Psychological Review
1997, Vol. 104, No. 4, 801-838

0033-295X/97/$3.00

Lexical Access in Aphasic and Nonaphasic Speakers

Gary S. Dell
University of Illinois

Myrna F. Schwartz
Moss Rehabilitation Research Institute

Nadine Martin and Eleanor M. Saffran
Temple University School of Medicine

Deborah A. Gagnon
Moss Rehabilitation Research Institute

An interactive 2-step theory of lexical retrieval was applied to the picture-naming error patterns of aphasic and nonaphasic speakers. The theory uses spreading activation in a lexical network to accomplish the mapping between the conceptual representation of an object and the phonological form of the word naming the object. A model developed from the theory was parameterized to fit normal error patterns. It was then "lesioned" by globally altering its connection weight, decay rates, or both to provide fits to the error patterns of 21 fluent aphasic patients. These fits were then used to derive predictions about the influence of syntactic categories on patient errors, the effect of phonology on semantic errors, error patterns after recovery, and patient performance on a single-word repetition task. The predictions were confirmed. It is argued that simple quantitative alterations to a normal processing model can explain much of the variety among patient patterns in naming.

Difficulty in word retrieval is the most pervasive symptom of language breakdown in aphasia. As with other symptoms of brain damage, word retrieval is subject to graceful degradation (Marr, 1982; Rumelhart & McClelland, 1986): Unsuccessful attempts at retrieval generally resemble the target, either in sound or in meaning. The similarities between these aphasic "paraphasias" and normal slips of the tongue have been apparent to most students of aphasia, among them Sigmund Freud, who asserted that "the paraphasia in aphasic patients does not differ from the incorrect use and the distortion of words which the healthy person can observe in himself in states of fatigue or divided attention" (Freud, 1891/1953, p. 13).

We call this claim the *continuity thesis*. In this article we examine the continuity thesis in the context of word retrieval in production, using a spreading activation model to predict the distribution of error types in normal speakers and fluent aphasic speakers in a picture-naming task. A major objective was to extend the range of observations that bear on the architecture of the speech production system. Demonstrating that the model can handle error frequencies and distributions that represent significant departures from normality would serve not only to validate the continuity thesis but also the assumptions of the model.

Error types, defined by their intrinsic (e.g., semantic, phonological) or contextual (e.g., distance between target and error source) relation to the target, have provided important clues to the architecture of the normal production system (e.g., Fromkin, 1971; Garrett, 1975, 1980). Accordingly, here we focus on the distribution of error types in aphasic naming. We examine the relation between the distribution of types of naming errors in individual patients and the error distributions generated by perturbations (implemented as departures from normal parameter settings) in the spreading activation model. Our study therefore differs from previous assessments of the continuity thesis (e.g., Buckingham, 1980; Kohn & Smith, 1990), which have been concerned with the detailed structural similarities between normal and aphasic speech errors. In particular, we do not seek to account for the precise character of phonological errors, although the approach we set out is compatible with a number of current proposals (e.g., Dell, 1988; Levelt & Wheeldon, 1994; Sevald & Dell, 1994; Shattuck-Hufnagel, 1979, 1987; Stemberger, 1990).

In this article, we present background information on the aphasia syndromes and the classification of paraphasias. We then discuss the two-step theory of lexical retrieval, which provides a framework for explaining paraphasias, and the implemented model that combines the two-step notion with an interactive activation retrieval mechanism. In subsequent sections, we use this model to fit the naming data from a group of nonaphasic

Gary S. Dell, Beckman Institute, University of Illinois; Myrna F. Schwartz and Deborah A. Gagnon, Moss Rehabilitation Research Institute, Philadelphia, Pennsylvania; Nadine Martin and Eleanor M. Saffran, Center for Cognitive Neuroscience, Temple University School of Medicine.

This research was supported by National Institutes of Health Grants DC-00191 and DC-01924 and the National Science Foundation Grant SBR 93-19368. A preliminary version of some of this work was presented in Dell, Schwartz, Martin, Saffran, and Gagnon (1996).

We are grateful for the invaluable assistance of A. Roach, A. Brecher, R. Grewal, the advice and comments of K. Bock, H. Buckingham, R. Berndt, J. Cutting, S. Doane, V. Ferreira, Z. Griffin, B. Gordon, S. Garnsey, P. Gupta, J. Huitema, J. McClelland, D. Plaut, and C. Sevald, and the technical contributions of D. Adams, A. Govindjee, J. Lowery, and L. May.

Correspondence concerning this article should be addressed to Gary S. Dell, Beckman Institute, University of Illinois, 405 North Mathews Avenue, Urbana, Illinois 61801. Electronic mail may be sent via Internet to gdell@s.psych.uiuc.edu.

control participants and the individual data from 21 fluent aphasic patients. Finally, we use the model's characterization of each patient to predict other, previously unexamined aspects of their behavior.

Paraphasias in Aphasic Speech and Naming

Aphasia Subtypes

A major distinction in the classification of aphasia is between the conditions that affect motor speech production, and thereby fluency, and those that do not. This fluent–nonfluent distinction has its basis in the neuroanatomical organization of language in the left hemisphere: Nonfluent aphasia arises from damage to the anterior speech zone, including Broca's area; fluent aphasia arises from more posterior lesions, including Wernicke's area and the inferior parietal lobule. Although all of the types of paraphasias that are of interest to us here can be found in the naming efforts of both nonfluent and fluent aphasic individuals (Howard, Patterson, Franklin, Orchard-Lisle, & Morton, 1984), our study concerned only fluent aphasic individuals. We have found that the articulatory-phonetic impairment in nonfluent aphasia makes reliable coding of sublexical paraphasias highly problematic.

Fluent aphasia subsumes a number of diagnostic syndromes that are distinguished from one another by, among other things, the predominance of one or more types of paraphasias in connected speech. Historically, paraphasias have been divided into two broad classes: (a) *lexical errors* (verbal paraphasias), in which one word is substituted for another, and (b) *sublexical errors* (e.g., phonemic paraphasias, neologisms), in which the phonological structure of a recognizable word is deformed by substitution, addition, deletion, or transposition of segments or syllables (Blumstein, 1973; Lecours & Lhermitte, 1969). Lexical errors may relate to the target in meaning (semantic errors; e.g., van → bus), in sound (formal error; e.g., train → tree), or in both meaning and sound (mixed error; snail → snake), or they may bear no relation to the target (unrelated error; e.g., banana → drum). Sublexical errors are categorized as phonemic paraphasias when their relation to a target is obvious (ghost → /goθ/) and neologisms when the relation is more remote (cane → /tΛ).

The speech of the individual with Wernicke's aphasia is peppered with lexical and sublexical paraphasias of all types, giving it a jargonlike quality.[1] The other varieties of fluent aphasia manifest more selectivity. Anomic aphasic patients produce occasional semantic errors in the context of overt word searches and circumlocutions. Conduction aphasic patients produce primarily phonemic paraphasias, with self-interruption and repeated attempts to repair the deviant utterance. Our sample included patients from all these fluent aphasia subtypes.

It is interesting that on picture-naming tasks these differences among the aphasia subtypes are greatly minimized, with lexical and sublexical errors occurring in all of them (Howard et al., 1984; Kohn & Goodglass, 1985; Mitchum, Ritgert, Sandson, & Berndt, 1990). No doubt this reflects the different cognitive and linguistic demands of naming and spontaneous speech, but an additional factor seems to be that connected speech lends itself to avoidance strategies in a way that naming does not. That is, the aphasic speaker can and often does choose to circumvent problematic vocabulary, such as by altering the character of the message or by using shorter and more familiar words. This is not possible in naming, and thus naming tasks may reveal difficulties with word retrieval or phonological encoding beyond what is evident in speech. This is one reason why we chose to focus on naming. Another is that the input to the naming task provides no sublexical clues to the target's phonology, unlike reading and repetition. Furthermore, it is considerably easier to identify targets for paraphasias generated in naming, compared with spontaneous speech.

Although the naming task homogenizes the syndrome-specific error profiles that emerge in spontaneous speech, dissociations at the level of the individual patient do occur. Thus, there are some patients in whom the predominant or exclusive type of commission error in naming is the semantic error (e.g., Berndt, Basili, & Caramazza, 1987; Caramazza & Hillis, 1990; Hillis & Caramazza, 1995; LeDorze & Nespoulous, 1989), whereas in other patients it is the sublexical error (e.g., Caplan, Vanier, & Baker, 1986; Ellis, Miller, & Sin, 1983; Kay & Ellis, 1987). Such dissociations are well accounted for by models in which retrieval of the semantic word form and the phonological word form are distinct processes, supported by spatially distinct regions of the brain and hence subject to selective damage (Cappa, Cavallotti, & Vignolo, 1981). Our model maintains the distinctiveness of semantic and phonological retrieval but offers a new account of how the dissociations in aphasics naming come about.

Varieties of Paraphasias

Sublexical paraphasias. As noted earlier, the speech of conduction aphasic individuals is marred by frequent phonemic paraphasias and successive attempts to repair them. These repairs tend to bring the patient closer to the target (Joanette, Keller, & Lecours, 1980; Valdois, Joanette, & Nespoulous, 1989), but rarely are they completely successful (Kohn, 1984). During this "conduite d'approche," the patient generally produces most, if not all, the phonemes of the target word in varying combinations (Butterworth, 1992).

These facts have fostered the view that phonemic paraphasias represent faulty encoding for articulation of the phonological representation retrieved from the lexicon. The many parallels between these errors and segmental errors in normal speech supports this view (for reviews, see Buckingham, 1980; Garrett, 1984; Schwartz, 1987). For example, like normal slips of the tongue, phonemic paraphasias are restricted to open class vocabulary and are nearly always phonotactically well formed (Buckingham & Kertesz, 1976; Lecours, 1982). When the error involves misordering, the interacting elements tend to be related

[1] The same description applies to transcortical sensory aphasia, which is distinguished from Wernicke's aphasia by the relative sparing of repetition. Patterns of performance across language tasks (e.g., speaking, repeating, comprehending) are at least as important in syndrome classification as the quality of connected speech (Goodglass & Kaplan, 1983; Kertesz, 1982).

phonetically and by their structural position in the word (Blumstein, 1973; Lecours & Lhermitte, 1969), and the phonemic content that precedes or follows the interacting segments also tends to be similar (Lecours & Lhermitte, 1969).

Not all the facts about phonemic paraphasias parallel the normal findings, however: A large proportion of phonemic paraphasias have no source in the immediate context (Blumstein, 1973; Talo, 1980), and many others have their source within the same word as the error segment (Blumstein, 1973; Lecours & Lhermitte, 1969; Schwartz, Saffran, Bloch, & Dell, 1994; see especially Pate, Saffran, & Martin, 1987). For those that do involve between-words movement errors, the expected preference for word onset involvement (Shattuck-Hufnagel, 1987) is not reliably observed (Kohn & Smith, 1990; however, see Blumstein, 1973).

As noted earlier, phonemic paraphasias are distinguished from other phonologically deviant responses by their greater phonological proximity to the target. The criteria used to identify phonemic paraphasias vary across studies. Consequently, some investigators (following Butterworth, 1979) have opted to cast a wider net for sublexical errors, eschewing the traditional distinction between phonemic paraphasias and neologisms and instead grouping errors into two categories: the target-related neologism, in which there is evidence of phonological overlap with a conceivable target, and the non-target-related (or abstruse) neologism, in which the target is unidentifiable. This has led to a number of important discoveries.

1. Neologisms in spontaneous speech are preceded by hesitation pauses greater than those before well-formed targets and well-formed word substitutions (Butterworth, 1979, 1985; Ellis et al., 1983). Abstruse neologisms are preceded by significantly longer pauses than target-related neologisms (Butterworth, 1979).

2. Neologisms in naming tend to preserve the word shape of the target, that is, its stress pattern and number of syllables (Ellis et al., 1983; Miller & Ellis, 1987), as well as the target's initial consonant (Gagnon et al., in press). This is just the type of information that remains available in tip-of-the-tongue (TOT) states (for a review, see Brown, 1991).

3. Neologisms occur more often to low- than high-frequency targets (Caramazza, Berndt, & Basili, 1983; Ellis et al., 1983; Kay & Ellis, 1987), another characteristic they share with TOT states (Burke, MacKay, Worthley, & Wade, 1991).

Whereas studies focusing exclusively on phonemic paraphasias have implicated postlexical phonological encoding, the data derived from the more inclusive grouping suggest a problem in lexical-phonological retrieval. Both accounts may be true: The retrieval deficit may explain the more remote deviations and the encoding deficit the simpler errors that more closely resemble speech errors (Buckingham, 1987; Kohn, 1993; Kohn & Smith, 1994a). In particular, it may be the case that with sufficiently severe retrieval blocks, a special "neologism generator" comes into play that supplies phonological context de novo (Butterworth, 1979, 1985; see also Buckingham, 1987).

Nevertheless, our model does not contain such a device, and it accounts for all sublexical paraphasias in the same way. Such errors occur during a phonological access step that involves the selection of a phonological frame, phonological segments, and the insertion of segments into frame slots (e.g., Berg, 1988; Dell, 1988; Levelt, 1989; Meyer & Bock, 1992; Shattuck-Hufnagel, 1979; Stemberger, 1990). These processes can be more or less disrupted, resulting in neologisms that differ greatly from the target at one extreme and errors that differ in only a single sound at the other extreme (for related accounts, see Ellis, 1985; Miller & Ellis, 1987; Schwartz et al., 1994).

Lexical paraphasias. The most studied type of lexical paraphasia is the substitution of a word that bears a semantic relation to the target. These semantic errors have long been of interest to aphasiologists for what they reveal about the organization of semantic knowledge (Buckingham & Rekart, 1979; Rinnert & Whitaker, 1973) and how this enters into speech planning (Garrett, 1992). In addition to supporting a central semantic component that is shared by all language tasks—expressive and receptive (Butterworth, Howard, & McLoughlin, 1984), spoken and written (Caramazza & Hillis, 1990)—these studies also have provided some support for word production models in which the semantic component transmits activation in parallel to the phonological forms of all words that approximate the semantic specification (Caramazza & Hillis, 1990; Howard & Orchard-Lisle, 1984). In two-step models of lexical retrieval, in which there is a mediating lexical representation between semantics and phonology, this can be achieved by feedforward, or "cascading," activation (Humphreys, Riddoch, & Quinlan, 1988; McClelland, 1979). Ours is a two-step model that incorporates both feedforward and feedback activation, in short, interaction.

As we explain, the model requires interactive feedback to account for lexical errors that bear a purely phonological relation to the target. Aphasiologists debate whether these actually occur in aphasic speech. Certainly, Wernicke's and conduction aphasic individuals produce form-related word errors in their connected speech, but the low frequency of these in comparison to form-related nonword errors raises the possibility that what appear to be lexical errors that overlap in form are in reality sublexical errors that happen by chance to create real words (what Butterworth, 1979, called "jargon homophones"; see also Lecours, Deloche, & Lhermitte, 1973). Contrary to this are reports of patients who, on naming tasks, produce more formal errors than is predicted by their rate of production of neologisms and estimates of chance creation of words (Best, 1996; Blanken, 1990; Martin & Saffran, 1992). Moreover, Gagnon et al. (in press), in a separate analysis of naming data from this study, found that patients' formal errors showed effects of word frequency and grammatical class, lexical factors that should exert no effect on jargon homophones. One of the challenges we have posed for ourselves in this study was to predict, from the distribution of their naming errors, which patients would produce true formal errors in addition to jargon homophones.

Among the types of lexical errors that concern us here, mixed (semantic + formal) errors are perhaps the most important theoretically because they can be interpreted as evidence for the interaction of meaning and form information during lexical access. We postpone further discussion of mixed errors to first develop the context in which the debate over interactivity has taken shape (viz. the two-step lexical access hypothesis).

Stages in Naming: The Two-Step Lexical Access Hypothesis

Picture naming, roughly speaking, involves the translation of a visual stimulus into a conceptual representation, the retrieval of the name of the picture, and the articulation of that name (e.g., Glaser & Glaser, 1989; La Heij, 1988; Potter & Faulconer, 1975; Theios & Amrhein, 1989). Our concern is with name retrieval, or lexical access. The input to name retrieval is a conceptual specification of the pictured object, which we take to be a set of features, and the output is the phonological form of the object name, a set of phonemes specified for order, and syllabic and metrical organization.

Detailed accounts of lexical access in production recognize that the mapping from a conceptual representation to a phonological form is complex. It is complex in at least two senses: First, it involves the coordination of different kinds of information: conceptual, pragmatic, syntactic, and phonological information (see chaps. 6 and 9 in Levelt, 1989). Second, the mapping between concepts and phonological form is a mapping between two unrelated spaces. Aside from morphologically related words and isolated cases of phonetic symbolism, there is no tendency for words that are similar in form to be similar in meaning.

Most theories of lexical access deal with the complexity of the meaning to sound mapping by assuming that access involves two steps (Butterworth, 1989; Dell, 1986; Fay & Cutler, 1977; Fromkin, 1971; Garrett, 1975; Kempen & Huijbers, 1983; Levelt, 1989; Levelt et al., 1991a; Roelofs, 1992; see Starreveld & La Heij, 1996, for an opposing view). The first step, *lemma access,* consists of a mapping from a conceptual representation to a *lemma,* a nonphonological representation of a word. The lemma is associated with semantic and grammatical information, such as major syntactic category, and other features that are part of the grammar of the language (e.g., number for nouns). The second step, *phonological access,* consists of the mapping from the lemma to the phonological form of the word.

Why two steps? Why not go straight from a conceptual to a phonological representation? There are both functional and empirical reasons. On the functional side, the arbitrary relation between form and meaning motivates an intermediate step if the mapping is carried out by spreading activation. Specifically, if the meaning and phonological form of a word are each activation patterns across a set of nodes, a one-step mapping, realized by direct connections between meaning nodes and form nodes, would not be possible.[2] Table 1 illustrates this impossibility by providing hypothetical specifications of the meanings of *mother, father, woman,* and *man* and considering the phonological feature of whether each word begins with /m/. Notice that the two /m/-initial words are dissimilar in meaning. *Man* is neither female nor parent, and *mother* is both. The mapping is formally equivalent to the exclusive-OR function. For a network to compute this function, it needs an intermediate layer of nonlinear "hidden" nodes (Minsky & Papert, 1969; Rumelhart, Hinton, & Williams, 1986). Lemmas serve the function of the required intermediate layer.

Table 1
Hypothetical Specifications for Four Words Showing Formal Equivalence to the Exclusive-OR Relation

	Semantic input		Phonological output
Word	Female	Parent	Begins with /m/
Mother	1	1	1
Father	0	1	0
Woman	1	0	0
Man	0	0	1

Lemmas are more than just hidden units, however. They are further motivated by the fact that the syntactic structure of a sentence is not the same as its conceptual or phonological structure. Sentence production requires the manipulation of words as syntactic entities according to purely syntactic considerations (Bock & Eberhard, 1993; Bock & Loebell, 1990; Garrett, 1975; Levelt, 1989). Lemmas are the units that guide these manipulations.

Empirical support for two steps in lexical access comes from several sources: speech errors in aphasic and nonaphasic speakers (e.g., Garrett, 1984, 1975, 1980), analysis of the time course of lexicalization (e.g., Schriefers, Meyer, & Levelt, 1990), experimental studies of the production of multiword utterances (Ferreira, 1993; Kempen & Huijbers, 1983; Meyer, 1994; Schriefers, 1992), and the TOT phenomenon (e.g., Meyer & Bock, 1992).

We have distinguished between speech errors that involve whole words (lexical errors) and errors that involve the sounds of words (sublexical errors). The former can be associated with lemma access and the latter with phonological access. Importantly, whole-word substitution and exchange errors often involve words of the same syntactic category (Fay & Cutler, 1977; Garrett, 1975, 1980; MacKay, 1982). Nouns replace nouns, verbs replace verbs, and so on. At the same time, these errors, particularly the exchanges, frequently involve words that are not similar in sound (Garrett, 1975). This syntactic category constraint in the absence of strong phonological relations is good evidence that a lemmalike representation is being manipulated. One also can see the two steps at work in complex errors. For example, Garrett cited the error "The skreeky gwease gets the wheel" for "The squeaky wheel gets the grease." It looks like the lemmas for "wheel" and "grease" exchanged—they are both nouns—and this was followed by a phonological error involving "squeaky" and "grease," which became adjacent after the lemma exchange. One also can interpret some naming errors made by aphasic patients in terms of two steps. Errors such as *unicorn* spoken as "house" can be seen as a lemma access error *unicorn* → *horse,* followed by a phonological access error that transforms *horse* into "house" (Martin, Dell, Saffran, & Schwartz, 1994).

Several experiments represent attempts to test the two-step

[2] We have been assuming that the conceptual representations of pictures are sets of features. An alternative (Roelofs, 1992) is that these representations each consist of a single nondecomposed unit. If this is the case, then the mapping from nondecomposed units to phonological form is not complex. Roelofs, nonetheless, adhered to the two-step conception.

notion by determining when semantic and phonological information are active during the course of picture naming. In these studies, participants have the primary task of naming a picture. While they are engaged in this task, participants see a word that may be either semantically or phonologically related to the picture name. In some experiments (Schriefers, Meyer, & Levelt, 1990), the participants simply try to ignore the word and its interference with picture naming is assessed. In other experiments, they must respond to the word either by naming it (Peterson & Savoy, in press) or making a lexical decision to it (Levelt et al., 1991a), and the response time to the word is the main dependent measure. All of these studies manipulate the time of presentation of the word relative to picture onset. They show that semantically related words exert an influence early in the process and that phonologically related words have an effect later on. Assuming that the semantic effects are occurring during lemma access (for evidence, see Schriefers et al., 1990) and the phonological effects during phonological access, these data are consistent with the two-step hypothesis.

The TOT phenomenon also supports a two-step view of lexical access because a TOT state can be characterized as a failure of phonological access in conjunction with successful lemma access. The speaker knows that a word exists—the lemma has been retrieved—but cannot access the word's sounds. More strikingly, speakers often know the gender of TOT words in languages such as Italian that have grammatical gender (Badecker, Miozzo, & Zanuttini, 1995; Vigliocco, Antonini, & Garrett, in press). These results support the association of grammatical information with lemmas.

A Two-Step Theory Implemented by Interactive Activation

Our model combines the two-step notion with an interactive activation retrieval mechanism. Lexical knowledge is embedded in a network of three layers—a semantic, a word (lemma), and phoneme layer—as shown in Figure 1. The semantic layer contains units that represent the concept of the pictured object.

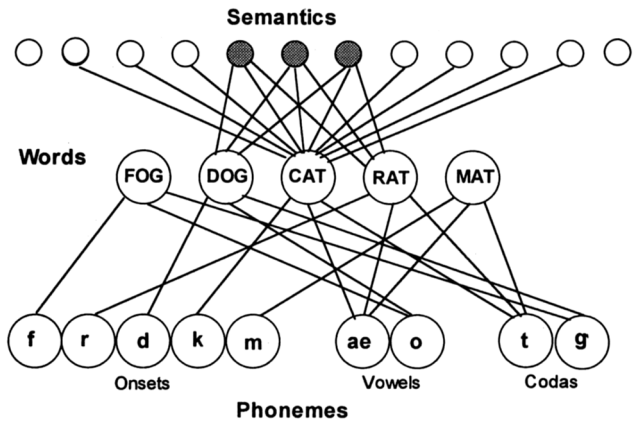

Figure 1. Illustration of a lexical network for the interactive two-step model of naming. Connections are excitatory and bidirectional. The common semantic features of *cat, dog,* and *rat* are shaded in gray.

Specifically, we assume that each object corresponds to 10 units. The 10 units for each object connect to the object's word node by excitatory bidirectional connections. Each word node also connects to its phoneme nodes, also through bidirectional excitatory connections. There are no inhibitory connections. The strength of each connection is assumed to be a product of learning and recent experience.

Because connections run in both top-down and bottom-up directions, the model is a form of interactive activation model. Connections allow activation to spread from semantics to word to phoneme units and from phoneme to word to semantic units. In this respect, it is similar to a variety of interactive models of lexical access in production (Berg, 1988; Dell, 1985, 1986; Harley, 1984; MacKay, 1987; Stemberger, 1985, 1990, 1991), and it differs from modular two-step models of lexical access (e.g., Butterworth, 1989; Fay & Cutler, 1977; Garrett, 1980; Levelt et al., 1991a). This combination of interaction with two distinct steps is one of the most important aspects of the model, enabling it to account for the data that motivate the two-step approach as well as a variety of error phenomena that suggest interaction among processing levels. To show how the model can do this, we provide an example and then use the example to illustrate error phenomena. First lemma access and then phonological access are described.

