The selective impairment of phonological processing in speech production

Alfonso Caramazza,[1] Costanza Papagno,[2] and Wheeler Ruml[1]
[1]Harvard University, [2]Universita' di Palermo

Address correspondence to:
Alfonso Caramazza
Cognitive Neuropsychology Laboratory
William James Hall
Harvard University
33 Kirkland St.
Cambridge, MA 02138
USA
e-mail: caram@wjh.harvard.edu

Abstract

We report the naming performance of a patient (D.M.) with a fluent progressive aphasia who made phonological errors in all language production tasks. The pattern of errors in naming was strikingly clear: D.M. made very many phonological errors that resulted almost always in nonword responses. The complete absence of semantic errors and the very low ratio of formal errors relative to nonword errors (1.6%:30.3%) in D.M.'s performance are discussed in the context of recent claims about the nature of naming deficits in fluent aphasics. We argue that D.M.'s performance makes highly improbable the claim that fluent aphasia results from global lesions affecting all levels of the lexical access system equally.

Patients who make errors in naming or in spontaneous speech rarely make only one type of error – for instance, only semantic or only phonological paraphasias. The typical patient encountered in the clinic presents with a complex pattern of error types, including semantic, phonological and formal paraphasias, morphological substitutions, 'no responses,' circumlocutions, perseverations, and fragmentary responses. What are the causes of the various types of errors? Recently, Dell, Schwartz, Martin, Saffran, and Gagnon (1997) argued that the different patterns of naming errors produced by fluent aphasics can be explained as the result of two possible forms of damage, both of which occur uniformly at all levels of lexical processing, but that interact to give various patterns of damage – the so-called globality assumption. (The two forms of damage involve perturbations in the decay and connection strength parameters in a network model of the lexicon, as we discuss in more detail below). On this view, different patterns of naming errors do not reflect differential damage to different levels of the lexical access system. Thus for example, a fluent aphasic who makes almost exclusively semantic errors and one who makes almost exclusively phonological errors would both be assumed to have damage at all levels of the word production system, but to different degrees for the two possible forms of damage.

The claim that seemingly different forms of aphasia are really points on a continuum rather than qualitatively different types of impairment to components of a processing system is not novel – it was the dominant view in the 1950s and '60s (e.g., Schuell & Jenkins, 1959). However, there are two important features of Dell et al.'s proposal that sets it apart from previous forms of this claim. First, their proposal is based on a detailed theory of lexical access that was originally developed to account for a wide range of facts about normal speech production (Dell, 1986). The theory has been used to examine phenomena as diverse as speaking rate (Dell, 1986), priming (Dell & O'Sheaghdha, 1992), and word repetition (Dell et al., 1997). The other important feature of Dell et al.'s claim is that it is associated with a computational model, thereby allowing explicit and quantitative predictions about the distribution of error types that are expected following damage to the system. This makes it possible to determine how good a fit there is between individual patients' naming performance and the model. Dell et al. have shown that the simulation model seems to have the impressive property that when appropriately "lesioned" it can generate error profiles similar to those obtained with fluent aphasic patients.

An alternative explanation for the occurrence of error mixtures is that at least in some cases they reflect qualitatively different forms of damage to different levels of the system. For example, semantic errors might arise from damage to the semantic system and phonological errors from damage to the segmental layer. The co-occurrence of semantic and phonological paraphasias might, then, reflect damage to both levels of processing[1]. This is not to say, of course, that mixtures of error types necessarily imply damage to multiple levels of the system. Rather, the claim is much weaker, but more realistic: certain mixtures of errors sometimes result from different forms of damage to different levels of the word production system; sometimes they result from (more or less) uniform damage to all levels of the system; and sometimes from damage to a single level of the system. For example, it could be that damage to phonological lexical nodes results both in semantic and phonological errors. If it could be demonstrated that different error types have different loci in at least some patients, then we would have to reject the assumption that different mixtures of error types (in patients of a given clinical category) always result from homogeneous damage to the speech production system.

---

[1] Of course, the type of error does not directly reflect the locus of impairment. For example, we know that naming errors that are classified as semantic paraphasias can result either from damage to the semantic system (e.g., Hillis, Rapp, Romani, & Caramazza, 1990; Howard & Franklin, 1988) or from damage to modality-specific lexical nodes (e.g., Caramazza & Hillis, 1990).

