Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Memory and Language 54 (2006) 228–264
Journal of Memory and Language
www.elsevier.com/locate/jml

A case-series test of the interactive two-step model of lexical access: Evidence from picture naming ☆

Myrna F. Schwartz [a,*], Gary S. Dell [b], Nadine Martin [c], Susanne Gahl [b], Paula Sobel [a]

[a] *Moss Rehabilitation Research Institute, Albert Einstein Healthcare Network, 1200 West Tabor Road, Philadelphia, PA, USA*
[b] *Beckman Institute, University of Illinois, Urbana, IL, USA*
[c] *Department of Communication Sciences, Temple University, Philadelphia, PA, USA*

Received 13 June 2005; revision received 12 September 2005
Available online 15 December 2005
Communicated by Kathyrn Bock

Abstract

Many facts about aphasic and nonaphasic naming are explained by models that use spreading activation to map from the semantics of a word to its phonology. The implemented model of picture naming discussed here achieves this by coupling interactive feedback with two selection steps. The model's structure and default parameters were set up to match the normal naming profile; and aphasic naming is simulated by altering "lesionable" parameters away from default settings. Past studies within this framework have used different sets of lesionable parameters. Here, in the largest and most representative case-series ever modeled, we show the superiority of the version of the model that allows lesions to weaken semantic and/or phonological connections. Pairing this approach to lesions with the assumptions of interactivity and two-step selection equips the model to explain the wide variation in individual naming response profiles and key facts about lexical and sublexical errors.

Keywords: Lexical access; Aphasia; Picture naming; Computational model; Weight-decay; Semantic–phonological; Case-series; Interactive two-step model

Introduction

The early history of language production research was dominated by speech-error studies (Baars, 1976; Baars, Motley, & MacKay, 1975; Dell, 1986; Dell & Reich, 1981; Fromkin, 1971; Garrett, 1975, 1980; Harley, 1983, 1984; Levelt, 1983; Martin, Weisberg, & Saffran, 1989; Shattuck-Hufnagel, 1979; Stemberger, 1985). The focus on errors led naturally to investigations of aphasic speech, where speech errors are commonplace (Blumstein, 1973; Buckingham, 1980, 1987; Buckingham & Kertesz, 1976; Butterworth, 1979; Ellis, 1985; Garrett, 1982, 1984; Harley & MacAndrew, 1992; Kohn & Smith, 1990; Lecours, 1982; Nickels, 1995; Saffran,

☆ This project is supported by a grant from the NIH: RO1 DC000191 (M.F. Schwartz). The authors owe a special debt of gratitude to all who participated in this study (with invariable patience and good cheer) and to the speech–language pathologists of the Center for Communication Disorders of MossRehab and other Philadelphia-area facilities who generously took the time to refer these individuals to us. We also thank Adelyn Brecher, who contributed to patient recruitment and testing and helped the project along in countless other ways, Megan Bartlett-Williams and Esther Lee, who provided assistance with data management, data analysis, and manuscript preparation, and Kay Bock, Merrill Garrett, Matt Lambon Ralph, and an anonymous reviewer who made helpful suggestions on an earlier draft.
* Corresponding author. Fax: +1 215 456 5926.
E-mail address: mschwart@einstein.edu (M.F. Schwartz).

0749-596X/$ - see front matter
doi:10.1016/j.jml.2005.10.001

1982; Schwartz, 1987; Schwartz, Saffran, Bloch, & Dell, 1994) and to aphasic picture naming, which has the additional advantage that the relation between target and error can be precisely defined (Berndt, Basili, & Caramazza, 1987; Garrett, 1992; Howard & Orchard-Lisle, 1984; Kay & Ellis, 1987; Martin & Saffran, 1992; Mitchum, Ritgert, Sandson, & Berndt, 1990). Contemporary researchers who study speech production have a larger array of methods to choose from, including chronometric paradigms measuring response times or eye gaze duration, and functional neuroimaging (e.g., Belke, Meyer, & Damian, 2005; Bock, 1996; Buchsbaum, Hickok, & Humphries, 2001; Glaser & Dungelhoff, 1984; Griffin & Bock, 2000; Indefrey et al., 2001; Indefrey & Levelt, 2000; Meyer, 1996; Meyer, Sleiderink, & Levelt, 1998; Schriefers, Meyer, & Levelt, 1990; Smith & Wheeldon, 1999). Still, naming studies with patients remain an important source of data for testing competing models of speech production, and, in particular, lexical access.

Two recent developments in this area have stimulated renewed interest and debate. The first is the application of connectionist, spreading activation logic to explain how aphasic errors arise (Harley & MacAndrew, 1992) and to simulate the quantitative error patterns in actual patient data (Laine, Tikkala, & Juhola, 1998; Martin, Dell, Saffran, & Schwartz, 1994; Martin, Saffran, & Dell, 1996). The second development is the introduction of case-series methods, in which numbers of individual patients are studied with uniform procedures, with the aim of explaining the variation across patients on the task of interest (Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Goodglass et al., 1997; Lambon Ralph, McClelland, Patterson, Galton, & Hodges, 2001; Lambon Ralph, Moriarty, & Sage, 2002; Nickels, 1994; Nickels & Howard, 1995a).

Combining the computational and case-series approaches for the first time, Dell et al., 1997 fitted the naming data from 21 patients with aphasia to the interactive two-step model of lexical access in production (Dell, 1986, 1988; Dell & O'Seaghdha, 1991). That paper garnered considerable attention from aphasia researchers, who used the model to fit new patients (Caramazza, Papagno, & Ruml, 2000; Croot, Patterson, & Hodges, 1998; Hanley, Dell, Kay, & Baron, 2004; Hillis, Boatman, Hart, & Gordon, 1999; Rapp & Goldrick, 2000; Ruml & Caramazza, 2000; Ruml, Caramazza, Capasso, & Miceli, 2005; Ruml, Caramazza, Shelton, & Chialant, 2000; Schwartz & Brecher, 2000), sometimes with the goal of challenging its generality or its processing assumptions. (See especially Rapp and Goldrick, 2000; Ruml et al., 2005, and the exchange between Ruml and Caramazza, 2000; and Dell, Martin, Saffran, Schwartz, and Gagnon, 2000; also Laine et al., 1998.) To facilitate the application of the model to new data, an automated data-fitting program was developed, which can be accessed from http://langprod.cogsci.uiuc.edu/cgi-bin/webfit.cgi.

The present study extends the computational case-series approach to a new and much larger series of cases. Earlier patient samples were small and nonrepresentative. The largest English-language sample (Dell et al., 1997, 21 patients) did not include subjects with articulatory difficulties or those who made more than 15% omission errors. The largest sample (Ruml et al.'s study of 50 Italian patients, 2005) over-represented patients with semantic deficits. Moreover, as we explain later, our research group has proposed two versions of the model that make different assumptions about aphasic deficits. Although these versions are conceptually different, they make fairly similar predictions for naming, and thus the smaller scale studies that have compared them have not generally found a reliable superiority of one version over the other (e.g., Foygel & Dell, 2000; Ruml et al., 2000). By evaluating the models in a large, representative sample, we hope to provide a definitive account of their particular strengths and weaknesses. On a more general level, as in Dell et al. (1997) we aim to provide support for a theory of lexical access that incorporates interactivity and a two-step selection process. We also aim to enumerate what a successful model of lexical access must contain in the way of processing characteristics and lesion effects in order to successfully capture the full range of empirical findings.

Overview of the interactive two-step model

Fig. 1 illustrates the structure of the model's network. Units for semantic features, words, and phonemes are arranged in layers, with bi-directional connections linking semantic features and words, and words and phonemes. The model's two retrieval steps, word retrieval and phonological retrieval, are each achieved by activation spreading within this network. We use "word retrieval" synonymously with "lemma access," the term "lemma" referring to a wholistic lexical representation that is associated with grammatical information and ultimately links up with a syntactic frame or procedures that control how words are sequenced and inflected (e.g., Dell et al., 1997; Kempen & Huijbers, 1983; Levelt, Roelofs, & Meyer, 1999). There is no level containing morphological representations, or other wholistic word-form representations (e.g., "lexemes"), although we acknowledge that a more complete model might require this (e.g., Caramazza, 1997; Dell, 1986; Levelt et al., 1999). In this model, knowledge of word form is represented in connections from word nodes to phoneme nodes in the segmental layer that are labeled for syllable position. No other aspects of phonological structure are represented.

The word retrieval step begins with a jolt of activation to the semantic features of the intended word (e.g., CAT). The activation spreads throughout the

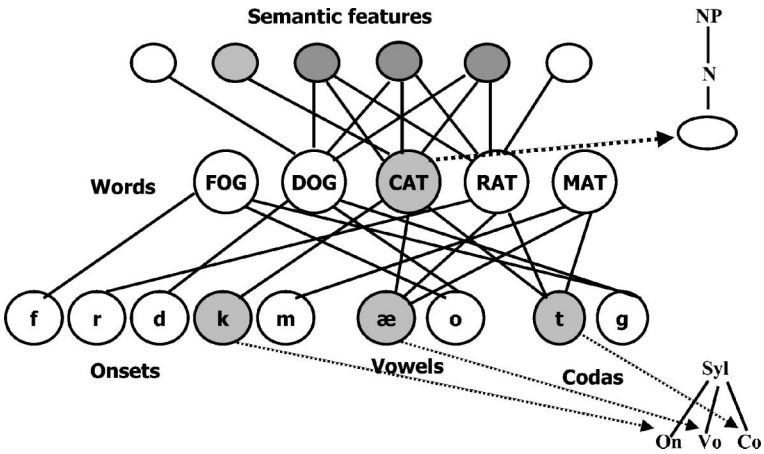

Fig. 1. Structure of the model.

network—both downward and upward—and is concluded by the selection of the most active word unit from the proper grammatical category, which we assume to be nouns for naming pictures of objects. The phonological retrieval step starts with a jolt of activation to this selected word. Activation again spreads through the network culminating in the selection of the most activated phonemes. Errors can occur during either step and result because nontarget words and phonemes gain activation as well as correct ones. Nontarget units can become highly active if they are connected to activated target units and, also, because the activation of all units is subject to random noise. For example, during the word retrieval step, semantically related words (DOG), formally related words (MAT), mixed (semantic plus formal) words (RAT), and even unrelated words (LOG) could be more active than the target word, thus leading to a lexical error. Phonological retrieval provides a further opportunity for errors, including the possibility of selecting a set of phonemes that does not correspond to a word (e.g., LAT).

As its name implies, the interactive two-step model's principal characteristics are that it is *interactive*—information flows up from phonology and down from semantics—and that it has two distinct retrieval steps. These two characteristics are, at least on the surface, strange bedfellows. By its very nature, interaction makes boundaries between processing modules or stages more permeable (Dell & O'Seaghdha, 1991; Farah, 1994). In the model, interaction allows phonological information to affect how semantic information is mapped onto lexical representations (the word retrieval step), and it allows semantic information to impact how phonological representations are processed (the phonological retrieval step). But the distinction between the steps is maintained because, among other things, selection at the first step limits the effect of interactivity at the second step; the boost to the selected word is large in comparison to residual activation generated by interactivity. In this way, the model occupies the middle ground between discrete-stage theories, which do not allow any cross-stage influences (e.g., Levelt et al., 1999), and highly interactive models in which the mapping between meaning and the output of lexical forms is achieved in a single settling step (e.g., Plaut and Shallice, 1993; see Rapp & Goldrick, 2000).

The next section briefly reviews the key qualitative findings from previous empirical studies and shows how the interactive two-step model accounts for them. These are divided into effects that apply to all speakers (1–6), and those that characterize aphasia specifically (7–9).

Lexical-access error phenomena

1. Lexical–sublexical distinction

Some errors appear to be replacements of one entire word with another, while others are characterized by sublexical distortion. It is simply not possible to explain lexical errors as fortuitous combinations of sublexical errors; lexical errors are far too numerous. Hence, any account of impaired lexical access must have a mechanism for both types of error. In our model, it is the two-step assumption that provides the mechanism. Lexical errors result when an incorrect word unit is selected during word retrieval, while sublexical errors require the misselection of phoneme units during phonological retrieval. The model does not deal with other sublexical levels of organization, such as morphology or syllabic constituents, but these are clearly important determinants of the actual form of sublexical errors (Berg, 2005; Dell, 1986; Fromkin, 1971; Garrett, 1975).

2. Semantic errors

The typical lexical error is the semantic substitution (e.g., *elbow* → "knee"; *orange* → "apple"). These are the most common errors made by unimpaired speakers in picture naming (Dell et al., 1997), and every aphasic patient that we have tested on the Philadelphia Naming Test has made at least one. It is thus important that a model generate semantic errors. In our model, semantically related words share feature units. When the target's features are active, they send activation to the word units for semantic neighbors of the target as well as the target itself. If there is enough noise in the system, they can be selected during the word retrieval step instead of the target.

3. Distinction between lexical and sublexical phonological errors, and the dual nature of formal errors

Errors that are phonologically similar to the target can be either words (e.g., *population* → "pollution") or nonwords (e.g., *castle* → "kaksel"). Nonword errors arise during the processing of sublexical units, although they reveal some lexical influences (Schwartz, Wilshire, Gagnon, & Polansky, 2004). Phonologically similar word errors—what we call formal errors—are more complex. Some appear to be caused by the same mechanisms that lead to nonword errors, that is, the movement, deletion, addition, or substitution of sublexical units (e.g., *castle* → "cackle," a perseveratory substitution of /s/ by /k/). These processes just happened to result in a word instead of a nonword (cf., Butterworth's "jargon homophone," 1979). However, a formal error like *population* → *pollution* would require several segment slips to occur if it was generated by segmental transformations. If each such transformation occurred independently, the error likelihood would be miniscule. Of course, if segmental transformations are nonindependent and biased to create lexical items (e.g., phonological attractors in models such as Plaut & Shallice, 1993), then it is conceivable that some more distant formal errors such as *population* → *pollution* could be akin to other sublexical errors. However, an important fact about formal errors that differ from the target in multiple respects is that they often obey the *grammatical category constraint*: the error belongs to the same grammatical category as the target (Fay & Cutler, 1977; Gagnon, Schwartz, Martin, Dell, & Saffran, 1997). Our model explains this by assuming that the mechanism that causes these formal errors is lexical, rather than sublexical.

We label the hypothesis that formal errors can occur at either the lexical or sublexical level the *dual nature of formals*. In our model, both the two-step and the interactive assumptions are required to explain this dual nature. Sublexical formals and nonwords occur during phonological retrieval, and lexical-level formals occur during word retrieval. Lexical-level formals are made possible by interactive feedback from the phonological to the lexical level. During word retrieval, activation of the target spreads to its phonemes. The activated phonemes, in turn, send activation to phonologically similar nontarget words, thus giving their lexical units a boost and a chance at being selected. Because selection at the lexical level is constrained by grammatical category, lexical-level formals (but not sublexical formals) are required to be of the same category as the target.

4. Mixed-error effect

Mixed errors are lexical substitutions that resemble the target both semantically and phonologically (e.g., *start* → "stop," *snail* → "snake"). The mixed-error effect is the finding that these errors are much more likely than would be expected if they were either semantic or phonological. Another way to say this is that the contributions of semantic and phonological similarity to the error process are super-additive, demonstrating that mixed errors are not simply semantic errors that happen to be phonologically similar, or formal errors that happen to be semantically similar. The mixed-error effect is robust in collections of normal speech errors (e.g., Dell & Reich, 1981; Harley, 1984; Laine et al., 1998) and in some (Martin, Gagnon, Schwartz, Dell, & Saffran, 1996), but not all (Dell et al., 1997; Laine et al., 1998) collections of aphasic naming errors. Any account of lexical access must explain the mixed-error effect and its variability in aphasia. Our model locates mixed-errors at the word retrieval step, but requires phoneme-to-word interaction to generate the mixed-error effect. A mixed neighbor of the target gains activation from both shared semantic features and shared phonemes (via interactive feedback). Hence, semantic and phonological similarity jointly trigger the substitution. Disruptions to the bottom-up flow of activation from phonemes to words can reduce the mixed error effect, and this helps explain its absence in some aphasic individuals.

5. Influence of lexical variables

Naming accuracy is affected by word length, word frequency, age of acquisition, and concreteness. Short, common, early-learned, concrete words are, on average, more accurately produced than long, uncommon, later-learned, abstract ones (e.g., Caplan, Vanier, & Baker, 1986; Dell, 1990; Ellis & Morrison, 1998; Martin, Saffran et al., 1996; Nickels, 1995; Nickels & Howard, 2004). Although our model does not directly represent word frequency or age of acquisition, their influence can be attributed to learning mechanisms (not simulated) that set connection weights in the network. Words

that are produced more often or were experienced particularly early during training have stronger weights (e.g., Plaut, McClelland, Seidenberg, & Patterson, 1996; Zevin & Seidenberg, 2002). The model also does not simulate length effects but, again, it is quite consistent with their existence. The more phonemes a word has, the greater the chance that at least one of them will fail to be accessed during the phonological retrieval step. As for concreteness effects, the model can handle these provided that concreteness impacts directly or indirectly the semantic input to the lexical level (Hanley et al., 2004; Martin, Saffran et al., 1996). For example, if concrete concepts are associated with more features than abstract ones (e.g., Plaut & Shallice, 1993), their lexical units will gain more activation and thus be more resistant to error.

While the main effects of frequency, length, age of acquisition, and concreteness are unsurprising, there is one variable whose influence on lexical access in production is counterintuitive and difficult to explain. This is phonological neighborhood density, operationalized as the number of words that differ from a target by one phoneme. We know that words that inhabit phonologically dense neighborhoods are less accurately recognized than those from sparse neighborhoods (Vitevitch & Luce, 1998). This effect is intuitive, reflecting the natural competition involved in word recognition. If the task is word production, however, targets in dense neighborhoods are retrieved more accurately and rapidly, the opposite of what occurs during recognition (Gordon, 2002; Vitevitch, 2002). Our model's interactive assumption provides a possible explanation for the effects in production (Dell & Gordon, 2003). During both word and phonological retrieval, lexical units for phonological neighbors become active because of phonological-to-word feedback. Target phonemes present in the neighbors then boost the activation of the target word, preventing semantic errors during word retrieval, and substitutions of nontarget phonemes during phonological retrieval. It is true that the neighbors' nontarget phonemes become active, as well, and these can compete with targets for selection at phonological retrieval. The benefit that arises from the neighbors' phonemes that are shared with the target outweighs this competition, though. Because neighbors, by definition, share all but one phoneme with the target, a neighbor contributes on average more support for target phonemes than nontarget ones, and a large set of neighbors will distribute this support to all the target phonemes.

