 4–3) provides a robust basis for knowledge of sequences and sublexical entities because of the sequence knowledge stored in the acoustic–articulatory motor pattern associator. However, the direct concept representations–articulatory motor pattern associator (pathway 1–2) does not contain much knowledge of sequences and sublexical entities because it translates spatially distributed patterns of activity corresponding to concepts into temporally distributed sequences of activity corresponding to articulated words. This spatial–temporal translation precludes significant acquisition of sequence knowledge and makes this substantially a whole word pathway. The existence of this direct, whole word naming route finds support in studies of subjects

with repetition conduction aphasia: some appear to have lost most phonologic sequence knowledge (pathways 3, 4, and 7; resulting in a severe deficit in auditory verbal short-term memory) but can speak quite well, producing few if any phonologic paraphasic errors, can repeat real words (with evidence of influence by semantic attributes but little influence of word length), and are severely impaired in repeating nonwords and functors (Caramazza et al. 1981; Friedrich, Glenn, and Marin 1984; Saffran and Marin 1975; Warrington and Shallice 1969). It also finds some support in reports of subjects with conduction aphasia who are able to repeat words better than nonwords (Caramazza, Miceli, and Villa 1986; Friedrich, Glenn, and Marin 1984; McCarthy and Warrington 1984; Saffran and Marin 1975) and who are able to repeat words better when they are given in a sentence context than when given as a single word (thereby increasing the likelihood of engaging concept representations; McCarthy and Warrington 1984). However, a model in which the only link from the concept representations domain to the articulatory motor domain is the direct one (pathway 1–2) cannot account for observations that normal subjects exhibit phonologic slips of the tongue, and aphasic subjects produce phonemic paraphasias in naming and internally generated spoken language quite comparable to those produced during repetition. To explain these observations, one must posit access from concept representations to phonologic sequence knowledge, as indicated in pathway 4–3 of the model. Thus, this PDP model predicts that there should be two pathways enabling naming of concepts.

Further evidence of two pathways supporting naming of concepts has been provided by a subject who, depending upon type of verbal cue provided, could be induced to use either the whole word (direct) naming route or the phonologic (indirect) naming route (Roth et al. 2006). This left-handed subject had a Broca's aphasia stemming from a massive infarct involving the entire left middle cerebral artery (MCA) territory. His language was largely limited to single words, which he produced quite readily and with good articulation. He tended to pursue a semantic *conduite d'approche*, which however was successful only about 10% of the time. He made very rare phonologic paraphasic errors. When he was asked to name an object—for example, a faucet—and given either no cue or a semantic cue, a typical response would be "dishes … chairs … dishwasher … shut … water … ready to go … water … shut … hot … cold … sink … water … heavy … water … heavy … washer … tub." However, when given the phonemic cue "faus," he replied: "fauwash … fau … fau … fauswah … thafaush … fallshine … fallsha … fallshvine … fallswash … fallsh." These patterns of response to bedside testing suggested that he normally used a whole word route to confrontation naming (as well as in internally generated language; pathway 1–2; see figure 2.1) but that by providing him with a phonemic cue, we could induce him to employ a phonologic route—a route that engaged sublexical representations implicit in his stores of phonologic sequence knowledge

(pathway 4–3; see figure 2.1). He actually was able to successfully name objects 30% of the time using this pathway, but at the cost of producing large numbers of relatively undesirable nonword errors. The dual-route naming hypothesis was further tested and validated with systematic cued naming studies.

As discussed, lexical knowledge corresponds to patterns of connectivity between the domains of concept representations and the acoustic–articulatory motor pattern associator. Because of the critical but counterintuitive nature of this principle, I will consistently use the term semantic–phonologic (lexical) to refer to the knowledge encoded in this connectivity.

The Representation of Knowledge in Auto-Associator and Pattern Associator Networks: Attractor Basins, Attractor Trenches, and Quasi-Regular Domains

A major theme of later sections in this book will be the robustness or redundancy with which certain knowledge is represented in neural connectivity and hence the degree to which that knowledge is susceptible to lesion effects. It is therefore important to have a clear idea of how robustness of knowledge is instantiated in connectionist networks. Put most succinctly, robustness is instantiated in terms of the relative strength of connectivity representing that particular knowledge throughout the network. However, it may be worth enlarging on this point.