How can we find out whether or not a patient's errors in naming result from a global lesion? The approach taken by Dell et al. is straightforward. They argued that if by "globally lesioning" their model of lexical access they could reproduce the patterns of error mixtures found in fluent aphasics, then we would be able to conclude that the error profiles observed in these patients provide support for the core assumptions of the computational model used to reproduce the error patterns. And since the model consists of assumptions about the structure of the normal lexical access process and assumptions about the ways in which the process is altered by brain damage, the success of the model in simulating patient performance would extend support to both sets of assumptions.

The model used by Dell et al. is an interactive activation model consisting of three levels of representations: semantic, lexical, and phonological (Figure 1). The model simulates lexical access by allowing activation to propagate from the semantic layer through the system and selecting, after a specified number of time steps, the most highly activated lexical node and, then, after a specified number of further time steps, the most highly activated phonological nodes. A crucial assumption of the model is that activation flows both downward from semantic to lexical to phonological nodes and back up from lexical to semantic and from phonological to lexical to semantic nodes. The globality assumption is represented in the model by changing the value of two parameters at all levels of the network. Dell et al. chose to vary the *decay* and *connection strength* parameters, which affect, respectively, the maintenance (or representational integrity) and the flow of activation (or activation transmission) in the network. By varying the values of these two parameters, the distribution of errors produced by the model changes. It is possible, therefore, to ascertain whether there are values of these parameters such that the resulting distribution of errors matches the error profile of a given patient. Dell et al. tested their model against the performance of 21 fluent aphasics and concluded that the model is able to fit these patients' profiles of naming errors. They further concluded that the results support both the interactivity assumption in their model of normal lexical access and the globality assumption of naming deficits in fluent aphasia.

FIGURE 1

Dell et al.'s conclusions, especially those concerning the globality assumption, have been challenged on several grounds (Rapp & Goldrick, in press; Ruml & Caramazza, in press; Ruml, Caramazza, Shelton, & Chialant, in press). For example, Ruml & Caramazza (in press) carried out an exhaustive investigation of the space of possible error patterns that could be generated by changing the connection strength and decay parameters in Dell et al.'s model. The results showed that a sizeable proportion of the patients tested by Dell et al. produced error profiles that fall outside the range generated by the model. Figure 2 shows examples of the overlap between the model's range of possible predictions and the patient data. Each small dot represents a combination of error type frequencies that was obtained from the model using some setting of the two parameters that Dell et al. assume to be damaged in fluent aphasia. Circles represent the actual performance of patients reported by Dell et al., while triangles represent patients tested by Ruml et al. (in press) using the same methodology. Each panel in the figure reflects only two response categories, thus nearby points represent error profiles that have similar frequencies of those two categories, but that might differ in the remaining four types. The area in each panel covered by dots represents the range of frequencies for those error types that can be obtained by systematically trying many settings of the two parameters (see Ruml & Caramazza, in press, for details). The varying density of dots in each panel merely reflects the systematic method used to cover the model's entire space of possible patterns---only the outline of the model's region is relevant for our purposes. (The dense regions happen to correspond to the model's random error opportunities, which were set by Dell et al. to correspond to values they determined for English speakers.)

Many of the patients' error profiles fall well outside the space of possible patterns than can be generated by the lesioned model. For example, the top leftmost panel shows the relation of semantic errors to correct performance. As can be seen, patient E.A. produced many more semantic errors (~27%) than are expected by his overall level of correct performance: for a performance level of around 40% correct, the model predicts a rate of semantic errors between 7 and 14%. For another example, consider the middle panel of the bottom row, which shows the relation between mixed and nonword errors. Patient J.F. produced many more mixed errors (13%) than are expected for a patient who makes virtually no nonword errors: for a nonword error rate of around 0% the model's expected rate of mixed errors is 0-4%. Generally, the model associates nonword error frequency with correctness (middle right) and has difficulty dissociating formal and semantic errors (bottom right) or semantic and nonword errors (top middle).

FIGURE 2

Dell et al. acknowledged that patients have been reported whose performance seems to be incompatible with the globality assumption (see also Foygel & Dell, in press; Saffran, Dell, & Schwartz, 2000). They note that patients who make (almost) exclusively semantic (Caramazza & Hillis, 1990; Hillis, Rapp, Romani, & Caramazza, 1990) or phonological errors (Caplan, Vanier, & Baker, 1986; Hillis, Boatman, Hart, & Gordon, 1999; Wilshire & McCarthy, 1996) cannot be fit by their model when lesioned globally. However, Dell et al. raise various methodological reasons as to why the performance of some of these problematic cases for their model should be discounted. For example, they note that Caramazza & Hillis (1990) and Hillis et al. (1990) used different scoring procedures from those used in their study. They also note that in the case of the patient R.L. reported by Caplan et al. (1986) only a 60-item test was used. Since the model's expected rate of semantic errors for the overall level of performance of the patient tested by Caplan et al. was 7% (4 items on a 60-item test) the absence of this error type may simply reflect random variation. Furthermore, Dell et al. note that the test items used by Caplan et al. were somewhat more familiar than the ones they used and that this may also have contributed to the paucity of semantic errors in the naming performance of Caplan et al.'s patient[2].