6. *Time-course of lexical access*

Although our focus is on error phenomena, there is a large literature investigating lexical access through manipulations that affect the speed of retrieval (see Levelt et al., 1999, for review). For our purposes, the most important conclusion from these studies concerns the time-course of the process. Its initial phases can be affected by semantic manipulations, while its later phases are more affected by phonological manipulations (Schriefers et al., 1990). There also seems to be a period of time in the middle of the process in which both semantic and phonological manipulations have effects (Cutting & Ferreira, 1999; Levelt et al., 1991; Peterson & Savoy, 1998). Our model is consistent with this time course because the earlier word retrieval step is mostly (but not entirely) under the influence of semantic factors, whereas the later phonological retrieval step is mostly (but not entirely) affected by phonological variables.

Aphasic error effects

7. *Interactions between severity and error types*

To us, one of the most telling findings from our previous studies of aphasic naming errors was the varied relation between overall correctness, or severity, and the probability of particular kinds of error (Dell et al., 1997; Schwartz & Brecher, 2000). Some error types increase with severity, as one would expect. This is the case with nonword errors, especially those that are many sounds removed from their targets, formal errors, and unrelated word errors (e.g., *pig* → "brain"). Each of these error types is uncommon in unimpaired and mildly impaired subjects, but increases dramatically with impairment such that each is maximally likely when correctness is at a minimum. Other error types have little relation to correctness. Semantic and mixed errors are, by far, the most common naming error types in normal speakers; but, on average, these types show little tendency to increase with severity.

There are many reasons for the severity/error-type interactions in the model. But the most important is what we have called the *continuity thesis*. The original statement of this thesis was by Freud, who asserted that speech errors and paraphasic errors did not differ qualitatively, only quantitatively (Freud, 1891/1953; p. 13). Our modeling efforts can be thought of as a modern instantiation of this view. Specifically, implementing continuity involved four steps (Dell et al., 1997). First, a normal version of the model was set up so that it matched normal error proportions. Then, an estimate was made of the expected proportion of error types in an individual who generated responses by selecting phonemes at random. To determine this random pattern, we assumed that random responses are nonetheless phonologically legal and approximate the length of patient responses in our naming tasks (i.e., one or two syllables). With these assumptions, random responses would mostly form nonwords. But there is some chance that they

would be real-word errors, even formally or semantically related words. Specifically, we identified the random pattern as 80% nonwords, 10% unrelated words, 9% formals, 1% semantic, and 0.4% mixed. This is what one would expect if a speaker did our picture-naming test but generated responses at random. The third step is to set up the model so that when it is maximally inaccurate, its error pattern is similar to the random pattern. We did this by first setting up a very simple phonology. All words were CVC, with a small set of possible onsets, vowels, and codas. These could be independently combined yielding the model's set of phonologically legal syllables. Next, two different lexical neighborhoods were defined, each with a designated target word and formal, semantic, and (in one neighborhood) mixed neighbors of the target. Fig. 1 shows most of the words of one neighborhood. Overall, the set of possible words was small in comparison to the set of legal syllables, thus mimicking the relative sparsity of words in English phonological space. Finally, it was stipulated that one neighborhood—the one with a mixed neighbor—would be sampled 10% of the time and the other would be sampled 90% of the time. (See Dell et al., 1997, for the actual lexicons used, and Rapp and Goldrick, 2000 and Ruml et al., 2005, for other examples of this approach to setting up error opportunities.) This resulted in the model's simulation of the random pattern: 75% nonwords, 9% unrelated words, 8% formals, 4% semantic, and 0.4% mixed, which is reasonably similar to the estimated random pattern above. (These proportions do not add to 100% because the model's lexicon is very small and there is 4% chance of generating the correct word with a random response. In a real lexicon, this chance would be miniscule.)

At this point, both a normal and a random version of the model had been defined. The final step in the process was to gradually transform the correct version of the model into the random version by altering its parameters. The intermediate versions would, then, represent the model's set of possible aphasic error patterns, or its *error-pattern space*. Hence, this transformational process is a way of "lesioning" the model. Dell et al. (1997) did this by assuming that lesions corresponded to global, that is, network-wide, reduction in connection weight (*weight* parameter) and increase in activation–decay rate (*decay* parameter). This form of the model will henceforth be called the *weight–decay model*. The model's error-pattern space is a two-dimensional surface—one dimension for each lesionable parameter—running from the normal to the random pattern.

In Foygel and Dell (2000) the lesioning was done differently. The weight parameter was split into separate *semantic* (*s*) and *phonological* (*p*) weights, which could be differentially lesioned, and the *decay* parameter was no longer lesionable. This is the *semantic–phonological model*. This model's error-pattern space also is a two-dimensional surface—one for *s* and one for *p*—running from the normal to the random pattern. However, this model allows for more variation near the random pattern, because *s*- and *p*-weight lesions produce different effects in this region of error space. We will show that this property of the model is important in accounting for patients' error patterns (see also Ruml et al., 2005).

As we said before, the continuity thesis, as implemented in these models, is the most important component of our explanation for the severity/error-type interactions. The error types that increase dramatically with severity—nonwords, formals, and unrelated words—are much more likely at the random pattern than at the normal pattern, and the types that are more weakly related to severity—semantic and mixed errors—are actually slightly more common in the normal than in the random pattern. Thus, these severity interactions fall out naturally from the simple assumption that aphasic error patterns lie between normality and randomness.

8. Mechanisms that can differentially affect lexical and sublexical errors

Although a great deal of the variation in aphasic naming-error patterns is due to severity/error-type interactions that we attribute to the continuity thesis, there are differences among patients that are independent of severity. Consider the error patterns of two patients from Dell et al. (1997), shown in Table 1.

Although these two patients are at the same level of severity, 69% correct, their error distributions differ markedly. L.H. makes mostly nonword errors, whereas I.G.'s errors are predominately semantic. One can characterize L.H., a conduction aphasic, as having a deficit that leads largely to sublexical errors, and I.G., an anomic patient, as having a tendency for lexical errors. Interestingly, both make a fair number of formal errors, which accords well with the dual nature of formals; per-

Table 1
Response category proportions from two patients with equal levels of correctness in naming, but different error patterns (from Dell et al., 1997)

Patient	Response category						
	Correct	Semantic	Formal	Mixed	Unrelated	Nonword	Others
L.H.	.69	.03	.07	.01	.02	.15	.03
I.G.	.69	.09	.05	.03	.01	.02	.10

haps L.H.'s formals are sublexical and I.G.'s are lexical. More generally, it is clear that patients exhibit variation in the extent to which their errors are lexical (typically semantic) and sublexical (typically nonwords), and this variation cannot be attributed just to severity/error-type interactions (e.g., Caplan et al., 1986; Caramazza & Hillis, 1990; Caramazza et al., 2000; Cuetos, Aguado, & Caramazza, 2000; Howard & Orchard-Lisle, 1984; Laine et al., 1998; Nickels, 1995; Pate, Saffran, & Martin, 1987; Rapp & Goldrick, 2000). Consequently, models of aphasic lexical access must address this variation.

Our models explain the lexical–sublexical error distinction through their common assumption about two-step selection; they explain patient variation in lexical and sublexical errors through their assumptions about possible lesions. In the semantic–phonological model, reduction of the semantic weights promotes lexical errors and particularly semantic errors with mild lesions, while reduction of phonological weights leads to phonological errors, particularly nonwords. The weight–decay model also accounts for this variation, but less transparently. Increasing the decay rate tends to increase semantic, mixed, and formal errors, while decreasing the weight parameter increases nonword errors more than any other type (see Foygel and Dell, 2000 for formal analysis of how lesion types affect error probabilities).

9. Complex errors

Collections of aphasic naming errors include examples of complex errors, such as *unicorn* → "house," which can be analyzed as a transformation of *unicorn* to *horse*, and *horse* to "house" (Martin & Saffran, 1992). One never sees such slips in normal naming, thus raising the question of whether it is wise to characterize aphasic error patterns as augmentations of normal error tendencies. Despite their anomalous appearance, these errors are easy to understand in a two-step lexical access process. An error occurred on both steps, e.g., a semantic substitution of *horse* for *unicorn* during word retrieval, and a phonological error (here a sublexical formal error) during phonological retrieval. The chance of error on each step is sufficiently large in aphasic speakers that the double error becomes possible. We do not explicitly model the proportions of double errors; they are too unlikely to provide good modeling targets. We simply note that their occurrence in aphasia does not require the addition of assumptions to the model and that, more generally, they provide direct support for a two-step access process.

Overview of the study

This study, like our prior ones (Dell et al., 1997; Foygel & Dell, 2000), tests the assumptions of our approach against data from aphasic picture naming. The patient sample is new, and represents the largest case-series ever studied computationally. One major goal is to compare the weight–decay and semantic–phonological versions of the model for their success in fitting the naming-error profiles of the individual patients. Both versions incorporate the continuity thesis, the most important factor in fitting the error data, so both are expected to fit the individual patterns reasonably well. The semantic–phonological version produces a purer separation between lexical and sublexical errors; so to the extent that these error types dissociate in patients in the sample, the semantic–phonological model is predicted to fit the naming-error patterns better than the weight–decay model. This turns out to be the case.

Given this, a second major goal is to determine whether the semantic–phonological model generates accurate predictions regarding the characteristics of formal, semantic, and unrelated errors in the corpus overall and the data from individual patients. These predictions (e.g., grammatical category constraint in formal errors; mixed error effect) issue directly from the model's major processing assumptions (two-step selection; interactivity). Thus, successful tests of the predictions will constitute support for these processing assumptions.

Along the way, we will also be presenting evidence relating to severity/error type interactions, the characteristics of patients who are not well fit by the semantic–phonological model, and that model's ability to capture significant clinical generalizations. The general discussion will review the findings in relation to prior studies and some alternative models.

Methods

Participants

Participants were recruited mostly in the years 1997 through 2002, from inpatient and outpatient rehabilitation programs and home care settings. Some participants were identified through direct referrals from speech–language pathologists, neuropsychologists, and physicians. Others were identified through a search of the Philadelphia Cognitive Rehabilitation Research Registry, a consent-based database of research volunteers (Schwartz, Brecher, Whyte, & Klein, 2005).

Inclusion criteria were set broadly, with the aim of maximizing the diversity of the sample. We sought to enroll individuals with clinically significant post-acute or chronic aphasia, where the precipitating incident was a left hemisphere cerebrovascular accident (CVA). In cases where the clinical or CT/MRI record revealed evidence of prior left hemisphere stroke(s), we included the patient unless there was multi-focal damage throughout the left hemisphere.

Exclusion criteria were as follows: age below 18 or over 80; native or dominant language other than English; symptomatic lesions outside the left hemisphere; confounding diagnosis of traumatic brain injury, dementia, or mental illness; significant and uncorrected visual or hearing impairment; speech rendered unscorable by articulatory distortion or jargon; Aphasia Quotient above the upper limit for aphasia (as measured on the Western Aphasia Battery, WAB; Kertesz, 1982).

The final sample comprised 94 patients. The mean (and range) for age was 59 (22–86). (Due to oversight, three participants exceeded the upper age limit, at 81, 84, and 86 years.) Mean and range for years of education was 13.2 (7–24); and for months post-onset, 38.2 (1–195). Forty percent of the sample was female; 39% was African American. According to the WAB classification scheme, 33% had Broca's aphasia; 31% anomic aphasia, 19% conduction aphasia; 15% Wernicke's aphasia; and 2% had transcortical sensory aphasia.

All participants were tested under a research protocol approved by the Institutional Review Boards of Albert Einstein Healthcare Network (AEHN) and Temple University. Participants gave written informed consent and were paid for their participation. Human subjects policy at AEHN and Temple University now precludes the identification of participants by their actual initials. All identifiers used here are codes. Future reports that issue from our institution(s) that involve any of these same participants will identify them by the same codes. A few participants in this study participated in earlier published studies, where their true initials identified them. Within the restrictions of confidentiality policies, information regarding prior participation will be supplied by the authors upon request. None of the patients in this study appeared in the original computational case-series (Dell et al., 1997; Foygel & Dell, 2000).

Procedures

Philadelphia Naming Test

The Philadelphia Naming Test is a 175-item single word picture-naming test developed for collecting a large corpus of naming responses from a standardized set of items (Dell et al., 1997; Roach, Schwartz, Martin, Grewal, & Brecher, 1996). The pictured items were selected from original and published collections on the basis of their familiarity, name agreement, and good image quality (minimal complexity and confusability). Target names range in length from 1 to 4 syllables, and in noun frequency from 1 to 2110 tokens per million of printed English test (Francis & Kûcera, 1982). Length and frequency are not balanced across the items. Low-frequency targets (1–24 per million) and targets of one syllable predominate. A control group of 60 non-brain-injured, non-language-impaired, native English speakers, ranging in age from 40 to 75 years, produced the correct name on 97% of trials (Dell et al., 1997).

In this study, the Philadelphia Naming Test was administered and scored in accordance with published guidelines (Dell et al., 1997; Roach et al., 1996). Pictures were digitized and presented to participants on a Macintosh computer using MacLaboratory for Psychology experiment running software (Chute, 1990). The software controlled the stimulus presentation; trial initiation was experimenter controlled. The procedure incorporated feedback and a trial duration limit; each trial ended with the experimenter announcing the target name, and trials were terminated after 30 s. if the participant had not responded. Sessions were scored on-line by an experienced speech–language pathologist and were also audiotaped. Both sources were used to generate the final transcription of the session; disagreements were generally resolved in favor of the on-line scoring.

Scoring

As is standard for the Philadelphia Naming Test, we scored only the first complete response produced on each trial. For most subjects, a response was scored correct only if it exactly matched the designated target. An exception was made, however, for patients with clinically obvious articulatory–motor impairments. Such patients were excluded from our 1997 study because their speech distortions are sometimes difficult to distinguish from phonological errors (McNeil, Robin, & Schmidt, 1997). Here, we chose to include them and to score their responses leniently. We ignored minor distortions that were consistent for that patient; and we scored as correct responses that deviated by the addition, deletion, or substitution of a single consonant or consonant cluster. All other responses were classified into error categories, the first five of which correspond to the basic error types that the model is designed to explain (for additional details, see Roach et al., 1996):

(1) *Semantic* is a synonym of the target, or a coordinate, superordinate or subordinate member of its category. Noun associates are also included in the semantic error category (e.g., bride → *wedding*), whereas non noun associates are not; they are considered non naming responses and coded in the category *description/circumlocution* (e.g., bride → *getting married*; or *marrying*). Semantic errors were not further classified as to type of relation (e.g., category coordinate vs. associated) or whether target and error were visually, as well as semantically, related. These are certainly important matters (cf. Humphreys, Riddoch, & Quinlan, 1988; Vitkovitch, Humphreys, & Lloyd-Jones, 1993), but they are outside of the limited

scope of the model, which reduces semantic similarity relations to featural overlap.

(2) *Formal* is any word response (excluding proper nouns) that meets the Philadelphia Naming Test's phonological similarity criterion. This criterion requires that target and error start or end with the same phoneme; or have a phoneme in common at another corresponding syllable or word position, aligning words left to right; or have more than one phoneme in common in any position (excluding unstressed vowels). This criterion is deliberately liberal, allowing for further analysis of differences in the type or degree of phonological overlap that occurs across patients, targets, or errors (e.g., Gagnon et al., 1997; Schwartz et al., 2004). In this study, the only additional analysis that we performed with the formal errors examined target–error phonological similarity in all nonsemantic word errors (formals and unrelateds). This analysis, which is discussed later, did not depend on the criterion for formals because it collapsed across the formal and unrelated categories.

(3) *Mixed* is a response that meets both semantic and formal similarity criteria.

(4) *Unrelated* meets neither criterion and is not visually related to the target. Included in this category are unrelated responses that repeat a name produced earlier in the list (i.e., perseverations).

(5) *Nonword* is a neologism that is not also a blend. This includes nonwords that meet the phonological criterion (i.e., target-related nonwords) and others that do not. While distinctions among nonwords are certainly relevant to theories of lexical access (e.g., Schwartz & Brecher, 2000; Schwartz et al., 2004), they are ignored in this study, as in our prior modeling studies, for reasons of simplicity and consistency. It is worth noting, however, that more than three-quarters of the nonwords in this study (77.3%) are target related, according to the phonological similarity criterion.

The remaining responses are classified with one of three additional codes: *Description/circumlocution*; *no response*; *and miscellaneous error*. The miscellaneous category includes names of objects whose only relation to the target is visual.

Normalized response proportions

The model-fitting routine takes as input the individual subject's response proportions for correct, semantic, formal, mixed, unrelated, and nonword. Responses assigned to the remaining three codes are grouped together into what we here call "omissions." In Dell et al. (1997) and Foygel and Dell (2000), omissions were ignored and the six response proportions were calculated in relation to the total number of trials. This required the exclusion of patients who made numerous omissions, which probably biased the sample away from the more severely impaired namers (e.g., Mitchum et al., 1990; Schwartz & Brecher, 2000).

Our current approach to modeling includes all patients, even those who make many omissions. As proposed by Ruml et al. (2000), omissions are treated as events that are outside of the model and independent of it. Dell, Lawler, Harris, and Gordon (2004) compared this independence account of omissions to alternative threshold and lexical-editor accounts. The former account generates an omission if the activation of the selected word unit is below a threshold, and the latter generates an omission with a certain probability if the potential response is not lexical. The threshold and independence accounts were both successful in explaining the naming error patterns of patients who often fail to produce naming attempts. In fact, these two accounts led to almost identical predictions. The lexical-editor account fared much worse. Here, we have chosen to use the independence account rather than the threshold account because it is computationally simple. Omission responses are subtracted out, and the remaining categories are normalized, that is, they are expressed as proportions of the modified total (all responses minus omissions). The normalized proportions are then fit to the model as we describe later.

Supplementary test battery

Patients also performed a battery of language tests, aimed primarily at measuring auditory–phonological input processing, lexical–phonological processing, and lexical–semantic processing. The composition of this supplementary test battery evolved over the course of the study; patients tested early in the study did not receive all tests. Relevant results from supplementary testing will be discussed in connection with certain of the models' fits and predictions.

Fitting the models to the data

Each patient's normalized error proportions were applied to both the weight–decay and the semantic–phonological versions of the model. The fitting process consisted of assigning free parameters to the models (*weight* and *decay* for the weight–decay model, and *s* and *p* for the semantic–phonological model) to make the model's error proportions as close as possible to those of the patient. Specifically, the chosen parameters minimized the Chi Squared goodness of fit value. The search for these parameters was carried out using a pre-stored variable-resolution map of the parameter space (see Dell et al., 2004, for description of the fitting routine, and http://langprod.cogsci.uiuc.edu/cgi-bin/webfit.cgi to use the routine). The web-based version of the routine allows the user to control the percentage of trials in

which the model uses a network that has a mixed (semantic–formal) neighbor to the target. Dell et al. (1997) had used the network with a mixed neighbor 10% of the time. Dell et al. (2004), however, showed that 20% was a better value, and we use the 20% value here, effectively doubling the opportunities for mixed errors. (See Rapp and Goldrick, 2000, who also recommended using more opportunities for mixed errors than Dell et al., 1997 did.)