Lemma access. Assume that a picture of a cat is presented. Visual processes that are outside of the model identify the picture and the 10 semantic nodes corresponding to the *cat* concept are each given a jolt of activation. The size of the jolt was arbitrarily set at 100, which, when divided among 10 nodes, gives a jolt of 10 to each semantic node for *cat.* This activation then spreads throughout the lexical network for n time steps according to a linear activation function:

$$A(j, t) = A(j, t - 1)(1 - q) + \sum_i w(i, j)A(i, t - 1) + \text{noise}, \quad (1)$$

where $A(j, t)$ is the activation of node j at time step t, q is a decay parameter, $w(i, j)$ is the connection weight from node i to node j. For the implemented model, it is assumed that each of these weights is the same, designated by p. Each node's activation also is perturbed by normally distributed noise during each time step. This noise is the sum of two components. One component, *intrinsic noise,* has a mean of zero and standard deviation $SD1$; the second, *activation noise,* has a mean of zero and standard deviation $SD2 * A(j, t)$. The more active a node is, the greater the noise. However, a node with zero activation still has some noise. Because noise can result in an activation level less than zero, one further assumption is required: A source node with a negative activation level sends no activation.

Equation 1 applies to every node in the network, regardless of its layer, and it applies during every time step in both lemma access and phonological access. This means that *cat's* semantic nodes will be subject to decay, input from neighbors, and noise. It also means that phonological nodes will gain some activation during lemma access. As activation spreads from semantic to word units, it will continue down to phonological units. This is the feedforward (cascading) activation mentioned earlier. The

assumption that connections are bottom-up as well as top-down creates positive feedback, and so semantic units receive input from activated word units and word units get feedback from phonological units.

The existence of feedback in the network means that during lemma access the word nodes of phonological neighbors of *cat,* such as *mat, sat,* or *can,* will become activated. This is in addition to semantic neighbors such as *dog,* which obtain activation from shared semantic nodes. Thus, the most activated word nodes at this step are the target and its semantic and formally related neighbors.

Lemma access is concluded by a selection process. The most highly activated word node of the proper syntactic category is selected. During the production of a sentence, selection entails the linkage of a word to a slot in a syntactic frame. Frame and slot approaches to grammatical encoding in production have ample empirical support (Bock & Loebell, 1990; Garrett, 1975; Levelt, 1989; see Bock & Levelt, 1994, for a review). In the case of object picture naming, we assume a degenerate frame consisting of a slot for a single noun. Therefore, in our implementation of the naming task, the most highly activated noun is selected.

Phonological access. The second step of lexical retrieval begins when the selected word node, *cat,* is given a large jolt of activation. This is also 100 units' worth, the same as the initial jolt to the semantic level. When a sentence is being produced, the jolt to a word occurs when the syntactic frame says that it should occur. In a single-word naming task, it occurs immediately on selection because there is only a single noun slot in the frame.[3]

This large jolt to the selected word is important because it introduces a nonlinearity and hence allows the word nodes to act as a useful hidden layer. This, in turn, allows the meaning-to-form mapping to be achieved. The jolt makes the word node *cat* much more active than any of its competitors at the beginning of phonological access. This function of enhancing the "winner" is often carried out by lateral inhibition among competitors (e.g., Feldman & Ballard, 1982; Grossberg, 1982; Harley, 1990; McClelland & Rumelhart, 1981) or by an absolute threshold that, when crossed, boosts the activation of a node (e.g., MacKay, 1987). The jolt to the selected word in our model is similar to these mechanisms. However, we are specifically tying it to syntactic processes. The source of the jolt is the syntactic slot that the selected word is linked to (e.g., Berg, 1988; Dell & O'Seaghdha, 1991; Eikmeyer & Schade, 1991; MacKay, 1982, 1987; Stemberger, 1985).

After the jolt to *cat,* activation spreads for n more time steps. As was true for lemma access, activation spreads both upward and downward, and nodes other than those connected to *cat* can become activated. The goal of the spreading process during phonological access, however, is to retrieve the phonemes of *cat.* After the n time steps, the most highly activated phoneme nodes are selected and linked to slots in a phonological frame, a process analogous to the linking of the selected word to a syntactic slot in lemma access. A phonological frame represents the structure of a word—its number of syllables and their stress pattern and the sequence of consonants and vowels within each syllable. Most current theories of production hypothesize that phonological access consists of the retrieval of phoneme-size units and their insertion into frame slots, although there are differences among theorists with respect to the nature of the frame (for reviews, see Levelt, 1992; Meyer & Bock, 1992; Shattuck-Hufnagel, 1992). The evidence for phonological frames comes from speech errors (e.g., Shattuck-Hufnagel, 1979; Stemberger, 1990) and experimental studies showing that frame structures can be primed (Meijer, 1994; Romani, 1992; Sevald, Dell, & Cole, 1995).

One simplification of the implemented model is that it only has a frame for single-syllable consonant–vowel–consonant (cvc) words. Each phoneme node is labeled according to whether it is an onset consonant, a vowel, or a coda consonant. Selection consists of picking the most highly activated onset, vowel, and coda and associating them with the corresponding slots in the frame. In the case of *cat,* the nodes for /k/-onset, /ae/-vowel, and /t/-coda would likely be selected. Thus, phonological selection is guided by categories, such as onset or vowel, in the same way that lemma selection is guided by syntactic categories. Categorical selection in phonological access is suggested by patterns of sound substitutions in phonological speech errors (e.g., MacKay, 1970, 1972; Shattuck-Hufnagel, 1979).

The simplified phonological frame of the implemented model limits its ability to make detailed predictions about phonological errors in picture naming. For example, we cannot model the influence of word length or stress pattern or explain the particular difficulties associated with consonant clusters. However, we should emphasize that models similar to ours have been developed for phonological access in normal language production and that these do handle detailed phonological error patterns with some success (Berg & Schade, 1992; Dell, 1986, 1988; Eikmeyer & Schade, 1991; Hartley & Houghton, 1996; MacKay, 1987; Stemberger, 1990, 1991). Moreover, the approach is generally compatible with the dominant frame-based view of phonological access.

Although the model does not deal with subsequent stages in production, we assume that after a word's phonemes have been selected and linked to frame slots, each phoneme is given a jolt of activation to enhance the translation of the phonemes into articulatory codes. We further assume that these subsequent stages proceed without error in nonaphasic speakers and the fluent aphasic patients that make up our sample.

Errors in lexical access. The model allows for the basic kinds of errors that occur during lexical access. We distinguish five error categories: semantically related word errors, formally related word errors, mixed semantic and formal errors, unrelated word errors and nonwords or neologisms. A semantic error, such as *dog* for *cat,* occurs because their concepts share semantic nodes. During lemma access, the word node for *dog* is activated directly by the shared semantic nodes of *cat.* For *dog* to be selected, though, its activation must exceed that of every other noun, including *cat.* This can happen because of the noise in activation levels.

A form-related error, such as *mat* for cat, can arise either as

[3] The jolt sets the activation of the selected word to 100 rather than adding 100 to the current level. One can think of this as, first, a postselection deactivation setting the current activation to zero, as in Dell (1986), followed by the addition of 100 units from the jolt.

an error of lemma access or one of phonological access. During lemma access, nodes for words that share phonemes with the target become activated by feedback from the target phonemes to the word layer. Hence, *mat* and *sat* would gain some activation. Of these, any noun, such as *mat,* could be selected if the noise in the system has made it the most activated. Form-related errors that arise at lemma access would therefore be expected to be nouns in a picture-naming task and, more generally, should follow the syntactic class constraint. It is known that the vast majority of normal speech errors that are thought to be form-related lexical errors, sometimes called *malapropisms,* obey this constraint. In fact, it is one of the challenges to production theory to explain why errors that are so clearly phonological in nature also obey the syntactic class constraint (Fay & Cutler, 1977). The model's answer to this challenge is that these errors are truly errors of lemma access and so are syntactically governed. Lemma access, however, does involve the activation of phonologically related word competitors because of the interactive nature of the model. One of the predictions that we examine in this article concerns the extent to which aphasic patients' form-related errors are truly errors of lemma access.

The second way that a form-related word error can arise is during phonological access. The correct word, *cat,* may have been selected during lemma encoding, but, because of interference from other activated words or from noise, one or more phonemes of *cat* may be replaced by other phonemes. If the resulting string of phonemes is a word such as *mat,* the error would be in the form-related category. The difference is that a nonnoun, form-related error such as *sat* can occur during phonological access. The phonological selection procedure is indifferent to the syntactic category to which the string of selected phonemes belongs. In fact, the selection procedure is sensitive only to the phonotactic constraints dictated by the phonological frames and, of course, the activation levels of the phoneme nodes. By assuming that phonological selection is guided only by phonological variables, the model exhibits the insulation of general phonological knowledge from grammatical categories that has been emphasized in theories of production since the seminal work of Garrett (1975). Therefore, although spreading activation makes the lexicon interactive, the selection processes associated with each step are modular.

Mixed semantic-formal errors, such as *rat* for *cat,* are particularly important because they may reveal the joint effects of semantic and phonological similarity and hence speak to the model's assumption that semantic and phonological information are active at the same time. In the model, the *rat* word node obtains activation directly from shared semantics and from feedback from shared phonemes. The combination of top-down and bottom-up information gives *rat* a much better chance of occurring as an error than a purely semantic or purely formal neighbor. Specifically, in the model, mixed errors are expected to be more likely than semantic errors that happen to be phonologically related or formal errors that happen to be semantically related. That is, semantic and formal influences should not be independent. Analyses of normal speech error collections (Dell & Reich, 1981; del Viso, Igoa, & Garcia-Albert, 1991; Harley, 1984) and experimental studies of normal speakers' picture naming (Bredart & Valentine, 1992; Martin, Weisberg, & Saffran, 1989) have consistently shown that there is a true mixed-error effect (i.e., semantic and formal influences are not independent). Our model attributes this nonindependence to interaction among network layers during lemma access. Alternative accounts, which do not hypothesize phonological activation during lemma access, attribute mixed-error effects to the action of late editorial processes in production (e.g., Levelt et al., 1991a).

Unrelated word errors are word substitutions that are neither semantically nor formally related to the target. This category includes errors such as *log* for *cat,* in which one can find a distant relationship (*cat* is related to *dog,* which is related to *log*), as well as errors in which distant relations are not so apparent.[4] Unrelated errors occur at lemma access and are attributable to noise and any small amounts of activation obtained from distant relations to the target. In principle, unrelated word errors, like any word error, also can occur during phonological access because of the remote possibility that a correct word selected at lemma access is phonologically encoded as an unrelated word. Alternatively, an unrelated error may reflect trouble at both lemma and phonological access. In the case of *log* for *cat,* a semantic error at lemma access, *dog,* could be phonologically encoded as *log,* resulting in an unrelated word error.

The final error category allowed in the model is for neologisms, or nonwords. A nonword such as *lat* or *cag* indicates a problem at phonological access. Noise combines with interference from other activated words, resulting in the replacement of one or more target phonemes. A nonword that resembles the target, or target-related neologism, would likely reflect correct lemma selection followed by incorrect phonological access. A nonword that does not resemble the target, an abstruse neologism, could arise from either a severe disruption of phonological access or difficulties at both lemma and phonological access.

Although the model's five error categories can code most of the error responses reported in the literature, the categories do not reflect some influences that are sometimes coded. These influences are observed in errors in which the phonemes of the target word are misordered and errors in which previously spoken words or their sounds perseverate into the target. Because of the model's assumptions about the association of phonemes to slot positions, it does not allow for the misordering of sounds within a syllable, such as *cat* spoken as *act* or *tack.* This is probably correct for normal speech errors (e.g., Dell, 1986), but not for aphasic speakers, who, as noted earlier, do exhibit misordering within the syllable or word. In addition, because the model assumes that each naming attempt is independent of other attempts, perseverative effects do not occur. In principle, a spreading activation model can account for such effects through persistence of activation or connection weight changes (Plaut & Shallice, 1993a), but we did not implement such a mechanism.

Implementation. Our goal is to fit the model to nonaphasic picture-naming data and then lesion it to fit patient data. For such an implementation, both the network structures and the

[4] With the model's current set of parameters, words that are distantly related to the target gain little activation from that relation, but such activation can be detected in experiments; see Dell and O'Seaghdha (1991) and O'Seaghdha and Marin (1997) for a discussion.

spreading activation parameters have to be specified. Our strategy for implementation has three phases. In the first phase, we specify the network structure, making sure that the model's network preserves essential features of the domain to which it is applied: the English lexicon. The second phase consists of testing nonaphasic speakers of English on the Philadelphia Naming Test (PNT) and then parameterizing the model so that it fits the resulting error data. The chosen parameters also will have to be consistent with facts about the time course of picture naming, specifically that the activation pattern initially includes semantic, but not phonological, neighbors of the target and then later includes phonological, but not semantic, neighbors (Peterson & Savoy, in press; Schriefers et al., 1990). The final phase of modeling involves applying the parameterized model to patient data.

Network structure. Network models of lexical processes must work with a vocabulary that is small in comparison to lexical knowledge. Even in the best cases (e.g., Seidenberg & McClelland, 1989), the network includes only a fraction of what would be expected in a native user's mental lexicon. Because a model's behavior depends greatly on the subset of knowledge that it contains, it is important to consider the relation between that subset and the set that it models.

Our strategy is to use small networks—required because we do extensive explorations of the model's parameter space—whose characteristics preserve what we call the error *opportunities* of the English lexicon, particularly as it is used in picture-naming experiments. The error opportunities afforded by a system are the distribution of error types that would occur if output is "random" (i.e., if output is not affected at all by lexical retrieval). We assume that if output is random, it nonetheless is phonologically legal. This assumption, which is justified because speech errors are, for the most part, phonologically legal (Buckingham & Kertesz, 1976; Wells, 1951), will enable us to make claims about error opportunities in English.

To estimate the error opportunities, we first determined the opportunities for nonwords (i.e., the likelihood that a legal string is a nonword). Dell and Reich (1981) and Best (1996) took sets of words from picture-naming studies and speech error collections and replaced a single phoneme in each word with another phoneme creating a legal string. The proportion of nonwords in these sets ranged from around .55 to .80, in which wordhood meant being listed in a college-level dictionary. This technique essentially determines the chance that phonologically legal neighbors of real words are nonwords. However, because it involves looking at the neighbors of real words, and words tend to clump in phonological space, it may underestimate the error opportunities for nonword outcomes. For this reason, we chose the most conservative value, .80, as the estimated error opportunity for nonwords. We also used the substitution technique on the target words for the PNT and found an estimate of .74 nonwords using the dictionary criterion, which made us feel reasonably comfortable with the chosen .80 value.

The next step is to determine opportunities for the various kinds of word outcomes, that is, how the remaining .20 is divided among the semantic, formal, mixed, and unrelated categories. To do this, we borrowed the analysis of Martin et al. (1994). They took a set of error–target pairs from a patient picture-naming study in which each error was a word and then randomly paired errors and targets and determined how often the resulting pairs were formally or semantically related. Because Martin et al. used definitions for semantic and formal relations that were highly similar to our definitions, their random pairing estimated the relative opportunities for our lexical error categories. Spreading these over the 20% of opportunities reserved for words led to estimates of .10 for unrelated, .09 for formal, and .01 for semantic errors. The opportunity for a mixed error was near zero (.004). Therefore, most of the time (.80) a random output for a given target word should be a nonword, and, if it is a word, it is more likely to be unrelated than in some other category. It may seem surprising that formally related word errors have nearly as many opportunities as unrelated words. However, our definition of a formal relation, which is given in the *Method* section of Experiment 1, is not strict and so many word errors count as formal errors. Semantic, and especially mixed, errors have few opportunities, as one would expect.

We configured the network so that its error opportunities were similar to the estimated opportunities. We identified two six-word neighborhoods, as shown in Table 2. Each neighborhood was made into a network in which each word connects to 10 semantic nodes. The phonology of each network allows for 24 legal strings (6 of which are words). Each neighborhood has a target word, which, for convenience, is called *cat* in both cases. The first of these has two formal neighbors (which each shared two phonemes with the target), one semantic neighbor, and two unrelated neighbors. The second neighborhood is the same as the first, except that one of the formal neighbors is replaced by a mixed neighbor. In both neighborhoods the semantic neighbor, and in the second neighborhood, the mixed neighbor as well, shared three semantic nodes with the target. (The mixed, semantic, and target words shared the same three semantic nodes.) We further assume that the first neighborhood is sampled 90% of the time and the second 10% of the time. Under these conditions, the error opportunities afforded by the model are similar to our estimates for English. If model outputs are random, the string will be a nonword 75% of the time. Our estimate for the real opportunities for nonwords was 80%. The next most likely error type in the model with random outputs is an unrelated word (8.3%), which is slightly more likely than a formally related word (7.9%). The real opportunities for these latter two cases were estimated to be 10% and 9%, respectively.

Table 2
Network Structures Used in the Model

Neighborhood 1	Neighborhood 2
1 target (e.g., cat)	1 target (e.g., cat)
1 semantic (e.g., dog)	1 semantic (e.g., dog)
2 formals (e.g., hat, mat)	1 formal (e.g., mat)
2 unrelated (e.g., log, fog)	1 mixed (e.g., rat)
	2 unrelated (e.g., log, fog)

Note. For phonology, each neighborhood defines 24 legal strings (6 onsets × 2 vowels × 2 codas). Of these 6 are words and 18 are legal nonwords (e.g., cag, fot). For semantics, each word connects to 10 features. Semantically related words share three features with the target.

The least likely outcomes in the model are for semantically related words (model opportunities = 4%, real opportunities estimated at 1%) and mixed words (model and real opportunities = 0.4%). The opportunities for semantic errors in the model are somewhat greater than they should be, but, on the whole, the model's opportunities are much like the estimated real opportunities.

It is crucial that the model's error opportunities resemble real opportunities because the model's error probabilities are determined by its opportunities, particularly when error rates are high. As will become clear, the model must accurately characterize what would happen under total breakdown to handle the range of patient error patterns. That is why we have concentrated on these opportunities. By distilling the complexities of a real lexicon into opportunities for five error categories, however, we are admittedly failing to reproduce many other aspects of what the English lexicon affords. Our model, like any other, attempts to be faithful to critical characteristics of the modeled domain at the expense of other aspects. We claim that, for a model of naming errors, the error opportunities represent the critical characteristics.

The next phase of implementation involves setting model parameters so that its performance is similar to that of nonaphasic speakers in a picture-naming task. In Experiment 1 we gathered these data.

Experiment 1: Control Data

The Philadelphia Naming Test

The data used in the parameter-setting phase of implementation were derived from the PNT, a new 175-item test of confrontation naming developed for collecting a large corpus of naming responses from a standardized set of items. Description of the development of the test, testing procedure, and scoring protocol follows (for additional details, see Roach, Schwartz, Martin, Grewal, & Brecher, 1996).

Participants. Two control groups participated in the study. The groups each consisted of 30 non-brain-injured, non-language-disordered, native English speakers. Participants were selected to provide a demographic match to our aphasic patient population. They lived in the Philadelphia Metropolitan Area, and they ranged in age from 40 to 75 years and in education from the sixth grade (with skills training) to the doctoral level. Two thirds of the control sample were female; three quarters were White. The groups differed from one another only with respect to the particulars of the procedures they underwent. All were reimbursed $15 for their participation.

Method. Naming responses from the first group of control participants were used to determine which test items from a set of 277 to include in the PNT. These pictured items were selected from original and published collections on the basis of their familiarity and good image quality (minimal complexity and confusability).

Pictures were digitized and presented to participants on a Macintosh LCIII (Apple Computer, Inc., Cupertino, CA) computer using MacLaboratory for Psychology experiment running software (Chute, 1990). The software controlled stimulus presentation (although the trial initiation was experimenter controlled) and on each trial recorded a response latency, measured from stimulus onset (marked by a brief tone) to response onset. Because our concern in this study was solely with accuracy and the nature of errors, the latency data are not discussed further.

All Group 1 participants saw the items in the same order. They were instructed to provide a one word label (e.g., "well," not "wishing well") as quickly as possible for each item; otherwise, speed was not emphasized. Sessions were tape-recorded for later scoring.

Test items for which there was a high degree of naming agreement across participants were identified. From the original set of 277 items, 175 proved acceptable. Each of these 175 items was named with the designated target label by at least 25 of the 30 participants, and none elicited the same erroneous response from more than 3 participants. One hundred thirty-six items elicited the target name in all 30 participants. The 175 selected words ranged in length from one to four syllables and in noun frequency of occurrence from 1 to 2,110 tokens per million of printed English text (Francis & Kucera, 1982). Length and frequency were not balanced across the items: Low-frequency targets (1–24 per million) and targets of one syllable predominated (target names are provided in Appendix A).

The testing procedure for the second group of control participants was the same as that for the first with the following exceptions: Only the 175 selected items were tested. These items were split into two blocks that were administered in succession. The order of the blocks alternated across participants; ordering within a block was constant. Ten nontest practice items were presented before the test blocks.

Scoring. The system used to score the control participants' responses was the system we developed for use with aphasic patients. Participants' responses to each PNT item were phonetically transcribed by two speech-language pathologists using loose International Phonetics Association transcription rules. Transcription discrepancies were resolved by a third judge. When there was more than one attempt at a response (a situation rare in nonaphasic speakers but common in aphasic speakers), we extracted the "first complete response" for scoring. This was the first minimally CV or VC attempt to name the object in which the vowel was not reduced and that was followed by a noticeable pause or had clear downward intonation or upward or questioning intonation. Utterances that preceded the first complete response that did not meet these criteria were considered fragments and, for purposes of this study, were ignored. Monosyllabic responses to a multisyllabic (three or more syllables) targets were also considered to be fragments.

First complete responses were scored as correct only if they exactly matched the designated target. All other responses were coded as errors. The first five error categories corresponded to the basic error types that the model was designed to explain: Semantic applied to a synonym, category coordinate, category superordinate, category subordinate, or associate of the target. Formal applied to any word response (excluding proper nouns) that met our criteria for phonological similarity. This was the case when target and error started or ended with the same phoneme; had a phoneme in common at another corresponding syllable or word position, aligning words left to right; or had more than one phoneme in common in any position (excluding unstressed vowels). Plural morphemes do not enter into the assessment of phonological similarity. Mixed applied to a response that met both semantic and formal criteria, excluding multimorphemic responses with target overlap in a stem or affix (e.g., microscope → telescope was coded as semantic rather than mixed). Unrelated met neither semantic nor formal criteria and was not visually related to the stimulus. A subset of the items in this category are perseverated responses to previous test items. Perseverated responses were coded according to their relation to the target; the majority were unrelated. All nonword responses except blends were coded as neologisms, but we tracked whether the neologism was phonologically related to the target, to a semantic or mixed alternative to the target, or to an unrelated or indeterminate word.

These five categories did not exhaust the responses generated by aphasic speakers or, for that matter, control speakers. The following categories captured the remaining responses: (a) descriptions—gener-

ally, multiword responses that characterize the object or explain its function or purpose, but single adjectives or adverbs also may be coded as descriptions; (b) no response—the speaker indicates, either verbally or nonverbally, that he or she is unable to name the object; and (c) miscellaneous—includes responses that name part of the object as depicted (e.g., *veil* for *bride*), other visual errors, blends of two near synonyms (e.g., */klæb/* for lobster), and morpheme omissions (e.g., *cheer* for *cheerleader*).

Results. The responses to the 175 items of the PNT were examined for both control groups. For Group 1 participants, these were extracted from the larger initial set of 277 test items. A total of 82 responses (0.7%) could not be scored because of equipment or experimenter error. Table 3 shows the remaining data. The level of accuracy and the pattern of errors were highly similar across the two groups; thus, we used the data from all 60 participants to set the parameters of the model.