Patients such as those described by Caplan et al. (1986; see also Hillis, et al., 1999), Caramazza & Hillis (1990), and Hillis et al. (1990) are especially important in this context because they would appear to be the best candidates for hypothesizing highly localized (as opposed to global) functional lesions to the lexical access system (Foygel & Dell, in press; Ruml & Caramazza, in press). It is important therefore to clearly document the performance of such cases with tests and scoring procedures that are not open to the objections raised by Dell et al. Here we report the naming performance of a fluent aphasic whose distribution of naming errors is similar to that of R.L. reported by Caplan et al.

Case report

D.M. was 73 years old at the time of testing (1999). He is a right-handed Italian man with a university degree in engineering. In January 1998 he began producing phonemic substitutions in spontaneous speech, the severity of which progressively increased in the following months. Apart from this problem, he did not complain of other deficits. No previous neurological disease was reported, but the patient was

[2] The basis for this argument is not clear to us since "familiarity" is not a factor in Dell et al.'s model and therefore it is not clear why this factor would lead to differential rates of semantic as opposed to other error types.

diabetic and had suffered an acute myocardial infarction two years earlier. On neurological examination, no motor weakness, sensory impairment or visual field defect could be detected.

A SPECT-scan performed in June 1998 (about six months after onset) revealed hypoperfusion in both temporo-parietal regions, while an MRI in September 1998 showed an almost symmetrical bilateral temporal lobe atrophy, more severe in the middle and posterior parts. No focal lesions were found.

General neuropsychological examination.

D.M. was first seen as an outpatient in the Neurological Department of the Legnano Hospital in July 1998, six months after onset of signs of language impairment. He participated in neuropsychological evaluations in September 1998 and in February 1999. Here we report the results of the second evaluation. D.M. showed no signs of oral apraxia in a 10-item test. His articulation rate in counting and reciting the alphabet, as timed with a stop-watch, was entirely normal (as compared to a matched control subject). D.M. produced spontaneous speech at a lower volume than normal (the patient reported that he was "ashamed" about the errors he produced and therefore tended to speak quietly). However, his speech was informative, fluent, well articulated, and with a normal variety of syntactic forms. There was no tendency to omit closed-class words or inflectional forms: articles, pronouns, auxiliaries and noun and verb inflections were produced correctly. D.M. also produced complex sentence constructions with subordinate clauses. The salient feature of his impairment was the presence of a considerable number of phonemic substitutions in naming and repetition, with conduite d'approche. Occasional word-finding difficulties were also observed. However, as already noted, his language production ability was otherwise normal. A speech sample follows. The patient is describing how he shaves. Errors are in italics and the target is given in parentheses.

…Prendo il *resoio*, *risoio*, *rasello* (target word: rasoio), la *lamella* (lametta). Prima mi insapono con la *spuma*, la *spiuma*…,la *spuma* (schiuma); poi passo la *lamella* (lametta) e dopo mi *sciocchio*, *schiocchio* (sciacquo), mi sciacquo e alla fine metto il dopo*barsa*…*barda* (dopobarba), no!…
(I take the razor, the razor blade. First I put on shaving cream; then I shave with the razor; then I rinse and put on after-shave lotion.)

Apart from this language production impairment, his everyday behavior was normal and he did not show signs of cognitive deterioration.

D.M. obtained normal scores on a standard examination for dementia (Brazzelli et al., 1994), and the Raven's Coloured Progressive Matrices. His performance on the Block Tapping Task (Orsini et al., 1987), Benton's Line Orientation Test (Benton et al., 1983), and Short Story (Novelli et al., 1986) proved to be normal, while digit span (Orsini et al., 1987) was slightly below the cut-off (3, n.v.$\geq$3.75). No signs of buccofacial or limb apraxia were found. The results of the general neuropsychological examination are given in Table 1.