Results

There were 18 patients with clinically obvious articulatory–motor impairment, who qualified for the lenient scoring (17 Broca's, 1 conduction). Transcripts from one-third of these were independently scored by a second experienced speech–language pathologist. Averaging across the six transcripts, inter-rater agreement was 97.5% for correct vs. error (range 94–100%) and 92.3% for error code assigned (range 86–98%).

The 94 patients produced a total of 6757 errors, including 2377 (35%) omissions. Most of the omissions were non naming behaviors, that is, null responses and descriptions (49 and 40% of omissions, respectively). The rest fell in the miscellaneous category.

Severity interactions

Correct scores ranged from .03 to .98 both before and after normalizing for omissions. The large severity range is appropriate for testing the validity of the continuity thesis. The relevant data are shown in Fig. 2, which plots individual error scores against correct scores (all normalized) for each error type.

The phonological errors, nonwords and formals, along with unrelated errors (panels A–C), show a steep drop-off as correctness gets higher. At levels of correctness above 95%, these errors are largely nonexistent. The semantically related errors, semantic and mixed (panels D–E), do not show the same drop-off with increasing correctness. The weak relation that is present is curvilinear, these errors being more likely in moderately impaired individuals than they are in either mildly or severely impaired subjects. At very high correctness, semantic and mixed errors, unlike nonwords and formals, remain above zero. Fig. 3 shows that these semantically related errors on average constitute a large proportion of the errors in mild patients (as they do in the normal pattern), but a small proportion of errors in the severe patients (as in the random pattern). Nonword, formal, and unrelated errors, of course, show the reverse pattern. These interactions with severity are precisely in line with the continuity thesis. Very low correctness approximates the random pattern—many nonwords, formals, and unrelateds. At increasing levels of correctness, patients come to approximate the normal naming pattern of exclusively meaning-related errors, semantic and mixed.

There are exceptions to these generalizations. In particular, in Fig. 2A, the lower left quadrant represents patients who made many errors (low correctness) but few in the nonword category. We will return to these cases shortly.

Model fits

An informal comparison of the weight–decay and semantic–phonological models was done using the mean of the uncorrected root mean squared deviation[1] and total variance accounted for. Root mean squared deviation is an intuitive measure of the deviation of the model for an individual patient: a value of .03, for example, means that the predicted proportions differed from the observed proportions by an average of .03. Total variance accounted for compares the squared deviations between each observed response proportion and the across-subject mean for that proportion with the squared deviations between the observed proportions and the proportions predicted by the model. For example, a value of 94% means that the squared deviations between each model point and each data point in the study were only 6/100ths of the squared deviations between each data point and the mean of the data for that response category. The fits of both models to the individual subject data are shown in Table 2. Figs. 4 and 5 show the deviations between obtained and expected proportions for all subjects, for both models.

For the weight–decay model, mean root mean squared deviation is .034 and total variance accounted for, 87.0%. While most deviations are small (clustering around zero), Fig. 4 shows that the model has a tendency to underpredict unrelated errors and overpredict nonwords. For the semantic–phonological model, the mean root mean squared deviation is .024 and the total variance accounted for, 94.4%. As Fig. 5 shows, deviations are tightly clustered around zero and the systematic category deviations identified for weight–decay are reduced (for unrelated) or eliminated (for nonwords). Only in the mixed category are the deviations asymmetric around zero, revealing a small but consistent tendency for underprediction by the model. A prior study with many

[1] When we say that root mean squared deviation is uncorrected, we mean that it does not correct for the loss of degrees of freedom for free parameters. Such a correction would increase the reported values by a factor of 1.41. If root mean squared deviation is uncorrected, the sum of squared deviations is divided by the number of observations, here six per subject. If it is corrected, three degrees of freedom are lost, two for free parameters, and one for the fact that the six response proportions are constrained to add to 1.0.

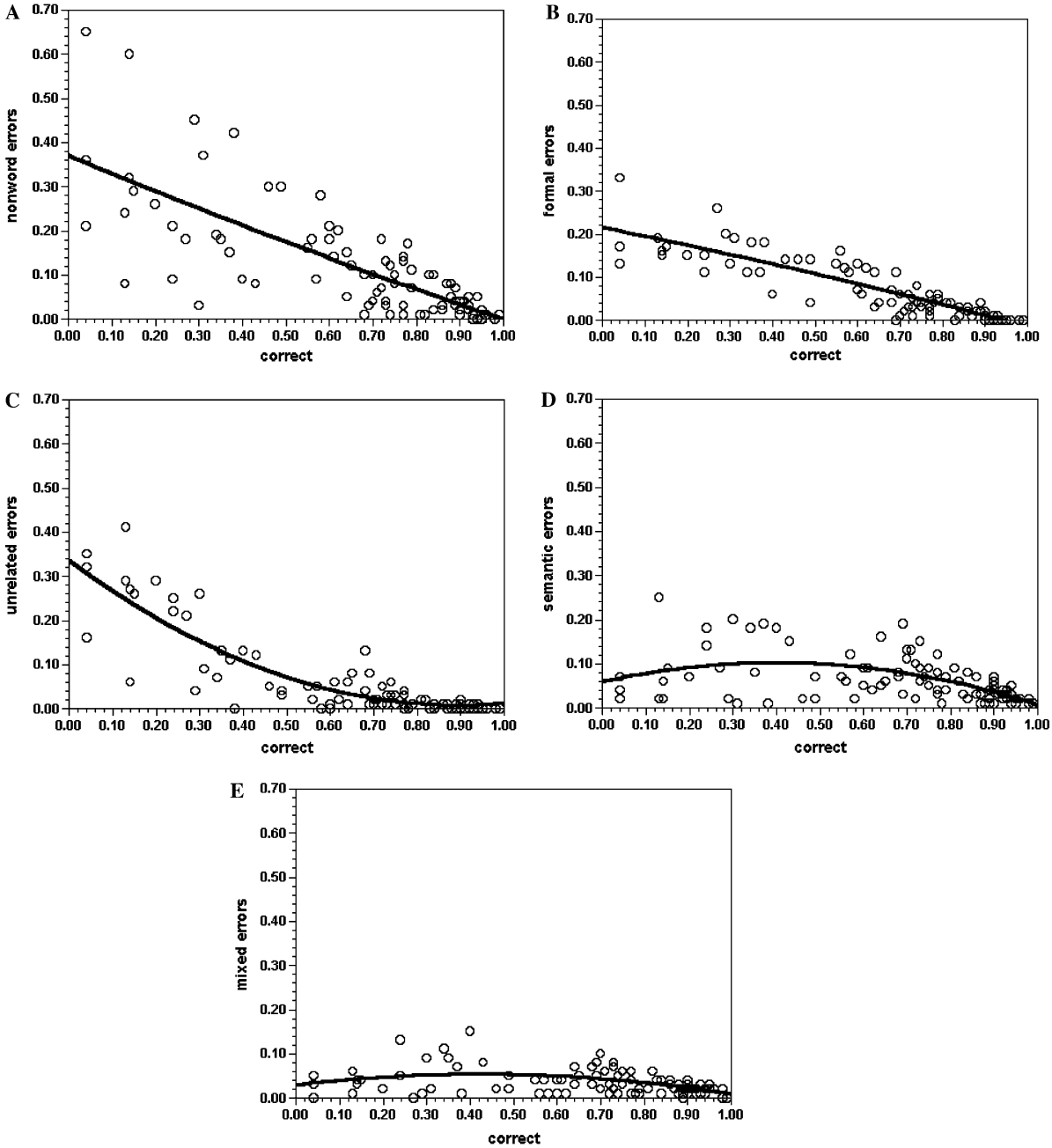

Fig. 2. Normalized response proportions for each error category are plotted against the normalized proportion correct. Each point represents one participant's data. (A) Nonwords, (B) formal errors, (C) unrelated word errors, (D) semantic errors, and (E) mixed errors.

fewer subjects also found that mixed errors were underpredicted by the semantic–phonological model (Foygel & Dell, 2000). We return to this topic in the discussion of deviating patients, below.

The weight–decay and semantic–phonological models were statistically compared using an unpaired, two-tailed t test, with root mean squared deviation as the dependent variable. The difference in means is statistically reliable ($t(df = 93) = 4.575$; $p < .0001$).

The weight–decay model has a tendency to overpredict nonwords because it is unable to fit patients who make few nonword errors but substantial numbers of word errors. The semantic–phonological model does better because the differential lesioning of s and p

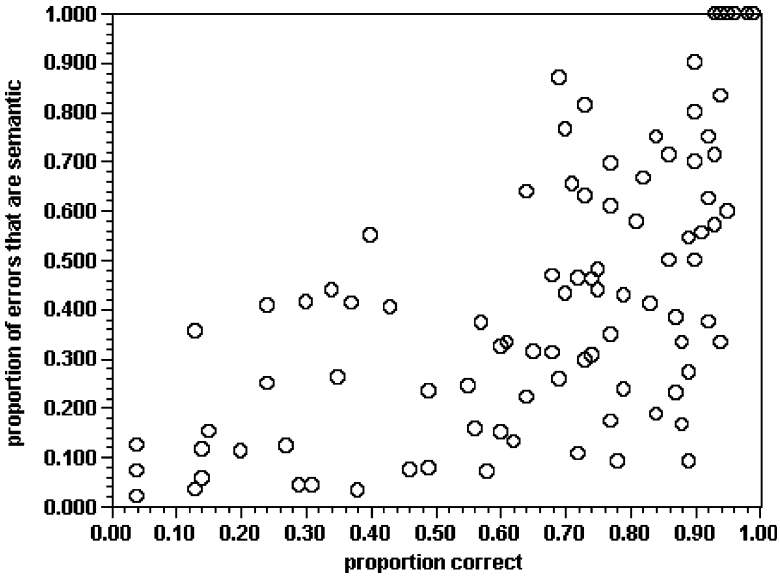

Fig. 3. The ordinate represents the errors that bear a semantic relation to the target (i.e., were classified as semantic or mixed) expressed as a proportion of total errors. The abscissa represents the proportion correct. Each point represents data from a single participant.

weights allows word and nonword errors to vary independently. This pattern in exemplified by patient NAC (second entry in Table 2), whose .26 errors represent all word categories (.07 semantic, .08 formal, .05 mixed, .06 unrelated) with only .01 nonwords. The relatively high proportion of word errors indicates, in the weight–decay model, both weakened activation transmission (lower global weight) and disrupted representational integrity (higher decay) ($w = .0452$; $d = .6896$). However, this lesion overpredicts nonwords by a factor of 10. When the model's global parameters are consistent with the production of many word errors, they necessarily lead to many nonword errors. This is particularly true for variation in the weight parameter (see Dell et al., 1997, for further characterization of the different functions of the weight and decay parameters). The semantic–phonological model accommodates the observed lexical–phonological dissociation by assigning the patient a lesion that affects semantic connection weight more than phonological connection weight ($s = .0165$; $p = .0382$). This fitted model gives better results in the nonword category (predicted .00; observed .01) and in the overall root mean squared deviations (.018, compared with .053 for the weight–decay model).

Deviating patients

It is apparent from the occasional standout lines in Fig. 5 that even the semantic–phonological model's predictions are off in some cases. To identify patients who might be considered "deviating" under the semantic–phonological model, we calculated the sum of squared deviations (ssd) from the rounded response proportions shown in Table 2 and took as deviating all those with ssd > .010, which corresponds to root mean squared deviation > .041, give or take some rounding error. This conservative criterion identified 16 patients, who happened to fall into three distinct subgroups, as described next.

Subgroup 1: The WR pattern

A subgroup of four patients exemplifies the "WR pattern," named after a patient from the original case series (Dell et al., 1997; see also Foygel & Dell, 2000). The members of this subgroup are FAH, KK, KAM, and BT. Like WR, these patients have in common that they produce many more unrelated errors than the semantic–phonological model predicts (see Fig. 6). Unrelateds, in our coding system, are whole word responses that bear no relation to the target. This can include perseverations, i.e., responses that repeat a word produced on an earlier trial. Dell et al. (1997, p. 819, footnote 8) observed that WR's high proportion of unrelateds stemmed from his strong tendency to perseverate. The same applies to the present four cases, which turn out to have the highest perseveration proportions of any patient in the sample (.13 to .21, relative to total number of trials; +2.7 to +4.7 standard deviations above the sample mean). The WR pattern of deviation, then, reflects the model's failure to account for perseverations.

Table 2
Fits of both models to the individual subject data. Each participant is represented by coded initial and aphasia subtype (A(nomic), B(roca's), C(onduction), W(ernicke's), T(rans)C(ortical)S(ensory)). Key to symbols appears below the last table entry

Source	Type	w/s	d/p	Correct	Semantic	Formal	Mixed	Unrelated	Nonword	χ^2	RMSD
QB	A										
$N = 172$											
Naming				0.94	0.01	0	0.01	0	0.05		
Weight–decay model		0.0089	0.5014	0.9	0.04	0.01	0.01	0	0.03	9.87	0.022
Semantic–phonological model		0.0581	0.0164	0.93	0.02	0	0.01	0	0.04	2.57	0.008
NAC	TCS										
$N = 168$											
Naming				0.74	0.07	0.08	0.05	0.06	0.01		
Weight–decay model		0.0452	0.6896	0.67	0.1	0.07	0.04	0.03	0.1	23.71	0.053
Semantic–phonological model		0.0165	0.0382	0.73	0.1	0.08	0.03	0.06	0	5.5	0.018
EC	B										
$N = 120$											
Naming				0.9	0.07	0.01	0.01	0.01	0.01		
Weight–decay model		0.0355	0.6257	0.89	0.05	0.02	0.02	0	0.02	2.82	0.009
Semantic–phonological model		0.0254	0.0352	0.9	0.05	0.02	0.02	0.01	0.01	1.51	0.007
KCC	A										
$N = 174$											
Naming				0.95	0.01	0	0.02	0	0.02		
Weight–decay model		0.0392	0.6296	0.92	0.04	0.01	0.02	0	0.01	7.06	0.017
Semantic–phonological model		0.0978	0.0199	0.93	0.04	0	0.01	0	0.01	4.92	0.015
MBC	A										
$N = 171$											
Naming				0.93	0.03	0.01	0.01	0.01	0.01		
Weight–decay model		0.0257	0.5852	0.92	0.04	0.01	0.01	0	0.02	1.27	0.006
Semantic–phonological model		0.029	0.0329	0.93	0.04	0.01	0.01	0	0.01	1.41	0.005
BAC	B										
$N = 175$											
Naming				0.89	0.01	0.04	0	0	0.07		
Weight–decay model		0.0076	0.5024	0.84	0.05	0.03	0.01	0.01	0.06	12.27	0.027
Semantic–phonological model		0.0355	0.0192	0.85	0.03	0.02	0.01	0	0.09	9.79	0.021
MAC	A										
$N = 156$											
Naming				0.9	0.05	0	0.03	0.02	0.01		
Weight–decay model		0.0342	0.6268	0.87	0.06	0.02	0.02	0.01	0.03	11.66	0.02
Semantic–phonological model		0.0235	0.0362	0.88	0.06	0.02	0.02	0.01	0.01	5.46	0.014
$\wedge$ *BBC*	W										
$N = 163$											
Naming				0.29	0.02	0.2	0.01	0.04	0.45		
Weight–decay model		0.003	0.5074	0.25	0.07	0.12	0.02	0.09	0.45	22.71	0.048
Semantic–phonological model		0.0186	0.0089	0.26	0.05	0.1	0.02	0.05	0.52	19.52	0.05
EBC	A										
$N = 156$											
Naming				0.95	0.02	0	0.03	0	0		
Weight–decay model		0.0922	0.6025	0.95	0.02	0	0.03	0	0	0.42	0.003
Semantic–phonological model		0.0828	0.077	0.96	0.02	0	0.02	0	0	1.25	0.006
$\wedge$ *KAC*	C										
$N = 172$											
Naming				0.38	0.01	0.18	0.01	0	0.42		
Weight–decay model		0.0034	0.5001	0.35	0.08	0.11	0.02	0.08	0.36	35.36	0.059
Semantic–phonological model		0.0318	0.0084	0.39	0.02	0.09	0.02	0.01	0.48	23.63	0.046
EAC	A										
$N = 137$											
Naming				0.7	0.11	0.06	0.02	0.01	0.1		

Table 2 (*continued*)

Source	Type	w/s	d/p	Correct	Semantic	Formal	Mixed	Unrelated	Nonword	χ^2	RMSD
Weight–decay model		0.0429	0.677	0.71	0.09	0.06	0.03	0.02	0.08	2.87	0.013
Semantic–phonological model		0.0216	0.0214	0.71	0.08	0.06	0.02	0.03	0.11	4.39	0.017
CAC	A										
$N = 163$											
Naming				0.72	0.02	0.06	0.01	0.01	0.18		
Weight–decay model		0.0056	0.501	0.68	0.07	0.06	0.02	0.03	0.14	11.98	0.034
Semantic–phonological model		0.0319	0.0155	0.72	0.03	0.04	0.01	0	0.19	2.94	0.012
MD	B										
$N = 170$											
Naming				0.93	0.04	0	0.03	0	0		
Weight–decay model		0.074	0.6958	0.93	0.03	0	0.03	0	0	0.6	0.003
Semantic–phonological model		0.0882	0.066	0.95	0.03	0	0.02	0	0	2.19	0.011
ND	B										
$N = 145$											
Naming				0.77	0.08	0.01	0.06	0.04	0.03		
Weight–decay model		0.0542	0.7156	0.71	0.1	0.06	0.04	0.02	0.07	14.87	0.037
Semantic–phonological model		0.0197	0.0291	0.77	0.09	0.06	0.02	0.04	0.03	16.22	0.024
XD	B										
$N = 123$											
Naming				0.15	0.09	0.17	0.04	0.26	0.29		
Weight–decay model		0.086	0.8953	0.18	0.09	0.18	0.05	0.11	0.4	28.19	0.075
Semantic–phonological model		0.0019	0.0165	0.15	0.12	0.2	0.02	0.22	0.29	3.64	0.024
KD	C										
$N = 174$											
Naming				0.78	0.01	0.04	0.01	0	0.17		
Weight–decay model		0.0058	0.5006	0.71	0.07	0.05	0.02	0.03	0.13	20.54	0.044
Semantic–phonological model		0.0449	0.0126	0.75	0.02	0.03	0.01	0	0.18	2.87	0.012
WAD	B										
$N = 175$											
Naming				0.98	0.02	0	0	0	0		
Weight–decay model		0.0343	0.5813	0.97	0.02	0	0.01	0	0	1.65	0.004
Semantic–phonological model		0.0433	0.0403	0.97	0.02	0	0.01	0	0	1.66	0.004
DD	B										
$N = 117$											
Naming				0.57	0.12	0.12	0.04	0.05	0.09		
Weight–decay model		0.0502	0.7152	0.58	0.11	0.1	0.04	0.04	0.13	1.84	0.017
Semantic–phonological model		0.0152	0.0228	0.57	0.11	0.11	0.03	0.09	0.11	3.6	0.018
KE	C										
$N = 171$											
Naming				0.9	0.01	0.02	0.04	0	0.04		
Weight–decay model		0.0414	0.652	0.86	0.06	0.02	0.02	0.01	0.03	9.84	0.027
Semantic–phonological model		0.0305	0.0246	0.89	0.04	0.02	0.01	0	0.04	11.8	0.016
ME	B										
$N = 172$											
Naming				0.98	0.01	0	0.02	0	0		
Weight–decay model		0.0723	0.5586	0.97	0.02	0	0.01	0	0	1.22	0.005
Semantic–phonological model		0.0557	0.0858	0.97	0.01	0	0.02	0	0	0.34	0.003
* *EE*	B										
$N = 150$											
Naming				0.73	0.15	0.01	0.07	0.01	0.03		
Weight–decay model		0.0861	0.8149	0.73	0.11	0.04	0.07	0.01	0.05	8.11	0.022
Semantic–phonological model		0.0188	0.0285	0.74	0.09	0.07	0.02	0.05	0.04	33.75	0.041