Assuming for the moment that the semantic network is undifferentiated (rather than consisting of multiple separate components in various association cortices), then the activity pattern of the auto-associator neural network supporting semantic knowledge corresponds to a function in N-dimensional feature hyperspace. I will refer to a concept representation corresponding to a particular activity pattern in this unitary hyperspace as a "granular distributed concept representation" to signify its dependence on the activity of units corresponding to underlying features. Later, I will recapture the idea that a concept representation is distributed across multiple association cortices by using the term "multifocal granular distributed concept representation" (MFGDCR).

By taking a three-dimensional "slab" of the network activity function, we can achieve some insight into the order in the patterns of activity (Garrard et al. 2001; Woollams et al. 2008; O'Connor, Cree, and McRae 2009). Let us take a slab corresponding to mammal knowledge (see figure 2.3). Because we are dealing with an auto-associator network, network activity tends to settle into attractor basins, one corresponding to mammals in general, the central, lowest energy point of which corresponds to the "centroid" of mammal knowledge—the representation of a creature that best defines our sense of mammalness. Within the mammal basin there are innumerable attractor subbasins corresponding to specific mammals. Very close to the centroid are subbasins corresponding to mammals likely to be very close to the

centroid representation (e.g., dogs, cats, cows, horses, etc.). Distant from this centroid are subbasins corresponding to mammals that are quite atypical (e.g., platypuses and whales). Distance from the centroid is defined by the degree of atypicality, which is defined by feature and feature combination frequency within the semantic domain (mammals in this example; Kiran, Ntourou, and Eubank 2007; Patterson 2007; Woollams et al. 2008). Within any given subbasin, there may be sub-subbasins—for example, corresponding to types of dogs. Within these sub-subbasins may be sub-sub-subbasins corresponding to types within types, or perhaps the Labrador that was our pet and generic Labradors. The depth of a basin relative to that of its subbasins reflects the degree to which features are shared by the subbasins within that basin. For example, the mammal basin is deeper than the tool basin because mammals, by and large, share many more features than tools.

The topography of the various attractor basins within the semantic field reflects regularities in the knowledge of the relationships between concepts. This knowledge, instantiated in neural connections strengths, reflects the impact of regularities in experience (e.g., the mammal attractor basin of an Australian is likely to be somewhat different from that of an American), frequency of exposure, and age of acquisition effects.

The purpose of this metaphor of basins and subbasins is to provide insight into the constraints on the patterns of activity exhibited by all the units in the network. All other things being equal, the activity pattern of the network supporting semantics is more likely to "settle" into deep basins and basins situated near the centroid. However, the state of this network is quite dynamic. Its connectivity is susceptible to recent learning effects: for example, the behavior the network exhibits in settling may be somewhat different if we just spent the afternoon at the zoo. Its state of background activity, and to a minor extent, its connectivity, is susceptible to the influence of our current situation. For example, the network will likely behave a little differently if we are at an aquarium watching killer whales perform. And finally, most importantly, the network's settled activity state will be strongly influenced by the specific input features, which in most circumstances will absolutely define the subbasin or sub-subbasin into which the network settles (its position in state space), all the other factors exerting their major influence either on response latencies or the occasional errors. Errors will consist of slips into nearby subbasins. This settling in response to input features instantiates *content-addressable memory*. Some inputs, rather than leading to settlement into a particular subbasin, may serve to define broad subbasins within the mammal basin, corresponding, for example, to mammals of North America, mammals of the Kalahari, or mammals of particular classes.