It is evident from Table 3 that nonaphasic speakers named these pictures highly accurately. Moreover, the vast majority of the errors that occurred exhibited semantic relations to the target. Vitkovitch and colleagues obtained similar results using a speeded naming paradigm (Vitkovitch & Humphreys, 1991; Vitkovitch, Humphreys, & Lloyd-Jones, 1993). However, there also may be a tendency for formal relations in the errors, as seen in the relatively large number of semantic errors that are also formally related: the mixed category. In fact, there was a true mixed-error effect in these data. Martin, Gagnon, Schwartz, Dell, and Saffran (1996) took this set of semantic and mixed errors (i.e., all errors with some clear semantic relation) and determined the proportion of cases in which the error matched the target with respect to initial, second, and third phonemes. (Morphologically related target–error pairs were excluded.) These proportions were, respectively, .127, .183, and .076. By chance, one would expect match proportions of .055, .068, and .060, in which chance was determined by the techniques of Dell and Reich (1981). The difference was significant for the first two positions, and, generally speaking, the match proportions were similar to those found in demonstrations of the mixed-error effect in natural speech error collections (Dell & Reich, 1981; del Viso et al., 1991; Harley, 1984). Hence, Martin, Gagnon, et al. concluded that formal relations play a role in the generation of semantic errors made by nonaphasic speakers in a picture-naming task.

Table 3
Number and Percentage of Responses of Each Code Type for First Complete Naming Attempts by Group 1, Group 2, and Groups 1 and 2 Control Participants

	Responses					
	Group 1		Group 2		Collapsed	
Code	*n*	%	*n*	%	*n*	%
C	5,037	97.4	5,057	96.4	10,094	96.9
S	48	0.9	72	1.4	120	1.2
F	0	0.0	6	0.1	6	0.1
N	2	0.0	3	0.1	5	0.0
M	45	0.9	45	0.9	90	0.9
U	16	0.3	12	0.2	28	0.3
D	1	0.0	7	0.1	8	0.1
NR	6	0.1	8	0.2	14	0.1
Miscellaneous	16	0.3	36	0.7	52[a]	0.5

Note. C = correct; S = semantic; F = formal; N = neologism; M = mixed semantic and formal; U = unrelated; D = description; NR = no response.
[a] Forty-eight of the miscellaneous errors are picture-part names.

The analysis of Martin, Gagnon, et al. (1996) is important for the model for two reasons: First, it provides another demonstration of the mixed-error effect. This effect can be taken as evidence for the simultaneous activation of semantic and phonological information, as hypothesized in the model. Second, it justifies our use of a mixed category. We cannot, of course, claim that every error placed in this category is a "real" mixed error. However, at least we do know that there are formal relations present in the set of semantic errors above expected chance levels. The mixed category separates out those that are the most likely to reflect formal as well as semantic factors.

Model Fit to Control Data

Using the neighborhood structure described earlier, we explored the parameter space of the model to fit the control data. The chosen parameters led to error probabilities reasonably close to those in the data. Table 4 shows these parameters and the simulated error probabilities, which were based on 100,000 trials. In both the data and the model, naming was highly accurate, semantically related errors predominated, and formal relations were present only in conjunction with a semantic relation.

The chosen parameters define a model of normal performance in which errors occur much more often in lemma access than in phonological access and in which the errors in lemma access are overwhelmingly semantic. In this respect, the model is behaving like a modular two-step model. The principal competitors to a target lemma are semantic. The feedback from the phoneme layer is not strong enough to lead to pure formal errors at lemma access. In fact, we see only the influence of this feedback indirectly, through the occurrence of mixed errors. The model, like the nonaphasic speakers, exhibits a true mixed-error effect. This can be shown by using the second neighborhood in Table 2 and directly comparing the probability of selecting the mixed lemma with that of the semantic lemma. With the chosen normal parameters, the mixed word node is 5.7 times more likely to be selected than the semantic one lacking the formal relation. The only reason that mixed errors in Table 3 are, overall, less likely in the model than semantic errors is that there are many fewer opportunities for them. The second neighborhood, which has a mixed word, is sampled from less often than the first neighborhood, which has no opportunity for a mixed error.

The model, with this set of parameters, also exhibits a particular time course for the activation of semantic and formal neighbors. Using the first neighborhood in Table 2 and eliminating any noise, we determined the relative activation of semantic and formal competitors at each time step. Specifically, the proportion of the total activation in the word nodes other than the target was calculated for the formal and

Table 4
Naming Data From 60 Control Participants and Simulated Probabilities

Data source	Response category					
	Correct	Semantic	Formal	Nonword	Mixed	Unrelated
Controls	.969	.012	.001	.000	.009	.003
Simulated probabilities[a]	.966	.021	.000	.001	.012	.000

[a] Chosen parameters: $p = .1$, $q = .5$, $SD1 = 0.01$, $SD2 = 0.16$; $n = 8$.

the semantic word nodes.[5] These proportions are shown in Figure 2. The figure includes the eight time steps for lemma access (from the semantic jolt to word selection), eight for phonological access (from word jolt to phoneme selection), and eight more following a jolt to the selected phonemes. In the initial phases of lemma access, only semantic competitors are active, and these remain more active than formal competitors throughout lemma access, although the formal competitors do become active. During phonological access, the relative activation in the formally related words gradually builds compared with the semantic competitor's activation. The advantage for formal over semantic increases after the jolt to the selected phonemes, although the semantic activation does not go away completely.

The time course shown in Figure 2 is similar to that shown in studies of the response time to semantic or formally related probe words during picture naming (Levelt et al., 1991a; Peterson & Savoy, in press). Figure 3 shows the data of Peterson and Savoy. Early on, that is, when the probe occurs close to picture onset, only semantic probes have an influence compared with unrelated ones. After that, there is a period in which both semantic and formal probes have an effect. Finally, there is a late phase in which only formal probes are primed. The model's time course resembles this pattern, with one exception: There is some semantic activation in the model at the latest time steps. If it is assumed that relatively small amounts are not detectable in the behavioral measures, then the model's time course is a reasonable match to the pattern shown in the studies.

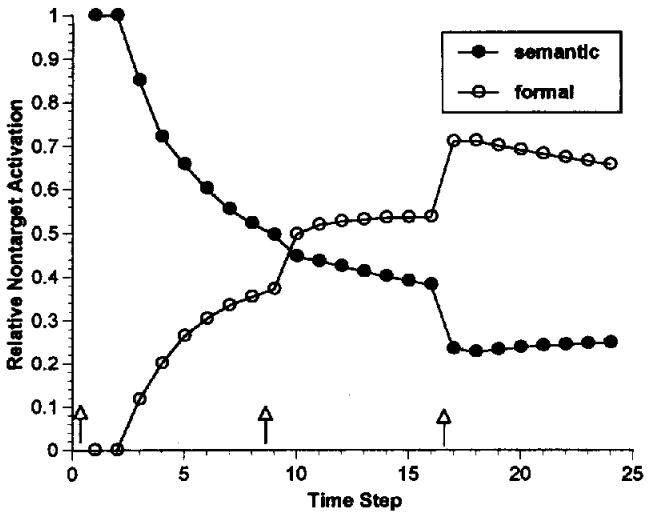

Figure 2. The proportion of nontarget activation in the word nodes that lies in the semantic neighbor (*dog*) and the formal neighbors (*mat* and *hat*) as a function of time. Neighborhood 1 is used for the network, and the parameters are set at normal values. Each arrow indicates when an activation jolt is delivered.

The evolution of activation from semantic to phonological occurs only in the model under certain parameter values. In particular, the activation that is generated in the network must be small in proportion to the external jolts that are supplied to the semantics, to the selected word, and to the selected phonemes. The generation and maintenance of activation attributable to spreading is tied to connection weight, p, decay rate, q, the amount of time between each jolt, n, and the connectivity of the network. We have chosen values for these so that activation levels are small when each jolt is given.

The lesson of the time-course experiments, and particularly the discussion by Levelt et al. (1991a, 1991b) of these studies, is that processing evolves from semantics to sound. Our model does this through its serially ordered jolts. At the beginning of each access step, a jolt is delivered to the appropriate nodes on the appropriate level, and this jolt dominates the residual activation in the network. Therefore, although spreading activation in the network is interactive and tends to bring in information from other layers, this interaction is countered by ordered jolts that impose seriality on the process and a degree of modularity to each step. Dell and O'Seaghdha (1991) referred to these kinds of theories as *globally modular* and *locally interactive*. In this way, the model exhibits characteristics of both modular stage theories and interactive activation approaches.

Next, we apply the model to naming deficits in aphasia. If the continuity thesis is correct, the model should characterize these deficits without a great deal of added complexity. The application to aphasia involves testing a group of patients on the PNT, developing an account of which aspects of the model are altered in aphasia, and fitting the model to the error data of individual patients.

[5] The proportion for formals is based on the total activation in the two formal nodes divided by the sum of the activation of all nontarget word nodes. For semantics, the numerator is the activation in the single semantic node. These proportions can be taken as indexes of the relative amount of form-related and semantic-related activation in the network.

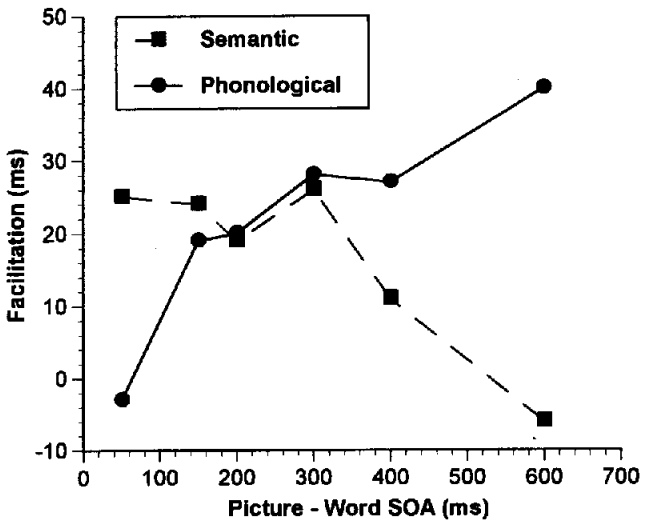

Figure 3. Facilitation (unrelated − related) in naming times to a probe word as a function of the nature of its relation (semantic or phonological) to a target picture and picture-word stimulus onset asynchrony (SOA). The data are from Peterson and Savoy (in press).

Application to Naming Deficits

Experiment 2: Patient Data

Participants. Aphasic participants were recruited from clinical and research programs throughout the Philadelphia area. Criteria for inclusion were deliberately broad: any variety of fluent aphasia (e.g., anomia, Wernicke's, conduction, transcortical sensory) in which the precipitating event was a left hemisphere cardiovascular accident (CVA). All patients were postdischarge from acute hospitalization, but, apart from this, the time after onset was not a factor. Neither site nor distribution of lesion were factors either, although the restriction to fluent forms of aphasia would have ruled out most patients with anterior damage. The only exclusion criteria were etiology different from left CVA, presence of a motor speech impairment (verbal apraxia or dysarthria), or phonemic jargon with no comprehensible speech. In casting such a broad net, we were aiming to capture the range of variation that has been reported in the literature with respect to both severity and error pattern. As we show, we came close to achieving this goal, although our sample did not turn out to include any patients who produced only semantic errors or only phonological errors.

The sample comprised 23 patients who ranged in age from 28 to 83 years (*Mdn* = 59.6) and in education from Grade 11 to doctoral level. All were right-handed. Eleven of the 23 were female, and 13 were White. Individual biographical and language profiles are shown in Table 5. Participants were reimbursed for their participation at a rate of $15 per session.

Methods and scoring. A trial duration limit and feedback were incorporated into the testing procedure for aphasic patients. The experimenter terminated the trial after 30 s if the participant had not provided a final response and provided the target name (regardless of whether the target had been produced) after each trial. The purpose of this was to minimize frustration and the likelihood that the search for an unavailable word would interfere with subsequent trials. For some patients, the task was excessively time-consuming or tiring. When this was the case, the two blocks were tested on separate days. Other than these accommodations to the needs of aphasic patients, the procedure was the same as it had been for the Group 2 control participants.

The scoring protocol was also the same as described earlier. Responses were coded by two or three members of the research group independently of one another. Aphasic participants made many more errors than the control participants, providing an opportunity to test the reliability of the scoring protocol. Coding agreement across scorers was computed for the first 6 aphasic participants tested on the PNT. For the identification of the first complete response, the percentage of trials per participant on which the coders agreed ranged from 87 to 98. For assignment of codes, the percentage of agreement ranged from 86 to 100. Thus, it appears that our scoring protocol was reliably applied.

Results. Table 6 shows the proportions for the six response categories allowed by the model (correct plus five error types) as well as a seventh category representing all other errors. For most of the patients, the preponderance of responses fell in the modeled categories. This is important because the model's six categories were constrained to add to 100%, and if the sum of these categories in the patients was much lower than this, the fits would necessarily be off. Two patients presented problems in this regard: G.B. produced an unusually high number of descriptions (.22), and V.P. produced many descriptions and nonresponses (.12 and .14, respectively). We therefore made no attempt to fit these patients to the model.

There was considerable variation in the performance of the aphasic participants, from a near normal 95% to 8% correct. For some patients neologisms predominated, whereas for others semantic or formal errors were the most common. Neologisms came in several varieties. In this study 80% were phonologically related to the target, another 7% were phonologically related to a semantic neighbor of the target, and 13% were unrelated to the target (i.e., abstruse neologisms).

A number of general patterns can be observed in the data, and these are illustrated in Figure 4, which shows the relation between each error category and overall severity, which we indexed by the proportion of correct responses. The fitted curves are polynomials of degree 2. Semantic errors occurred in all patients, but, interestingly, they did not increase with severity. Mixed errors showed a similar pattern but at a lower overall rate. By contrast, the remaining error categories, formals, unrelated, and nonwords, clearly increased with severity. It appears that the categories with the greatest number of error opportunities became relatively more prevalent as the system became more disordered. Consider, for example, the difference between I.G. and G.L. in Table 6. I.G. was much more accurate than G.L. (69–28%), but the difference was due entirely to many more errors committed by G.L. in the nonword, formal, and unrelated categories.

However, the variation among the patients was not all due to severity. Patients with the same overall degree of correctness can have different patterns. For example, both L.H. and I.G. correctly named 69% of the targets. L.H. made nearly four times as many nonword errors as the total of semantic and mixed errors. I.G.'s pattern was the opposite, with six times as many semantic and mixed errors as nonwords.

The error patterns present a rich set of constraints for modeling. The error categories differed among themselves in how they varied with severity, and each category's proportions could vary among patients at the same severity level. Although there were several different patterns that occurred, there were also ones

Table 5
Biographical Information and Language Profiles of the Aphasic Patients

Patient	Age	Education (years)	Months post onset	BDAE:[a] Diagnostic subgroup and mean percentiles from subtest summary profile: Aphasia type	Auditory comprehension	Phrase repetition	BNT[b] (% correct)
J.A.	55	12	2	Transcortical sensory	58	80	42
J.B.	36	12+	20	Conduction	85	NA	52
G.B.	83	12	4.1	Wernicke's	48	60	15
L.B.	77	16	10	Anomic	87	60	78
H.B.	65	12	3.2	Conduction	81	0	NA
V.C.	53	11	4	Anomic	83	90	20
N.C.[c]	28	11	23	Conduction	35	NA	43
J.F.	44	16+	1.2	Anomic	69	80	27
A.F.	77	12	1.5	Wernicke's	35	70	53
J.Fr.	70	12	11	Wernicke's	73	50	91
E.G.	75	12	2.1	Anomic	86	80	75
I.G.	75	12+	72	Anomic	70	NA	53
J.G.	76	16+	1	Conduction	83	75	35
L.H.	47	12	1.9	Conduction	89	60	55
J.L.	36	12	1.5	Conduction	86	55	43
G.L.	45	12	8.6	Wernicke's	70	55	23
B.Me.	61	16	100	Anomic	78	NA	75
B.Mi.	72	12	5	Anomic	90	80	68
V.P.	64	12+	60	Anomic	65	NA	25
W.R.	61	12	2.5	Wernicke's	29	75	5
G.S.	68	12+	1	Wernicke's	36	20	43
T.T.	47	14	2	Anomic	83	90	85
W.B.	56	12	1.8	Wernicke's	46	50	81

Note. NA = not available.
[a] Boston Diagnostic Aphasia Examination (Goodglass & Kaplan, 1983).
[b] Boston Naming Test (Kaplan, Goodglass, & Weintraub, 1983).
[c] In prior studies, we characterized N.C. as a Wernicke's aphasic patient with severe impairments in auditory input processing and short-term memory (Martin, Dell, Saffran, & Schwartz, 1994; Martin & Saffran, 1992). At the onset of the current study, his auditory comprehension was still impaired, but the quality of his speech and the disproportionate difficulty repeating isolated words qualified him for a diagnosis of conduction aphasia.

that did not occur. For example, we did not find a patient with a low accuracy whose dominant error type was semantic. J.F., with .56 correct, came closest to this. In general, we did not observe examples of complete dissociations: Large numbers of neologisms in the absence of semantic (or formal) errors or vice versa. We attempt to explain the full range of findings—what occurred and what did not occur—in terms of the model. It may turn out, however, that the nonoccurrence of extreme dissociations was an accident of sampling rather than a characteristic of the population of interest. We address this possibility in the General Discussion section.

Theory of Naming Deficits

We apply the model to the patient data by starting with the model for the control data and altering its parameters in an attempt to fit each patient's performance pattern. Our goal is to make concrete the central idea behind the continuity thesis: Errors arising from language pathology are not qualitatively different from those made by nonaphasic speakers. Our theory of naming deficits has two components: Brain damage reduces the ability to transmit activation between levels in the network, it reduces the integrity of the representations at each level, or both. The first of these, activation transmission, is tied to connection weight, p. A strong connection ensures that representations (i.e., activation patterns on each level) are consistent with one another. For example, if the word node for *cat* is strongly active, then the sounds /k/-onset, /ae/-vowel, and /t/-coda also should be highly activated. Strong connections between *cat* and its sounds make this happen. One possible effect of brain damage is that activation transmission between levels is impaired, with the result that the information provided by each jolt does not effectively get where it should. When this happens, noise has a greater influence and errors increase. To model this impairment, we decrease p. The second hypothesized component of naming deficits, the integrity of a given representation, can be associated with a number of parameters: the decay rate (q), the amount of intrinsic noise in the system $(SD1)$, or even the size of the jolt.[6] Each of these parameters reflects how well an activation

[6] Technically, jolt size is not a separate parameter from intrinsic noise. It is arbitrary and just sets the activation scale. Therefore, jolt size gives the standard deviation of intrinsic noise its meaning. One can double the noise simply by halving the jolt size.

Table 6
Naming Data From 23 Fluent Aphasic Patients

Patient	Aphasia type	Naming response						
		Correct	Semantic	Formal	Nonword	Mixed	Unrelated	All others
W.B.	Wernicke's	.94	.02	.01	.01	.01	.00	.01
T.T.	Anomic	.93	.01	.01	.00	.02	.00	.04
E.G.	Anomic	.93	.03	.00	.01	.02	.00	.01
J.Fr.	Wernicke's	.92	.01	.01	.02	.02	.00	.02
V.C.	Anomic	.87	.02	.01	.03	.01	.00	.05
B.Me.	Anomic	.84	.03	.01	.00	.05	.01	.07
B.Mi.	Anomic	.83	.05	.01	.01	.02	.01	.07
L.B.	Anomic	.82	.04	.02	.09	.01	.01	.01
J.A.	Transcortical sensory	.78	.04	.00	.02	.03	.01	.13
J.B.	Conduction	.76	.06	.01	.05	.02	.01	.09
J.L.	Conduction	.76	.03	.01	.06	.03	.01	.10
A.F.	Wernicke's	.75	.02	.03	.07	.06	.04	.03
N.C.	Conduction	.75	.03	.07	.08	.01	.00	.07
G.S.	Wernicke's	.70	.02	.06	.15	.01	.02	.03
L.H.	Conduction	.69	.03	.07	.15	.01	.02	.03
I.G.	Anomic	.69	.09	.05	.02	.03	.01	.10
H.B.	Conduction	.61	.06	.13	.18	.02	.01	.01
J.F.	Anomic	.56	.14	.01	.02	.11	.01	.15
J.G.	Conduction	.55	.06	.08	.18	.04	.03	.05
G.B.	Wernicke's	.39	.07	.09	.08	.01	.03	.32
G.L.	Wernicke's	.28	.04	.21	.30	.03	.09	.05
V.P.	Anomic	.28	.07	.11	.04	.05	.17	.28
W.R.	Wernicke's	.08	.06	.15	.28	.05	.33	.06

pattern at a given level is maintained. For the sake of simplicity, we treat decay rate as the relevant parameter; damage is assumed to increase the value of q. Appendix B shows that variation in decay and intrinsic noise have highly similar effects on the model. Hence, we are not claiming that decay rate is the parameter that underlies variation in the function that we have called representational integrity. We are manipulating decay, as opposed to other parameters, for convenience.

The globality assumption. Another simplifying assumption of the patient modeling is that variation in connection weight and decay is global. To model a particular patient, we do not set different values of p and q for semantic, lexical, and phonological parts of the network. Rather, each patient is assigned a p that applies to all connections and a q that characterizes all the nodes. Pragmatically speaking, it is easier to work with a small parameter space. It is easier to search the space and to understand its properties. Importantly, the globality assumption enables us to focus attention on the differences between lesions in activation transmission and representational integrity without the added complication of differential involvement of network levels.

The globality assumption also can be treated as a substantive claim, namely, that the population of interest (fluent aphasic patients) can be modeled without assumptions of differential involvement of processing levels. One might be tempted to reject this substantive claim out of hand on the basis of the existence of error patterns favoring one type of error over another (e.g., patients who make many phonological but few semantic errors). However, as will become evident, the properties of the model are such that the severity of damage and the nature of damage—whether to weight or decay—interact in such a way as to generate dissociations among the error types (for a related discussion, see Dunn & Kirsner, 1988). Globality as a substantive claim turns on whether the patterns observed in our sample are compatible with the model and whether the observed patterns represent a fair sampling of those present in the population at large.

Our use of the globality assumption does not imply that we believe that damage is uniform across the vocabulary, only that the damage involves all the levels in the lexical network. Thus, when values of p and q are determined for particular patients, they should be viewed as averages across the vocabulary being tested. In actual patients, some word classes are likely to be affected more than others (e.g., Goodglass, Klein, Carey, & Jones, 1966; Goodglass, Wingfield, Hyde, & Theurkauf, 1986; Hillis & Caramazza, 1995; Rapp, Benzing, & Caramazza, 1995).

The globality assumption also does not assert that the affected values of p and q apply to all tasks that use words. Whether the lexical network used in producing a spoken word is the same as that used in listening, reading, and writing is a theoretical issue that is often the subject of research in the cognitive neuropsychology of language (e.g., Allport & Funnell, 1981; Caramazza, 1988; Morton & Patterson, 1980; Nickels & Howard, 1995b). It is an issue, though, that we set aside, with one exception. The exception concerns the relation between naming and single-word repetition. In principle, if we can develop an effective characterization of a patient's naming deficit, performance on any task that is hypothesized to use some of the same knowledge as naming should be predictable from the model.