(*continued on next page*)

Table 2 (*continued*)

Source	Type	w/s	d/p	Correct	Semantic	Formal	Mixed	Unrelated	Nonword	χ^2	RMSD
TE	B										
$N = 141$											
Naming				0.6	0.09	0.07	0.04	0.01	0.18		
Weight–decay model		0.0302	0.6423	0.6	0.09	0.09	0.03	0.05	0.15	5.33	0.019
Semantic–phonological model		0.0206	0.0187	0.62	0.08	0.07	0.02	0.04	0.17	5.81	0.017
SE	C										
$N = 168$											
Naming				0.62	0.04	0.12	0.01	0.02	0.2		
Weight–decay model		0.0053	0.5075	0.59	0.08	0.08	0.02	0.05	0.19	10.79	0.029
Semantic–phonological model		0.0226	0.0167	0.61	0.06	0.07	0.02	0.02	0.22	9.92	0.025
* *CE*	B										
$N = 131$											
Naming				0.37	0.19	0.11	0.07	0.11	0.15		
Weight–decay model		0.0873	0.863	0.4	0.12	0.16	0.07	0.06	0.19	11.49	0.0422
Semantic–phonological model		0.0101	0.0222	0.41	0.13	0.16	0.03	0.15	0.13	15.34	0.042
OE	A										
$N = 155$											
Naming				0.9	0.06	0	0.03	0	0.01		
Weight–decay model		0.0562	0.6875	0.9	0.05	0.01	0.03	0	0.01	1.83	0.006
Semantic–phonological model		0.0978	0.0199	0.93	0.04	0	0.01	0	0.01	3.81	0.017
FG	C										
$N = 168$											
Naming				0.77	0.04	0.02	0.04	0.01	0.13		
Weight–decay model		0.0113	0.547	0.74	0.07	0.05	0.02	0.02	0.11	12.51	0.025
Semantic–phonological model		0.0272	0.0196	0.77	0.05	0.04	0.01	0.01	0.12	6.88	0.013
MG	A										
$N = 158$											
Naming				0.77	0.12	0.03	0.04	0.03	0.01		
Weight–decay model		0.0537	0.7097	0.74	0.1	0.05	0.04	0.01	0.06	12.45	0.027
Semantic–phonological model		0.0193	0.035	0.79	0.09	0.05	0.02	0.04	0.01	6.86	0.021
FAG	B										
$N = 147$											
Naming				0.49	0.07	0.04	0.05	0.04	0.3		
Weight–decay model		0.0278	0.6479	0.44	0.09	0.11	0.03	0.07	0.25	15.08	0.044
Semantic–phonological model		0.0199	0.0156	0.52	0.07	0.08	0.02	0.04	0.26	12.33	0.03
TG	A										
$N = 169$											
Naming				0.73	0.06	0.05	0.02	0.01	0.13		
Weight–decay model		0.0061	0.5012	0.73	0.07	0.05	0.01	0.03	0.12	1.85	0.009
Semantic–phonological model		0.0242	0.0192	0.71	0.06	0.05	0.02	0.02	0.15	0.54	0.01
# *FAH*	C										
$N = 136$											
Naming				0.04	0.04	0.33	0.03	0.35	0.21		
Weight–decay model		0.0898	0.9102	0.18	0.08	0.18	0.05	0.11	0.4	118.46	0.152
Semantic–phonological model		0.0012	0.0205	0.16	0.14	0.24	0.03	0.26	0.18	32.03	0.084
FBH	W										
$N = 174$											
Naming				0.46	0.02	0.14	0.02	0.05	0.3		
Weight–decay model		0.0071	0.5443	0.43	0.08	0.11	0.02	0.07	0.29	11.9	0.033
Semantic–phonological model		0.0184	0.0149	0.47	0.07	0.09	0.02	0.05	0.3	10.73	0.029
* *HH*	B										
$N = 122$											
Naming				0.69	0.19	0	0.08	0.01	0.03		
Weight–decay model		0.096	0.849	0.71	0.12	0.04	0.07	0.01	0.06	11.95	0.036
Semantic–phonological model		0.0155	0.0299	0.65	0.11	0.1	0.03	0.08	0.03	41.51	0.065

Table 2 (*continued*)

Source	Type	w/s	d/p	Correct	Semantic	Formal	Mixed	Unrelated	Nonword	χ^2	RMSD
NH	B										
$N = 166$											
Naming				0.89	0.02	0.04	0.01	0.01	0.04		
Weight–decay model		0.0101	0.5177	0.87	0.05	0.02	0.01	0.01	0.04	3.92	0.013
Semantic–phonological model		0.0278	0.0256	0.87	0.05	0.02	0.01	0.01	0.04	4.41	0.015
BI	W										
$N = 140$											
Naming				0.14	0.06	0.16	0.04	0.27	0.32		
Weight–decay model		0.0783	0.8685	0.17	0.08	0.18	0.04	0.11	0.42	35.95	0.078
Semantic–phonological model		0.0015	0.0161	0.14	0.12	0.2	0.02	0.22	0.31	8.22	0.035
KI	B										
$N = 106$											
Naming				0.24	0.14	0.15	0.05	0.22	0.21		
Weight–decay model		0.094	0.915	0.22	0.1	0.19	0.06	0.1	0.33	21.32	0.073
Semantic–phonological model		0.005	0.0196	0.23	0.13	0.2	0.03	0.21	0.2	2.76	0.022
SI	A										
$N = 112$											
Naming				0.68	0.08	0.04	0.07	0.13	0.01		
Weight–decay model		0.0485	0.7095	0.57	0.11	0.1	0.04	0.04	0.13	39.82	0.081
Semantic–phonological model		0.0141	0.038	0.65	0.11	0.11	0.03	0.09	0.01	16.13	0.041
* *NI*	B										
$N = 94$											
Naming				0.34	0.18	0.11	0.11	0.07	0.19		
Weight–decay model		0.0983	0.9059	0.37	0.12	0.17	0.07	0.07	0.19	6.79	0.04
Semantic–phonological model		0.0094	0.0206	0.36	0.13	0.16	0.03	0.16	0.17	29.98	0.057
FJ	W										
$N = 147$											
Naming				0.68	0.07	0.07	0.03	0.04	0.1		
Weight–decay model		0.0337	0.6493	0.67	0.09	0.07	0.03	0.03	0.11	1.41	0.012
Semantic–phonological model		0.0192	0.0225	0.67	0.09	0.07	0.02	0.05	0.1	1.97	0.011
CK	A										
$N = 170$											
Naming				0.75	0.05	0.04	0.06	0.01	0.1		
Weight–decay model		0.0924	0.8414	0.69	0.12	0.05	0.07	0.01	0.06	12.09	0.04
Semantic–phonological model		0.0225	0.0232	0.76	0.07	0.05	0.02	0.02	0.08	24.09	0.025
# *KK*	W										
$N = 122$											
Naming				0.13	0.02	0.19	0.01	0.41	0.24		
Weight–decay model		0.0783	0.8685	0.17	0.08	0.18	0.04	0.11	0.42	114.25	0.145
Semantic–phonological model		0.0013	0.0199	0.16	0.14	0.23	0.03	0.25	0.2	27.19	0.083
BAL	A										
$N = 91$											
Naming				0.81	0.09	0.04	0.02	0.02	0.01		
Weight–decay model		0.0411	0.6617	0.79	0.08	0.04	0.03	0.01	0.05	3.99	0.02
Semantic–phonological model		0.0207	0.0347	0.82	0.08	0.04	0.02	0.03	0.01	0.29	0.006
FL	W										
$N = 143$											
Naming				0.43	0.15	0.14	0.08	0.12	0.08		
Weight–decay model		0.0873	0.8638	0.4	0.12	0.16	0.07	0.06	0.19	17.95	0.054
Semantic–phonological model		0.0106	0.0251	0.45	0.13	0.15	0.03	0.15	0.08	16.79	0.029
SL	A										
$N = 169$											
Naming				0.9	0.05	0	0.02	0.01	0.02		
Weight–decay model		0.0355	0.6258	0.89	0.05	0.02	0.02	0	0.02	3.58	0.008
Semantic–phonological model		0.0284	0.0276	0.89	0.05	0.02	0.01	0	0.03	4.56	0.008

(*continued on next page*)

Table 2 (*continued*)

Source	Type	w/s	d/p	Correct	Semantic	Formal	Mixed	Unrelated	Nonword	χ^2	RMSD
EL	B										
$N = 172$											
Naming				0.84	0.02	0.03	0.01	0	0.1		
Weight–decay model		0.0069	0.5001	0.81	0.06	0.03	0.01	0.01	0.08	9.41	0.024
Semantic–phonological model		0.04	0.0165	0.83	0.02	0.02	0.01	0	0.11	1.13	0.005
BL	C										
$N = 155$											
Naming				0.65	0.06	0.04	0.05	0.08	0.12		
Weight–decay model		0.0348	0.6564	0.63	0.09	0.08	0.03	0.04	0.13	12.88	0.028
Semantic–phonological model		0.0175	0.022	0.62	0.1	0.09	0.02	0.06	0.11	12.67	0.03
* *EAL*	W										
$N = 101$											
Naming				0.3	0.2	0.13	0.09	0.26	0.03		
Weight–decay model		0.0914	0.8941	0.29	0.11	0.19	0.06	0.09	0.26	63.15	0.125
Semantic–phonological model		0.0048	0.0322	0.29	0.16	0.24	0.03	0.25	0.03	15.31	0.052
GAL	A										
$N = 174$											
Naming				0.92	0.03	0	0.02	0	0.03		
Weight–decay model		0.0283	0.597	0.91	0.04	0.01	0.01	0	0.02	6.29	0.01
Semantic–phonological model		0.0993	0.0127	0.92	0.04	0	0.01	0	0.03	1.6	0.006
UL	W										
$N = 140$											
Naming				0.61	0.09	0.06	0.04	0.06	0.14		
Weight–decay model		0.0325	0.6509	0.6	0.09	0.09	0.03	0.05	0.14	2.22	0.012
Semantic–phonological model		0.0173	0.0205	0.58	0.09	0.09	0.02	0.06	0.14	4.1	0.017
# *KAM*	W										
$N = 151$											
Naming				0.04	0.07	0.17	0.05	0.32	0.36		
Weight–decay model		0.0741	0.8566	0.15	0.08	0.17	0.04	0.11	0.45	72.21	0.103
Semantic–phonological model		0.0011	0.0141	0.12	0.11	0.18	0.02	0.2	0.37	24.97	0.061
SM	A										
$N = 165$											
Naming				0.9	0.04	0.01	0.03	0	0.01		
Weight–decay model		0.0514	0.6747	0.89	0.05	0.01	0.03	0	0.01	0.87	0.006
Semantic–phonological model		0.0263	0.0348	0.91	0.05	0.02	0.01	0.01	0.01	4.53	0.008
FAM	B										
$N = 163$											
Naming				0.89	0.04	0.02	0.02	0	0.03		
Weight–decay model		0.0359	0.6305	0.88	0.06	0.02	0.02	0	0.02	2.53	0.01
Semantic–phonological model		0.0285	0.026	0.88	0.05	0.02	0.01	0	0.04	2.95	0.007
SAM	A										
$N = 166$											
Naming				0.86	0.07	0.02	0.03	0.01	0.02		
Weight–decay model		0.051	0.6843	0.85	0.07	0.02	0.03	0	0.02	0.42	0.003
Semantic–phonological model		0.0245	0.0305	0.87	0.06	0.03	0.02	0.01	0.02	2.38	0.008
FM	C										
$N = 153$											
Naming				0.49	0.02	0.14	0.02	0.03	0.3		
Weight–decay model		0.0041	0.5007	0.46	0.08	0.1	0.02	0.07	0.28	12.66	0.037
Semantic–phonological model		0.0212	0.0142	0.5	0.06	0.08	0.02	0.03	0.3	10.21	0.029
DN	A										
$N = 114$											
Naming				0.82	0.06	0.04	0.06	0.02	0.01		
Weight–decay model		0.061	0.7283	0.78	0.09	0.04	0.04	0.01	0.04	6.51	0.023
Semantic–phonological model		0.0207	0.0347	0.82	0.08	0.04	0.02	0.03	0.01	10.2	0.019

Table 2 (*continued*)

Source	Type	w/s	d/p	Correct	Semantic	Formal	Mixed	Unrelated	Nonword	χ^2	RMSD
HN	C										
$N = 164$											
Naming				0.35	0.08	0.18	0.09	0.13	0.18		
Weight–decay model		0.0905	0.8848	0.33	0.12	0.18	0.06	0.08	0.23	9.95	0.034
Semantic–phonological model		0.0094	0.0206	0.36	0.13	0.16	0.03	0.16	0.17	24.95	0.035
HAN	A										
$N = 143$											
Naming				0.69	0.03	0.11	0.05	0.08	0.03		
Weight–decay model		0.0458	0.696	0.62	0.1	0.09	0.04	0.04	0.12	22.95	0.056
Semantic–phonological model		0.0155	0.0298	0.65	0.11	0.1	0.03	0.08	0.03	11.44	0.038
* *CBN*	A										
$N = 88$											
Naming				0.4	0.18	0.06	0.15	0.13	0.09		
Weight–decay model		0.0976	0.9	0.4	0.13	0.16	0.07	0.06	0.18	25.23	0.073
Semantic–phonological model		0.0088	0.0273	0.41	0.14	0.18	0.03	0.18	0.06	50.32	0.075
∧ *DAN*	C										
$N = 142$											
Naming				0.31	0.01	0.19	0.02	0.09	0.37		
Weight–decay model		0.0311	0.6763	0.29	0.09	0.14	0.03	0.09	0.37	12.26	0.039
Semantic–phonological model		0.0113	0.0134	0.28	0.09	0.13	0.02	0.11	0.38	13.75	0.041
NN	A										
$N = 174$											
Naming				0.94	0.03	0	0.02	0.01	0		
Weight–decay model		0.0369	0.6257	0.91	0.05	0.01	0.02	0	0.01	7.19	0.016
Semantic–phonological model		0.0261	0.0554	0.93	0.03	0.01	0.02	0	0	3.37	0.007
HO	C										
$N = 168$											
Naming				0.79	0.04	0.06	0.01	0.01	0.11		
Weight–decay model		0.0063	0.5001	0.76	0.06	0.05	0.01	0.02	0.1	5.61	0.019
Semantic–phonological model		0.027	0.0195	0.76	0.05	0.04	0.01	0.01	0.12	3.93	0.016
MO	B										
$N = 167$											
Naming				0.87	0.01	0.01	0.02	0.01	0.08		
Weight–decay model		0.0078	0.5045	0.84	0.05	0.03	0.01	0.01	0.06	10.96	0.025
Semantic–phonological model		0.0293	0.0216	0.83	0.04	0.02	0.01	0	0.08	8.88	0.022
BQ	B										
$N = 138$											
Naming				0.73	0.09	0.03	0.08	0.03	0.04		
Weight–decay model		0.0653	0.753	0.7	0.11	0.06	0.05	0.01	0.07	7.97	0.025
Semantic–phonological model		0.0188	0.0285	0.74	0.09	0.07	0.02	0.05	0.04	24.61	0.029
NQ	B										
$N = 147$											
Naming				0.74	0.07	0.04	0.01	0.03	0.12		
Weight–decay model		0.0069	0.5102	0.73	0.07	0.05	0.01	0.02	0.11	0.93	0.006
Semantic–phonological model		0.0216	0.0214	0.71	0.08	0.06	0.02	0.03	0.11	2.25	0.016
XQ	C										
$N = 172$											
Naming				0.88	0.01	0.02	0.01	0	0.08		
Weight–decay model		0.0076	0.5006	0.85	0.05	0.03	0.01	0.01	0.05	11.25	0.025
Semantic–phonological model		0.0458	0.0164	0.87	0.02	0.01	0.01	0	0.08	2.06	0.007
MQ	C										
$N = 166$											
Naming				0.58	0.02	0.11	0.01	0	0.28		
Weight–decay model		0.0051	0.5103	0.53	0.08	0.09	0.02	0.05	0.23	20.92	0.046
Semantic–phonological model		0.0336	0.0113	0.56	0.02	0.06	0.02	0.01	0.33	7.69	0.027

(*continued on next page*)

Table 2 (*continued*)