The effect of lesions (focal or diffuse) will be to produce graceful degradation of network performance (Hodges et al. 1992; Shallice 1988; Warrington 1975;
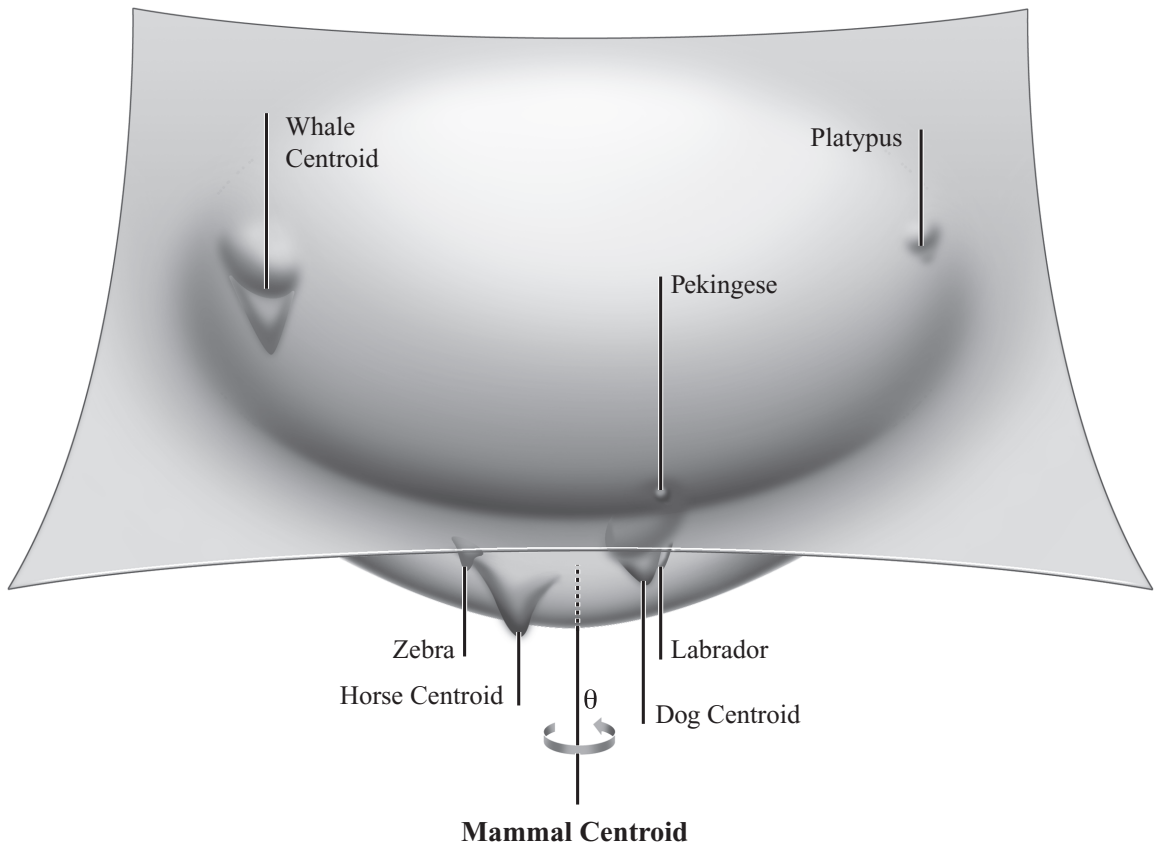
Figure 2.3
The topography of the semantic network activity function in the vicinity of the mammal attractor basin. Each point corresponds to an activity level of all features in an N-dimensional feature hyperspace, in mathematical terms, a state space. Because all features are connected to all other features (or some approximation thereof), the network supports a capacity for settling of the activity function into attractor basins, subbasins within basins, and sub-subbasins, as illustrated here. The point of maximal typicality is represented by the centroid of a basin. Distance from the centroid reflects degree of atypicality. The value of θ defines the manner in which atypically is defined. For example, whales and platypuses are both atypical, but in very different ways.