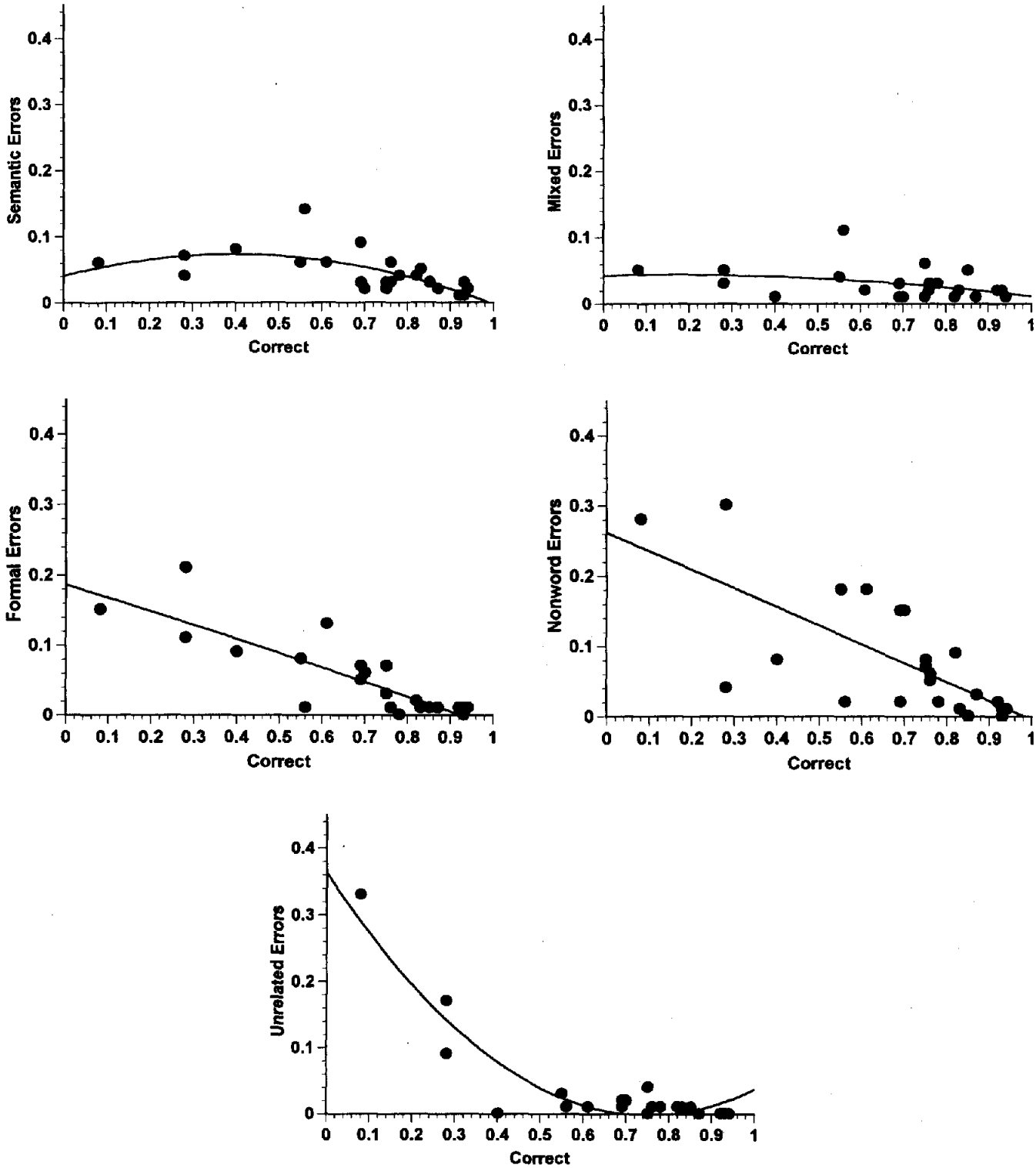

Figure 4. The relation between error proportions and correctness for semantic, formal, nonword, mixed, and unrelated errors for the 23 patients tested in Experiment 2. The fitted functions are polynomials of a degree of 2.

We show this for repetition, assuming that repetition and naming share the phonological access step in production. For this step, naming and repetition should use the same values of p and q.

Error patterns in the model. A loss in activation transmission (lower p) or representational integrity (higher q) leads to errors. Activation levels get small and intrinsic noise has a greater influence. Figure 5 shows a contour map of error probability in the model as a function of p and q. The normal point on the map is in the lower right corner at $p = .1$ and $q = .5$. As p decreases or q increases, errors become more likely. There is a band, the *aphasic region,* in which errors are frequent, and a large area, in which performance is completely dominated by noise, the *random region.* In our modeling exercise we attempt to place patients in this parameter space.

To gain an understanding of the model's behavior, we simulated many combinations of p and q and examined how error types are related to the parameters. We found that the most useful way to see these relations was with a map of error–category space rather than a parameter–space map such as Figure 5. In error–category space, dimensions are probabilities of particular error categories. The set of error patterns that can be created in the model by varying p and q can be placed in the space. Figure 6 shows a 3-D error category space representing the nonword, unrelated word, and related word categories, with the latter category being made up of the sum of the semantic, mixed, and formal categories. The related errors were combined to make a viewable space that nonetheless includes all errors and because they are, to some extent, alike in how p versus q lesions affect them. The region that looks like a twisted piece of metal contains the patterns allowed by the model. At the left is the model's normal point, a position where there are few errors, nearly all of which are related words. At the other end of the region is the random point, the error pattern that occurs when noise is overwhelming. This point corresponds to the random region in parameter space. The random point is determined entirely by error opportunities; nonword errors are the most likely, and unrelated errors are common. The permissible error region is defined by two edges, with each running between the normal and random points. The nearer edge in the figure labels the set of error patterns that result from varying the decay rate, q, from its normal level of .5 to an extreme level of .97 while holding connection weight, p, constant at its normal level of .1. This line therefore defines the error patterns that arise from pure decay rate lesions. The far edge in the figure is derived analogously from varying connection weight, holding q constant at the normal value of .5. Here the random point is reached with p around .001. The surface between the boundaries represents patterns that arise from lesions in both p and q.

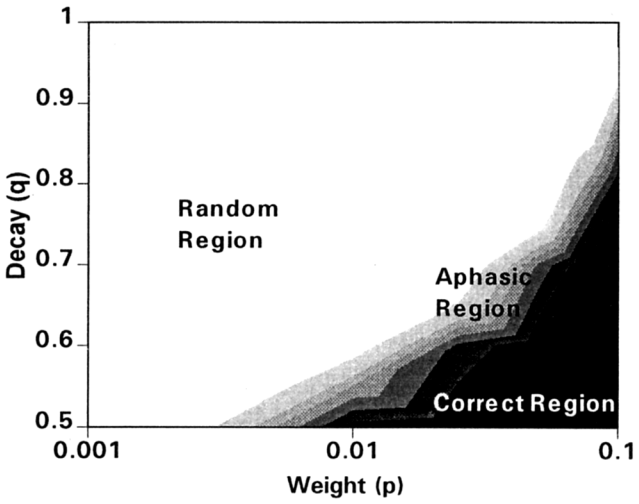

Figure 5. A contour map of overall error probability in the model as a function of connection weight, p, and decay rate, q. Other parameters retain their normal values. The connection weight is plotted with a log scale.

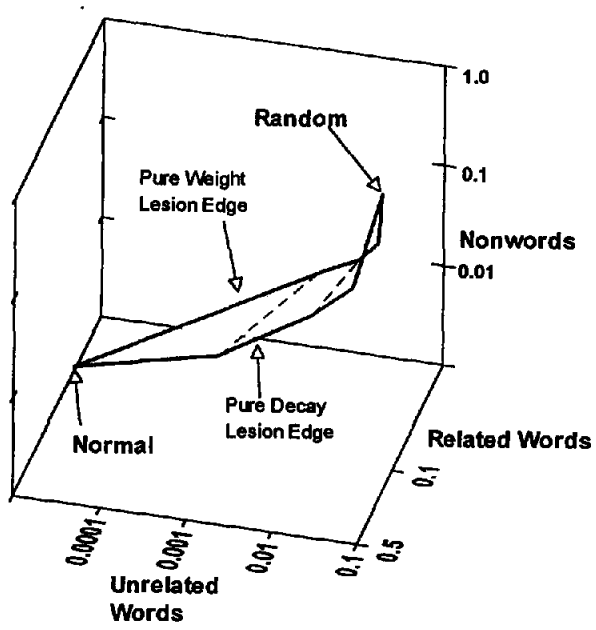

Figure 6. Error patterns that can occur in the model as a result of varying connection weight and decay rate. Patterns fall in the subregion of the potential error space that lies between a normal point in which errors are rare and related and a random point in which errors are completely determined by error opportunities. The dotted lines connect points of each edge that have an equal degree of correctness. Each dimension is plotted with a log scale.

Figure 6 shows that the error patterns allowed by the model are restricted. Many logically possible patterns cannot occur. In addition, the figure suggests that we can describe the model's error space along two dimensions. The first is overall severity. As we move from the normal to the random point, severity increases. The second dimension concerns the degree to which p or q is affected. Lesions that affect activation transmission do something a bit different than lesions that affect representational integrity. Of these two dimensions, severity is the most important. In fact, when performance is nearly normal or nearly random, whether the lesion is primarily p or q hardly matters. The error region spreads out only along the p versus q dimension in the middle of the severity range.

Nonetheless, there are clear differences in the model as a function of whether p or q is affected. Notice in Figure 6 that the pure decay lesion edge is lower and more to the front than the pure weight lesion edge. This means that, relatively speaking, a pure decay lesion promotes more related word errors, whereas a pure weight lesion causes more nonwords. The differences between the pure lesion types can be seen more clearly in Figures 7 and 8, which present detailed pictures of the pure connection weight and pure decay lesion edges, respectively. Each figure shows the model's probability for each error type as correctness varies from the normal level to about 30% correct.

There are some common features between Figures 7 and 8. In both, semantic and mixed errors do not consistently increase as correctness decreases, whereas nonwords, formals, and unrelated errors do consistently increase. Notice that this pattern is exactly what was illustrated by Figure 4 in the patients. In the model the increase in nonwords, unrelated errors, and formal errors with severity is attributable to their relatively greater opportunities, and we hypothesize that the same is occurring with the patients. The difference between the decay and weight lesions is that the former is associated with relatively more formal, semantic, and mixed errors and the latter with more nonword and unrelated errors. For example, at about 30% correct, the weight lesion has 41% nonword, 12% formal, 10% unrelated, 7% semantic, and 1% mixed errors; for decay it is 26% nonword, 20% formal, 7% unrelated, 13% semantic, and 3% mixed errors. These differences reflect the different functions of activation transmission and representational integrity. In the model, reducing weight diminishes the extent to which different levels of representation are consistent with one another. This promotes what we might call "stupid" errors. Nonwords arise when the phoneme layer is not consistent with the word layer. Unrelated word errors occur when the semantic and word layers are inconsistent. When decay rate is increased but connection weights are still strong, many errors occur, but they show evidence of successful activation transmission among the levels through a higher occurrence of mixed, formal, and semantic errors. These are "smart" errors, in which the word level is consistent with the semantic level, phonological level is consistent with the word level, and the chosen word bears some resemblance to the target.

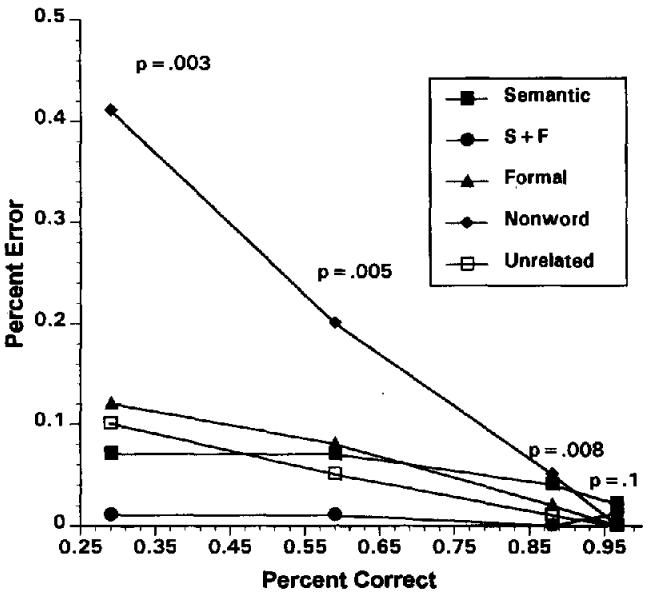

Figure 7. The effect of varying connection weight lesions on error proportions in the model. The decay rate is held constant at .5. S = semantic; F = formal.

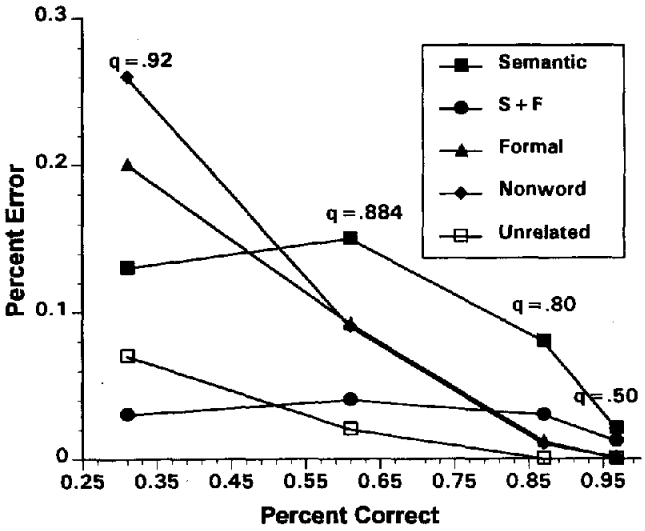

Figure 8. The effect of varying decay lesions on error proportions in the model. Connection weight is held constant at .1. S = semantic; F = formal.

The combination of variation in severity and type of lesion allows the model to simulate dissociations between patients who produce mostly semantic errors and those who make mostly phonological errors. Decay lesions that are not severe lead to patterns in which nearly all the errors have a semantic component, such as, the point where $q = .80$ in Figure 8. The opposite pattern occurs for severe lesions in connection weight, such as $p = .003$ in Figure 7, where 89% of the errors are nonsemantic. The dissociation arises in the model because only semantic and mixed errors predominate near the normal point, and, because these errors are "smart," they tend to be promoted with mild decay lesions. Nonsemantic errors, particularly nonwords, dominate with severe lesions of both types because their greater opportunities have an increasingly large impact as performance breaks down. Nonwords and unrelated words are further promoted if the lesion is in connection weight because the lesion reduces the correspondence between the activation patterns on different levels.

We should reemphasize, though, that the permitted error patterns in the model are limited. Variation in severity and lesion type can produce some interactions among error types, but not all interactions can be produced and those that can cannot be produced at all severity levels. Whether the model's limitations are like those in patient data can be be seen only by attempting to fit patterns of a number of patients by lesioning (altering the parameter settings of) the model.

Table 7
Philadelphia Naming Test 1 Results: Naming Data and Predictions of the Model

Patient and parameter settings	Naming response						RMSD
	Correct	Semantic	Formal	Nonword	Mixed	Unrelated	
W.B.	.94	.02	.01	.01	.01	.00	
p = .02, q = .56	.93	.04	.01	.02	.01	.00	.010
T.T	.93	.01	.01	.00	.02	.00	
p = .02, q = .56	.93	.04	.01	.02	.01	.00	.015
J.Fr.	.92	.01	.01	.02	.02	.00	
p = .02, q = .56	.93	.04	.01	.02	.01	.00	.014
V.C.	.87	.02	.01	.03	.01	.00	
p = .02, q = .57	.88	.05	.02	.04	.01	.00	.014
L.B.	.82	.04	.02	.09	.01	.01	
p = .007, q = .5	.82	.04	.03	.08	.01	.02	.007
J.B.	.76	.06	.01	.05	.02	.02	
p = .0065, q = .5	.78	.06	.04	.08	.01	.03	.021
J.L.	.76	.03	.01	.06	.03	.01	
p = .025, q = .6	.83	.06	.03	.06	.01	.01	.033
G.S.	.70	.02	.06	.15	.01	.02	
p = .0057, q = .5	.69	.07	.06	.14	.01	.03	.022
L.H.	.69	.03	.07	.15	.01	.02	
p = .0057, q = .5	.69	.07	.06	.14	.01	.03	.018
J.G.	.55	.06	.08	.18	.04	.03	
p = .045, q = .7	.57	.10	.11	.16	.02	.04	.025
E.G.	.93	.03	.00	.01	.02	.00	
p = .1, q = .6	.95	.03	.00	.00	.02	.00	.009
B.Me.	.84	.03	.01	.00	.05	.01	
p = .1, q = .82	.85	.09	.01	.02	.03	.00	.028
B.Mi.	.83	.05	.01	.01	.02	.01	
p = .055, q = .7	.84	.08	.02	.03	.02	.01	.016
J.A.	.78	.04	.00	.02	.03	.01	
p = .058, q = .7	.89	.07	.01	.02	.02	.00	.047
A.F.	.75	.02	.03	.07	.06	.04	
p = .1, q = .85	.77	.11	.03	.05	.04	.00	.043
N.C.	.75	.03	.07	.08	.01	.00	
p = .1, q = .85	.77	.11	.03	.05	.04	.00	.041
I.G.	.69	.09	.05	.02	.03	.01	
p = .1, q = .86	.73	.13	.04	.05	.04	.01	.027
H.B.	.61	.06	.13	.18	.02	.01	
p = .05, q = .713	.59	.11	.11	.14	.02	.03	.030
J.F.	.56	.14	.01	.02	.11	.01	
p = .1, q = .86	.73	.13	.04	.05	.04	.01	.077
G.L.	.28	.04	.21	.30	.03	.09	
p = .079, q = .85	.27	.11	.20	.29	.03	.10	.030
W.R.	.08	.06	.15	.28	.05	.33	
p = .1, q = .94	.18	.09	.20	.37	.03	.13	.102

Note. The first row for each patient shows the proportions obtained and the second row the proportions predicted. RMSD = root mean squared deviation.

Model Fit to the Patient Data

The model's fit to the error category proportions for 21 individual patients is shown in Table 7. (Recall that 2 patients were not fit because they had too many nonnaming responses.) The table is organized by type of lesion (predominantly p vs. predominantly q) and, within that, by degree of accuracy. The fitting process was informal. For each patient, we identified combinations of p and q that give a level of correctness similar to that patient. Each combination was simulated using 1,000 trials with each of the two neighborhoods in the model and averaging them with weights of .9 and .1 for the first and second neighborhoods, respectively. We then chose the combination that led to the best match between patient and model, specifically the one with the smallest value of chi-square, using the model error proportions to determine expected values.[7] If a patient could be reasonably well fit with a combination that we had

[7] When the expected proportion for a category was zero, which sometimes happened with the unrelated category, we simply left it out of the chi-square calculation. If a comparison between two potential fits required a comparison between chi-squares based on a different number of categories because one was left out due to zero expectation, the comparison used the root mean squared deviation instead of chi-square to arbitrate.

already used, we just used that combination rather than try to fine tune the fit.

Because the model's predictions are based on stochastic simulations, there is a small chance that a simulated proportion can deviate from the true model proportion by as much as .03 when simulations are based on 1,000 trials per neighborhood. Consequently, after a good fit was found for each patient, we checked the fit by testing 10,000 trials per neighborhood with the fitted parameters. If the two sets of model proportions differed by a root mean squared deviation (RMSD) of .01 or greater, the patient was refit. Three patients required refitting.

The fit between the model and patients was good. Across the 21 patients, the RMSD between model and patient proportions ranged from .007 to .102 (*Mdn* = .026). To evaluate the fit in RMSD terms, we generated 10 pseudopatients, each with random proportions assigned to the six response categories (constrained to add to one) and attempted to fit the model to them. The median RMSD for the pseudopatients was .220. The fits for the real patients were substantially better, even in the case of the poorest fit, W.R. In every error category, the deviations were small. In fact, there was only a single point out of 126 in which the deviation was greater than .15, the unrelated proportion for patient W.R.[8] The only error category, as a whole, in which there appeared to be a systematic deviation between the model and the data was the semantic category, in which the model overpredicted somewhat. This overprediction may be traced to the initial fit to normal data, in which there also was a small overprediction of the semantic category.

Consider the model's fit to the 3 patients whom we used earlier to illustrate the general patterns in the data. These patients, G.L., I.G., and L.H., had fits by the RMSD measure around the median, .030, .027, and .018, respectively. First, compare G.L. with I.G. Their data show how error patterns varied with severity, with G.L.'s naming being more disrupted (.28 correct) than I.G.'s (.69 correct). The differences in the model's proportions lay in the nonwords (G.L. = .29, I.G. = .05), the formals (G.L. = .20, I.G. = .04), and the unrelateds (G.L. = .10, I.G. = .01). The model's semantic (G.L. = .11, I.G. = .13) and mixed (G.L. = .03, I.G. = .04) proportions were similar for them and, in fact, were slightly more likely in the less severe patient. The actual data also exhibit these patterns. In the model, these interactions are caused by relatively more phonological errors (nonwords and formals) and, generally, by a greater influence of categories with many opportunities with more extreme lesions.

Both G.L. and I.G. were characterized by lesions in decay more than in connection weight. I.G.'s lesion, in fact, was a pure decay lesion. The effect of type of lesion can be seen by comparing the fit of I.G. to L.H., who was fit by a loss in connection weight but whose naming performance was the same as I.G.'s (.69 correct). The weight lesion promoted nonwords (L.H. = .14, I.G. = .05) and unrelateds (L.H. = .03, I.G. = .01) at the expense of related word errors, particularly semantic (L.H. = .07, I.G. = .13) and mixed (L.H. = .01, I.G. = .04). These were exactly the differences between I.G.'s and L.H.'s actual error patterns. As we said earlier, the weight lesion exaggerates the errors associated with a lack of consistency among representational levels, whereas the decay lesion's errors tend to maintain some consistency.

A good way to understand the model's fit is to scale error proportions relative to error opportunities. For example, earlier we estimated that the opportunities for a formal error for the patients was .09. That is, a random phonologically legal string of about the length that is typically produced in naming tasks would result in a word that is formally related to the target (by our definition) approximately 9% of the time. Similarly, in the model, random outputs would create a formal error slightly less than 8% of the time. One can ask of each patient and the model's fit of each patient whether each error type occurs more or less often (and by how much more or less often) than these opportunities. Ideally, the patients and the fits would agree.

We can quantify the error proportion relative to opportunities as the natural log transform of the ratio of the actual patient error proportion to the estimated real opportunities. The same can be done for the fit to each patient by taking the same transform of the model's predicted proportion over the model's opportunities. For both the patients and the model, a positive number indicates that that error type occurred more than expected by chance and a negative number that it occurred less often than chance opportunities. Table 8 shows these values for each patient and for the model for the five error categories, and Figure 9 shows the relation between the patient and model values. The correlation between predicted and obtained values was .94, with a slope of 1.2 and an intercept of .03. (A perfect fit would have correlation of 1.0, a slope of 1.0, and a zero intercept.) There were only 3 (out of 92) points in which predicted and obtained values had a different sign. The good fit using the values scaled to opportunities is to be expected given the overall good fit between predicted and obtained error proportions and the fact that the model's opportunities are similar to the estimated real opportunities.[9]

The opportunities transform allowed us to look more closely at how error patterns change as a function of severity and hypothesized lesion type. We placed some of the patients into five groups: (a) high correctness, mostly weight lesions (W.B., T.T., J.Fr., V.C., and L.B.); (b) high correctness, mostly decay lesions (E.G. and B.Me.); (c) medium correctness, mostly weight (J.L., G.S., and L.H.); (d) medium correctness, mostly decay (A.F., I.G., N.C., and J.F.); and (e) low correctness, mostly decay (G.L. and W.R.). Not all of the patients were included for two reasons: First, we wanted relatively "pure" lesions. None of

[8] W.R. often perseverated his word responses. Our coding scheme simply categorized these depending on their relation to the target, which was typically unrelated. This was why he had so many unrelated responses (see Table 6). Because the model does not deal with perseverative influences, it could not handle this effect.

[9] The largest discrepancy between the model and the real opportunities lay with the semantic category, in which the model's opportunity was .04 and our estimate was .01. Consequently, the effect of the opportunities transform is to reduce the model's prediction for the semantic category relative to the patient data. Recall that semantic error proportions were on average less than predicted by the fits. Therefore, when fits are scaled by opportunities, this discrepancy is corrected and, in fact, is now slightly reversed: Patients make more semantic errors than predicted on a per opportunity basis.

the designated decay-lesion patients had fitted weights below .07, and none of the weight-lesion patients had decay rates above .6. Second, we wanted to equate the average correctness levels for the lesion types. Average correctness for the high categories was .896 and .885 for the weight- and decay-lesion groups, respectively. For the medium categories, the averages were .717 and .688, respectively. The low-correctness group, which contained only decay lesions, averaged .180 correct.

Figures 10 and 11 show the error patterns on a per opportunities basis for the five patient groups and the model's fits. In these figures, each point is an average of the opportunities transforms for the patients in the group. (When a predicted or obtained value was undefined because the input to the log transform was zero, the value was set to −4.90, which was the transform of the smallest nonzero ratio.) Clearly, the model was producing the general patterns found in these groups. Consider, in particular, how patterns change with severity by comparing high-, medium-, and low-correctness patterns for the decay lesion (see Figure 10). As severity increased, the values tended to approach zero, that value in which errors equaled opportunities. (A profile of zero for all errors represented the random point, the point of maximum severity.) The effect of lesion type also is apparent from these figures: Decay lesions tended to promote semantic and mixed errors (the "smart" errors) rather than nonword and unrelated errors (the "stupid" errors), which were more in evidence with connection weight lesions (see Figure 11).