Source	Type	w/s	d/p	Correct	Semantic	Formal	Mixed	Unrelated	Nonword	χ^2	RMSD
FQ	B										
$N = 146$											
Naming				0.71	0.13	0.02	0.06	0.02	0.06		
Weight–decay model		0.076	0.788	0.7	0.11	0.05	0.06	0.01	0.06	4.51	0.0157
Semantic–phonological model		0.0203	0.0254	0.74	0.09	0.06	0.02	0.04	0.06	19.74	0.033
* *SS*	TCS										
$N = 89$											
Naming				0.24	0.18	0.11	0.13	0.25	0.09		
Weight–decay model		0.0996	0.928	0.26	0.11	0.19	0.06	0.09	0.29	49.96	0.116
Semantic–phonological model		0.004	0.0266	0.25	0.16	0.24	0.03	0.25	0.07	35.96	0.067
HS	B										
$N = 157$											
Naming				0.64	0.05	0.11	0.03	0.01	0.15		
Weight–decay model		0.0288	0.636	0.62	0.09	0.08	0.03	0.04	0.14	7.62	0.025
Semantic–phonological model		0.0206	0.0187	0.62	0.08	0.07	0.02	0.04	0.17	8.12	0.024
BS	W										
$N = 109$											
Naming				0.64	0.16	0.03	0.07	0.06	0.05		
Weight–decay model		0.0661	0.7637	0.64	0.11	0.08	0.05	0.02	0.09	12.77	0.037
Semantic–phonological model		0.015	0.0293	0.64	0.12	0.1	0.03	0.08	0.04	18.14	0.042
FT	C										
$N = 167$											
Naming				0.14	0.02	0.15	0.03	0.06	0.6		
Weight–decay model		0.0928	0.9427	0.11	0.06	0.15	0.03	0.11	0.55	10.58	0.036
Semantic–phonological model		0.0149	0.0066	0.16	0.05	0.1	0.02	0.06	0.62	9.53	0.026
ST	C										
$N = 169$											
Naming				0.87	0.03	0.01	0.02	0	0.08		
Weight–decay model		0.0076	0.5006	0.85	0.05	0.03	0.01	0.01	0.05	8.09	0.017
Semantic–phonological model		0.0769	0.0087	0.86	0.03	0	0.01	0	0.09	1	0.007
KT	B										
$N = 167$											
Naming				0.96	0.02	0	0.02	0	0.01		
Weight–decay model		0.0487	0.6418	0.95	0.03	0	0.02	0	0	1.84	0.006
Semantic–phonological model		0.0922	0.0389	0.94	0.04	0	0.01	0	0.01	1.89	0.01
ET	W										
$N = 165$											
Naming				0.55	0.07	0.13	0.04	0.05	0.16		
Weight–decay model		0.046	0.7047	0.53	0.1	0.11	0.04	0.05	0.16	2.38	0.017
Semantic–phonological model		0.0164	0.0193	0.54	0.1	0.1	0.02	0.07	0.17	6.82	0.022
# *BT*	B										
$N = 127$											
Naming				0.2	0.07	0.15	0.02	0.29	0.26		
Weight–decay model		0.086	0.8953	0.18	0.09	0.18	0.05	0.11	0.4	45.15	0.094
Semantic–phonological model		0.0048	0.0176	0.21	0.12	0.19	0.03	0.2	0.26	8.84	0.045
CT	B										
$N = 160$											
Naming				0.86	0.03	0.03	0.04	0.01	0.03		
Weight–decay model		0.0422	0.6571	0.85	0.07	0.03	0.02	0.01	0.03	5.22	0.017
Semantic–phonological model		0.0247	0.0288	0.86	0.06	0.03	0.02	0.01	0.03	7.12	0.015
TT	B										
$N = 142$											
Naming				0.56	0.06	0.16	0.01	0.02	0.18		
Weight–decay model		0.035	0.6659	0.53	0.1	0.11	0.03	0.06	0.18	11.34	0.035
Semantic–phonological model		0.017	0.0189	0.54	0.09	0.1	0.02	0.06	0.18	12.71	0.036

Table 2 (*continued*)

Source	Type	w/s	d/p	Correct	Semantic	Formal	Mixed	Unrelated	Nonword	χ^2	RMSD
TAT	A										
$N = 171$											
Naming				0.94	0.05	0	0.01	0.01	0		
Weight–decay model		0.0306	0.6026	0.92	0.04	0.01	0.01	0	0.01	6.36	0.011
Semantic–phonological model		0.027	0.0406	0.93	0.04	0.01	0.01	0	0	3.77	0.008
KAT	A										
$N = 171$											
Naming				0.92	0.04	0.01	0.02	0	0.02		
Weight–decay model		0.0339	0.6137	0.92	0.04	0.01	0.01	0	0.01	1.22	0.004
Semantic–phonological model		0.0313	0.0279	0.92	0.04	0.01	0.01	0	0.02	1.29	0.004
BAT	A										
$N = 97$											
Naming				0.27	0.09	0.26	0	0.21	0.18		
Weight–decay model		0.0949	0.9157	0.24	0.1	0.2	0.06	0.1	0.31	24.79	0.079
Semantic–phonological model		0.005	0.0196	0.23	0.13	0.2	0.03	0.21	0.2	6.48	0.036
MT	A										
$N = 166$											
Naming				0.91	0.03	0	0.02	0	0.04		
Weight–decay model		0.0147	0.5419	0.9	0.04	0.02	0.01	0	0.03	6.06	0.011
Semantic–phonological model		0.0942	0.0096	0.91	0.04	0	0.01	0	0.04	0.57	0.004
NU	A										
$N = 170$											
Naming				0.92	0.02	0	0.01	0.01	0.05		
Weight–decay model		0.0085	0.5003	0.89	0.04	0.02	0.01	0.01	0.04	4.89	0.015
Semantic–phonological model		0.0305	0.0246	0.89	0.04	0.02	0.01	0	0.04	5.6	0.016
FU	W										
$N = 112$											
Naming				0.04	0.02	0.13	0	0.16	0.65		
Weight–decay model		0.0583	0.832	0.07	0.05	0.11	0.02	0.1	0.66	10.27	0.033
Semantic–phonological model		0.0013	0.007	0.07	0.06	0.11	0.01	0.12	0.63	8.06	0.03
EW	A										
$N = 172$											
Naming				0.88	0.01	0.02	0.03	0	0.05		
Weight–decay model		0.0271	0.6036	0.85	0.06	0.03	0.02	0.01	0.04	10.27	0.023
Semantic–phonological model		0.031	0.0235	0.88	0.04	0.02	0.01	0	0.05	8.23	0.013
BW	B										
$N = 174$											
Naming				0.93	0.02	0.01	0.03	0.01	0.01		
Weight–decay model		0.0464	0.6567	0.91	0.05	0.01	0.02	0	0.01	5.22	0.014
Semantic–phonological model		0.029	0.0329	0.93	0.04	0.01	0.01	0	0.01	6.31	0.01
HW	B										
$N = 174$											
Naming				0.99	0.01	0	0	0	0.01		
Weight–decay model		0.0136	0.5005	0.98	0.02	0	0	0	0	2.42	0.007
Semantic–phonological model		0.0499	0.0293	0.97	0.02	0	0.01	0	0	2.97	0.01
* *KX*	W										
$N = 132$											
Naming				0.7	0.13	0.01	0.1	0.02	0.04		
Weight–decay model		0.0846	0.8164	0.7	0.12	0.05	0.07	0.01	0.06	10.5	0.025
Semantic–phonological model		0.017	0.0281	0.69	0.1	0.08	0.02	0.06	0.04	43.86	0.048
SAX	A										
$N = 159$											
Naming				0.84	0.08	0.01	0.04	0.01	0.02		
Weight–decay model		0.0545	0.6998	0.83	0.08	0.03	0.03	0	0.03	3.65	0.009
Semantic–phonological model		0.0233	0.0311	0.85	0.07	0.03	0.02	0.01	0.02	5.57	0.012

(*continued on next page*)

Table 2 (*continued*)

Source	Type	w/s	d/p	Correct	Semantic	Formal	Mixed	Unrelated	Nonword	χ^2	RMSD
MX	B										
$N = 157$											
Naming				0.83	0.03	0	0.04	0	0.1		
Weight–decay model		0.0256	0.6078	0.79	0.07	0.04	0.02	0.02	0.07	17.75	0.033
Semantic–phonological model		0.1	0.0056	0.84	0.04	0	0.02	0	0.1	4.16	0.01
DX	C										
$N = 171$											
Naming				0.6	0.05	0.13	0.01	0	0.21		
Weight–decay model		0.0049	0.5023	0.57	0.08	0.08	0.02	0.05	0.2	16.25	0.033
Semantic–phonological model		0.0219	0.0164	0.59	0.07	0.07	0.02	0.03	0.23	14.74	0.028
QX	B										
$N = 165$											
Naming				0.79	0.07	0.05	0.02	0.01	0.07		
Weight–decay model		0.0282	0.6169	0.79	0.07	0.04	0.02	0.01	0.07	1.1	0.005
Semantic–phonological model		0.0237	0.0237	0.79	0.07	0.04	0.02	0.02	0.07	1.26	0.005
NX	C										
$N = 131$											
Naming				0.72	0.1	0.03	0.03	0.05	0.07		
Weight–decay model		0.0331	0.6444	0.7	0.09	0.06	0.03	0.03	0.1	6.42	0.023
Semantic–phonological model		0.0203	0.0254	0.74	0.09	0.06	0.02	0.04	0.06	3.92	0.018
KAX	C										
$N = 157$											
Naming				0.77	0.03	0.06	0.01	0	0.14		
Weight–decay model		0.0063	0.5033	0.74	0.07	0.05	0.01	0.02	0.11	10.14	0.027
Semantic–phonological model		0.0335	0.016	0.75	0.03	0.04	0.01	0	0.17	3.48	0.016
* *KBX*	W										
$N = 48$											
Naming				0.13	0.25	0.19	0.06	0.29	0.08		
Weight–decay model		0.0989	0.9328	0.22	0.1	0.2	0.06	0.1	0.33	38.7	0.146
Semantic–phonological model		0.0012	0.0247	0.18	0.15	0.26	0.03	0.28	0.1	6.29	0.055
DAX	A										
$N = 158$											
Naming				0.75	0.09	0.03	0.03	0.03	0.08		
Weight–decay model		0.033	0.6339	0.75	0.08	0.05	0.02	0.02	0.08	3.83	0.013
Semantic–phonological model		0.0215	0.0245	0.76	0.08	0.05	0.02	0.03	0.07	4.06	0.015

Symbols appearing to the left of the coded initials represent patterns of deviation from the semantic–phonological model (see text): #, WR pattern; *, JF pattern; ∧, others. The first row of numbers for each particant shows the observed response proportions; subsequent rows show the parameter values assigned by the best fitting WD (weight–decay) and SP (semantic–phonological) models and the response proportions predicted by each. Default (control) value is .6 for parameter *d*, and .1 for parameters *w*, *s*, and *p*. RMSD, root mean squared deviation.

Subgroup 2: The JF pattern

The larger deviating subgroup ($N = 9$; Fig. 7) shows the "JF pattern," named after the patient in our earlier case series who exemplified this pattern and who produced the worst fit to the semantic–phonological model (Foygel & Dell, 2000). The members of this subgroup are EE, CE, HH, NI, EAL, CBN, SS, KX, and KBX. The JF-pattern of deviation reflects the difficulty that the semantic–phonological model (as well as the weight–decay model) has in fitting an extreme dissociation featuring low correctness and an error pattern that is purely semantic (i.e., all or nearly all errors falling in the semantic and mixed categories). There are two aspects to the problem. The first concerns the combination of low correctness and few nonwords. In presenting the severity plot for nonwords (Fig. 2A), we mentioned this in connection with the continuity thesis, which associates low correctness with the random pattern, implying many nonwords. It turns out that the data points in the problematic lower left quadrant of Fig. 2A mostly represent JF-pattern patients. As we discussed earlier in connection with patient NAC, the semantic–phonological model's ability to restrict the connection weight lesion to just the semantic connections gives it an advantage

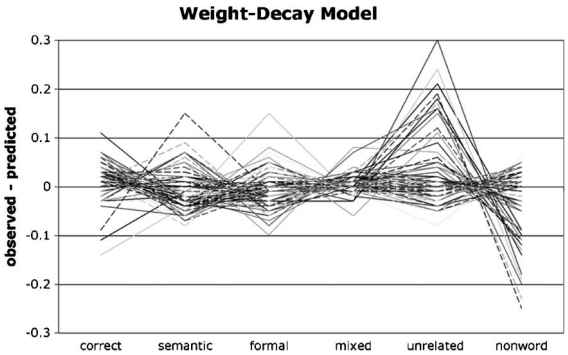

Fig. 4. The deviations between observed and expected proportions for the weight–decay model. Each line represents data from a single patient.

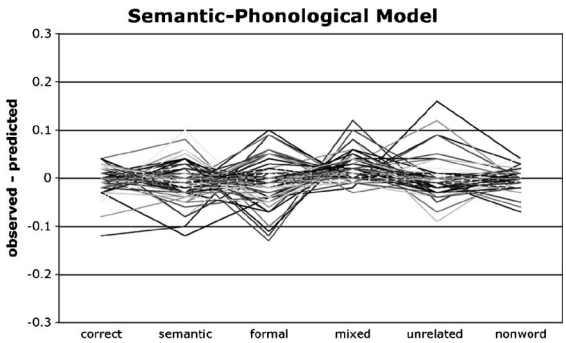

Fig. 5. The deviations between observed and expected proportions for the semantic–phonological model. Each line represents data from a single patient.

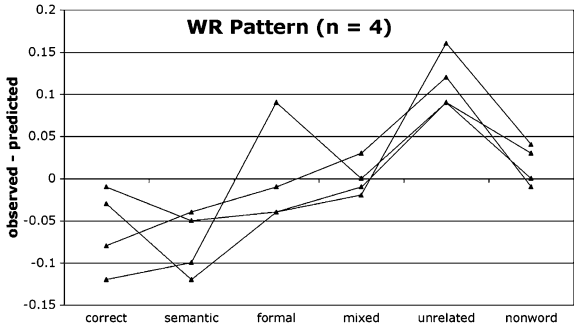

Fig. 6. The deviations between observed and expected proportions for the patients who exhibited the WR-pattern, relative to the predictions of the semantic–phonological model.

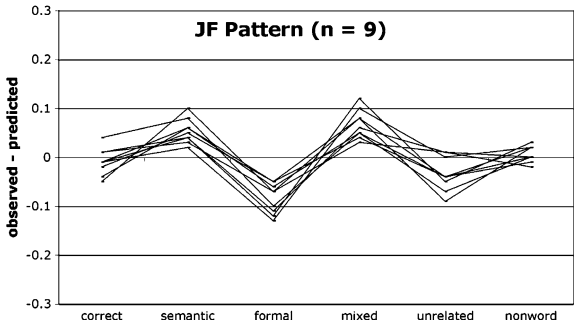

Fig. 7. The deviations between observed and expected proportions for the patients who exhibited the JF-pattern, relative to the predictions of the semantic–phonological model.

over the weight–decay model in accommodating an error pattern with many word errors but few nonwords. Inspection of Table 2 shows that for all the JF-pattern patients except NI, the best-fitting weight–decay lesion overpredicts nonword errors, and the semantic–phonological model improves the fit to this category by assigning a lesion that is almost purely semantic. But here we come to the second aspect of the problem: A semantic lesion severe enough to fit the low correctness tends to predict too many formals and unrelateds to make for a good fit. This explains the pattern of deviation shown in Fig. 7: the best fitting semantic–phonological model underpredicts semantic and mixed errors while overpredicting formals and, to a lesser extent, unrelateds.

Earlier we commented on the semantic–phonological model's tendency to underpredict mixed errors. This turns out to be largely due to rates of mixed errors in the JF-pattern patients, for whom the observed–expected deviations for mixed range from .03 to .12, with observed values up to 5 times the predicted values. The mixed-error deviations are particularly large (compared to semantic) because the low opportunities for mixed errors make the semantic–phonological model unable to predict a rate greater than .03. By sampling from the network with a mixed neighbor 20% of the time, we effectively doubled the number of mixed-error opportunities in this, compared to prior studies (Dell et al., 1997; Foygel & Dell, 2000). However, the opportunities are still low (about 0.8%), and they offer little flexibility to match the relatively high mixed-error proportions in these patients. There is more flexibility for predicting semantic errors, and hence the deviations expressed in proportional terms for this category are not as dramatic as those for the mixed errors (underpredicted by a factor of 2–3 for the semantic category compared to as much as a factor of five for the mixed).

The difficulty that the pure semantics patients pose for both the weight–decay and semantic–phonological models has been discussed before (Dell et al., 1997; Foygel & Dell, 2000) and has garnered these patients considerable attention in discussions of the pros and cons of these models. Dell et al. (1997, p. 832) noted that the pure semantics pattern tends to co-occur with high rates of omissions; and that is clearly the case in the present

sample (Fig. 8). The significance of this lies in the possibility that some or all of these omissions may in reality be suppressed nonwords or other semantically unrelated errors (i.e., that more types of error are generated than are emitted). This would imply that our treatment of omissions is wrong, or at least incomplete, in failing to recognize incipient error suppression as a mechanism for omissions (for evidence to this effect, see Schwartz and Brecher, 2000 and Mitchum et al., 1990; and for evidence of an association between omissions and semantic impairment in a different case-series investigation, Lambon Ralph et al., 2002).

The masking of other errors by omissions may account for the pure semantics pattern in some patients, but it cannot be the whole story, because patients have been described who exhibit the pure semantics pattern in conjunction with few omissions. In one well-studied patient, language testing revealed the presence of a central semantic deficit, suggesting that naming errors arose from faulty semantic input to lexical access, and not from lexical access itself (Hillis & Caramazza, 1995; Hillis, Rapp, Romani, & Caramazza, 1990; Rapp & Goldrick, 2000). For the current set of pure semantics patients, we have limited background testing that speaks to this issue (Table 3); and it appears that four of the nine are candidates for a central semantic deficit, as indexed by low standard scores on these receptive tests for lexical- and conceptual–semantic processing. These four are identified in the table by bold.

Patients HH, KX, and EE do not score in a range that we would consider indicative of a central semantic deficit; and their production of omissions is not unusually high (52, 41, and 21 omissions, respectively, corresponding to z scores of +1.13, +0.69, and +0.01). These three are reminiscent of two other patients who exhibit the pure semantics pattern in naming in conjunction with preserved central semantics and a low proportion of omissions (PW, reported in Rapp & Caramazza, 1997, 1998; Rapp & Goldrick, 2000), and DP, reported in Cuetos et al., 2000. The challenge these patients pose for the model will be taken up in the general discussion.

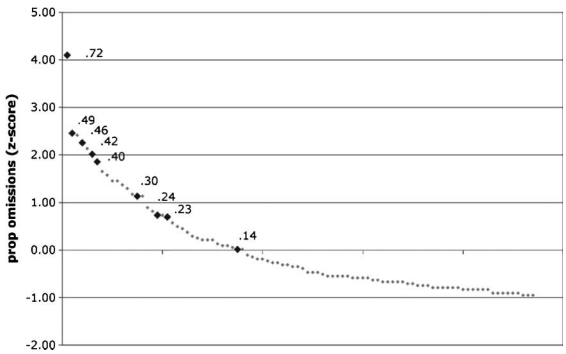

Fig. 8. The ordinate represents omissions proportional to total responses, standardized against the current sample ($n = 94$). For the JF-pattern patients (represented by darker symbols) the raw proportion is given alongside the respective symbol.

Subgroup 3: Underprediction of formals

The three patients represented in Fig. 9 (BBC, KAC, and DAN) are only mildly deviating ($\leqslant$.10 deviation for each category of error; root mean squared deviation $\leqslant$ .05), but there is some consistency in the patients' response proportions, as well as in the model's deviations. Each has low correctness in conjunction with a predominantly phonological error pattern (nonwords and formals). However, the proportion of formals they produce is high, relative to the other error types, so that the assigned lesion undershoots the mark. Because the deviations here are small and there are only three such patients, their apparent systematicity may be illusory. We will return to these cases in the general discussion.

Continuity thesis as instantiated in the semantic–phonological model

We stated above that the semantic–phonological model does a better job of fitting patients with low correctness and few nonwords, whom it characterizes as having semantic connections weak enough to seriously compromise word-level selection and phonological connections strong enough to ensure that the selected words are phonologically encoded. From this it follows that the semantic–phonological model also accommodates a severe deficit in which response proportions are different from the random pattern, which features nonwords as the predominant category.