Warrington and Shallice 1984). Deep basins will become shallower, and subbasins, particularly those that are shallower and more distant from the centroid—corresponding to more atypical exemplars—will disappear (Woollams et al. 2008). The deeper and therefore more resilient basins are defined by knowledge that is represented in neural connectivity with the greatest degree of redundancy, as a result of stronger connection strengths, which reflect frequency, age of acquisition, and familiarity effects (Hinton and Shallice 1991; Hodges, Graham, and Patterson 1995; Rogers et al. 2004). As subbasins become shallow or disappear, responses will reflect the settling of the network into surviving neighbors located nearer the centroid—neighbors of higher typicality (yielding coordinate errors, e.g., horse in lieu of donkey), the parent basin (yielding superordinate errors, e.g., animal in lieu of donkey), or failure to settle at all, yielding omission errors, particularly likely to occur with more atypical exemplars that lack near neighbors. This is precisely what has been observed in semantic dementia (Woollams et al. 2008).

Recent studies contrasting the behavior of subjects with semantic dementia with that of subjects with large strokes and global aphasia appear, at least at first, to raise some questions about this model. This is a serious matter because the model described represents one of the foundations of the theories discussed in this book. Expectably, in word–picture matching tests, subjects with semantic dementia are most accurate in matching superordinate names to pictures (e.g., animal), somewhat less accurate with category names (e.g., dog), and least accurate with subordinate names (e.g., Labrador; Crutch and Warrington 2008). This response pattern is entirely consistent with the effects of degradation of the substrate for semantic representations as discussed above. On the other hand, subjects with large strokes and global aphasia tend to respond with the most accuracy with subordinate names, with intermediate accuracy with category names, and with the least accuracy with superordinate names, notwithstanding that category names are less frequent and subordinate names least frequent (Crutch and Warrington 2008). Crutch and Warrington (2008) offered a number of potential explanations, but the most straightforward one that is fully congruent with the neuroanatomy is that in stroke, damage to white matter connectivity between the substrate for concept representations and perisylvian cortex supporting phonology makes the major contribution to language impairment and, to a varying degree, language is dependent on underdeveloped semantic–phonologic connectivity in the right hemisphere. These white matter connections reflect the frequency with which a particular person associates a particular phonologic word form with a particular concept—something referred to as the basic level effect, which reflects the impact of experience (and with it, expertise) on our choice of name for a given entity (Rogers and McClelland 2004). Many people, on seeing a dog, would call it a "dog." However, many would more often produce the name of the type of dog (e.g., "poodle" or "German

shepherd"). Very few would respond "animal." Basic level names (e.g., "dog" or "poodle") do not bear a consistent relationship to frequency in word databases. This is because they are substantially defined by personal experience and expertise. The basic level terms employed by a veterinarian might involve, for example, specific types of poodles. Basic level effects are also strongly reflective of age of acquisition (see above; Lambon Ralph and Ehsan 2006). Basic level effects interact with the properties of the semantic system. Atypical exemplars (e.g., platypus) receive weak support from semantic attractor basin effects but might receive stronger support from basic level effects wired into semantic–phonologic (lexical) connectivity in someone whose favorite cuddly toy as a child was a platypus. On the other hand, with semantic network damage, superordinate categories (e.g., animals) will have an advantage and will be preserved even as semantic subcategories are lost, however strong the connectivity instantiating basic level effects. Basic level effects also represent, in part, properties of the semantic system. Expertise affects not just frequency of use of particular terms but also degree of differentiation in semantic networks.

In the Crutch and Warrington study (2008), when there was damage to semantic–phonologic connectivity, the most redundantly represented connections linking the substrate for concepts to the substrate for category or subcategory name representations (high basic level effect) were more likely to survive than those linking concepts to superordinate names (lower basic level effect). Why the stroke subjects performed best when supplied subordinate names is less clear. Normal subjects show reliably faster reaction times for category names (e.g., dog) and slower responses to subordinate (e.g., Labrador) and superordinate names (e.g., animal; Crutch and Warrington 2008). The faster response to category names relative to superordinate names is likely to represent patterns instantiated in semantic–phonologic connectivity (the basis for the basic level effect) whereas the faster response to category names relative to subordinate names likely reflects the implicitly hierarchical nature of semantic attractor basin representations. Thus, in the word–picture matching comprehension test, performance level reflects the interactions of regularities in semantic representations, regularities in semantic–phonologic connectivity, and lesion geography, and the results pose no threat to the auto-associator attractor basis conceptualization of semantics introduced above.