Table 8
ln(Errors/Opportunities) for Patients and Model Fits

	Error category				
Patient	S	F	N	M	U
W.B.	0.53	−2.10	−4.23	0.41	—
	0.00	−2.07	−3.62	0.92	—
T.T.	−0.51	−2.71	−4.90	1.45	—
	0.00	−2.07	−3.62	0.92	—
J.Fr.	−0.51	−2.10	−3.55	1.45	—
	0.00	−2.07	−3.62	0.92	—
V.C.	0.83	−2.10	−3.16	1.01	—
	0.22	−1.37	−2.93	0.92	—
L.B.	1.39	−1.67	−2.17	1.01	−2.81
	0.00	−0.97	−2.24	0.92	−1.42
J.B.	1.84	−2.10	−2.75	1.75	−2.81
	0.41	−0.68	−2.24	0.92	−1.02
J.L.	1.22	−2.10	−2.64	1.95	−2.81
	0.41	−0.97	−2.53	0.92	−2.12
G.S.	0.83	−0.36	−1.65	1.01	−1.77
	0.56	−0.28	−1.68	0.92	−1.02
L.H.	1.22	−0.27	−1.68	1.01	−1.77
	0.56	−0.28	−1.68	0.92	−1.02
J.G.	1.84	−0.12	−1.48	2.30	−1.08
	0.92	0.33	−1.54	1.61	−0.73
E.G.	1.10	—	−4.38	1.61	—
	−0.29	—	—	1.61	—
B.Me.	1.06	−2.71	—	2.44	−2.81
	0.81	−2.07	−3.62	2.01	—
B.Mi.	1.63	−2.71	−4.90	1.75	−2.21
	0.69	−1.37	−3.22	1.61	−2.12
J.A.	1.39	—	−3.85	1.98	−2.21
	0.56	−2.07	−3.62	1.61	—
A.F.	0.53	−1.13	−2.38	2.66	−0.92
	1.01	−0.97	−2.71	2.30	—
N.C.	1.22	−0.27	−2.37	0.41	—
	1.01	−0.97	−2.71	2.30	—
I.G.	2.15	−0.55	−3.55	2.15	−2.21
	1.18	−0.68	−2.71	2.30	−2.12
H.B.	1.74	0.34	−1.51	1.75	−2.81
	1.01	0.33	−1.68	1.61	−1.02
J.F.	2.66	−2.10	−3.85	3.31	−2.21
	1.18	−0.69	−2.71	2.30	−2.12
G.L.	1.25	0.87	−1.00	1.98	−0.08
	1.01	0.93	−0.95	2.01	0.18
W.R.	1.74	0.50	−1.06	2.44	1.20
	0.81	0.93	−0.71	2.01	0.45

Note. If a patient's or the model's error proportion is zero, the ln transform does not exist and is indicated by dashes. The first line for each patient is the data, and the second is the model's fit. S = semantic; F = formal; N = nonword; M = mixed; U = unrelated.

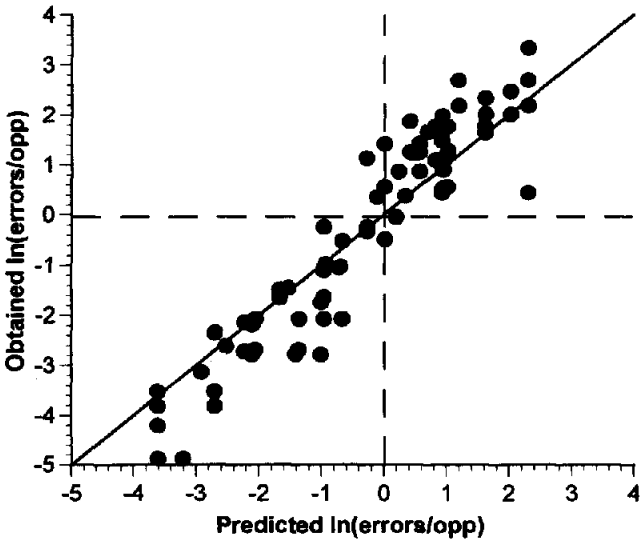

Figure 9. The relation between ln(errors/opportunities [opp]) for 21 patients and the model's fit to these patients for all error types.

The good fit between the patient data and the model suggests three conclusions. First, it extends support for the interactive two-step approach to naming. A model that successfully characterized normal performance could be applied to the range of performance that fluent aphasic individuals exhibit. Although only a restricted set of error patterns is allowed by the model, the patients' patterns appeared to fall within that set. Second, the good fit supports the continuity thesis. A large component of disordered naming can be linked to general severity. More severe aphasic patients have an error pattern that is closer to the error opportunities afforded by the lexicon, whereas less severe aphasic patients have a pattern that is similar to the normal pattern. Finally, the fit supports the hypothesis that variation in patient error patterns can be associated with global lesions in activation transmission, representational integrity, or both.

Predictions From the Model

If the model's characterizations of each patient are valid, we ought to be able to use the fits to predict other previously

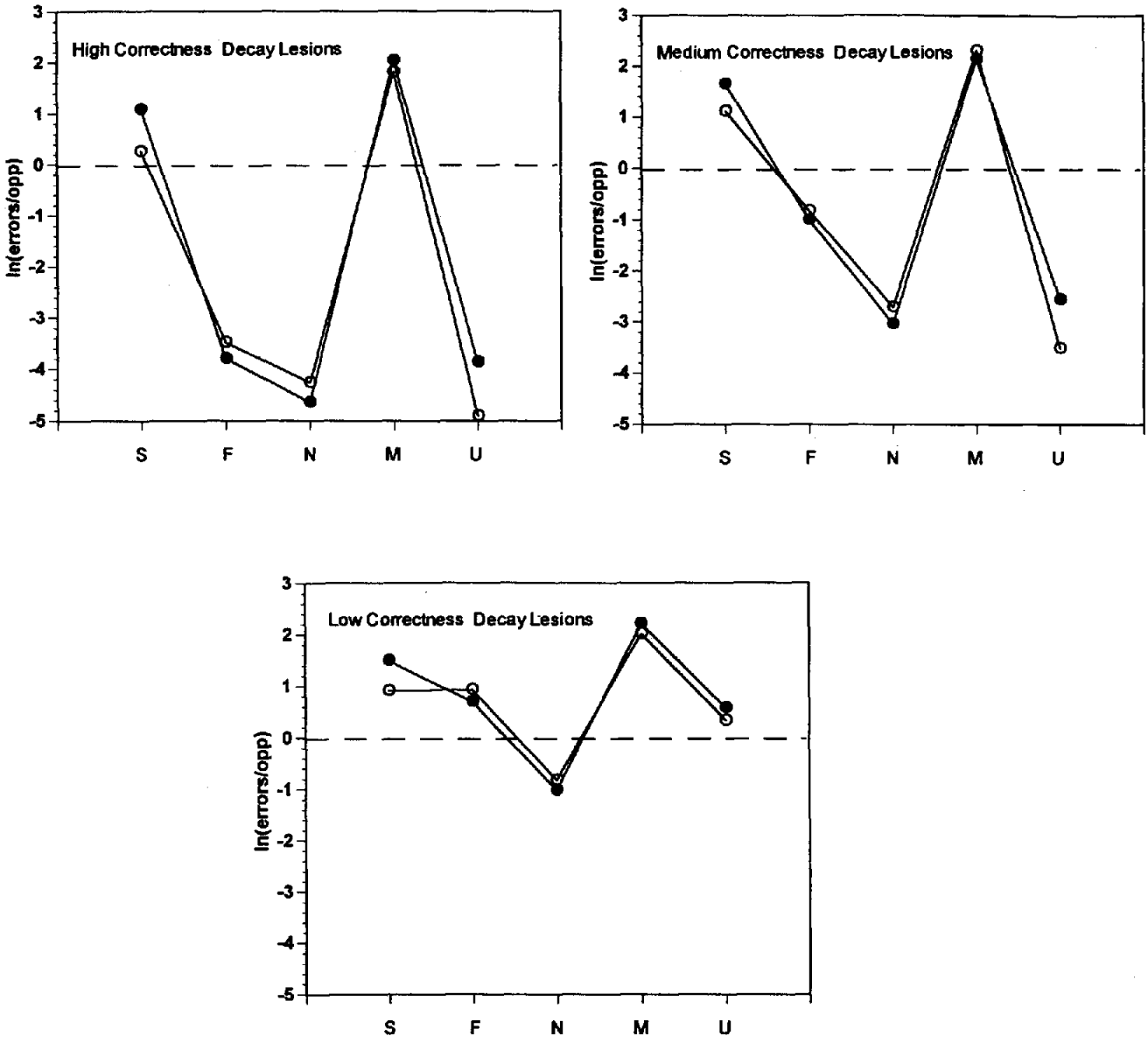

Figure 10. Patient (black dots) and model (open dots) error proportions expressed as ln(errors/opportunities [opp]) for patients with relatively pure decay lesions grouped by correctness. S = semantic; F = formal; N = nonword; M = mixed; U = unrelated.

unexamined aspects of their performance. We have done this with respect to four effects: the influence of syntactic category on formal errors, the extent to which semantic errors are influenced by phonology, the naming error pattern in the patients after some recovery, and the patients' ability to repeat auditorily presented words.

The most novel aspect of the model's application to aphasia is the claim that the differences in error pattern, at a given level of severity, are simply attributable to whether the lesion involves activation transmission or representational integrity. Three of the predictions that we tested addressed this claim directly. As a preliminary to testing these predictions, though, we first needed to go back to the model and study its behavior more closely, particularly the conditions under which the model allowed for interaction among its processing levels.

Because the model's connections run in both top-down and bottom-up directions, phoneme nodes are activated during lemma access, and their activation affects that of the word nodes. The flow of activation from phonemes to words at lemma access produces true malapropisms (i.e., form-related word substitutions) and the mixed-error effect (i.e., the tendency for formal similarity to augment semantic substitutions). The occurrence

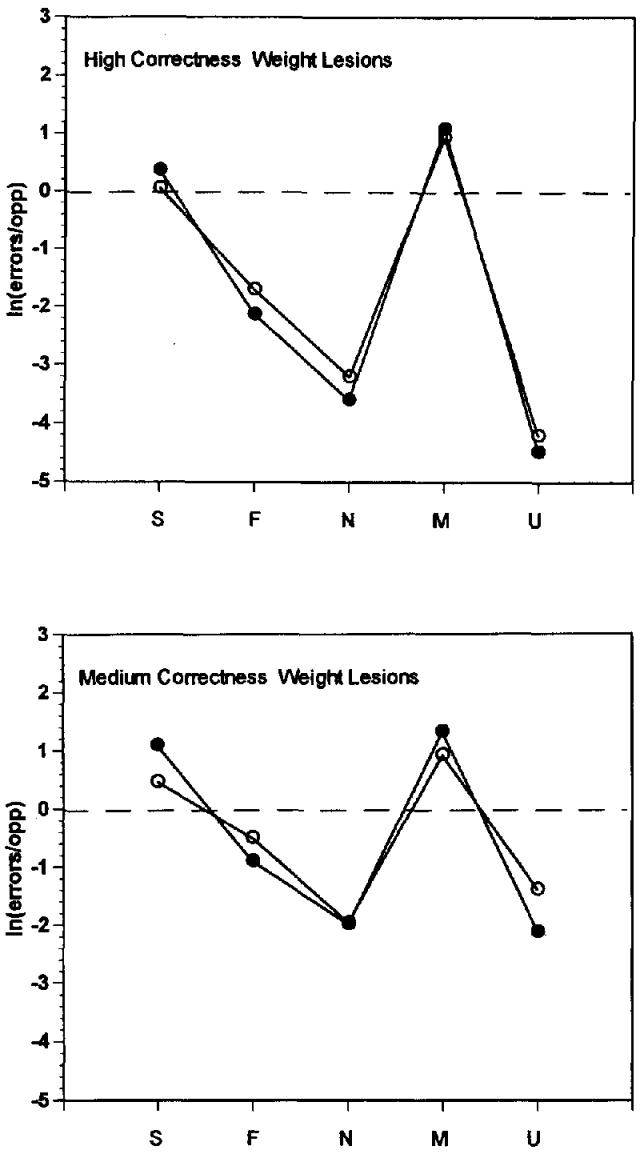

Figure 11. Patient (black dots) and model (open dots) error proportions expressed as ln (errors/opportunities [opp]) for patients with relatively pure connection weight lesions grouped by correctness. S = semantic; F = formal; N = nonword; M = mixed; U = unrelated.

of these interactive effects depends heavily on model parameters. To study these dependencies, we developed a measure of the strength of interaction in the model. Recall that we showed earlier that there was a mixed-error effect in the model with the normal parameter settings by comparing the probability of selecting a mixed neighbor at lemma access with that of a semantic neighbor. The neighborhood that had exactly one semantic and one mixed neighbor was used in this comparison. The extent to which the mixed word's lemma has a greater selection probability than the purely semantic one is an index of the impact of bottom-up activation from the phonemes on lemma activation.

We varied p and q throughout the normal and aphasic regions of the parameter space and ran 1,000 trials of lemma access for each combination using the neighborhood described earlier. The measure of strength of interaction was a z-score test of the difference between the number of mixed and semantic selections, assuming a normal approximation to the binomial. The larger the z score the greater the interactiveness. Figure 12 is a contour map of the size of this measure. It shows dramatically how the potential for interaction is limited to parameter combinations with near-normal connection weights. Specifically, there is no real tendency for interaction unless the weight, p, is .05 or larger. In the aphasic region, this constraint delineates an area in which decay is much greater than normal. If there are near-normal weights, then there must be an abnormally high decay rate to create aphasic performance.

If interaction is confined to models with a p of at least .05, one can partition the patients into those whose connection weights are high enough to create interactive effects and those that are not. This partition, reflected in Table 7, creates 11 high- and 10 low-weight patients. To increase the number of patients, V.P. and G.B., the ones who were not modeled because of their high proportion of nonresponses, were evaluated according to whether they should be considered high or low weight. Because both had many fewer nonwords than related word errors, which is the principal feature of high-weight (or decay-lesion) patients, they appear to belong in the high-weight group. We verified this intuition by fitting the model to V.P. and G.B. in two ways: First, we fit the raw proportions for the six categories in the same way as for the other patients. This will necessarily lead to imperfect fits because of the large number of missing

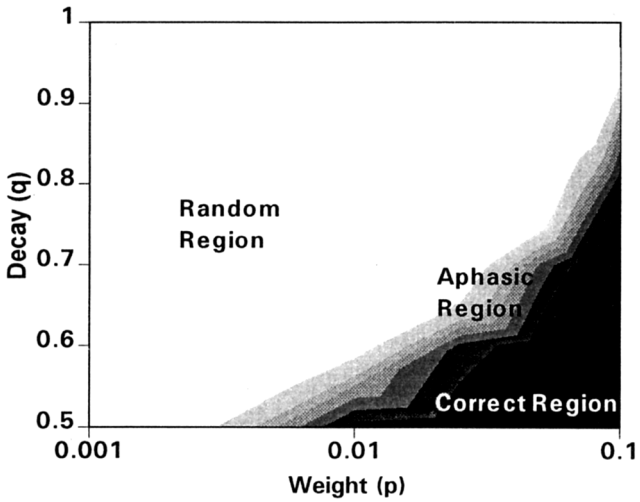

Figure 12. Contour map of the model's parameter space expressing the difference between the number of selections (out of 1,000) of a mixed lemma and a purely semantic lemma with Neighborhood 2. The difference is expressed as a z-score test of the difference between mixed and semantic proportions with a sample size of 1,000, assuming a binomial approximation to the normal. The darker the region, the greater the evidence that mixed lemmas are stronger competitors than purely semantic lemmas. The connection weight is plotted with a log scale.

observations. The parameters assigned by this method placed both in the high-weight group (for V.P., $p = .1$, $q = .93$, RMSD = .128; for G.B., $p = .051$, $q = .72$, RMSD = .080). Second, we rescaled these patients' proportions for the modeled categories so that they summed to 100%. This effectively treated the failures to respond as missing data. Again, both were placed in the high-weight category (for V.P., $p = .1$, $q = .92$, RMSD = .110; for G.B., $p = .051$, $q = .72$, RMSD = .016). Because both ways of fitting the 2 patients led to the same conclusion, we therefore assigned them to the high-weight group, increasing its size to 13. In the next three sections of this article, we describe tests of the predictions related to the partition between high- and low-weight patients.

Syntactic Category Effects on Formal Errors

Formal errors that happen at lemma access should, according to the general theory behind the model, obey the syntactic class constraint. In a picture-naming task, the target category is noun and therefore lemma access errors should be nouns. If the model's parameters for a patient allow for interactive feedback, as in the high-weight group, there can be many formal errors at lemma access. Hence, the set of formals made by these patients will be nouns in excess of what would be expected by chance. If the characterization of a patient does not allow for interaction, which is true for the low-weight group, the patient's formal errors would arise largely during phonological access. Because phonological selection is indifferent to syntactic category, the formal errors made by low-weight patients should be nouns only at chance levels.

Gagnon et al. (in press) studied the formal errors made by 9 aphasic patients in this study and found that these errors do create nouns in excess of chance. Our prediction here, though, is more specific because the model identifies which patients should contribute to the effect and which should not.

We examined all formal errors of the high- and low-weight groups and categorized them as "noun" if the word's most frequent entry in Francis and Kucera (1982) was a noun and as "nonnoun" otherwise. Mixed errors were not counted as formals because they are likely to be nouns as a consequence of their semantic relation to the target.

The result was a clear difference between the groups, as predicted. For the high-weight group, 80.4% of the 148 formal errors were nouns. The low-weight group generated only 58.5% nouns ($n = 53$). To establish the reliability of the difference, we first determined chance expectations. Martin et al. (1994), using a technique developed by Blanken (1990), created a pseudocorpus of word errors of the same structure and word frequency as aphasic patients' formal errors in picture naming. Of the words created by this method, 60% were nouns by our definition. Gagnon et al. (in press) generated a complete set of CVC legal strings and found that of the words in that set, 64% were nouns. To be conservative, we adopted the larger figure as the chance estimate. For each patient we determined the expected number of nouns (.64 × number of formals) and then subtracted the expected number from the actual number. The resulting noun scores of the high- and low-weight groups were then compared to each other and to chance. The high-weight group had a greater noun score than the low-weight group, $t(21) = 3.32$, $p < .004$, and its score also was significantly greater than chance, $t(12) = 3.87$, $p < .003$. The low-weight group's score was not distinguishable from chance ($t < 1$).

The analysis of the difference between the high- and low-weight groups' formals with respect to syntactic category supported both the model's characterization of the patients and the association of syntactic category with lemma access and not with phonological access (contrary to Caramazza & Hillis, 1991). The association between lemmas and syntax is arguably the most important aspect of two-step theories of lexical access (Garrett, 1980; Jescheniak & Levelt, 1994; Levelt, 1989). The model, however, takes this association further and explains how a syntactically specified lemma can be influenced by phonological relations.

Note that the division of patients into high- and low-weight groups was based on data that were independent of the noun–nonnoun distinction, and hence the analysis of this distinction offers independent validation of the model. The noun–nonnoun differences between the patient groups' formal errors was linked to differences in the proportions of other error types, exactly as predicted. More generally, the analysis showed that patients differed in their propensity to produce formal errors at lemma access: true malapropisms. The model's account of this difference is that the patients differ in how much activation transmission takes place between the phoneme and word level.

Phonological Effects on Semantic Errors

The mixed error is another variable in which differences between high- and low-weight groups are predicted. Patients with strong weights should exhibit a true mixed-error effect for the same reason that these patients should make malapropisms; activation can flow from target word nodes to phonemes and then to mixed neighbors. Mixed-error effects are typically evaluated by taking all word errors that are semantically related to the target and then assessing whether the errors also are phonologically related to the targets by comparing how often targets and errors match on phonological dimensions with chance expectations (Dell & Reich, 1981; del Viso et al., 1991). Using this method on the error data from most of the patients in this study, Martin, Gagnon, et al. (1996) showed that, overall, there was a mixed-error effect. Again, though, our expectation was more specific: The mixed-error effect should be confined to patients in the high-weight group.[10]

Like Martin, Gagnon, et al. (1996), we selected all errors categorized as semantic or mixed (except for morphological relatives), and counted how often their first, second, and third phonemes matched their targets. By chance, these proportions

[10] Unlike the prediction for nouns and nonnouns, this mixed-error prediction is not entirely independent of the numbers used to derive the fits. Recall that there is a mixed-error category used in the proportions that guide the fitting.

should be .057, .060, and .056, respectively.[11] The high-weight group showed a substantial mixed-error effect, about the same size as that shown by nonaphasic speakers. In the 198 relevant errors, the match proportions were .121, .126, and .106 for the first through third phoneme positions, respectively. Each of these was significantly different from chance.

The low-weight group's match proportions were lower and did not differ from chance. The proportions, based on 74 errors, were .081, .122, and .041 for the three positions, respectively. The second position proportion was nearly as large as that for the high-weight group, but, because of the smaller number of errors in the low-weight group, it was within the 95% confidence limit around the chance value of .060. In general, there was not a great deal of power for detecting a small mixed-error effect in the low-weight group. Pooling across the three phoneme positions, the expected number of matches was 13 for the low-weight group, and the obtained number of matches was 18. Contrast this with the high-weight group's 34 expected and 70 obtained matches. Therefore, although we could definitely conclude that the high-weight group had a mixed-error effect, we had to suspend judgment for the low-weight group. On the whole, though, the data are consistent with model predictions.

The association of a mixed-error effect with the high-weight group bolsters the findings with syntactic category in the formal errors. The results speak to the validity of the model's fits and, in general, support the view that errors of lemma access are influenced by phonological factors.

Experiment 3: Naming After Recovery

The fit of the model identifies the deficit for each patient. If this fit is useful, it should characterize the patient's future as well as present behavior. We found this to be true in prior studies of Patient N.C., assuming that recovery entails resolution of abnormal parameter values toward the normal state. In N.C.'s case, reducing the abnormally high decay rate used to fit his error patterns in the acute stage captured changes associated with recovery in both naming and repetition of single words (Martin et al., 1994) and in repetition of word pairs (Martin, Saffran, & Dell, 1996).

We are not the first to look to recovery for confirmation of a model of lexical-phonological deficits. Those who explain neologistic speech as anomia (i.e., failed word retrieval) masked by random generation of phonological content often cite as evidence the recovery pattern that starts from Wernicke's aphasia and evolves into anomic aphasia (Kertesz & Benson, 1970). The notion is that the return of self-monitoring blocks the neologism generator from going into operation and unmasks the underlying word retrieval deficit (Butterworth, 1979). Alternative accounts of neologisms make different predictions about recovery (Buckingham, 1987). For example, on the basis of their theory that neologisms arise from faulty activation of lexical-phonological forms, Kohn and Smith (1994b) predicted a progression from complex neologistic distortion of targets toward simpler, less remote deviations. They also observed this in a longitudinal study of naming in a patient with Wernicke's aphasia.

From our perspective, the most intriguing finding from Kohn and Smith's (1994b) study concerns the changes in lexical errors in naming over time. At 6 weeks' postonset, the profile was dominated by unrelated word errors (and neologisms). Over time, unrelated errors were, relatively speaking, replaced by formally related errors; semantic errors remained constant. Kohn and Smith hypothesized a breakdown in the addressing mechanism in which lexical-phonological entries are contacted. The ability to read the phonological address evolves from total failure (yielding nonresponses), to severe misreading of the address (unrelated errors), to partially correct reading of the address (formal errors). This deficit, and its recovery, is presumed to be independent of the lexical-phonological activation deficit that produced the neologisms. In addition, although they do not discuss this, there would have to be a third deficit that explains the occurrence of semantic errors and their resistance to recovery over time.

In our model, recovery can be characterized as the movement of affected parameters toward normal values. In other words, the factors that explain differences across patients also explain differences within patients over time. Given this premise, the model is consistent with many of Kohn and Smith's (1994b) recovery findings, including the switch from unrelated to formal errors and the relative stability of semantic errors. Like Kohn and Smith, we do not invoke devices or systems external to the model to account for the presence or elimination of symptoms. However, where their account postulates multiple deficits, each subject to its own principles of recovery, we hypothesize alterations in the connection weight, decay parameters, or both, each of which is subject to a single principle of recovery: movement toward the normal value. If our account is valid, naming error patterns after recovery should fit the model just as well as the original tests. Moreover, if recovery consists of a normalization of altered parameters, it follows that the character of the fit, whether the patient's lesion is primarily in decay or weight, should not change. For example, if a particular patient is characterized as having a high decay but a normal weight, the characterization after recovery should not be one of a low weight and normal decay. The decay parameter may or may not improve, but at least the patient should not change from the high- to the low-weight category. Thus, for each retested patient, we fit their error data and then determined whether they fell into the high- or low-weight group. If there is some stability to the model's fits, patients should not change groups.