To further illustrate this characteristic of the semantic–phonological model, we conducted an analysis, based on Ruml et al. (2005), in which we measured, for each patient, the difference between the response proportions predicted by the best-fitting semantic–phonological model and the response proportions at the random point. The difference was expressed using root mean squared deviation, with larger values corresponding to larger differences. The results are plotted in Fig. 10A, with correctness on the x-axis and deviation from the random point (root mean squared deviation among proportions) on the y-axis. Next, we performed the same analysis, substituting the observed proportions for each patient for the proportions predicted by the model. The resulting plot, shown in Fig. 10B, is almost identical to what the model produced: for levels of correctness between .50 and 1.0, the difference from random gets smaller as correctness gets lower, instantiating continuity; however, below .50, the points spread out somewhat. Ruml et al. observed the same spread of points at low correctness in their case series of Italian aphasics and concluded that the continuity thesis is incorrect. Our view, in contrast, is that the data require a weaker

Table 3
Scores for the JF-pattern patients on background tests of semantic and lexical–semantic processing

ID	Subtype	Picture–name verification—semantic foils[a]	Noun–verb synonymy judgments[b]	Pyramids and palm trees test[c]–all pictures version	
		z score	*z* score	Raw score	*z* score
CE	**Broca**	**−0.57**	**−0.69**	**73**	**−2.27**
EE	Broca	0.49	0.11	NT	
KX	Wernicke	−0.25	0.11	NT	
EAL	Wernicke	NT	−1.04	NT	
KBX	**Wernicke**	**−2.7**	**NT**	**NT**	
NI	**Broca**	**−2.81**	**−1.21**	**NT**	
HH	Broca	0.12	−0.12	88	−0.22
SS	**TCS**	**−1.42**	**−2.18**	**73**	**−2.27**
CBN	Anomic	NT	−1.04	NT	

[a] Each of the 175 pictures from the Philadelphia Naming Test appears three times, pseudo-randomly, with the target, a close semantic foil, and a remote semantic foil. The task is to indicate whether the spoken stimulus names the picture.
[b] A subtest of the *Philadelphia Comprehension Battery*; Saffran et al., 1988 ($n = 30$ test items), this task requires a decision about which two of three nouns or verbs (written and spoken) are most similar in meaning.
[c] A published test of semantic-conceptual knowledge, involving picture match-to-sample based on an associative relationship (Howard & Patterson, 1992).

version of the continuity thesis, and the semantic–phonological-model instantiates that version: To the extent that phonological lesions are present, the data and the model show the gradual transformation from normal to the random point, defined as before in terms of random selection of phonologically legal strings. The lower edge of the scatter of points represents this transformation. In the absence of phonological lesions, however, the transformation from normal to random is to a different random point, one associated with the *correct* production of a randomly selected *word*. The points with low proportion correct but greater deviation from the (phonologically defined) random point are for patients whose phonological lesions are not particularly strong. The more phonologically intact these patients are, the greater the upward sweep from the bottom edge of the scatter.

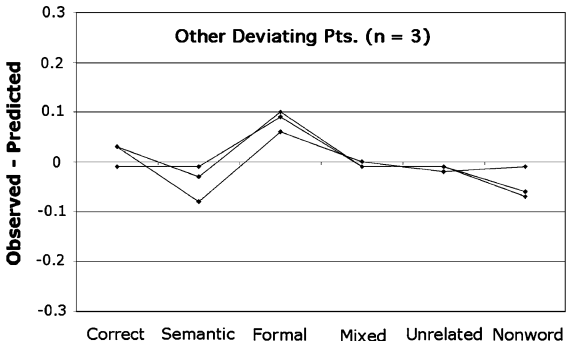

Fig. 9. The deviations between observed and expected proportions for the remaining patients who were considered deviating, relative to the predictions of the semantic–phonological model.

Aphasia subtypes in model space

It is well known that all clinical subtypes of aphasia are accompanied by naming difficulties, and that particular error types are not unique to any one subtype (e.g., Howard, Patterson, Franklin, Orchard-Lisle, & Morton, 1984; Kohn & Goodglass, 1985; Mitchum et al., 1990). To determine whether the semantic–phonological model's analysis distinguishes the different groups on the basis of naming responses proportions, we examined how patients from different groups cluster in semantic–phonological parameter space (Foygel & Dell, 2000). The analysis was restricted to patients with less than .90 correct responses, because the model needs a fair number of errors to reliably determine whether a lesion is primarily semantic or primarily phonological. Results are shown in Fig. 11.

The patients with conduction aphasia fall in the part of space with lower phonological weights than semantic weights. Clinically, this subtype is defined by poor repetition, frequent phonemic paraphasias in speech, and relatively good comprehension; the low phonological weights may explain the first two symptoms and the relatively high semantic weights may explain the latter. (This assumes that the same semantic–lexical connections are utilized in comprehension as in production; e.g., Roelofs, 2003; Dell, Martin, and Schwartz, submitted.) The patients with Wernicke's aphasia tend to have lower semantic than phonological weights. Clinically, this subtype is defined by poor lexical comprehension; the low semantic weights would explain this. The model's characterization of the anomic patients is like that of the Wernicke's patients, but with higher semantic and phonological weights. This is consistent with the

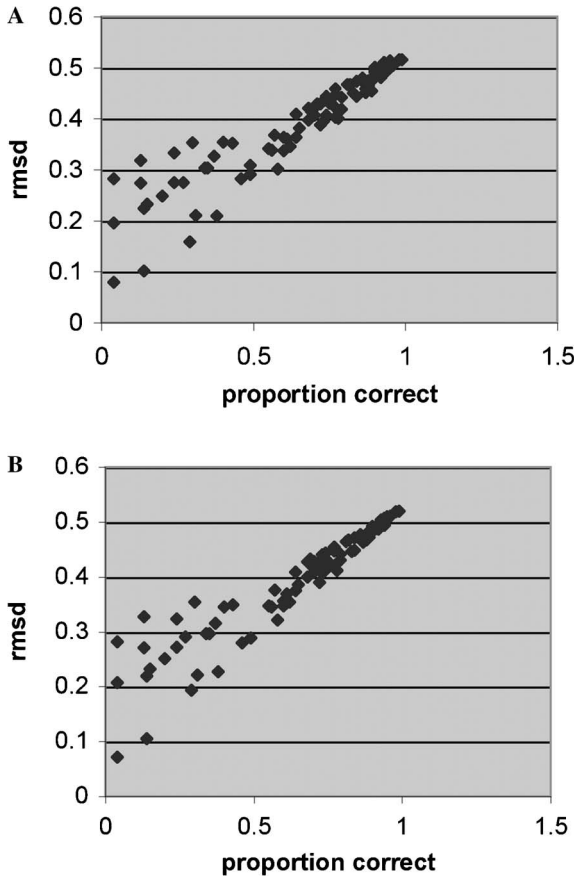

Fig. 10. The top graph (A) shows the difference between the semantic–phonological model's predicted response proportions and the response proportions at the random point (expressed as the root mean squared deviation) in relation to the subject's naming accuracy. The bottom graph (B) shows the difference between the subject's actual response proportions and those at the random point.

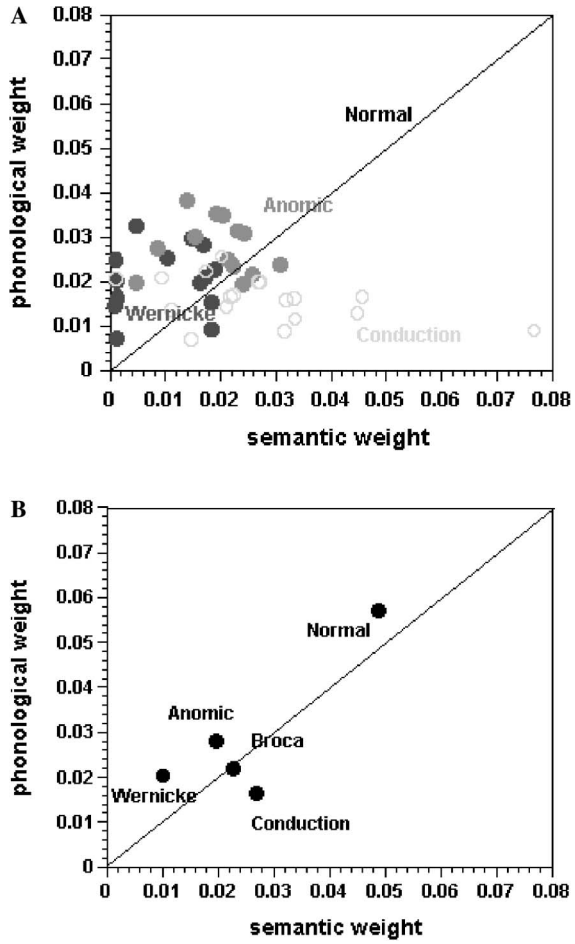

Fig. 11. This figure shows how the classical aphasia subtypes cluster in semantic–phonological parameter space. The top panel graphs parameters for individual patients; the bottom panel graphs average parameters for patients of the respective subtype.

clinical evidence that Wernicke patients often recover to an anomic profile (Crary & Kertesz, 1998). Indeed, in Fig. 11B, the alignment of Wernicke, Anomic, and non-aphasic speakers along a diagonal indicates that the three categories lie along a continuum of severity. The semantic–phonological model's analysis of the Broca speakers is not particularly distinctive. The mean semantic and phonological weight for patients in this category lies almost exactly between the Anomic and Conduction groups (Fig. 11B). Interestingly, Broca's aphasia is not defined by symptoms that bear on the integrity of semantics or phonology; its hallmark symptoms are effortful, poorly articulated speech and agrammatism.

In conclusion, the model's analysis of picture naming clusters Wernicke, Anomic, and Conduction speakers in different nearby regions of semantic–phonological parameter space. Furthermore, its characterization of the average subtype lesion as being more strongly phonological or more strongly semantic provides a plausible account of other language behaviors associated with these conditions.

Predictions related to the dual nature of formals

The hypothesized dual nature of formals derives from the fact that each of the model's two retrieval steps is subject to selection errors that yield formals. Sublexical formals occur during phoneme selection; lexical formals occur during lexical selection, when cascaded activation to the target's phonemes sends reciprocal activation back to the words (target and nontarget) that contain these phonemes, making it possible for a phonological neighbor of the target sometimes to compete successfully for lexical selection.

The empirical reality of sublexical formals is not in dispute, and all production models are capable of generating these at the level of phoneme selection or encoding. The empirical reality of lexical formals is more controversial, and many production models are not capable of generating such errors. For example, Slipnet, a discrete two-step model implemented for the simulation of aphasic naming, generates formals only at the sublexical level, a small fraction of which are words by chance (Laine et al., 1998). The frequency of formal errors in the simulated data from 10 aphasic patients was low enough in that study that it did not pose a challenge to the model's architecture and the single-origin account of formals (see also Lambon Ralph et al., 2002; Lecours, Deloche, & Lhermitte, 1973; Nickels & Howard, 1995b; Ruml et al., 2005).

Without doubt, some aphasic patients generate formal errors at high enough rates to warrant postulating a second mechanism (e.g., Blanken, 1990; Martin et al., 1994). However, our group's advocacy of the dual nature of formals rests on an additional argument, based on the grammatical category constraint. Only lexical formals are predicted to preserve the nounness of (naming) targets more often than chance. In Gagnon et al. (1997), we showed that aphasics' formal errors exhibited the noun effect, and in Dell et al. (1997) we demonstrated that the weight–decay model could accurately predict which patients would exhibit the noun effect and which would not. Those who showed the noun effect had lesions characterized by globally strong connection weights but relatively high rates of decay. Foygel and Dell (2000) found that the semantic–phonological model produced essentially the same split among those patients; those with globally strong connection weights (and high rates of decay) tended to have strong phonological weights (and weak semantic weights), which is what the semantic–phonological model seemed to require as a basis for generating lexical formals.

Here, we performed a more extensive exploration of the semantic–phonological model's basis for generating lexical formals. We evaluated the model for 25 combinations of the s and p parameters, enough to make a map. For each parameter combination, we determined the percentage of the observed formals that occurred because the formal lexical unit was selected. It is only these true lexical formals that should be nouns. The results are plotted in Fig. 12.

The graph shows contour regions, defined by the extent to which the model produces lexical formals. 0–25% lexical means a region in which 0–25% of the formals are lexical, and so on. When the semantic and phonological parameters are equal (imagine a diagonal line from lower left to upper right) the expected percentage of lexical formals is in the 25–50% range, and when the phonological parameter is smaller than the semantic one, we quickly get into the region where few formals are lexical (0–25%). The fact that the contour lines are parallel with a slope close to one means that the lexicality percentage is predicted just by the difference between the p and s parameters. As $p-s$ increases, so does the percentage.

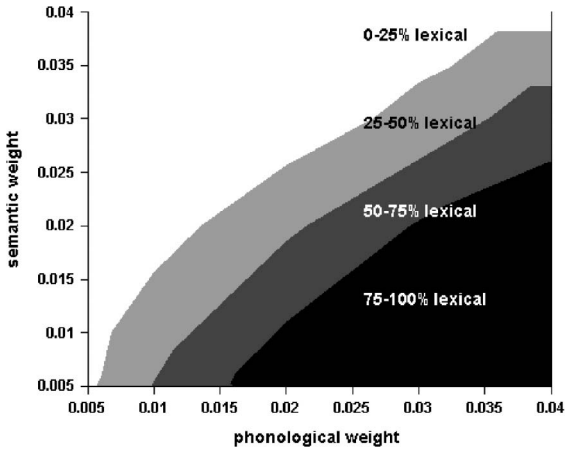

Fig. 12. Semantic–phonological parameter map showing the percentage of the model's formals that occurred because the formal word was selected.

The central finding here is that *for both low and high phonological weights*, the percentage of lexical formals is substantial only when the phonological weight is greater than the semantic weight. This might seems to contradict the claim that lexical neighbors require phonological feedback to become competitive. But if the feedback is weak (low phonological weight) this is compensated for when the semantic weight is even weaker, as this allows the weakly activated formal neighbors to have a chance against the correct and semantically related lexical units.

This finding motivates the prediction that the magnitude of the noun effect should be greater in patients whose modeled connection weight is stronger for p than s, and this should be true for those with relatively weak p weights as well as those whose p weights are relatively strong. Seventy-six patients produced at least one formal error and so were included in the analysis. Fitted p weights in this group ranged from a low of .007 to a high of .040, with many more patients at the low end. Applying a cut-off of .025 to subdivide the group allowed for a reasonable number to be included in the "high p" group ($p \geqslant .025 =$ "high," $N = 27$; $p < .025 =$ "low," $N = 49$).

As in prior work, we used 64% as the estimate of how often random phoneme substitution creates nouns by chance (Gagnon et al., 1997). Using this figure, we calculated the number of noun formals expected by chance for each patient and subtracted this from the number of nouns actually produced, to obtain an estimate of noun formals in excess of chance (noun effect). We then

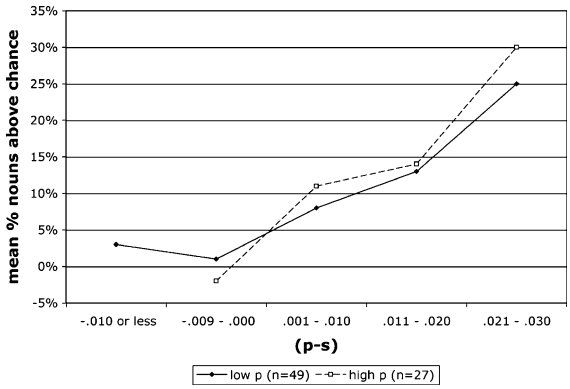

Fig. 13. For low- and high-p patients, the mean percentage of noun formals in excess of chance, binned for $p-s$ (the difference between fitted phonological and semantic weights).

averaged the estimated noun effect in patients binned by the $p-s$ difference and plotted the results separately for high and low p patients. The results, shown in Fig. 13, accord nicely with the prediction: a bigger noun effect with larger $p-s$, for both low-p and high-p patients.

We then tested the significance of the noun effect in each $p-s$ bin, averaging over high- and low-p patients. As shown in Table 4, the effect was significant only for the two highest bins, that is, for the bins with $p-s > .011$. It is also noteworthy that for patients in the two lowest bins, where the noun effect was negligible, the mean percentage of noun formals was 65.5%, which is remarkably close to the 64% estimate of chance.

Predictions related to the mixed error effect

The mixed error effect, like the dual nature of formals, speaks to the model's interactivity assumption. In the model, word retrieval errors, including semantically mediated retrieval errors, are influenced by phonological factors, reflecting feedback. To establish predictions from the semantic–phonological model regarding where the mixed error effect should be greatest, we explored its parameter space in a network that contains both RAT (the mixed-error distractor) and DOG (the semantic error distractor) and no other semantically related items. (This is the neighborhood we used for testing the mixed error effect in Dell et al., 1997.) The mixed error effect is quantified as the ratio of the number of selections of RAT to the number of selections of DOG during word retrieval. To the extent that RAT is chosen over DOG, that is, the ratio is greater than 1.0, the mixed error effect is present.

A contour-map display of the result is shown in Fig. 14. (The labeled points refer to patients and will be explained shortly.) For all parameter values examined, RAT is chosen more than DOG; but the ratio does not get much greater than 1.0 until the value of p gets quite large. Even then, the ratio remains close to 1.0 unless the value of s is between .03 and .07. The major result, then, is that both parameters need to be close to their normal values (e.g., s and $p > .05$) in order for the model to show a large mixed error effect. Why do both parameters need to be large? When p is large, there

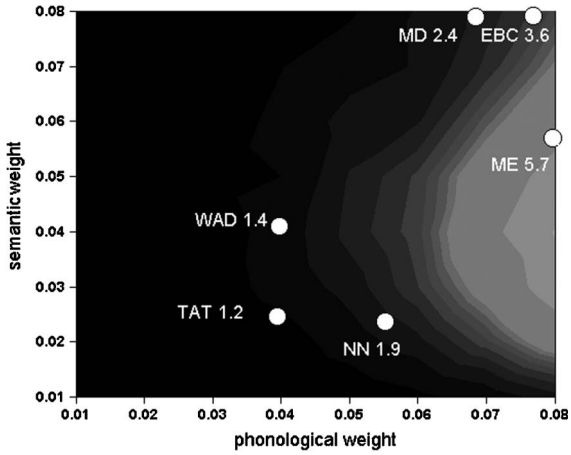

Fig. 14. Contour map depicting the variation in strength of the mixed error effect in semantic–phonological parameter space, measured as the ratio of the number of selections of RAT to the number of selections of DOG during word retrieval. Lighter shading represents a higher ratio (hence greater mixed error effect). Points represent the six patients with fitted s and p parameters of .04 or higher; numbers alongside coded initials show the patient's ratio of actual-to-chance target–error overlap for phoneme positions 1–3.