The attractor basin metaphor can also be applied to pattern associator networks incorporating auto-associator features, in which case the counterpart of an attractor basin is an *attractor trench*. For example, consider the pattern associator network of Plaut et al. (1996) supporting reading aloud. The network activity functional counterpart of "ust" words is a very simple trench, in effect a groove, linking "…ust" to /…ʌst/. The attractor trench corresponding to "…int" words is more complicated as it corresponds to a quasi-regular domain and incorporates two rather unequal

subtrenches, one corresponding to "…int"—/…Int/, and the other to "…int"—/…int/. The attractor trench corresponding to "…own" words includes two nearly equal subtrenches, one corresponding to words like grown, shown, and flown, the other to words like crown, gown, and down. Within the reading aloud pattern associator network activity function, there is a trench that includes subtrenches corresponding to tough, rough, and enough; cough and trough; though, dough, and borough; through and slough; plough; drought; and ought, fought, thought, bought, brought, sought, and wrought. To bring this conceptualization closer to the central themes of this book, there exists a complex attractor trench within the acoustic–articulatory pattern associator network corresponding to the past tense of English verbs. One subtrench, a particularly large one, corresponds to verbs that accept an "ed" to form the past tense (e.g., jump–jumped, groom-groomed, and hate–hated). Others correspond to variously "irregular" verbs—for example, swim–swam, sing–sang, bleed–bled, slide–slid, and go–went. The rich terrain of verb past tense formation has been studied extensively and will be the topic of a later section in this book (in chapter 4, "Disorders of Grammar in Aphasia," see the subsection "Verb Past Tense Formation" in the "Grammatic Morphology—Special Cases" section). Attractor trenches exist only to the extent that there are regularities in the pattern associator network domain (in pattern associators supporting orthogonal domains, trenches are reduced to grooves supporting single relationships, e.g., between word meaning and word sound).

Attractor basins (e.g., mammals) and trenches (e.g., "int" words) constitute the neural network counterparts of quasi-regular domains (see figure 2.4). They are regular because the exemplars have many features in common. They are only quasi-regular to the extent that exemplars have features that distinguish them. Subbasins and subtrenches correspond to competing regularities within these domains.

The general principles diagrammed in figure 2.4 presumably also characterize all the networks supporting grammatic function. Understood in this way, it becomes quite clear why the single most important influence on the pattern of language performance of an aphasic subject is the terrain of knowledge implicit in the connectivity of these networks that was established in the acquisition of his or her particular language. The locus and extent of the lesion have a somewhat more modest influence.

Semantic–Phonologic (Lexical) and Phonologic Impairment in Aphasias

Discussion of insights into the neural architecture of grammar that have been gained from studies of subjects with aphasia will generally be limited to chapter 4 ("Disorders of Grammar in Aphasia"). However, I will make a brief digression into aphasia at this point in order to link the model to brain anatomy and to
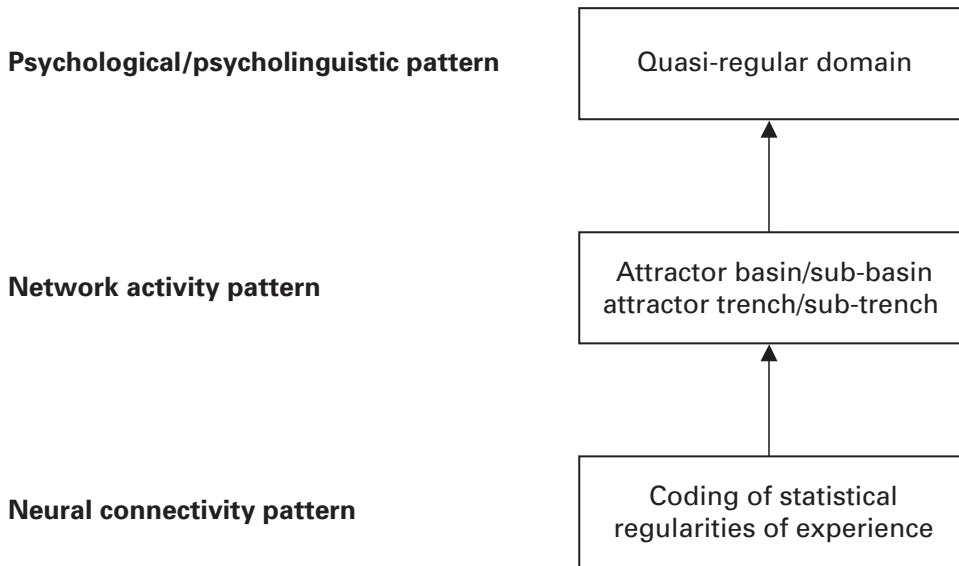
Psychological/psycholinguistic pattern   Quasi-regular domain