Method. Because we planned to measure recovery by representing the PNT at an interval of several months, we first conducted a small study that assessed practice and learning effects from repeat exposure at short intervals. In this practice study, 6 patients from the original study were retested on equivalent forms of the PNT, one form after an interval of 2 days from the initial test date and the other after 5 days. The administration procedures and scoring were identical to those used in the original study (PNT-1). Five of these 6 aphasic patients and 5 others from the original study were subsequently retested on the full PNT (PNT-2). The shortest "recovery interval" was 1.5 months and the longest was 9 months (*Mdn* = 3 months). All aphasic patients in

[11] These proportions are slightly different from the ones cited previously in the analysis of the mixed-error effect in the errors made by the normal speakers because the method of calculating chance (Dell & Reich, 1981) uses the actual target–error pairs of the particular participant group.

Table 9
Philadelphia Naming Test 2 Results: Recovered Naming and Predictions of the Model

Patient and parameter settings	Naming response						RMSD
	Correct	Semantic	Formal	Nonword	Mixed	Unrelated	
J.B.[a]	.87	.01	.01	.03	.03	.00	
p = .0085, q = .5	.89	.04	.01	.04	.01	.00	.018
A.F.[a]	.94	.01	.01	.02	.02	.01	
p = .01, q = .5	.94	.03	.01	.02	.00	.00	.012
G.S.[a]	.91	.00	.02	.05	.01	.00	
p = .009, q = .5	.91	.04	.01	.03	.01	.00	.019
L.H.	.76	.01	.09	.10	.02	.01	
p = .0065, q = .5	.78	.06	.04	.08	.01	.03	.032
J.G.	.90	.02	.01	.03	.03	.00	
p = .009, q = .5	.91	.04	.01	.03	.01	.00	.012
H.B.	.75	.05	.06	.09	.02	.01	
p = .054, q = .713	.75	.09	.05	.07	.02	.02	.019
J.F.	.74	.09	.01	.02	.09	.02	
p = .1, q = .85	.77	.11	.03	.05	.04	.00	.030
G.L.	.36	.02	.19	.32	.03	.03	
p = .051, q = .74	.35	.10	.17	.28	.02	.08	.043
W.R.[a]	.19	.08	.21	.19	.01	.26	
p = .1, q = .94	.18	.09	.20	.37	.03	.13	.092
J.L.[a]	.96	.02	.01	.01	.01	.00	
p = .1, q = .5	.97	.02	.00	.00	.01	.00	.007

Note. RMSD = root mean squared deviation.
[a] Participated in the practice study.

the recovery study had scores at or below .76 correct on PNT-1. The methods and scoring system were the same as in the original study.

Results. In the practice study, the median change score (repeat minus original) was zero after 2 days (range = −.05 to .14) and .06 after 5 days (range = −.05 to .12). On the expectation that practice effects would be maximal at the shortest interval, the absence of an effect at 2 days provided some assurance that any gains observed at longer intervals represented true recovery.

The 10 aphasic patients run in PNT-2 showed a median improvement of .16 (range = .07–.35). Those who participated in the practice study showed similar median change scores to the unpracticed group (.19 and .14, respectively). The relative ordering among the 10 patients was consistent from PNT-1 to PNT-2; the difference in ranks averaged 1.3. In summary, there were measurable gains from PNT-1 to PNT-2, which, given the minimal practice effects, we feel justified calling "recovery."

We next sought to test the hypothesis that the recovered data could be related to the model. The error patterns were fit as before, and, again, the fit was good (see Table 9). The RMSDs ranged from .007 to .092 (Mdn = .019). Once again, there was only a single point (out of 60) that deviated from the fit value by more than .15. Here it is the nonword value for Patient W.R., the same patient who had the worst fit in the first test.[12] To examine the fit, we expressed the patient and model error proportions on a per opportunities basis, as we did before. Figure 13 shows the relation between predicted and obtained values. The correlation was .92, with a slope of 1.1 and an intercept of 0.15.

Each patient was assigned to the high- and low-weight categories on the basis of their value of p, and this assignment was compared with the assignment derived from PNT-1.[13] The comparison showed considerable agreement between the tests. Of the 10 patients, 8 remained in the same group (4 high weights and 4 low weights), 1 (A.F.) changed from high to low weight, and 1 (J.L.) recovered to the normal level and hence was not assigned a group.

The single patient (A.F.) who changed groups recovered to a high level of correctness (.94). At this level, the difference between a high- and low-weight fit is small. For example, an alternative high-weight fit (p = .1, q = .7) was only slightly worse than the chosen low-weight fit, an RMSD of .014 for the high-weight fit compared with .012 for the chosen fit. Generally speaking, when there is a high level of correctness, there is little difference between kinds of fits. High- and low-weight fits are simply close to normal and hence are close to each other. Thus, it is not problematic that A.F. changed categories because the model, in principle, is not able to determine reliably the type of lesion for patients whose performance is near normal levels. We should therefore also not take much credit for correctly predicting which categories J.G. and G.S. belonged to because

[12] The fact that W.R. was not fit as well as the others in both tests was due to the underprediction of unrelated words and an overprediction of nonwords. As before, we attributed this to his tendency to perseverate words, which is not a feature of the model.

[13] When a patient is assigned to one of these categories, it is only after comparing the chosen fit to alternatives in the other category. If the best fits tended to be near to the original parameter values, there was no attempt to fine tune the fit by fiddling with both parameters. Either the original fit or an alteration of a single parameter was chosen if that alteration was better than the original.

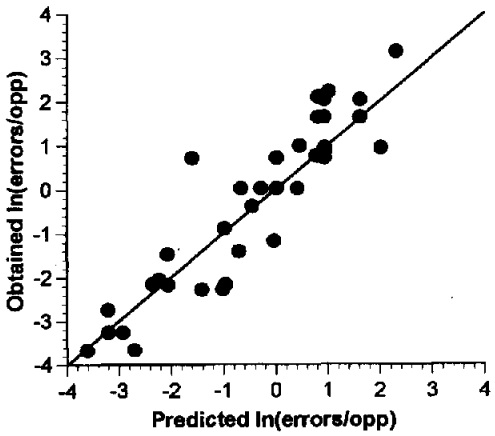

Figure 13. The relation between ln(errors/opportunities [opp]) for the 10 patients in the recovery study and the model's fit to these patients for all error types.

they also recovered to fairly high levels (.90 and .91, respectively). In a similar vein, we should perhaps discount the correct prediction for H.B., whose original (p = .05) and recovered (p = .054) values of connection weight were close to the adopted boundary between high- and low-weight categories. A recovered fit on just the other side of the boundary (p = .045, q = .68) yielded only a slightly worse fit: (χ^2 = 4.94, RMSD = .019 for the high-weight fit and χ^2 = 7.65, RMSD = .020 for the low-weight fit.) The same point can be made for G.L., whose recovered value of p was near the boundary. Discounting H.B., G.L., and the patients who recovered to 90% or better left 4 patients. All 4 were on one side or other of the high-low weight boundary at the first naming test, and their recovery placed them squarely on the same side of the boundary.

In summary, the study of naming recovery supported the model in two respects. First, it provided additional data sets that were adequately fit. Second, it showed that the general category that the model assigns a patient to on the first test predicts the patient's assignment on a test after recovery. That shows some degree of stability of the model's characterization of each patient. Broadly speaking, results of the recovery study support the continuity thesis. Recovery is the movement of pathological parameters toward normal values. Although the mechanism for this movement is not given by the model, the fact that recovery, or within-patients variation, takes place along the same dimensions as those that characterize between-patients variation provides a powerful constraint on theories of recovery. Both recovery patterns and individual differences show the continuity between normal and pathological states.

Application to Single-Word Repetition

Probably the most often-tested ability in aphasic patients is word repetition. The patient is told a word and simply repeats it. Repetition performance has figured importantly in aphasiology because it serves to differentiate certain aphasic syndromes. Preserved repetition distinguishes transcortical sensory aphasia from Wernicke's aphasia, and it is the occurrence of semantic errors in word repetition that defines the disorder known as *deep dysphasia.*

Because the language production system is ostensibly involved in repetition, one ought to be able to use the naming model to predict repetition. At the same time, though, one must recognize that naming and repetition are different tasks. To apply the naming model to repetition requires a theory of the relation between the tasks. Here we consider two contrasting theories of this relation: the single-network and perfect recognition views. We offer these for two reasons: First, we want to illustrate how one could use the naming model to understand other aspects of pathological language. In particular, we show that combining the naming model with a theory of the relation between naming and repetition allows one to characterize whether a patient's repetition is good or poor compared with their naming in a theoretically precise way. Second, we argue that one of the theories, the perfect-recognition view, actually may be correct for many patients. To anticipate our results, it will turn out that most of our patients' repetition error patterns were successfully predicted by the combination of the naming model and a perfect-recognition assumption.

Both the single-network and the perfect-recognition approaches share the assumption that single-word repetition consists of two stages, a word recognition stage and the phonological access stage of production. (Note that such a theory could not apply to the repetition of nonwords without additional assumptions. See Caplan & Waters, 1992, for alternative accounts of nonword repetition.) Therefore, both views say that the parameters derived from naming should play a role in predicting word repetition.

The single-network account asserts that the word recognition stage of repetition uses the same lexical network as the production system (e.g., Allport, 1984; MacKay, 1987). Martin and Saffran (1992) proposed such a view to account for the naming and repetition deficits of N.C., who, at the time, was diagnosed with deep dysphasia because he produced semantic errors in repetition. They suggested that N.C. suffered from an abnormally large decay of activation in a single network. Subsequently, Martin et al. (1994) showed that a model much like the one presented here could account for N.C.'s naming and repetition errors with a large decay lesion. In particular, the decay lesion explained N.C.'s semantic errors in repetition and formal errors in naming.[14]

A single-network view of word repetition, when combined with the naming model, can be implemented as follows: The input word produces a standard jolt of 100 distributed across the input phonemes. So, for a word like *cat,* each phoneme receives 33.3 units of activation. Activation then spreads for n time steps, and the most highly activated word node is selected. This completes word recognition. It is then followed by the phonological access stage in production. The selected word is jolted, and n time steps later the most activated phonemes are

[14] This Patient, N.C., was the same as the N.C. in our study. However, when tested for our study, his naming was much better and he no longer made semantic errors in repetition. We fit him with a pure decay lesion to be consistent with these earlier studies.

selected. With this view of repetition, errors occur during input (the wrong word could be recognized) and during output (the wrong phonemes could be retrieved). According to the single-network view, the same pathological parameters of the network apply during input and output. Hence, repetition can be predicted directly from naming.

The perfect-recognition view assumes that the patient recognizes the auditorily presented word correctly. The result of this recognition is that the correct word node in the naming model is given the standard jolt of activation. The network then attempts to produce this word by using the phonological access step of the naming model using parameters that were determined from the naming test. Therefore, the only errors that result are those of phonological access during output. Notice that the perfect-recognition view directly opposes the single-network one. The input and output phoneme layers and their attendant connections would have to be different to have perfect input along with disordered output. This view, however, is neutral on whether word and semantic layers are shared or separate.

Although the single-network and perfect-recognition views are much different, they are alike in that word repetition performance can be predicted from naming without estimating any additional parameters. In other words, the predictions are absolute, a big plus for both views. This benefit is paid for, though, by their extreme nature. Clearly, many patients have word recognition deficits, contrary to the perfect-recognition view. At the same time, many patients have relatively spared input systems coupled with disordered output, a challenge for a single-network view. Therefore, at the outset, we strongly suspect that neither extreme view will serve as a general theory of aphasic repetition. Nonetheless, it is of considerable interest to compare the predictions of both views with actual data, and we do so in Experiment 4.

Experiment 4: Repetition

Participants. Thirteen of the 21 patients who were modeled in Experiment 2 were also tested on a repetition version of the PNT, hereinafter called the Philadelphia Repetition Test (PRT). The remaining 8 patients were unavailable for testing within 1 month of the initial PNT, the established cutoff. For those tested, the interval between administration of the PNT and PRT ranged from 3 to 28 days (*Mdn* = 11).

Procedures. The PNT target names were randomized and recorded onto audiotape. Nine participants listened to the tape through headphones and repeated each word immediately after it was presented. Four participants were instead given a live voice presentation in which lip cuing was avoided by having the participant look away from the examiner. Participants' responses were tape-recorded and later transcribed. Accuracy of each transcription was agreed on by two examiners. The scoring procedures were identical to what was used for the PNT.

Results. The repetition data from 2 patients could not be modeled. G.S. produced a high proportion of nonresponses (.16), and G.L. was unable to perform the live voice version of the test without lip cues. The results for the remaining 11 patients are shown in Table 10. Repetition performance was considerably better than naming for these patients and showed less variation (.89–.98 correct). The typical error pattern also was different from the naming patterns. There were no pure semantic errors at all; the few errors with a semantic component occurred in the mixed category. Errors were confined mostly to the nonword and formal categories, with nonwords being slightly more likely. The presence of nonwords and a small

Table 10
Philadelphia Repetition Test Results: Single-Word Repetition and Predictions of the Model

Patient and parameter settings	Repetition response						
	Correct	Semantic	Formal	Nonword	Mixed	Unrelated	RMSD
T.T.	.98	.00	.02	.00	.00	.00	
p = .02, q = .56	.97	.00	.01	.02	.00	.00	.010
V.C.	.95	.00	.01	.04	.00	.00	
p = .02, q = .57	.95	.00	.01	.04	.00	.00	.000
L.B.	.91	.00	.03	.06	.00	.00	
p = .007, q = .5	.90	.00	.02	.08	.00	.00	.010
J.L.	.89	.00	.02	.03	.00	.00	
p = .025, q = .6	.92	.00	.02	.05	.00	.00	.015
J.G.	.91	.00	.02	.05	.01	.01	
p = .045, q = .7	.75	.00	.07	.17	.01	.00	.084
E.G.	.94	.00	.03	.01	.00	.00	
p = .1, q = .6	.99	.00	.00	.00	.00	.00	.024
B.Mi.	1.00	.00	.00	.00	.00	.00	
p = .055, q = .7	.95	.00	.02	.03	.01	.00	.025
J.A.	.90	.00	.02	.08	.00	.00	
p = .058, q = .7	.97	.00	.01	.02	.00	.00	.038
I.G.	.95	.00	.02	.02	.00	.01	
p = .1, q = .86	.89	.00	.03	.05	.02	.00	.029
J.F.	.94	.00	.02	.03	.01	.00	
p = .1, q = .86	.89	.00	.03	.05	.02	.00	.023
W.R.	.90	.00	.03	.06	.01	.00	
p = .1, q = .94	.36	.00	.19	.42	.03	.00	.273

Note. RMSD = root mean squared deviation.

number of formal errors, together with the absence of semantic errors, is exactly what one would expect if errors happened only at phonological access.

We then compared the obtained error proportions with those predicted by the naming model, augmented by either the single-network view or the perfect-recognition view. These predictions used the parameters derived from PNT-1 because the repetition test was given within 1 month of this test. The predictions from the single-network view turned out to be uniformly wrong. For every patient, the view severely underpredicted performance and, more important, predicted the wrong relative proportion of error types. The median RMSD was .210 and ranged from .182 to .332. Data from the patient with the median fit, J.A., illustrate the problems. Predicted repetition for J.A., based on his parameters derived from the naming test and assuming a single network, was .56 correct, .00 semantic, .40 formal, .01 nonwords, .03 mixed, and .00 unrelated. The obtained data were .90 correct, .00 semantic, .02 formal, .08 nonwords, .00 mixed, and .00 unrelated. Aside from the severe underprediction, most of the model's errors were incorrectly placed in the formal category instead of the nonword category. This was because most of these errors were in the word recognition step; the target word was mistakenly recognized as a similar one. We conclude that the single-network approach to repetition will not work, at least assuming our naming model. Other aphasiologists have reached this same conclusion on the basis of the poor correlation between input and output tasks (e.g., Monsell, 1987; Nickels & Howard, 1995b).

The predicted values from the model combined with the perfect-recognition view are shown in Table 10. Considering that there was no freedom to adjust parameters here, the fit was, with one exception, excellent. The median RMSD was .024. The model accurately predicted the level of correctness and the general features of the error pattern: slightly more nonwords than formals and no semantic and (almost) no unrelated errors. Moreover, the rare mixed errors occurred mostly in patients for whom mixed errors were predicted. It appears that for nearly all the patients tested, the naming model can account for word repetition with the assumption that word recognition is not disrupted. In passing, we note that this account of repetition predicts no errors with normal parameters in accord with the established view that the normal repetition of words is nearly error free.

There are two caveats, however, to this optimistic assessment. First, 1 patient, W.R., was way off. Repetition was predicted to be .36 and was actually .90. Either the naming model was not correct for W.R. (note that his fit for the PNT-1 and PNT-2 was not as good as the other patients), the combination of the naming model with the perfect-recognition view did not apply in his case, or both. Although he appeared to be an isolated exception, W.R. may in fact be representative of patients whose repetition is substantially better than their naming. Why was there a substantial number of phonological errors in his picture naming but few in repetition? Other repetition models have recourse to a separate, nonlexical pathway to phonology that can supplement the impaired lexical route (e.g., Caplan & Waters, 1992), and patients such as W.R. provide some support for this route.

The second issue concerns the perfect-recognition assumption. It would be good to have some evidence that it is in fact true for the patients. Of the 11 patients whose data are shown in Table 10, we have auditory lexical decision data for 4: J.L., E.G., J.F., and W.R. This provided some measure of their ability to map from auditory input to lexical entries. The patients were presented with 180 words taken from Kroll and Merves's (1986) listing of concrete and abstract words. Half were concrete (>5.5 on a scale of 1–7 for concreteness) and half were abstract (<4.0). There were 180 pronounceable nonwords created from the words by substituting one or two phonemes either in the initial, medial, or final positions. Two sets were created, each with 90 words and 90 nonwords, and items were assigned to each set so that words and their corresponding nonwords were not tested in the same set. The stimuli were presented auditorially and participants were required to judge whether each was a word.

All 4 patients showed good lexical-decision performance. Hit rates were .95, .98, .96, and .98 for J.L., E.G., J.F., and W.R., respectively. False alarms averaged .20. These results show that the assumption of perfect recognition is a reasonable approximation for these patients at least.

The fact that 10 patients' repetition was well predicted under the perfect-recognition assumption (and the 11th patient's repetition was actually better than predicted) does not mean that repetition will be generally as good as (or better than) predicted. Clearly, many aphasic patients have word recognition difficulties. When that is the case, the naming model/perfect-recognition view should overpredict repetition. To examine this issue, we turned to Patient N.C. Martin and Saffran (1992) found that he has difficulty in word input tasks. We verified this by giving N.C. the lexical-decision test described earlier; his hit rate was only .83. We then obtained a current profile of his naming performance by retesting him on the PNT. (This was 2 years after the original test, which is why he was not included in the recovery or repetition studies.) His performance was only marginally better (.81 correct) than on the earlier naming test (.75), and we found that the original parameters (p = .1, q = .85) gave a good fit to the more recent data (RMSD = .038). With these parameters, N.C.'s repetition under the perfect-recognition view was predicted to be .91 correct, .00 semantic, .02 formal, .05 nonword, .02 mixed, and .00 unrelated. Fourteen days later, he was tested on the PRT, and, as expected, his performance fell below the predicted values: .72 correct, .00 semantic, .12 formal, .11 nonword, .00 mixed, and .03 unrelated. The high proportion of formals is consistent with some targets being misperceived as similar words. This case illustrates how the model can illuminate when it fails to fit. It can say precisely when performance on some task is inconsistent with naming. In N.C.'s case, he did not repeat as well as he named. Also, here the explanation is clear: He had a deficit that affected word recognition.

Our extended discussion of the cases in which repetition was not accurately predicted should not obscure our main conclusions on this topic. First, it is possible to use the naming model to develop testable accounts of performance in related tasks. This requires a theory of the relation between the tasks. For repetition, we made the theories that were tested—the perfect-recognition and single-network views—simple so that parameter-free predictions were possible. Second, the data actually sup-

ported the perfect-recognition view to some extent. Our conclusion, which must be regarded as preliminary until more patients are tested, is that most patients' repetition patterns can be predicted from the assumption that word recognition is good and the claim that errors occur in a phonological access step shared with the naming task.

General Discussion

Results of our research have shown that aphasic naming deficits can be understood in terms of global parameter alterations to an interactive two-step model of lexical access in production. The implemented model was assigned network structure and parameters to account for the probability of various error categories in nonaphasic speakers and the time course of processing during lexicalization. More important, the network structure was set up to mimic the error opportunities available in naming studies of English. Global lesions in activation transmission and representational integrity enabled the model to fit the range of patterns in the fluent aphasic patients who were tested. The model's characterization of each patient was further shown to predict the syntactic category of his or her formal errors, the extent to which semantic errors showed phonological influences, the pattern of errors in a second naming test administered some months later, and performance on a single-word repetition task.

The most important dimension underlying performance in the model is the severity of the damage. Error patterns fall along a continuum between normal performance and a random pattern defined solely by error opportunities. The data that we have collected for fluent aphasic patients indicate that a similar continuum applies to them. This is true for both variation among patients and for the within-patients changes as a result of recovery. The model therefore instantiates the continuity thesis and the data support it.

Aside from general severity, the model postulates variation among patients with regard to the kind of damage. One lesion type creates an error pattern in which nonwords and unrelated words are relatively more common. With this pattern, we hypothesize that the transmission of activation among the levels in the network, implemented by the connection weight parameter, is weak. Consequently, errors that reflect a lack of concurrence among the levels predominate. The alternative lesion created errors under conditions that preserved the network's ability for activation flow. This lesion was implemented by changes in the decay rate, but it could just as easily have been implemented by variation in intrinsic noise or jolt size. We termed the affected dimension in these cases "representational integrity." Here, there were relatively more errors that reflected an interactive flow of activation: mixed, semantic, and formal errors.

The principal source of support for the distinction between representational integrity and information transmission came from the model's categorization of patients into high- and low-weight groups. The high-weight group had lost mostly representational integrity and the low-weight group's loss was primarily in transmission. Group membership appeared to be stable over recovery and, more important, predicted the extent to which error patterns showed phonological influences on lemma access. In particular, the high-weight group's formal errors tended to be nouns, and its semantic errors tended to exhibit phonological influences. The low-weight group showed neither of these effects.

Our general discussion of these results focuses, in turn, on four questions: What can be concluded about lexical access in production? How does our approach to the study of language pathology compare with other work? What are the limitations of our approach and findings? How does the model relate to the brain?

Lexical Access in Production

The fit of the model to normal and pathological naming and the successful tests of its predictions strengthen the case for an interactive two-step account of lexical access in production. The assumption that naming involves separate lemma and phonological access steps enables the model to explain the kinds of errors that occur and the time course of the retrieval process. To implement these two steps, the model assumes the existence of a layer of word nodes that is actively selected and controlled by syntactic processes. These processes also create a sizable nonlinearity in the network, enabling the word layer to act as a useful hidden layer. Such a layer is required because semantically similar words are not typically phonologically similar.

Although the model has two distinct steps in lexical access, it allows, during each step, the top-down and bottom-up flow of activation within the entire lexical network. Hence, phoneme nodes become active during lemma access and semantic units receive activation during phonological access. These assumptions make the model an interactive rather than a modular two-step theory. However, because of the way that activation from the serially ordered jolts dominates residual activation, the model is only locally interactive. An input at the semantic level has only mild effects at the phonological level and vice versa. Hence, the model occupies a middle ground between modular two-step theories of lexical access (e.g., Levelt et al., 1991a) and strongly interactive theories. In fact, we believe that the studies of the time course of lexical access (Levelt et al., 1991a; Peterson & Savoy, in press; Schriefers et al., 1990) provide evidence against strongly interactive theories of production.

We see a need for interaction, though, in explaining aphasic and nonaphasic error patterns. In particular, interaction was invoked to account for the mixed-error effect and the syntactic class congruency of formal errors. For both effects, the flow of activation from target words to phonemes and back to phonologically related words is the hypothesized mechanism.