Table 4
The mean number of noun formals (NF) and NF above chance, binned by $p-s$, the difference between the fitted phonological and semantic weights

$p-s$	Mn. NF	Mn. NF above chance	DF	t value	p value
.021–.030	0.932	0.292	4	9.766	0.0006
.011–.020	0.778	0.138	18	3.427	0.003
.001–.010	0.736	0.096	20	1.708	0.1031
−0.009–.000	0.645	0.003	19	0.07	0.945
−.010 or less	0.674	0.033	10	0.72	0.4878

is feedback to make the mixed competitor RAT more active than the semantic competitor DOG. When s is large, lexical errors of all types are rare, so only the most active competitors (e.g., RAT) have a fair chance of selection. So, when s and p are both near normal, the ratio of RAT/DOG (indexing the mixed error effect) is high, whereas when s is subnormal, lexical access becomes more error prone and the natural advantage of RAT over DOG diminishes.

The model's predictions for the patients, therefore, are that there will be a mixed-error effect overall, and that individual patients will show a big effect only if their fitted parameters are close to the normal region. However, those patients do not make many errors and their low error production makes it difficult to evaluate model predictions.

As in Dell et al., 1997, we examined the mixed-error effect in the group overall by selecting all errors categorized as semantic or mixed (excluding morphological relatives) and counting how often their first, second, and third phonemes matched their targets. We then compared the obtained overlap proportions to chance, as estimated in Martin, Gagnon et al.'s, 1996 study of aphasic performance on the Philadelphia Naming Test: .057 for the first phoneme, .060 for the second, and .056 for the third. For patients in the current study the average overlap proportions were, respectively, .16, .14, and .10. Each position differs reliably from chance (all p values $< .001$). Thus, the first of the model's predictions is confirmed—across all patients in the sample, there is a significant mixed error effect.

To evaluate the second of the model's predictions, we identified all cases in which the fitted p's qualified as near normal (.04 or higher). There were six such cases. Their data are plotted on the contour map of Fig. 14. The number alongside a patient's initials is the ratio of actual- to chance-overlap, for all three positions; ME, for example, has 5.7 times more phonologic overlap in his semantic errors than expected by chance. Three patients (ME, EBC, and MD) have fits that place them close to the "hot spot" on the map (the region of the model that yields the biggest mixed error effect), and these three patients have overlap scores that are at least twice chance. This large a mixed error effect was present in less than half the subjects in the full sample (44 of 94).

Thus, the model predicts that patients with fitted parameters close to the normal values will show strong mixed error effects; and Fig. 14 shows this to be true for the very small number of cases who qualify as having near-normal parameters. For the remaining patients in the sample, the model does not explain individual differences in the mixed error effect. Bivariate correlations between the size of the effect (expected minus chance overlap for positions 1–3) and fitted values of s, p, s–p, or $s + p$ fell well short of significance in all cases (Pearson $r < .20$).

Predictions related to similarity effects in nonsemantic word errors

According to the model, both mixed errors and lexical formals are products of lexical–phonological feedback. The logic that predicts above-chance overlap in errors classified as semantic or mixed also predicts above-chance overlap in nonsemantic errors generated at the lexical level, i.e., unrelated word errors and lexical formals. Of course, the dual nature of formals confronts us with the problem of selecting for the analysis just those formals that are most likely to have a lexical basis.

The model tells us that lexical formals are most likely to come from patients whose production of nonwords is low and whose formal errors are mostly nouns. These are patients with a high value of $p-s$. We therefore began the analysis by choosing a group of 20 patients with the highest values of $p-s$ and a proportion of noun formals in excess of chance. This group was further winnowed to 10 by selecting those with the fewest relative number of nonwords. The 10 patients selected were KBX, EAL, SI, SS, NAC, MG, FL, HAN, BS, and BAL. From these, we assembled a corpus of nonsemantic word errors, comprised of the errors that had been classified as formals ($N = 96$) or unrelateds ($N = 53$). Only noun responses were included. Five of the formals differed from the target word by one phoneme; as these might have been word errors by chance, we removed them from the corpus. In a further effort to avoid inflating target–error overlap by including errors that were in actuality phonological, we estimated from the patients' rate of nonword errors approximately how many word errors would be created by chance at the phonological level: The 10 subjects made a total of 41 nonword errors. On the assumption that phonological error processes create words 20% of the time (Dell et al., 1997), one would expect that beyond these 41, there would be an additional 10.2 phonological errors that happened to create words and thus ended up in the formals or unrelated categories. Of these 10.2 errors, we expect 64%, or 6.5 to be nouns. Had these been included in the nonsemantic error corpus, six or seven of the identified formal noun errors would have arisen at the phonological, not word level. Since these noun pseudo-formals are likely be quite similar to the target, they were probably already removed by our taking out the five that were one-phoneme off. To be conservative, though, we removed an additional seven formal errors at random from the analyzed set.

Removal of the five 1-off and seven additional formals left a total of 137 nonsemantic word errors (all nouns) for entry into the similarity analysis. We performed this analysis using the same procedures and estimates of chance overlap as in the mixed error analysis. Averaged across the 10 selected patients, the overlap proportions for first, second, and third phonemes were,

respectively, .22, .15, and .12. Each position differs reliably from chance (all p values $< .001$).

This analysis indicates that the errors identified as lexical-level formals are unlikely to be unrelated errors that are phonologically similar by chance. It should be noted that the ten subjects selected for this analysis all had lesions involving the *s* weight more than the *p* weight. (This is consistent with their pattern of producing more formals and unrelateds relative to nonwords.) These are the patients that the model predicts will demonstrate stronger phonological effects on their lexical errors, compared to subjects with other lesion types.

General discussion

This computational case-series is larger than any that has been studied to date. It represents an unbiased sample of aphasics with CVA etiology, who are post-acute, community dwelling, and willing and able to participate in language research. The demonstration by Dell et al. (2004) that omission errors can be accommodated in the model by treating them as largely independent from overt errors (Ruml et al., 2000) made it possible to include patients who produce numerous omissions; as a consequence, this sample includes patients with lower correctness than those we modeled in past studies (Dell et al., 1997; Foygel & Dell, 2000). We also developed a method for reliably scoring the errors of patients with motor speech impairments, which allowed us to include for the first time patients with Broca's aphasia.

Using this large and diverse case series, we found that the semantic–phonological version of the interactive two-step model outperformed the weight–decay version of the model in terms of variance accounted for, mean root mean squared deviation, reduction of systematic deviations in the nonword and unrelated word response categories, and fits to individual patients exhibiting dissociations of word vs. nonword errors.

Table 5 summarizes the findings from this and previous studies that used the weight–decay or semantic–phonological model to fit case-series data. The semantic–phonological model does uniformly better than the weight–decay model in every study, when total variance accounted for and average root mean squared deviation are considered. The values obtained in the present study match up well with most others. The larger differences across studies will be discussed shortly.

Foygel and Dell's (2000) study found no real advantage for either model in fitting naming response proportions from a 21-person case series (although the semantic–phonological model did better in fitting *repetition* response proportions; see also Dell et al., submitted). Foygel and Dell observed that a reason for preferring the semantic–phonological model is that its characterization of lesions as semantic and/or phonolog-

Table 5
This table shows how the weight–decay and semantic–phonological models performed in this and previous case-series investigations

Model	Patient group	*N*	Mn. RMSD	Total VAF	% sample with RMSD > .041
Weight/Decay	*Current sample*	*94*	*.034*	*.870*	*24%*
	Dell et al., 1997/Foygel and Dell, 2000[a]	21	.024	.901	10%
	Ruml et al., 2000[a]	13	.026	.658	15%
	Dell et al., 2004	14	.045	.816	29%
	Ruml et al., 2005 (Italian)—all patients	50	.042	.796	n.a.
	Ruml et al., 2005 (Italian)—Omissions ⩽ 15%	28	.028	.875	n.a.
Semantic/Phonological	*Current sample*	*94*	*.024*	*.945*	*17%*
	Dell et al., 1997/ Foygel and Dell, 2000[a]	21	.021	.929	14%
	Ruml et al., 2000[b]	13	.022	.802	15%
	Dell et al., 2004[c]	14	.030	.930	29%
	Ruml et al., 2005 (Italian)—all patients	50	.035	.857	n.a.
	Ruml et al., 2005 (Italian)—Omissions ⩽ 15%	28	.021	.919	n.a.

As much as possible, everything has been made commensurate with the current analysis, meaning that data from some studies were refit using the automated fitting routine, and substituting the "independence" treatment of omissions, 20% sampling from the mixed-neighbor network, and/or the .6 default value for decay (semantic–phonological model). See table notes below for details. Ruml et al.'s (2005) models have a neighborhood structure that is appropriate to Italian, and their semantic–phonological model has decay rate = .5; their data were not refit. As in the published report, we present their model fits for the full sample, and for the subset who produced no more than 15% omissions. RMSD, root mean squared deviation. VAF, variance accounted for.

[a] Individual response proportions (Dell et al., 1997; Table 7) were refit using the 20% mixed-neighborhood option and the automated fitting routine based on Foygel and Dell (2000).

[b] Individual response proportions (Ruml et al., 2000; Table 4) were refit using decay rate = .6; 20% sampling from the network with mixed neighbor; and the automated fitting routine based on Foygel and Dell (2000).

[c] Individual response proportions (Dell et al., 2004; Table 2) were refit using the normalization ("independence") routine to handle omissions.

ical meshes closely with current theories that identify these two subsystems as primary and distinct (Dell, 1986; Garrett, 1975; Lambon Ralph et al., 2002; Levelt et al., 1999; Martin et al., 1994; Plaut et al., 1996; Shelton & Caramazza, 1999). By comparison, the weight–decay model's characterization of lesions as globally compromised connection weight and/or decay rate is harder to integrate with theories of language and neurology. The current findings, and the consistent advantage for the semantic–phonological model documented in Table 5, provides a strong empirical basis for preferring the semantic–phonological model.

The analysis of current patients whose fits under the semantic–phonological model could be considered deviating revealed additional examples of two previously reported patterns (the WR and JF patterns). In addition, a possible new pattern of deviation was observed in three cases, involving small but consistent underprediction of formals. We noted that these three patients had a severe naming deficit characterized by mostly formal and nonword errors. They also had impaired input processing, including poor phoneme discrimination; and two (BBC and KAC) had word repetition scores well below the group mean (Dell et al., submitted). This deficit profile calls to mind a case recently reported by Wilshire and Fisher (2004), whose abnormality the authors characterized as a pathologically fast rate of decay within phonological representations (see also Martin et al., 1994; Martin & Saffran, 1992). A question for future research is whether it is worth the loss of parsimony to include phonological decay as a third lesionable parameter in the models.

The WR and JF patterns pose difficulties for both the weight–decay and semantic–phonological models. So do patients who produce many omissions. Looking at Table 5, one can see that the study of patients with many omissions (Dell et al., 2004) yielded root mean squared deviations that are somewhat larger. This is, at least in part, because the root mean squared deviations are calculated on the normalized response proportions. A deviation of .03 for a category in a patient with 50% omissions becomes a deviation of .06 once the proportions are normalized. One can also see this distortion in the difference in the root mean squared deviation for the Italian data. For the 28 patients with fewer than 15% omissions, the mean for the semantic–phonological model is a very respectable .021. When all 50 patients are included, the mean is considerably greater, .035. Not all of this increase is necessarily due to this distortion, though. As we said previously, the ability of the model to fit the normalized proportions may be poorer when there are many omissions. This may be because the independence assumption is overly general (see below). Also, it may be because in some patients the high number of omissions is part of the pure semantics pattern, which can be problematic for other reasons. The full sample in Ruml et al. (2005) likely over-represents pure semantics patients by virtue of the inclusion of two patients with semantic dementia and five with herpes virus encephalitis (14% of cases) and because, as the authors state, their patient population is generally biased toward those with semantic damage (p. 140).

It is worth noting in Table 5 that the root mean squared deviation and total variance accounted for sometimes present a different picture of the model's performance in a given study. The patients in Dell et al. (2004) are quite different from one another, apart from their tendency to produce many omissions. Here, where there is a good deal of variance to be explained, the semantic–phonological model explains much of it (93%), even though the fits are less than impressive. This contrasts with Ruml et al. (2000), where the 13 patients in the sample are all quite similar, the total variance explained by the weight–decay model is low (65.8%), but the mean root mean squared deviation (.026) is better than average for this model.

The criteria we used to enroll patients in this study were unlikely to have biased the sample towards or away from the response patterns that are problematic for the model, and they succeeded in producing a sample that was large and highly diverse. On the reasonable assumption that the incidence of deviating patients in this study is representative of their frequency in the population with persisting aphasia from LCVA, the current results for the weight–decay and semantic–phonological models provide an unbiased estimate of their ability to fit the naming patterns of this population.

There is another group of aphasics whose lexical access problems should be explainable by our models. In primary progressive aphasia of the nonfluent type, there is slow deterioration of lexical and syntactic processes with relative sparing of semantic memory. In at least some of these patients, the lexical access problem centers on phonological retrieval/encoding (e.g., Croot et al., 1998). We would expect the naming patterns in such patients to be well fit by a relatively pure p lesion or, in the weight–decay model, a relatively pure weight lesion (see Croot et al., 1998, footnote 1, for preliminary support). In contrast, we would not expect our models to fit the naming patterns associated with primary progressive aphasic of the *fluent* type (also known as semantic dementia). Patients of this type have naming problems as a consequence of damage to semantic representations (Lambon Ralph et al., 2001); as noted earlier, such damage falls outside the scope of the model.

Why the model works (to the extent that it does)

The model comparisons summarized in Table 5 reveal the superiority of the semantic–phonological model. However, in perusing Table 5, one is struck with

how similar the results are for these two different accounts of the lexical access disorder in aphasia. Evidently, much of the success of the models can be ascribed to the processing assumptions they share. To appreciate which assumptions are important and why, it is important to separate two aspects of the results: (1) the initial fits of the model to the six naming-response proportions and (2) the tests of predictions regarding formals, mixed, and nonsemantic errors. The next sections deal with each of these, in turn.

Initial fits to the six response proportions

For the initial fits, the model assumption that is most important is the continuity thesis. Any model that correctly fits the normal pattern, correctly characterizes the random pattern, and places patients on a continuum between these two patterns, will fit reasonably well. The weight–decay model characterizes the normal pattern as mostly semantic and the random pattern as mostly nonwords. The semantic–phonological model retains the characterization of the normal pattern but has a different random pattern for each of the two lesionable parameters. When there are phonological lesions present, the semantic–phonological characterization of random is the same as weight–decay; when the lesions affect only the *s*-weights, its characterization of random is different. In the current sample, most of the patients have some phonological involvement and so are placed on a similar continuum by the two models. Since impaired lexical–phonological access is commonplace in aphasia, this will be true in any relatively unbiased sample of patients (e.g., Lambon Ralph et al., 2002).

The second most important model assumption in producing good initial fits is the two-step assumption. With this assumption, the model distinguishes lexical and sublexical errors, allowing for the basic error types. When the two-step assumption is paired with an approach to lesioning that separates word and nonword errors, the model is capable of explaining much of the patient variability that is independent of overall severity. The weight–decay model accomplishes this separation to a degree, but the separation in the semantic–phonological model is purer, so the semantic–phonological model does a better job of matching patients whose errors are primarily nonwords or primarily words.

Lambon Ralph and colleagues have proposed an account of aphasic naming that has strong commonalities with the semantic–phonological model but lacks the two-step assumption (Lambon Ralph et al., 2002; also Lambon Ralph, 1998; Lambon Ralph, Sage, & Roberts, 2000; Lambon Ralph et al., 2001). Inspired by PDP-style connectionist models of word naming (e.g., Plaut & Shallice, 1993) and verb past-tense morphology (Joanisse & Seidenberg, 1999), their verbally specified model characterizes naming as resulting from the direct interplay between semantic and phonological representations, with no involvement of wholistic (localist) lexical representations. It ascribes aphasic naming entirely to damage to the semantic and/or phonological representations. In a case-series study ($N = 21$), these investigators used patients' performance on reading, repetition, and comprehension tasks to estimate their semantic and phonological abilities, and then entered these quantitative estimates into a regression model with picture naming accuracy as the dependent variable. The model explained an impressively large portion of the variance in naming accuracy. A second analysis revealed that the estimates of semantic and phonological abilities correlated significantly, and in the expected direction, with the frequency of semantic and nonword errors in naming. However, formal errors, which this characterization treats as phonological, produced a pattern of correlations that differed from nonwords and that defied explanation. This is predictable on the dual-origins account; one subset of formal errors reflects phonological weakness, while another subset reflects weakness in the lexical–semantic system. It remains to be seen whether a model built along the lines that Lambon Ralph et al. envision can explain this (see Rapp and Goldrick, 2000 for other challenges to a PDP model of naming). In any case, the differences between computational approaches should not obscure the impressive commonalities that are emerging between Lambon Ralph and colleagues' approach and our own.

To summarize, the most important assumption in initially fitting the six naming-response proportions is the continuity thesis; the next most important is the two-step assumption and an account of lesions that separates processes that act largely at the first step from those acting at the second step. The least important assumption in the initial fitting of naming patterns is the assumption of interactivity. Ruml et al. (2005) created versions of the semantic–phonological model with and without bottom-up feedback for use in fitting their Italian aphasics and found that the no-feedback model produced roughly comparable fits to the semantic–phonological model with feedback. The no-feedback version creates lexical-level formals when weakened *s*-weight lesions increase the contribution of noise and allow unrelated words to be selected; by chance, a portion of those will meet the similarity criterion for formals. This is indisputably one mechanism for generating lexical-level formals; the question, though, is whether formal neighbors to the target receive an added benefit as a result of bottom up feedback. We find Ruml et al.'s comparison of fits with and without feedback inconclusive on this matter. Probably on account of features of Italian and the patient sample, there are very few formal errors in their corpus and hence little variance across patients in this category. For this reason, neither semantic–phonological model—with or without feedback—fits the formals category better than simply guessing the mean. So it remains possible

that in a different sample, the addition of interactive feedback would improve the fit of the semantic–phonological model to the six response proportions. Nevertheless, we agree with Ruml et al.'s basic point that interactivity does not seem to play an important role in fitting those proportions—certainly not as important a role as the continuity and two-step assumptions.

Formal- and mixed-error predictions

In contrast to its minimal role in explaining the initial model fits, the interactivity assumption is critical for the tests of predictions that we have performed. Ruml et al.'s (2005) feedforward only version of the semantic–phonological model can not explain the success of the model predictions concerning the mixed error effect and the above-chance similarity in the selected nonsemantic word errors. It is also not obvious that it can explain the reliable association we found between $p-s$ and the noun effect in formals. In the model with interactive feedback, lexical-level formal errors are associated with phonological weights that are larger than semantic weights because phonological weights convey activation from the phonological level to the lexical level, thereby promoting lexical level formals. More generally, the success of the model's predictions for formal errors supports the two-step assumption (which makes possible both lexical and sublexical formal errors) and the interactivity assumption, which associates lexical formals with higher phonological weights. It is worth noting that this evidence is free of any taint of circularity, since what was predicted (nounness in formals) is independent of the data that formed the basis for the predictions (naming error proportions). The same applies to the prediction of repetition from naming proportions in Dell et al. (1997) and Dell et al. (submitted).