Network activity pattern   Attractor basin/sub-basin
attractor trench/sub-trench

Neural connectivity pattern   Coding of statistical
regularities of experience

Figure 2.4
The relationship between neural network connectivity, attractor basins–trenches, and patterns of
behavior.

introduce an important subtheme that will appear later in the book: the likely
importance of the nondominant hemisphere in accounting for much aphasic
phenomenology.

Phonologic Paraphasic Errors
Phonologic paraphasic errors are generally thought to reflect damage to dominant-
hemisphere networks supporting phonologic processing. In our model, this would
correspond to the dominant-hemisphere acoustic–articulatory motor pattern asso-
ciator network (pathway 7–3; see figure 2.1)—the repository of phonologic sequence
knowledge. Nonpropositional spoken language, which may be supported by the
nondominant hemisphere (Speedie et al. 1993), does reflect sequence knowledge,
but discrete phrasal, lexical, or sublexical elements of this knowledge cannot be
selected at will as they can in propositional language processing. The subject
described in the foregoing (Roth et al. 2006), who could be cued to use one or other
of the two concept naming routes (pathways 1–2 or 4–3; see figure 2.1), provides
new insight into the neural basis of phonologic paraphasic errors. His left hemi-
sphere was nearly completely destroyed. Thus, he must have been speaking with his
right hemisphere, which was undamaged, and his performance when naming after
phonemic cueing suggests that he was using discretely accessible phonologic

sequence knowledge represented in his right hemisphere. This suggests, in turn, that deficient development of networks instantiating this knowledge (as contrasted to damage to fully developed networks) can provide an alternative basis for phonologic paraphasic errors.

Factors Influencing Phonologic and Lexical Errors in Internally Generated Aphasic Language
In the two route naming model I have introduced, three factors may influence the pattern of errors observed in internally generated spoken language in perisylvian aphasias, and whether this pattern is marked exclusively by impaired word retrieval or, additionally, by phonemic paraphasic errors. First, given the likely anatomic representation of the network shown in figure 2.1, most dominant perisylvian lesions probably damage both the whole word and the phonologic output routes (see figure 2.5), and the pattern of spoken output may reflect the relative degree to which these two pattern associator networks are affected. This would explain why subjects with Wernicke's or conduction aphasia apparently do not have the option of relying entirely on the whole word route. Second, these two output pattern associator networks are likely to be differentially represented in the two hemispheres, the phonologic pathway being more frequently better developed in the dominant hemisphere and the whole word pathway more equally developed in the two hemispheres. Subjects with dominant-hemisphere perisylvian lesions almost invariably demonstrate impaired if not completely absent phonologic sequence knowledge but often exhibit partial sparing of semantic–phonologic (lexical) knowledge. Studies of subjects with callosal disconnection demonstrate that the disconnected right hemisphere has a phonologic input lexicon and conceptual semantic knowledge but impoverished phonologic processing (Zaidel et al. 2003). Third, there may be individual variability in the degree to which connectivity is developed in these two output routes, and this individual variability may vary as a function of hemisphere. Thus, in order to fully understand language production following a left-hemisphere lesion, a bihemispheric language model that incorporates both developmental attributes and the impact of the lesion must be considered (see figure 2.6). This will be a recurring theme throughout this book.