Accounting for error patterns is not the only motivation for interaction, however. Levelt et al. (1991a) tested the hypothesis that the phonological forms of semantic alternates to a target picture name are activated. For example, for a picture of a frog, are the sounds of *snake* activated? This is essentially a test for whether the system is cascaded, a prerequisite for an interactive system. Using a lexical-decision task, they found no evidence for cascaded activation and concluded that modular two-step theories of naming are to be preferred over interactive ones. Dell and O'Seaghdha (1991), however, claimed that an interactive system expects considerable activation in the phonological forms of semantic alternates only if alternates are strongly acti-

vated. Peterson and Savoy (in press), then, tested for activation of the sounds of both strong (e.g., *toad* for a picture of a frog) and weak (e.g., *snake*) alternates and found evidence for cascade with the strong alternates and, like Levelt et al. (1991a), no evidence with weak ones. Thus, there is support for the kind of cascading exhibited by our model.

Why should the lexicon be interactive? One possibility is that the production and word recognition systems use the same lexical and phonological units (e.g., MacKay, 1987; see the discussion in Levelt et al., 1991b). Then, units would necessarily have bidirectional connections allowing for interaction. At this time, though, we are inclined to disfavor this view. One conclusion from our application of the naming model to repetition deficits is that good word recognition occurred in concert with disturbed output during phonological access. This suggests, within the context of the model, that there are different input and output phoneme nodes, and so the existence of word recognition connections cannot be used to motivate the existence of phoneme-word feedback during production. A recent study by Nickels and Howard (1995b), exploring the relationship between input and output processing in aphasic individuals, supported the same conclusion. There is, however, a related motivation for interaction that lies entirely within the production system. The function of lexical access is to get from conceptual representations to phonological forms. It would be worthwhile if the decision about which word to choose at lemma access were to be informed about how retrievable that word's phonological form is. It is to the speaker's advantage to choose a lemma whose form will later be easy to find. In fact, there is some evidence that reducing the retrievability of a word's form influences decisions that are hypothesized to be mediated by lemma activation, namely, syntactic structure decisions (Bock, 1987). Interaction provides exactly the mechanism needed. Lemmas whose forms are accessible gain more activation (through feedback) than those whose are not. In short, interaction may prevent many TOT states, commitments to words whose forms are inaccessible.

Approaches to Language Pathology

The attempt to conceptualize pathological language as a quantitative change in the operation of normal language mechanisms is a recurring theme in the literature on aphasia. Contemporary efforts along these lines include the proposal that altered temporal dynamics are the source of impairments in sentence comprehension (e.g., Gigley, 1982; Haarmann & Kolk, 1991; Prather, Shapiro, Zurif, & Swinney, 1991) and production (Kolk, 1995), the work of Bates and her colleagues, which interprets comprehension deficits in terms of alterations in cue strength (e.g., Bates, Friederici, & Wulfeck, 1987), and recent proposals that attribute sentence processing impairments to limitations in working memory or processing capacity (Blackwell & Bates, 1995; Haarmann, Just, & Carpenter, 1994). In the same spirit, parallel distributed processing (PDP) models have been applied to pathological language, notably in the domain of reading (Hinton & Shallice, 1991; Patterson, Seidenberg, & McClelland, 1990; Plaut & Shallice, 1993b; Plaut, McClelland, Seidenberg, & Patterson, 1996). These models produce behavior ranging from normal to severely dyslexic by varying or damaging characteristics of the networks, such as hidden units or connection weights (see Farah, 1994). The thrust of these modeling efforts has been to simulate the patterns that define particular syndromes.

The major feature that distinguishes our approach from other modeling efforts is our attempt to simulate the performance of a large number of individual patients. To our knowledge, quantitative modeling of individual patient data has seldom been attempted (for two exceptions, see Bates, McDonald, MacWhinney, & Appelbaum, 1991; Plaut et al., 1996). Moreover, we have used the model's characterization of individuals to test predictions, in some cases by grouping patients according to parameters provided by the model. Thus, although our approach focuses on simulating individual performance patterns, the fits can serve as the basis for classifying patients and analyzing group data.

Our model (and data) assigns great importance to the severity dimension, which is in accord with much prior research on aphasic naming. Thus, regardless of their aphasia subtype, poor namers have been found to make proportionally more "remote" errors than those with milder impairment, resulting in the production of more unrelated words and nonwords and more nonnaming responses (Mitchum et al., 1990; Moerman, Corluy, & Meersman, 1983; Schuell & Jenkins, 1961). To our knowledge, our work is the first to link these errors of the more severe patients to the opportunities afforded by the structure of the lexicon.

In summary, our approach to language pathology has many antecedents in the literature on aphasia and the modeling of cognitive processes. At the same time, we would claim uniqueness for the combination of methods and pretheoretical assumptions that define our work, which demonstrates the utility of this approach to neuropsychological data.

Some Limitations

The success of the model in simulating a wide range of individual response patterns provides support for our approach to lexical access in production and the thesis that aphasic errors reflect the same mechanisms that underlie normal performance in the naming task. Our work, however, is subject to certain limitations. These limitations fall into two broad categories: the selection of data and simplifying assumptions made in the model.

Our analysis, like any other, focused on certain data at the expense of other data. These data selection decisions have ramifications for our conclusions about the continuity thesis. Continuity implies that the model should provide a complete account of aphasic naming performance, without recourse to mechanisms that have sometimes been invoked by aphasiologists, such as neologism-producing devices or editors that are not part of normal production. Although we believe that we have come far in meeting this goal, it is necessary to enter some caveats. First, in focusing on naming, particularly on the first attempt to name a picture, we have deliberately avoided contexts in which the motivation for such mechanisms is founded. Editing is often implicated in connection with repeated attempts to name a target, such as the *conduite d'approche* characteristic of conduc-

tion aphasic patients (e.g., Joanette et al., 1980), whereas devices that generate neologisms have been proposed to account for certain characteristics of the spontaneous speech of jargon aphasic patients (Butterworth, 1979). As we point out, the model would have to be modified to accommodate the characteristics of word production in spontaneous speech.

Another aspect of data selection concerns the choice of patients. In particular, we did not attempt to model the naming difficulties of patients with semantic loss secondary to diffuse or degenerative brain disease (so-called "semantic dementia"; see Hodges, Patterson, Oxbury, & Funnell, 1992). Whether these deficits reflect disturbances at the semantic-feature-level implemented in our model, as opposed to, for example, earlier processes that map from visual input to this level (for a discussion, see Chertow & Bub, 1990; Hodges, Patterson, & Tyler, 1994) remains to be seen. Thus, it seems premature to apply the model to these cases.

We also excluded CVA patients with nonfluent aphasia and, among the fluent patients, we did not attempt to fit those whose data included a significant number of trials that were not naming attempts. The latter condition eliminated 2 of 23 fluent patients. These restrictions limit our conclusions and raise the question of whether the excluded patients would challenge the model in a fundamental way. As mentioned previously, we did not consider nonfluent patients because of uncertainty in error coding. Determining whether an error is one of phonological encoding (and hence within the model's domain) or occurs at the articulatory level (and therefore outside the model) is extremely difficult. If this difficulty can be overcome, there is no reason why the model should not apply to these patients. With regard to the fluent patients who do not attempt to name many of the targets, the model does not apply because it does not implement an account of nonattempts. The production of no responses and descriptions raises the possibility that control processes are operating to block the selection of lemmas or phonological content that is remote from the target or the production of deviant forms such as nonwords. To the extent that this occurs, the model is incomplete in a significant respect. It may be that a simple decision rule, such as a threshold level of activation that is required for selection of a unit, could offer an account of failures to respond. However, because no responses and descriptions were seldom a sizable proportion of our patient data, we do not have an adequate database for investigating these events or incorporating a means of generating them into the model. We should point out, though, that the existence of control processes that abort output and edit out deviant forms in aphasic speech would not by itself challenge the continuity thesis. There is ample evidence of such processes in normal speech production (e.g., Baars, Motley, & MacKay, 1975; Garnsey & Dell, 1984; Levelt, 1983).

In addition to the lack of a mechanism for failures to respond, there are several other significant simplifications in the model. Chief among these is that, in its present form, the model does not deal with the precise character of phonological errors. A complete model will have to account for the differences across patients in the way in which target and error overlap in phonemic content and metrical and syllabic structure. To do this, it will have to incorporate into its vocabulary multisyllabic words comprised of syllables of different shapes. It also will have to account for the effects of variables such as word length and frequency. To deal with perseveratory responses, it will be necessary to broaden the influence of activation dynamics beyond that of a single trial (e.g., Plaut & Shallice, 1993a). Also, if the model is extended to multiword utterances and particularly to spontaneous speech, it will have to accommodate sequential order constraints. The general theory of phonological access that motivates the implemented naming model deals with some of these issues (Dell, 1986, 1988). Recent proposals for incorporating more complex words within the framework of a spreading activation production model include those of Hartley and Houghton (1996), Roelofs (1997), and Gupta and MacWhinney (in press).

Although lacking a complete implemented model, our exploration of the properties of the naming model provides the basis for some relevant speculations about phonological errors. First, the phonological overlap between target and error should diminish as the influence of noise becomes greater (e.g., as p decreases or q increases). We have observed this effect in the current model (viz. the generation of unrelated responses as the parameters diverge from normal levels). Second, as the effect of noise increases, one can expect a relaxation of the constraints that operate in the normal system. For example, the movement of phonemes in nonaphasic speech errors is constrained by serial position: Phonemes that are word initial interact with other initial segments to a greater extent than with segments in other structural positions (e.g., Shattuck-Hufnagel, 1992). The operation of this serial position constraint in nonaphasic speakers may reflect the fact that when the system is functioning efficiently, only the most powerful competing segments have a chance of being selected. With the lowered efficiency associated with patients, even weak competitors can exert an influence, and hence one would expect to see phonological errors at nonword-initial positions to a relatively greater extent.

Another drawback of our model is that it makes no allowance for variation in response time. Because we assumed that both lemma and phonological access take a fixed number of time steps, there is no way that time can vary as a function of the quality of information that is retrieved. Moreover, it is generally the case that patients vary among themselves (and from nonaphasic speakers) in the time taken to generate their first complete response. The model simply ignores these differences. There are at least three ways in which time differences could be incorporated into the model. One is that n could be treated as a variable. The second is that entire retrieval attempts (either phonological access, lemma access, or both) could be repeated if initially unsuccessful. Third, time could be associated with the difficulty in discriminating among activation levels during the process of selecting the most activated word and phoneme nodes. We are inclined to reject the first of these. Many of the desirable properties of the model are attributable to the size of the jolts relative to residual activation after n time steps for lemma and phonological access. If considerable variation in n is allowed, these properties are lost. Therefore, our inclination is toward the other two possibilities: treating time as the result of variation in number of retrieval attempts (which requires assumptions about what triggers another attempt) or selection decision difficulty (which

requires us to be more specific about competitive mechanisms in the model).

Of the simplifying assumptions in the model, those that probably stand out most are the particular semantic and phonological neighborhoods in the model's networks. As we noted before, our goal in choosing these neighborhoods was to match on error opportunities rather than other features of real neighborhoods. Our choice of 10 semantic features per word (with semantic neighbors matching on three of them) is arbitrary. Similarly, each word's formal neighbors matched on exactly two phonemes. Unrelated words matched on no semantic features and no phonemes. In real neighborhoods, semantic and phonological similarity is graded; some words are highly similar, some less so, and some not at all. Hence, the similarity structure in our neighborhoods is the roughest sort of discrete approximation to a continuous similarity structure.

The most controversial simplification in the model is the globality assumption: Parameter alterations affect all layers of the network equally and, hence, differences in error rates across patients are explicable without differential involvement of semantic, lexical, or phonological units. The good fits of the model to the data, which include a fairly wide range of error distributions, support this assumption of the model.

Although we found that the globality assumption worked for our sample, there are good reasons to reserve judgment about its general applicability. The problematic cases concern reports of patients with extreme dissociations (i.e., semantic errors without neologisms and other phonological errors or the reverse). For example, a series of articles by Caramazza and Hillis identified patients whose errors in oral naming were almost all of a semantic nature (e.g., Caramazza & Hillis, 1990; Hillis & Caramazza, 1995; Hillis, Rapp, Romani, & Caramazza, 1990). The model can handle such patients with a global lesion, provided that the level of correctness is reasonably high. However, these patients have low levels of correctness and therefore are significant challenges to the globality assumption. In response to this challenge, we make two points: First, we would argue that differences among laboratories in stimuli, procedures, and scoring preclude direct comparisons. For example, when patients self-correct, our practice is to take the first complete response, whereas some take the last or best response (e.g., Hillis & Caramazza, 1995; Nickels & Howard, 1995a; Mitchum et al., 1990). Moreover, we elected to count single phoneme errors as errors—either neologisms or formals depending on the lexical status of the outcome—in contrast to the alternative practice of requiring errors to deviate more from the target (e.g., Hillis et al., 1990; Nickels & Howard, 1995a). Therefore, before we conclude that a particular case is clearly contrary, there must be a greater concordance of methodology. It certainly may be true that these pure semantic patients would produce only semantic errors with our scoring and stimuli, but that remains to be seen. Second, we should point out that the pure semantic patients are often associated with high rates of failures to name: no responses and semantic descriptions. Recall that we did not fit patients with high rates of failures to respond because we had no account of these failures. For example, a pure semantic patient who produced .39 correct, .16 semantic and mixed errors, and no responses and semantic descriptions on the remainder of the trials would have been excluded from our study. When we leave such a patient out of the sample, we might be ignoring a type that requires a nonglobal lesion. However, this pattern does not necessarily require a nonglobal lesion. If nonattempts are construed as events in which the patient has retrieved a nonword or a word that is semantically unrelated to the picture, but has elected to suppress output, a global lesion is entirely consistent. The hypothetical error pattern with only semantic and mixed errors listed earlier would result from global parameters of $p = .1$ and $q = .91$ if the nonattempts are suppressed strings lacking any semantic relation to the pictured concept. In any case, it is apparent that a better understanding of nonnaming attempts is needed to evaluate the extent to which the globality assumption is challenged by patients whose errors are almost exclusively semantic.

The other challenging cases are patients whose errors are almost exclusively phonological (mostly neologisms), for example, the conduction aphasic patient reported by Caplan et al. (1986). This patient produced mostly phonemic errors and no semantic errors in a 60-item naming test. In general, the model with a global weight lesion is consistent with error patterns that are largely, but not exclusively, phonological provided that the level of correctness is not high. For example, at .34 correct, 77% of the errors can be classified as phonological. The patient of Caplan et al., however, had a higher level of correctness, .72, and thus appeared to be a patient who challenges the model's globality assumption. Again, though, methodological considerations may be important. At .72 correct, the model with a global weight lesion predicts only 7% semantic errors. On a 60-item test, this corresponds to around four semantic errors. Hence, their absence is not greatly inconsistent. Moreover, the test contained more familiar words, on average, than the PNT. Therefore, item differences also may be contributing to the relative lack of semantic errors.

At this point, we cannot fully endorse the globality assumption as a substantive claim about the functional basis of lexical retrieval disorders in aphasia. However, we do not find that the existing evidence compels rejection of the assumption either. Our view, based on the data reported here, is that the globality assumption works for a large enough segment of the population to merit further investigation.

Spreading Activation, PDP, and Neural Models

Models of language pathology based on connectionist or spreading activation principles naturally invite discussion about their relation to a neural level of analysis. Can components of the network and the model's processing assumptions be identified with parts of the brain and with neural parameters? Before answering this question, it is useful to consider the function of models of the sort that we have proposed.

Models in the cognitive neuropsychology of language can, at least in principle, range from purely functional information processing models to neural models. In our view, theory is advanced by the construction of models at many places along this range and by considering their relationships. Typically, models at the neural end are more limited in the sophistication of the behavior that is explained, whereas the information-processing

models allow for little to be said about neural implementation. Although it uses spreading activation principles, our model is much closer to the information-processing end of this range. It attempts to model behavioral patterns and makes no specific claims about how its structures and processing assumptions map onto neural structures and processes. For example, variation in the model's "decay rate" should not be construed as a global alteration of some time-dependent process inside neurons. Instead, we varied decay rate as a way of promoting a loss of representational integrity.

Rather than try to map our model directly to the neural level, it may be more useful to consider its relations to models that are a bit more neurally inspired, namely, PDP models. The PDP models of acquired language disorders—typically dyslexia—use multilayered networks with nonlinear activation functions to map from an input level (e.g., an orthographic representation for a reading task) to an output level (e.g., a phonological representation). One of the great advantages of these kinds of models is that connection weights are set by a learning algorithm, thus offering an explanation for the network structure from the task at hand. Learning in PDP models creates intermediate representations that are distributed (see, e.g., Plaut, 1996; Plaut et al., 1996). In the case of a model that maps from meaning to word form, these intermediate representations (the hidden units' activation patterns) would come to be sensitive to both meaning and form. For example, the distributed representation of *cat* would be similar to—share units with—*dog* and *mat*. Therefore, when attempting to retrieve *cat*, the model's activation pattern on its hidden units would be partially consistent with both semantic and formal neighbors. Our model does exactly the same thing, even though its intermediate representations, the word nodes, are not distributed. Because of its interactive assumptions, the actual activation patterns at the word level when *cat* is the intended word are similar to those of both *dog* and *mat*. In this way, the nondistributed model shares properties with the PDP approach. The "hidden units" of the two kinds of models function similarly. Although we have not explored the relationship between our model and the PDP approaches beyond these observations, we suggest that such explorations would be valuable. Therefore, rather than consider models of our type to be strict competitors with the PDP approach, one should consider how their advantages might complement one another. Drawing links between spreading activation and PDP models might also offer clues about how the functions that we lesioned in the spreading activation model might be neurally represented. That is, one may be able to draw analogous links between PDP models and neural models.

Conclusions

We conclude by returning to the continuity thesis. The claim that speech errors and paraphasias are generated by similar mechanisms can be evaluated only with respect to a model of the normal system. We have provided such a model and have attempted to apply it to pathological language. We found that aphasic error patterns may reflect extreme values of particular parameters of the normal system. This is the sense in which our work supports the continuity thesis. How well activation can be transmitted or maintained would be expected to vary within and among speakers. Our account is that aphasia simply extends this natural variation.

References

Allport, D. A. (1984). Speech production and comprehension: One lexicon or two? In W. Prinz & A. F. Sanders (Eds.), *Cognition and motor processes.* (pp. 209–228). Berlin: Springer-Verlag.

Allport, D. A., & Funnell, E. (1981). Components of the mental lexicon. *Philosophical Transactions of the Royal Society of London, 295(B),* 397–410.

Baars, B. J., Motley, J. T., & MacKay, D. (1975). Output editing for lexical status from artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior, 14,* 382–391.

Badecker, W., Miozzo, M., & Zanuttini, R. (1995). The two-stage model of lexical retrieval: Evidence from a case of anomia with selective preservation of grammatical gender. *Cognition, 57,* 193–216.

Bates, E., Friederici, A., & Wulfeck, B. (1987). Comprehension in aphasia: A crosslinguistic study. *Brain and Language, 32,* 19–67.

Bates, E., McDonald, J., MacWhinney, B., & Appelbaum, M. (1991). A maximum likelihood procedure for the analysis of group and individual data in aphasia research. *Brain and Language, 40,* 231–265.

Berg, T. (1988). *Die Abbildung des Sprachproduktionprozess in einem Aktivationsflussmodell* [A spreading activation model of language production]. Tübingen, Germany: Max Niemeyer.

Berg, T., & Schade, U. (1992). The role of inhibition in a spreading-activation model of language production: I. The psycholinguistic perspective. *Journal of Psycholinguistic Research, 22,* 405–434.

Berndt, R. S., Basili, A., & Caramazza, A. (1987). Dissociation of functions in a case of transcortical sensory aphasia. *Cognitive Neuropsychology, 4,* 79–107.

Best, W. (1996). When racquets are baskets but baskets are biscuits, where do words come from? A single case study of formal paraphasic errors in aphasia. *Cognitive Neuropsychology, 13,* 443–480.

Blackwell, A., & Bates, E. (1995). Inducing agrammatic profiles in normals: Evidence for the selective vulnerability of morphology under cognitive resource limitation. *Journal of Cognitive Neuroscience, 7,* 228–257.

Blanken, G. (1990). Formal paraphasias: A single case study. *Brain and Language, 38,* 534–554.

Blumstein, S. E. (1973). *A phonological investigation of aphasic speech.* The Hague, The Netherlands: Mouton.

Bock, K. (1987). An effect of accessibility of word forms on sentence structures. *Journal of Memory and Language, 26,* 119–137.

Bock, K., & Eberhard, K. M. (1993). Meaning, sound, and syntax in English number agreement. *Language and Cognitive Processes, 8,* 57–99.

Bock, K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). San Diego, CA: Academic Press.

Bock, K., & Loebell, H. (1990). Framing sentences. *Cognition, 35,* 1–39.

Bredart, S., & Valentine, T. (1992). From Monroe to Moreau: An analysis of face naming errors. *Cognition, 45,* 187–223.

Brown, A. S. (1991). A review of the tip-of-the-tongue experience. *Psychological Bulletin, 109,* 204–223.

Buckingham, H. W. (1980). On correlating aphasic errors with slips-of-the-tongue. *Applied Psycholinguistics, 1,* 199–220.

Buckingham, H. W. (1987). Phonemic paraphasias and psycholinguistic production models for neologistic jargon. *Aphasiology, 1,* 381–400.

Buckingham, H. W., & Kertesz, A. (1976). *Neologistic jargon aphasia.* Amsterdam: Swets & Zeitlinger.

Buckingham, H. W., & Rekart, D. M. (1979). Semantic paraphasia. *Journal of Communication Disorders, 12,* 197–209.

Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language, 30,* 542–579.

Butterworth, B. (1979). Hesitation and the production of verbal paraphasias and neologisms in jargon aphasia. *Brain and Language, 8,* 133–161.

Butterworth, B. (1985). Jargon aphasia: Processes and strategies. In S. Newman & R. Epstein (Eds.), *Current perspectives in dysphasia* (pp. 61–96). Edinburgh, Scotland: Churchill Livingstone.

Butterworth, B. (1989). Lexical access in speech production. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 108–135). Cambridge, MA: MIT press.

Butterworth, B. (1992). Disorders of phonological encoding. *Cognition, 42,* 261–286.

Butterworth, B., Howard, D., & McLoughlin, P. (1984). The semantic deficit in aphasia: The relationship between semantic errors in auditory comprehension and picture naming. *Neuropsychologia, 22,* 409–426.

Caplan, D., Vanier, M., & Baker, C. (1986). A case study of reproduction conduction aphasia: I. Word production. *Cognitive Neuropsychology, 3,* 99–128.

Caplan, D., & Waters, G. (1992). Issues arising regarding the nature and consequences of reproduction conduction aphasia. In S. Kohn (Ed.), *Conduction aphasia* (pp. 117–149). Hillsdale, NJ: Erlbaum.

Cappa, S., Cavallotti, G., & Vignolo, L. A. (1981). Phonemic and lexical errors in fluent aphasia: Correlation with lesion site. *Neuropsychologia, 19,* 171–177.

Caramazza, A. (1988). Some aspects of language processing revealed through the analysis of acquired aphasia: The lexical system. *Annual Review of Neuroscience, 11,* 395–421.

Caramazza, A., Berndt, R. S., & Basili, A. G. (1983). The selective impairment of phonological processing: A case study. *Brain and Language, 18,* 128–174.

Caramazza, A., & Hillis, A. E. (1990). Where do semantic errors come from? *Cortex, 26,* 95–122.

Caramazza, A., & Hillis, A. E. (1991). Lexical organization of nouns and verbs in the brain. *Nature, 349,* 788–790.

Chertow, H., & Bub, D. (1990). Semantic memory loss in dementia of the Alzheimer type. *Brain, 113,* 397–417.

Chute, D. L. (1990). *MacLaboratory for Psychology* [computer program]. Devon, PA: MacLaboratory.

Dell, G. S. (1985). Positive feedback in hierarchical connectionist models: Applications to language production. *Cognitive Science, 9,* 3–23.

Dell, G. S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review, 93,* 283–321.

Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language, 27,* 124–142.

Dell, G. S., & O'Seaghdha, P. G. (1991). Mediated and convergent lexical priming in language production: A comment on Levelt et al. *Psychological Review, 98,* 604–614.

Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior, 20,* 611–629.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1996). A connectionist model of naming errors in aphasia. In J. Reggia, R. Berndt, & E. Ruppin (Eds.), *Neural modelling of cognitive and brain disorders* (pp. 135–156). New York: World Scientific.

del Viso, S., Igoa, J. M., & Garcia-Albert, J. E. (1991). On the autonomy of phonological encoding: Evidence from slips of the tongue in Spanish. *Journal of Psycholinguistic Research, 20,* 161–185.

Dunn, J. C., & Kirsner, K. (1988). Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review, 95,* 91–101.

Eikmeyer, H.-J., & Schade, U. (1991). Sequentialization in connectionist language-production models. *Cognitive Systems, 3,* 128–138.

Ellis, A. W. (1985). The production of spoken words: A cognitive neuropsychological perspective. In A. W. Ellis (Ed.), *Progress in the psychology of language* (pp. 107–145). Hillsdale, NJ: Erlbaum.

Ellis, A. W., Miller, D., & Sin, G. (1983). Wernicke's aphasia and normal language processing: A case study in cognitive neuropsychology. *Cognition, 15,* 111–144.

Farah, M. J. (1994). Neuropsychological inference with an interactive brain: A critique of the "locality" assumption. *Behavioral and Brain Sciences, 17,* 43–104.

Fay, D., & Cutler, A. (1977). Malapropisms and the structure of the mental lexicon. *Linguistic Inquiry, 8,* 505–520.

Feldman, J. A., & Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive Science, 6,* 205–254.

Ferreira, F. (1993). The creation of prosody during sentence production. *Psychological Review, 100,* 233–253.

Francis, W. N., & Kucera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar.* Boston: Houghton Mifflin.

Freud, S. (1953). *On aphasia: A critical study* (E. Stengel, Trans.). Madison, CT: International Universities Press. (Original work published 1891)

Fromkin, V. A. (1971). The non-anomalous nature of anomolous utterances. *Language, 47,* 27–52.

Gagnon, D. A., Schwartz, M. F., Martin, N., Dell, G. S., & Saffran, E. M. (in press). The origins of form-related paraphasias in aphasic naming. *Brain and Language.*

Garnsey, S. M., & Dell, G. S. (1984). Some neurolinguistic implications of prearticulatory editing in production. *Brain and Language, 23,* 64–73.

Garrett, M. F. (1975). The analysis of sentence production. In G. Bower (Ed.), *Psychology of learning and motivation* (pp. 133–177). New York: Academic Press.

Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production* (pp. 177–220). New York: Academic Press.

Garrett, M. F. (1984). The organization of processing structure of language production: Application to aphasic speech. In D. Caplan, A. Lecours, & A. Smith (Eds.), *Biological perspectives on language* (pp. 172–193). Cambridge, MA: MIT Press.

Garrett, M. F. (1992). Disorders of lexical selection. *Cognition, 42,* 143–180.

Gigley, H. M. (1982). *Neurolinguistically constrained simulation of sentence comprehension: Integrating artificial intelligence and brain theory.* Unpublished doctoral dissertation, University of Massachusetts, Amherst.

Glaser, W. R., & Glaser, M. O. (1989). Context effects in Stroop-like word and picture processing. *Journal of Experimental Psychology: General, 118,* 13–42.

Goodglass, H., & Kaplan, E. (1983). *The assessment of aphasia and related disorders* (2nd ed.). Philadelphia: Lea & Febiger.

Goodglass, H., Klein, B., Carey, P., & Jones, K. J. (1966). Specific semantic word categories in aphasia. *Cortex, 2,* 74–89.

Goodglass, H., Wingfield, A., Hyde, M. R., & Theurkauf, J. (1986). Category-specific dissociation in naming and recognition by aphasic patients. *Cortex, 22,* 87–102.

Grossberg, S. (1982). *Studies of mind and brain: Neural principles of learning, perception, development, cognition, and motor control.* Boston: Reidel.

Gupta, P., & MacWhinney, B. (in press). Vocabulary acquisition and

verbal short-term memory: Computational and neural bases. *Brain and Language.*

Haarmann, H. J., Just, M. A., & Carpenter, P. A. (1994). Computational modelling of normal and aphasic sentence comprehension. *Brain and Language, 47,* 389–391.

Haarmann, H. J., & Kolk, H. H. J. (1991). A computer model of the temporal couse of agrammatic sentence understanding: The effects of variation in severity and sentence complexity. *Cognitive Science, 15,* 49–87.

Harley, T. A. (1984). A critique of top-down independent levels models of speech production: Evidence from non-plan-internal speech errors. *Cognitive Science, 8,* 191–219.

Harley, T. A. (1990). Paragrammatisms: Syntactic disturbance or breakdown of control? *Cognition, 34,* 85–91.

Hartley, T. A., & Houghton, G. (1996). A linguistically-constrained model of short-term memory for nonwords. *Journal of Memory and Language, 35,* 1–31.

Hillis, A. E., & Caramazza, A. (1995). Representation of grammatical categories of words in the brain. *Journal of Cognitive Neuroscience, 7,* 396–407.

Hillis, A. E., Rapp, B., Romani, D., & Caramazza, A. (1990). Selective impairment of semantics in lexical processing. *Cognitive Neuropsychology, 7,* 191–243.

Hinton, G. E. & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review, 98,* 74–95.

Hodges, J. R., Patterson, K., Oxbury, S., & Funnell, E. (1992). Semantic dementia. *Brain, 115,* 1783–1806.

Hodges, J. R., Patterson, K., & Tyler, L. K. (1994). Loss of semantic memory: Implications for the modularity of mind. *Cognitive Neuropsychology, 11,* 505–542.

Howard, D., & Orchard-Lisle, V. (1984). On the origin of semantic errors in naming: Evidence from the case of a global aphasic. *Cognitive Neuropsychology, 1,* 163–190.

Howard, D., Patterson, K. E., Franklin, S., Orchard-Lisle, V. M., & Morton, J. (1984). Consistency and variability in picture naming by aphasic patients. In F. C. Rose (Ed.), *Recent advances in aphasiology* (pp. 263–276). New York: Raven Press.

Humphreys, G. W., Riddoch, M. J., & Quinlan, P. T. (1988). Cascade processes in picture identification [Special issue]. *Cognitive Neuropsychology, 5*(1).

Jescheniak, J.-D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 824–843.

Joanette, Y., Keller, E., & Lecours, A. R. (1980). Sequences of phonemic approximations in aphasia. *Brain and Language, 11,* 30–44.

Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *Boston Naming Test.* Philadelphia: Lea & Febiger.

Kay, J., & Ellis, A. W. (1987). A cognitive neuropsychological case study of anomia: Implications for psychological models of word retrieval. *Brain, 110,* 613–629.

Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition, 14,* 185–209.

Kertesz, A. (1982). *The Western Aphasia Battery.* New York: Grune & Stratton.

Kertesz, A., & Benson, D. F. (1970). Neologistic jargon: A clinicopathological study. *Cortex, 6,* 362–386.

Kohn, S. E. (1984). The nature of the phonological disorder in conduction aphasia. *Brain and Language, 23,* 97–115.

Kohn, S. E. (1993). *Linguistic disorders and pathologies.* New York: de Gruyter.

Kohn, S. E., & Goodglass, H. (1985). Picture-naming in aphasia. *Brain and Language, 24,* 266–283.

Kohn, S. E., & Smith, K. L. (1990). Between-word speech errors in conduction aphasia. *Cognitive Neuropsychology, 7,* 133–156.

Kohn, S. E., & Smith, K. L. (1994a). Distinctions between two phonological output deficits. *Applied Psycholinguistics, 15,* 75–95.

Kohn, S. E., & Smith, K. L. (1994b). Evolution of impaired access to the phonological lexicon. *Journal of Neurolinguistics, 8,* 267–288.

Kolk, H. (1995). A time-based approach to agrammatic production. *Brain and Language, 50,* 282–303.

Kroll, J. F., & Merves, J. S. (1986). Lexical access for concrete and abstract words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 92–107.

La Heij, W. (1988). Components of Stroop-like interference in picture naming. *Memory & Cognition, 16,* 400–410.

Lecours, A. R. (1982). On neologisms. In J. Mehler, E. C. T. Walker, & M. F. Garrett (Eds.), *Perspectives on mental representation* (pp. 217–247). Hillsdale, NJ: Erlbaum.

Lecours, A. R., Deloche, G., & Lhermitte, F. (1973). Paraphasias phonemiques: Description et simulation sur ordinateur [Phonemic paraphasias: Description and computer simulation]. In *Colloquies IRIA: Informatique medical* (pp. 311–350). Rocquencourt, France: Institut de Recherche d'Informatique et d'Automatique.

Lecours, A. R., & Lhermitte, F. (1969). Phonemic paraphasias: Linguistic structures and tentative hypotheses. *Cortex, 5,* 193–228.

LeDorze, G., & Nespoulous, J.-L. (1989). Anomia in moderate aphasia: Problems in accessing the lexical representation. *Brain and Language, 37,* 381–400.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition, 14,* 41–104.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition, 42,* 1–22.

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991a). The time course of lexical access in speech production: A study of picture naming. *Psychological Review, 98,* 122–142.

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991b). Normal and deviant lexical processing: Reply to Dell & O'Seaghdha (1991). *Psychological Review, 98,* 615–618.

Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition, 50,* 239–269.

MacKay, D. G. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia, 8,* 323–350.

MacKay, D. G. (1972). The structure of words and syllables: Evidence from errors in speech. *Cognitive Psychology, 3,* 210–227.

MacKay, D. G. (1982). The problems of flexibility, fluency, and speed-accuracy trade-off in skilled behaviors. *Psychological Review, 89,* 483–506.

MacKay, D. G. (1987). *The organization of perception and action: A theory for language and other cognitive skills.* New York: Springer-Verlag.

Marr, D. (1982). *Vision.* New York: Freeman.

Martin, N., Dell, G. S., Saffran, E. M., & Schwartz, M. F. (1994). Origins of paraphasia in deep dysphasia: Testing the consequences of a decay impairment to an interactive spreading activation model of lexical retrieval. *Brain and Language, 47,* 609–660.

Martin, N., Gagnon, D. A., Schwartz, M. F., Dell, G. S., & Saffran, E. M. (1996). Phonological facilitation of semantic errors in normal and aphasic speakers. *Language and Cognitive Processes, 11,* 257–282.

Martin, N., & Saffran, E. M. (1992). A computational account of deep

dysphasia: Evidence from a single case study. *Brain and Language, 43,* 240–274.

Martin, N., Saffran, E. M., & Dell, G. S. (1996). Recovery in deep dysphasia: Evidence for a relationship between auditory-verbal STM capacity and lexical errors in repetition. *Brain and Language, 52,* 83–113.

Martin, N., Weisberg, R. W., & Saffran, E. M. (1989). Variables influencing the occurrence of naming errors: Implications for models of lexical retrieval. *Journal of Memory and Language, 28,* 462–485.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86,* 287–330.

McClelland J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: 1. An account of basic findings. *Psychological Review, 88,* 375–407.

Meijer, P. J. A. (1994). *Phonological encoding: The role of suprasegmental structures.* Unpublished doctoral dissertation, Nijmegen University, Nijmegen, The Netherlands.

Meyer, A. S. (1994). Timing in sentence production. *Journal of Memory and Language, 33,* 471–492.

Meyer, A. S., & Bock, K. (1992). The tip-of-the-tongue phenomenon: Blocking or partial activation? *Memory & Cognition, 20,* 715–726.

Miller, D., & Ellis, A. W. (1987). Speech and writing errors in "neologistic jargon aphasia": A lexical activation hypothesis. In M. Coltheart, G. Sartori, & R. Job (Eds.), *The cognitive neuropsychology of language* (pp. 235–271). Hillsdale, NJ: Erlbaum.

Minsky, M., & Papert, S. (1969). *Perceptrons.* Cambridge, MA: MIT Press.

Mitchum, C. C., Ritgert, B. A., Sandson, J., & Berndt, R. (1990). The use of response analysis in confrontation naming. *Aphasiology, 4,* 261–280.

Moerman, C., Corluy, R., & Meersman, W. (1983). Exploring the aphasiac's naming disturbances: A new approach using the neighbourhood limited classification method. *Cortex, 19,* 529–543.

Monsell, S. (1987). On the relation between lexical input and output pathways for speech. In A. Allport, D. MacKay, W. Prinz, & E. Sheerer (Eds.), *Language perception and production* (pp. 273–311). London: Academic Press.

Morton, J., & Patterson, K. (1980). A new attempt at an interpretation, or, an attempt at a new interpretation. In M. Coltheart, K. Patterson, & J. C. Marshall (Eds.), *Deep dyslexia* (pp. 91–118). London: Routledge & Kegan Paul.

Nickels, L., & Howard, D. (1995a). Aphasic naming: What matters? *Neuropsychologia, 33,* 1281–1303.

Nickels, L., & Howard, D. (1995b). Phonological errors in aphasic naming: Comprehension, monitoring and lexicality. *Cortex, 31,* 209–237.

O'Seaghdha, P. G., & Marin, J. W. (1997). Mediated semantic-phonological priming: Calling distant relatives. *Journal of Memory and Language, 36,* 226–252.

Pate, D. S., Saffran, E. M., & Martin, N. (1987). Specifying the nature of the production impairment in a conduction aphasic: A case study. *Language and Cognitive Processes, 2,* 43–84.

Patterson, K. E., Seidenberg, M. S., & McClelland, J. L. (1990). Connections and disconnections: Acquired dyslexia in a computational model of reading processes. In R. G. M. Morris (Ed.), *Parallel distributed processing: Implications for psychology and neuroscience* (pp. 131–181). London: Oxford University Press.

Peterson, R. R., & Savoy, P. (in press). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Plaut, D. (1996). Relearning after damage in connectionist networks: Toward a theory of rehabilitation [Special issue]. *Brain and Language, 52,* 25–82.

Plaut, D. C., & Shallice, T. (1993a). Perseverative and semantic influences on visual object naming errors in optic aphasia: A connectionist account. *Journal of Cognitive Neuroscience, 5,* 89–117.

Plaut, D. C., & Shallice, T. (1993b). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology, 10,* 377–500.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. E. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103,* 56–115.

Potter, M., & Faulconer, B. A. (1975). Time to understand pictures and words. *Nature, 253,* 437–438.

Prather, P., Shapiro, L., Zurif, E., & Swinney, D. (1991). Real-time examination of lexical processing in aphasics. *Journal of Psycholinguistic Research, 20,* 271–281.

Rapp, B., Benzing, L., & Caramazza, A. (1995, November). *The modality-specific representation of grammatical category.* Paper presented at the annual meeting of the Psychonomic Society, Los Angeles.

Rinnert, C., & Whitaker, H. A. (1973). Semantic confusions by aphasic patients. *Cortex, 9,* 56–81.

Roach, A., Schwartz, M. F., Martin, N., Grewal, R. S., & Brecher, A. (1996). The Philadelphia Naming Test: Scoring and rationale. *Clinical Aphasiology, 24,* 121–133.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition, 42,* 107–142.

Roelofs, A. (1997). *The WEAVER model of word-form encoding in speech production.* Manuscript submitted for publication.

Romani, C. A. (1992). *The representation of prosodic and syllabic structure in speech production.* Unpublished doctoral dissertation, Johns Hopkins University, Baltimore.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 318–364). Cambridge, MA: MIT Press.

Rumelhart, D. E., & McClelland, J. L. (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1). Cambridge, MA: MIT Press.

Schriefers, H. (1992). Lexical access in the production of noun phrases. *Cognition, 45,* 33–54.

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time-course of lexical access in production: Picture-word interference studies. *Journal of Memory and Language, 29,* 86–102.

Schuell, H., & Jenkins, J. J. (1961). Reduction of vocabulary in aphasia. *Brain, 84,* 243–261.

Schwartz, M. F. (1987). Patterns of speech production deficit within and across aphasia syndromes: Application of a psycholinguistic model. In M. Coltheart, G. Sartori, & R. Job (Eds.), *The cognitive neuropsychology of language* (pp. 163–199). Hillsdale, NJ: Erlbaum.

Schwartz, M. F., Saffran, E. M., Bloch, D. E., & Dell, G. S. (1994). Disordered speech production in aphasic and normal speakers. *Brain and Language, 47,* 52–88.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed developmental model of visual word recognition and naming. *Psychological Review, 96,* 523–568.

Sevald, C. A., & Dell, G. S. (1994). The sequential cuing effect in speech production. *Cognition, 53,* 91–127.

Sevald, C. A., Dell, G. S., & Cole, J. (1995). Syllable structure in speech production: Are syllables chunks or schemas? *Journal of Memory and Language, 34,* 807–820.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial-

ordering mechanism in sentence production. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 295–342). Hillsdale NJ: Erlbaum.

Shattuck-Hufnagel, S. (1987). The role of word-onset consonants in speech production planning: New evidence from speech error patterns. In E. Keller & M. Gopnik (Eds.), *Motor and sensory processes of language* (pp. 17–51). Hillsdale NJ: Erlbaum.

Shattuck-Hufnagel, S. (1992). The role of word structure in segmental serial ordering. *Cognition, 42,* 213–259.

Starreveld, P. A., & La Heij, W. (1996). Time-course analysis of semantic and orthographic context effects in picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 896–918.

Stemberger, J. P. (1985). An interactive activation model of language production. In A. W. Ellis (Ed.), *Progress in the psychology of language* (Vol. 1, pp. 143–186). Hillsdale, NJ: Erlbaum.

Stemberger, J. P. (1990). Wordshape errors in language production. *Cognition, 35,* 123–157.

Stemberger, J. P. (1991). Apparent anti-frequency effects in language production: The addition bias and phonological underspecification. *Journal of Memory and Language, 30,* 161–185.

Talo, E. S. (1980). Slips of the tongue in normal and pathological speech. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand* (pp. 81–86). New York: Academic Press.

Theios, J., & Amrhein, P. C. (1989). Theoretical analysis of the cognitive processing of lexical and pictorial stimuli: Reading, naming, and visual and conceptual comparisons. *Psychological Review, 96,* 5–24.

Valdois, S., Joanette, Y., & Nespoulous, J.-L. (1989). Intrinsic organization of sequences of phonemic approximations: A preliminary study. *Aphasiology, 3,* 55–73.

Vigliocco, G., Antonini, T., & Garrett, M. F. (in press). Grammatical gender is on the tip of Italian tongues. *Psychological Sciences.*

Vitkovitch, M., & Humphreys, G. W. (1991). Perseverant responding in speeded naming of pictures: It's in the links. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 664–680.

Vitkovitch, M., Humphreys, G. W., & Lloyd-Jones, T. J. (1993). On naming a giraffe a zebra: Picture naming errors across different object categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 243–259.

Wells, R. (1951). Predicting slips of the tongue. *Yale Scientific Magazine, 3,* 9–30.

Appendix A

Philadelphia Naming Test and Philadelphia Repetition Test Target Words

1. candle
2. ghost
3. dinosaur
4. tree
5. pen
6. scissors
7. cane
8. comb
9. thermometer
10. well
11. grapes
12. strawberries
13. bread
14. football
15. pig
16. apple
17. hand
18. towel
19. lion
20. glass
21. fork
22. plant
23. garage
24. can
25. table
26. waterfall
27. king
28. boot
29. foot
30. chair
31. banana
32. ring
33. dice
34. calendar
35. knife
36. vest
37. turkey
38. rake
39. balloon
40. duck
41. fireplace
42. pineapple
43. fan
44. window
45. lamp
46. drum
47. skull
48. bridge
49. eskimo
50. dog
51. iron
52. cheerleaders
53. snake
54. ambulance
55. carrot
56. sailor
57. book
58. bus
59. map
60. squirrel
61. microscope
62. bowl
63. van
64. helicopter
65. bottle
66. scarf
67. ball
68. frog
69. cow
70. beard
71. glove
72. owl
73. pipe
74. scale
75. tent
76. flashlight
77. camel
78. goat
79. fish
80. cannon
81. shoe
82. sandwich
83. spider
84. belt
85. toilet
86. wagon
87. ruler
88. tractor
89. queen
90. train
91. church
92. anchor
93. whistle
94. corn
95. pyramid
96. typewriter
97. rope
98. basket
99. letter
100. nose
101. chimney
102. horse
103. key
104. fireman
105. cross
106. crutches
107. bone
108. cat
109. kitchen
110. dragon
111. saddle
112. pie
113. snail
114. pirate
115. clock
116. pumpkin
117. sock
118. closet
119. hair
120. baby
121. bat
122. leaf
123. slippers
124. mountain
125. sun
126. mustache
127. ear
128. door
129. house
130. nail
131. binoculars
132. celery
133. vase
134. pencil
135. elephant
136. hose
137. bench
138. zebra
139. man
140. seal
141. wig
142. necklace
143. desk
144. bell
145. star
146. hammer
147. pillow
148. spoon
149. zipper
150. top
151. flower
152. kite
153. suit
154. cake
155. hat
156. crown
157. piano
158. stethoscope
159. bride
160. butterfly
161. heart
162. skis
163. clown
164. volcano
165. pear
166. octopus
167. saw
168. camera
169. bed
170. harp
171. broom
172. nurse
173. eye
174. cowboy
175. monkey

(*Appendixes continue*)

Appendix B

Relation Between Decay and Noise Lesions

Table B1
Error Probabilities With Various Lesions

Kind of lesion	Category proportions					
	Correct	Semantic	Formal	Nonword	Mixed	Unrelated
High level of correctness						
1. Decay $p = .1$, $q = .8$	.87	.08	.01	.01	.03	.00
2. Mostly decay $p = .058$, $q = .7$	.89	.07	.01	.02	.02	.00
3. Mostly weight $p = .02$, $q = .57$	.88	.05	.02	.04	.01	.00
4. Weight $p = .0085$, $q = .5$	.89	.04	.01	.04	.01	.00
5. Noise $SD1 = 0.5$	.88	.07	.01	.02	.02	.00
6. Mostly weight $p = .02$, $SD1 = 0.0257$	.89	.05	.02	.03	.01	.00
Medium level of correctness						
1. Decay $p = .1$, $q = .92$	.32	.13	.20	.25	.04	.07
2. Mostly decay $p = .056$, $q = .76$	.34	.11	.18	.27	.02	.09
3. Mostly weight $p = .0253$, $q = .75$	.33	.09	.14	.35	.01	.08
4. Weight $p = .0033$, $q = .5$	.34	.07	.12	.38	.01	.08
5. Noise $SD1 = 1.5$	.34	.12	.17	.26	.02	.09
6. Mostly weight $p = .02$, $SD1 = 0.09$	.32	.09	.14	.36	.01	.07
Low level of correctness						
1. Decay $p = .1$, $q = .94$	.18	.09	.20	.37	.03	.13
2. Mostly decay $p = .059$, $q = .8$	.18	.09	.17	.43	.02	.11
3. Mostly weight $p = .02$, $q = .65$	.18	.07	.13	.50	.01	.11
4. Weight $p = .0025$, $q = .5$	.19	.07	.11	.51	.01	.12
5. Noise $SD1 = 2.3$	.18	.08	.16	.43	.02	.13
6. Mostly weight $p = .02$, $SD1 = 0.10$	.19	.06	.15	.50	.01	.09

Increasing the decay rate q in the model has similar effects as increasing the standard deviation of intrinsic noise ($SD1$). We show this by presenting model error patterns at three levels of correctness, approximately .88, .33, and .18 (see Table B1). At each level, we show the pattern with a pure decay lesion (1), a lesion that is primarily decay (2), a lesion that is primarily weight (3), and a pure weight lesion (4). We then give a pure lesion in the noise parameter for each level of correctness (5). The noise lesion is equivalent to the pattern that is primarily decay, but not purely decay. We further show that the decay component to a lesion that is primarily in weight can be mimicked by replacing the decay component with a noise component (6). In summary, noise lesions are much the same as decay lesions. Note that each noise lesion, or noise lesion component, can be mimicked by a lesion in the jolt size. For example, a 50-fold increase in noise is equivalent to a 50-fold decrease in jolt size.

The six numbers after each lesion are proportions for, in order, correct, semantic, formal, nonword, mixed, and unrelated, respectively. For each group of models, the critical observation is that Cases 2 and 5 and Cases 3 and 6 are identical (within random error).

Received October 12, 1995
Revision received January 15, 1997
Accepted January 20, 1997 ■