The predictions regarding the mixed-error effect are also a product of interactivity, and the fact that the patients' errors as a whole did exhibit this effect is supportive. Here, though, the predictions for individual patients were not clearly confirmed, largely because weak effects were predicted for most patients and the patients for whom strong effects were expected had few errors.

Why the model does not work (to the extent that it does not)

The overall success of the model's fits to the error proportions and tests of predictions can be traced to its core assumptions—interactivity, two-step selection, and the continuity thesis—and speaks to the validity of these assumptions. However, it is important to acknowledge the small but systematic deviations between the model and the data. Most of these are due to simplifying assumptions whose generality was suspect at the outset: (1) perseverative naming responses can be treated like any other responses; (2) the semantic level is always intact in aphasia; (3) all omissions in naming can be captured by Ruml et al.'s (2000) independence assumption (which is functionally equivalent to the assumption of subthreshold activation, see Dell et al., 2004); and (4) damage is uniform throughout the vocabulary.

Simplifying assumption #1 is undermined by the WR pattern of deviation, in which the over-production of unrelated responses was traced to a high incidence of perseveration. The remaining assumptions are called into question by the JF deviation pattern. Regarding assumption #2, semantic-level damage was evident in several of those patients and likely explains why their errors were mostly of a semantic nature. Rapp and Goldrick (2000) have proposed a version of the model that rejects the assumption of an intact semantic level; it includes a third lesionable parameter capable of compromising input to the semantic features from a largely redundant conceptual representation. Ruml et al. (2005) adapted the three-parameter model for Italian and found that it accounted for 96.6% of the variance in their sample of Italian aphasics (compared with 79.6 and 85.7% for the weight–decay and semantic–phonological models, respectively). (Recall that this sample over-represents patients with semantic damage.)

The omissions assumption is likely to be another factor that contributes to the JF pattern of deviation. There is evidence that some aphasics omit responses as a consequence of self-editing (Mitchum et al., 1990; Schwartz & Brecher, 2000) and that omissions can sometimes result from failure to resolve competition for lexical selection (McCarthy & Kartsounis, 2000; Schnur, Schwartz, Brecher, Rossi, & Hodgson, in press; Schwartz & Hodgson, 2002; Wilshire & McCarthy, 2002). JF-pattern patients who make numerous omissions may be masking errors of other types, or they may be particular vulnerable to competitive interference.

We showed that a few patients with the JF deviation pattern did not have large numbers of omissions and had no apparent semantic-level damage; and there are others of this type in the published literature (e.g., Cuetos et al., 2000; Dell et al., 1997; Rapp & Goldrick, 2000; Ruml & Caramazza, 2000). Simplifying assumption #4 might provide an explanation. In patients who have relatively more damage to one semantic or grammatical category, the accuracy of the model's error opportunities will be affected. If the opportunities for one or more error categories are wrong, the model's proportions will be off. Mixed-error proportions could be especially affected, because while evidence shows that on average, semantically similar words do not resemble each other formally more than do other words, this varies considerably from one category to another (Martin, Gagnon et al., 1996). Thus, if a patient happened to suffer disproportionate damage in semantic categories that have sev-

eral semantic + formal neighbors (hence high mixed error opportunities) the mixed-error proportion might quickly exceed what the model can fit well.

So far, our proposed explanations for the JF deviation pattern center on noncore assumptions. However, it may be that a core assumption is flawed, in particular, the assumption of strong bottom-up feedback from the phonological to the lexical level (e.g., Rapp & Goldrick, 2000). Weak bottom up feedback could reduce the competitive advantage for formal and unrelated neighbors, relative to semantic and mixed, thereby producing a purer "semantic" pattern.

Conclusions

In this, the largest and most representative case-series ever modeled, we found that the semantic–phonological version of the interactive two-step model outperformed the weight–decay version. The semantic–phonological model accounted for the large share of the total variation in aphasic naming profiles (total variance accounted for = 94.5%), closely matched the individual response proportions (mean root mean squared deviation = .034), captured generalizations about clinical subtypes of aphasia, succeeded in fitting most patients exhibiting dissociations of word vs. nonword errors, and correctly predicted facts about formal and mixed errors.

On a more general level, the combination of the case-series method with computational modeling has provided an effective account of the variety of naming error patterns in aphasia. Although this method focuses on the individual by fitting the model to each case separately, it can also lead to theoretically motivated group studies. By reducing multidimensional error patterns into a small number of parameters, individuals can be characterized and then grouped for tests of model predictions, as we did for the predictions regarding lexical and sublexical formal errors. In a companion to this paper (Dell et al., submitted), we expand the prediction tests to another task, word repetition. In addition to naming words from pictures, aphasic participants hear and then repeat words. Model parameters based on naming data are then used to predict repetition performance, allowing for both a test of the viability of the model's parameters and a theory of the relation between naming and repetition.

Computational models and information-processing (box-and-arrow) models have often been viewed as competing frameworks (e.g., Harley, 2004, and subsequent commentaries). We, however, see the two-step interactive model as complementing, rather than competing with, neuropsychological information-processing models of lexical processing. Its representational levels (semantic, lexical, and phonological units) and their connections (lexical–semantic and lexical–phonological) correspond to a neuropsychological model with three boxes and four arrows, two in each direction. The semantic–phonological account of damage in the model further hypothesizes that deficits inhabit the arrows. What the computational model adds to the boxes, arrows, and deficit locations are explicit processing mechanisms. When these mechanisms are specified, we gain an understanding of the kinds of error that the mechanisms generate and how the errors are linked to damage sources. This, in turn, helps us understand and predict aphasic naming and, more generally, lexical processes in production.

References

Baars, B. J. (1976). Spoonerisms as sequencer conflicts: Evidence from artificially elicited errors. *American Journal of Psychology, 89*, 467–484.

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

Belke, E., Meyer, A. S., & Damian, M. F. (2005). Refractory effects in picture naming as assessed in a semantic blocking paradigm. *Quarterly Journal of Experimental Psychology A, 58A*, 667–692.

Berg, T. (2005). A structural account of phonological paraphasias. *Brain and Language, 94*, 104–129.

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

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*. Berlin, Germany: Mouton de Greyter.

Bock, K. (1996). Language production: Methods and methodologies. *Psychological Bulletin & Review, 3*, 395–421.

Buchsbaum, B. R., Hickok, G., & Humphries, C. (2001). Role of the left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive Science, 25*, 663–678.

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

Buckingham, H. D. (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 B.V..

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

Caplan, D., Vanier, M., & Baker, E. (1986). A case study of reproduction conduction aphasia: 1. Word production. *Cognitive Neuropsychology, 3*, 99–128.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology, 14*, 177–208.

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

Caramazza, A., Papagno, C., & Ruml, W. (2000). The selective impairment of phonological processing in speech production. *Brain and Language, 75*, 428–450.

Chute, D. L. (1990). *Maclaboratory for psychology* (Vol. 2.0). Devon, PA: MacLaboratory.

Crary, M., & Kertesz, A. (1998). Evolving error profiles during aphasia syndrome remission. *Aphasiology, 2*, 67–78.

Croot, K., Patterson, K., & Hodges, J. R. (1998). Single word production in nonfluent progressive aphasia. *Brain and Language, 61*, 226–279.

Cuetos, F., Aguado, G., & Caramazza, A. (2000). Dissociation of semantic and phonological errors in naming. *Brain and Language, 75*, 451–460.

Cutting, C. J., & Ferreira, V. S. (1999). Semantic and phonological information flow in the production lexicon. *Journal of Experimental Psychology: Learning, Memory and Cognition, 25*, 318–344.

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. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes, 5*, 313–349.

Dell, G. S., & Gordon, J. K. (2003). Neighbors in the lexicon: Friends or foes? In N. O. Schiller & A. S. Meyer (Eds.). *Phonetics and phonology in language comprehension and production: Differences and similarities* (Vol. 6, pp. 9–37). Berlin, Germany: Mouton de Greyter.

Dell, G. S., Lawler, E. N., Harris, H. D., & Gordon, J. K. (2004). Models of errors of omission in aphasic naming. *Cognitive Neuropsychology, 21*, 125–145.

Dell, G. S., Martin, N., & Schwartz, M. F. (submitted). *A case-series test of the interactive two-step model of lexical access: Evidence from spoken word repetition*.

Dell, G. S., & O'Seaghdha, P. G. (1991). Mediated and convergent lexical priming in language production: A comment on Levelt et al. (1991). *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., Martin, N., Saffran, E. M., Schwartz, M. F., & Gagnon, D. (2000). The role of computational models in neuropsychological investigations of language: Reply to Ruml and Caramazza (2000). *Psychological Review, 107*, 635–645.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review, 104*, 801–838.

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, N.J.: Erlbaum.

Ellis, A. W., & Morrison, C. M. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning Memory and Cognition, 24*(2), 515–523.

Farah, M. (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. *Linguistics Inquiry, 8*, 505–520.

Foygel, D., & Dell, G. S. (2000). Models of impaired lexical access in speech production. *Journal of Memory and Language, 43*, 182–216.

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

Freud, S. (1953). *On aphasia*. (E. Stengel, Trans.). New York City: International University Press.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language, 47*, 26–52.

Gagnon, D. A., Schwartz, M. F., Martin, N., Dell, G. S., & Saffran, E. M. (1997). The origins of formal paraphasias in aphasics' picture naming. *Brain and Language, 59*, 450–472.

Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 133–175). London: Academic Press.

Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production* (Vol. 1, pp. 177–220). London: Academic Press.

Garrett, M. F. (1982). Production of speech: Observations from normal and pathological language. In A. Ellis (Ed.), *Normality and pathology in cognitive functions* (pp. 19–76). London: 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*. Cambridge, MA: MIT Press.

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

Glaser, W. R., & Dungelhoff, F.-J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 640–654.

Goodglass, H., Wingfield, A., Hyde, M. R., Gleason, J. B., Bowles, N. L., & Gallagher, R. E. (1997). The importance of word-initial phonology: Error patterns in prolonged naming efforts by aphasic patients. *Journal of the International Neuropsychological Society, 3*, 128–138.

Gordon, J. K. (2002). Phonological neighborhood effects in aphasic speech errors: Spontaneous and structured contexts. *Brain and Language, 82*, 113–145.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science, 11*, 274–279.

Hanley, J. R., Dell, G. S., Kay, J., & Baron, R. (2004). Evidence for the involvement of a nonlexical route in the repetition of familiar words: A comparison of single and dual route models of auditory repetition. *Cognitive Neuropsychology, 21*, 147–158.

Harley, T. A. (1983). Phonological activation of semantic competitors during lexical access in speech production. *Language and Cognitive Processes, 8*, 291–309.

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. (2004). Does cognitive neuropsychology have a future? *Cognitive Neuropsychology, 21*, 3–16.

Harley, T. A., & MacAndrew, S. B. G. (1992). Modelling paraphasias in normal and aphasic speech. In *Proceedings of the 14th annual conference of the cognitive science society* (pp. 378–383). Hillsdale, NJ: Lawrence Erlbaum.

Hillis, A. E., Boatman, D., Hart, J., & Gordon, B. (1999). Making sense out of jargon: A neurolinguistic and computational account of jargon aphasia. *Neurology, 53*, 1813–1824.

Hillis, A. E., & Caramazza, A. (1995). Cognitive and neural mechanisms underlying visual and semantic processing: Implications from "optic aphasia". *Journal of Cognitive Neuroscience, 7*, 457–478.

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

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. (1992). *Pyramids and palm trees: A test of semantic access from pictures and words.* Bury St. Edmunds, UK: Thames Valley Test Company.

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, J., & Quinlan, P. T. (1988). Cascade processes in picture identification. *Cognitive Neuropsychology, 5*, 67–103.

Indefrey, P., Brown, C. M., Hellwig, F., Amunts, K., Herzog, H., Seitz, R., et al. (2001). A neural correlate of syntactic encoding during speech production. *Proceedings of the National Academy of Sciences, 98*, 5933–5936.

Indefrey, P., & Levelt, W. J. M. (2000). The neural correlates of language production. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed.). Cambridge, MA: MIT Press.

Joanisse, M. F., & Seidenberg, M. S. (1999). Impairments in verb morphology after brain injury: A connectionist model. *Proceedings of the National Academy of Sciences of the United States of America, 96*, 7592–7597.

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 lexicalisation process in sentence production and naming: Indirect election of words. *Cognition, 14*, 185–209.

Kertesz, A. (1982). *Western aphasia battery*. New York: Grune & Stratton.

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

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

McCarthy, R. A., & Kartsounis, L. D. (2000). Wobbly words: Refractory anomia with preserved semantics. *Neurocase, 6*, 487–497.

Laine, M., Tikkala, A., & Juhola, M. (1998). Modelling anomia by the discrete two-stage word production architecture. *Journal of Neurolinguistics, 11*, 275–294.

Lambon Ralph, M., McClelland, J., Patterson, K., Galton, C., & Hodges, J. R. (2001). No right to speak? The relationship between object naming and semantic impairment: Neuropsychological evidence and a computational model. *Journal of Cognitive Neuroscience, 13*, 341–356.

Lambon Ralph, M., Moriarty, L., & Sage, K. (2002). Anomia is simply a reflection of semantic and phonological impairments: Evidence from a case-series study. *Aphasiology, 16*, 56–82.

Lambon Ralph, M. A. (1998). Distributed versus localist representations: Evidence from the study of item consistency in a case of classical anomia. *Brain and Language, 63*, 339–360.

Lambon Ralph, M. A., Sage, K., & Roberts, J. (2000). Classical anomia: A neuropsychological perspective on speech production. *Neuropsychologia, 38*, 186–202.

Lecours, A. R. (1982). On neologisms. In J. E. Mehler, E. C. T. Walker, & M. F. Garrett (Eds.), *Perspectives on mental representations* (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.

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

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–75.

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

Martin, N., Dell, G. S., Saffran, E. M., & Schwartz, M. F. (1994). Origins of paraphasias 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 occurence of naming errors: Implications for models of lexical retrieval. *Journal of Memory and Language, 28*, 462–485.

McNeil, M. R., Robin, D. A., & Schmidt, R. A. (1997). Apraxia of speech: Definition, differentiation, and treatment. In M. R. McNeil (Ed.), *Clinical management of sensorimotor speech disturbances* (pp. 311–344). New York: Thieme Medical Publishers.

Meyer, A. S. (1996). Lexical access in phrase and sentence production: Results from picture-word interference experiments. *Journal of Memory and Language, 35*, 477–496.

Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. M. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition, 66*, B25–B33.

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

Nickels, L. (1994). A frequent occurrence? Factors affecting the production of semantic errors in aphasic naming. *Cognitive Neuropsychology, 11*, 289–320.

Nickels, L. (1995). Getting it right? Using aphasic naming errors to evaluate theoretical models of spoken word recognition. *Language and Cognitive Processes, 10*, 13–45.

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.

Nickels, L., & Howard, D. (2004). Dissociating effects of number of phonemes, number of syllables, and syllabic complexity on word production in aphasia: It's the number of phonemes that counts. *Cognitive Neuropsychology, 21*, 57–78.

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.

Peterson, R. R., & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24*, 539–557.

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

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

Rapp, B. C., & Caramazza, A. (1997). The modality-specific organization of grammatical categories: Evidence from impaired spoken and written sentence production. *Brain and Language, 56*, 248–286.

Rapp, B. C., & Caramazza, A. (1998). A case of selective difficulty in writing verbs. *Neurocase, 4*, 127–139.

Rapp, B. C., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review, 107*, 460–499.

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. (2003). Modeling the relation between the production and recognition of spoken word forms. In A. S. Meyer & N. O. Schiller (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 115–158). Berlin: Mouton de Gruyter.

Ruml, W., & Caramazza, A. (2000). An evaluation of a computational model of lexical access: Comment on Dell et al. (1997). *Psychological Review, 107*, 609–634.

Ruml, W., Caramazza, A., Capasso, R., & Miceli, G. (2005). Interactivity and continuity in normal and aphasic language production. *Cognitive Neuropsychology, 22*, 131–168.

Ruml, W., Caramazza, A., Shelton, J. R., & Chialant, D. (2000). Testing assumptions in computational theories of aphasia. *Journal of Memory and Language, 43*, 217–248.

Saffran, E. M. (1982). Neuropsychological approaches to the study of language. *British Journal of Psychology, 73*, 317–337.

Saffran, E. M., Schwartz, M. F., Linebarger, M. C., Martin, N., & Bochetto, P. (1988). *Philadelphia comprehension battery*. Unpublished test battery.

Schnur, T. T., Schwartz, M. F., Brecher, A., Rossi, N., & Hodgson, C. (in press). Semantic interference during blocked-cyclic naming: Evidence from aphasia. *Journal of Memory and Language*.

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

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). London: Erlbaum.

Schwartz, M. F., & Brecher, A. (2000). A model-driven analysis of severity, response characteristics, and partial recovery in aphasics' picture naming. *Brain and Language, 73*, 62–91.

Schwartz, M. F., Brecher, A. R., Whyte, J. W., & Klein, M. G. (2005). A patient registry for cognitive rehabilition research: A strategy for balancing patients' privacy rights with researchers' need for access. *Archives of Physical Medicine and Rehabilitation, 86*, 1807–1814.

Schwartz, M. F., & Hodgson, C. (2002). A new multiword naming deficit: Evidence and interpretation. *Cognitive Neuropsychology, 19*, 263–288.

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.

Schwartz, M. F., Wilshire, C. E., Gagnon, D. A., & Polansky, M. (2004). Origins of nonword phonological errors in aphasic picture naming. *Cognitive Neuropsychology, 21*, 159–186.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial ordering mechanism in speech production. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 295–342). Hillsdale, NJ: Erlbaum.

Shelton, J., & Caramazza, A. (1999). Deficits in lexical and semantic processing: Implications for models of normal language. *Psychonomic Bulletin & Review, 6*, 5–27.

Smith, M., & Wheeldon, L. (1999). High level processing scope in spoken sentence production. *Cognition, 73*, 205–246.

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: Erlbaum.

Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning Memory and Cognition, 28*, 735–747.

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in spoken word perception. *Psychological Science, 9*, 325–329.

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 Exper-*

imental Psychology: Learning, Memory, and Cognition, 19, 243–259.

Wilshire, C. E., & Fisher, C. A. (2004). "Phonological" dysphasia: A cross-modal phonological impairment affecting repetition, production, and comprehension. *Cognitive Neuropsychology, 21*, 187–210.

Wilshire, C. E., & McCarthy, R. A. (2002). Evidence for a context-sensitive word retrieval disorder in a case of nonfluent aphasia. *Cognitive Neuropsychology, 19*, 165–186.

Zevin, J. D., & Seidenberg, M. S. (2002). Age of acquisition effects in word reading and other tasks. *Journal of Memory and Language, 47*, 1–29.