   As the case of Roth et al. (2006) suggests, in the presence of an extensive left-hemisphere perisylvian lesion, deficient development of connectivity underlying nondominant-hemisphere concept representations may lead to the generation of anomia and semantic paraphasic errors, and deficient development of connectivity in the nondominant-hemisphere phonologic route may lead to the generation of phonemic paraphasic errors. The deficient development of both systems, particularly the phonologic one, may be a general characteristic of the right hemisphere, but deficient development of the phonologic route might also occur to one degree or
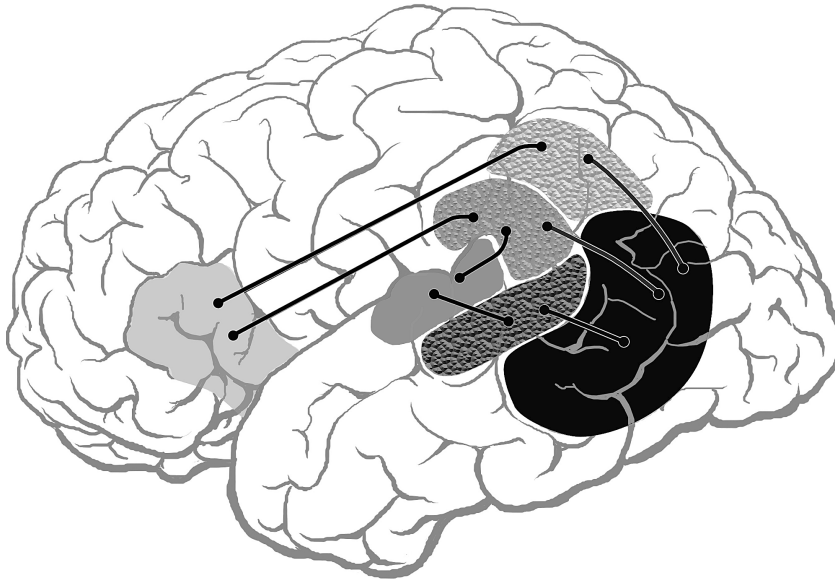
Figure 2.5
Illustration depicting the network of figure 2.1 mapped onto the brain. Shade coding is as in figure 2.1. Concept representations are assumed to be widely distributed across association cortices throughout the brain. In this illustration, only the region of presumed interface between concept representations and the remainder of the model is depicted. Given the paucity of information about the anatomic organization of the human perisylvian region, the mapping depicted here is, at best, approximate; the goal is primarily to demonstrate the feasibility of mapping a connectionist architecture of phonologic processing to cortical anatomy. However, recent magnetic resonance imaging diffusion tensor imaging tractographic studies of deep white matter pathways are shedding some light on the anatomic details. Catani, Jones, and Fytche (2005) have delineated two dominant perisylvian pathways linking Wernicke's and Broca's areas, a direct one, corresponding to the arcuate fasciculus (possibly corresponding to pathway 3 in our model), and an indirect one, projecting from Wernicke's area to inferior parietal cortex (Brodmann's areas 39 and 40), with apparent relay to Broca's area via what is likely component III of the superior longitudinal fasciculus (Makris et al. 2005; possibly corresponding to pathway 1–2 in our model). They concluded that the broad extent of origin and termination of these pathways and their large cross sections favored a connectionist account. (From Roth, H. L., S. E. Nadeau, A. L. Hollingsworth, A. M. Cimino-Knight, and K. M. Heilman. 2006. "Naming Concepts: Evidence of Two Routes." *Neurocase* 12:61–70.)
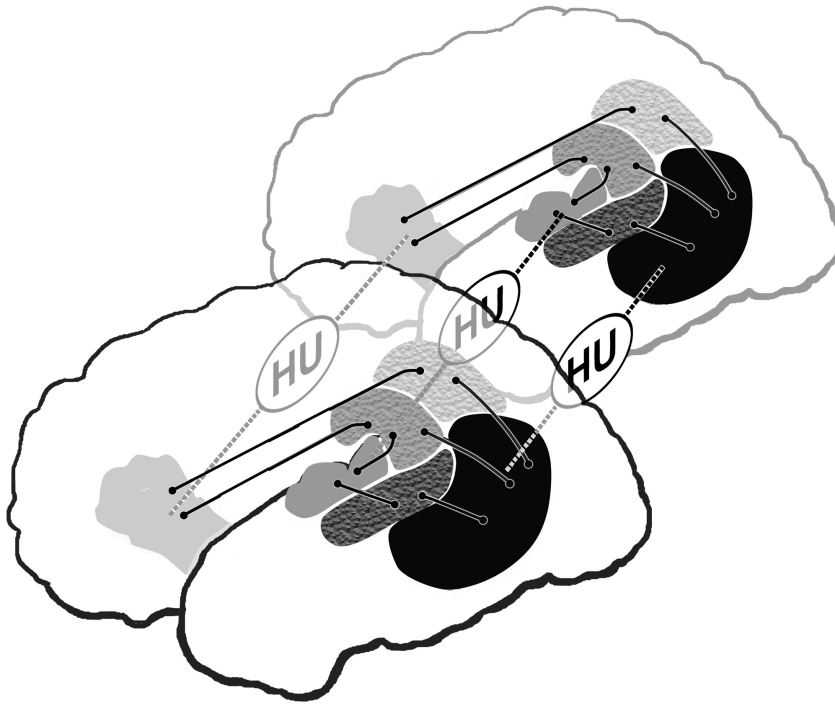
Figure 2.6
Illustration depicting mapping of a bihemispheric model to the brain. This provides the basis for a fuller explanation of the results of left-hemisphere lesions on language output in terms of bihemispheric contributions, the effect of the lesion, and the degree to which various networks are developed in each hemisphere. Shade coding is as in figures 2.1 and 2.3. HU = hidden units. (From Roth, H. L., S. E. Nadeau, A. L. Hollingsworth, A. M. Cimino-Knight, and K. M. Heilman. 2006. "Naming Concepts: Evidence of Two Routes." *Neurocase* 12:61–70.)

another in the left hemisphere as a result of normal variability in phonologic network ontogenesis (Goodglass 1993; Plaut et al. 1996).

In summary, with any given left-hemisphere lesion, the actual pattern of internally generated spoken language may reflect three factors: effects of the lesion on the two output routes (phonologic and whole word), the degree to which one or both of these routes is represented in the right hemisphere, and the degree of development of connectivity in each of the two routes in each hemisphere. If both naming routes are involved and the combined effect of the three factors differentially impacts the phonologic routes, the net result will be output marked predominantly by word retrieval deficits (with or without semantic paraphasic errors) as whole word routes become the predominant means of language production (as in repetition conduction aphasia). If whole word routes are differentially impacted, the net

result will be output marked by word retrieval deficits and a substantial incidence of phonemic paraphasic errors as damaged or inadequately developed phonologic routes become the predominant means of language production.

Subjects with reproduction conduction or Wernicke's aphasia predominantly use damaged or inadequately developed phonologic pathways. Wernicke's aphasia may reflect more severe impairment, hence greater difficulty with word retrieval and more profuse phonemic paraphasic errors than with conduction aphasia. Subjects with Wernicke's aphasia may also have damage to acoustic representations or the acoustic representations–concept representations pathways that enable verbal comprehension. Naming difficulty may arise through mechanisms discussed in the preceding paragraph or in two additional ways. First, damage to Brodmann's areas 37 and 39 in the dominant hemisphere may be associated with word retrieval deficits (Chertkow et al. 1997; Hart and Gordon 1990; Raymer et al. 1997; Whatmough et al. 2002). These areas may constitute the interface between association cortices throughout the brain supporting concept representations and the core language apparatus (proximal to the point at which the phonologic naming route provides differential access to sequence knowledge; Geschwind 1965). Second, dysfunction of conceptual (semantic) networks because of left-hemisphere damage and deficient development of right-hemisphere networks, as in the subject of Roth et al. (2006), would be expected to yield naming difficulty with production of semantic paraphasic errors.