TABLE 1

D.M. was also given a standard language examination and the short version of the Token Test (De Renzi & Faglioni, 1978). He showed normal ability to discriminate CV syllables. He performed virtually flawlessly in matching aurally and visually presented words to their corresponding pictures from an array that contained semantically and phonemically (or visually) related items. The single error in the auditory comprehension task involved a phonological error: he selected “treccia” [Italian for braid] instead of “freccia” [Italian for arrow]). D.M. also performed within normal limits in an auditory sentence-to-picture matching test and in a grammaticality judgment task. By contrast, he was impaired in repetition of single words (34/45 correct), nonwords (30/36 correct), and phrases and sentences (10/36 correct). There was a clear effect of stimulus length, with longer stimuli being more difficult to repeat than shorter ones. In oral naming of pictured objects and actions he correctly named 20/30 and 18/28, respectively. In both cases (object and action naming), errors consisted of phonemic substitutions (for instance "spella" instead of "spilla", or "stigare" instead of "stirare"). No other types of errors were produced. Performance on the short version of the Token Test was 26/36 (cut-off 29).

D.M. was also impaired in reading and in writing. On the *Batteria per la Valutazione dei Disturbi di Lettura* (Battery for the Evaluation of Reading Deficits, Burani, Laudanna & Miceli, unpublished), he only read aloud correctly 235/272 words (86.4%) and 61/122 (50%) nonwords. There were no effects of concreteness, frequency or grammatical class on word reading. And there was no effect of length. Incorrect responses in reading words consisted of nonword errors (25/37) or visually/phonologically similar word substitutions (12/37). In nonword reading, almost all errors were phonologically similar nonwords. In three cases he produced a visually/phonologically-related word ("chiediva" [nonword] "chiedeva" [he was asking]; "spreago" [nonword] "pregavo" [I was praying]; "tove" [nonword] "foce" [estuary]). In a writing-to-dictation task from the *Batteria per la Valutazione dei Disturbi di Scrittura* (Battery for the Evaluation of Writing Deficits, Burani, Laudanna, Capasso & Miceli, unpublished) he produced correctly 101/198 (51%) words and 17/80 (21%) nonwords. Errors were virtually all orthographically related nonword responses. The few exceptions (9/160) involved the production of an orthographically or morphologically similar word (for example: "oste" [host] for "ospe"; "soffro" [I suffer] for "soffre" [he suffers]). A length effect was present, with longer stimuli resulting in more errors.

In short, D.M. is a fluent aphasic, whose salient feature is the production of phonemic substitutions in naming, reading and repetition. Therefore, he is an ideal test case for the globality assumption. The remainder of the paper focuses on D.M.'s naming performance and the attempt to fit his profile of naming errors with Dell et al.'s model of naming.

Naming

D.M. was asked to name 314 pictured objects consisting of the pictures from the 130-item naming test *Batteria di Denominazione per Categorie Semantiche* (Naming Battery for Semantic Categories; Capasso & Miceli unpublished) and 184 pictures from the Snodgrass and Vanderwart (1980) picture set. He named 199 of the 314 items correctly (62.7%). Errors were classified following Dell et al.'s procedure (Dell et al., 1997), modified for Italian (Caramazza, Ruml, Capasso, Miceli, in preparation). In this procedure, only the first complete response is scored. Responses are classified as correct, semantic error, formal error, mixed error, nonword, no response, or other (multiword). An incorrect word that could be classified as a category superordinate, coordinate, subordinate or associate is scored as a semantic error. An incorrect response is scored as a formal error if it is a word and, excluding inflections, target word and response share (a) at least one third of their phonemes, (b) two phonemes in the first syllable, (c) the two initial phonemes, or (d) three consecutive phonemes[3]. A mixed error is an error that satisfies both the criteria for classification as a semantic and as a formal error. A nonword error is a response that does not correspond to an entry in the Italian lexicon, with or without formal

[3] The criteria used by Dell et al. (1997) for scoring a response as a formal error were as follows: the response starts or ends with the same phoneme as the target or shares more than one phoneme in any position; this category also includes instances in which target and error share one phoneme in the same within-syllable position. The application of this criterion to Italian resulted in most errors being scored as formal errors. We opted therefore for a more natural scoring procedure that resulted in a distribution of errors that approximates that found with English-speaking patients.

relation to the target. Finally, a response was scored as a no-response error when the patient failed to produce any sound at all.

The great majority of errors were nonwords, but there were also a few formal errors and no-responses. Erroneous responses were distributed as follows: 95 (30.3%) nonword errors, 5 (1.6%) formal errors (e.g., “nuota” [swims] instead of “suora” [nun]), and 15 (4.8%) no-responses. Errors were distributed roughly equally across semantic categories, though there was a slight trend for poorer performance with living than non-living things (70% versus 76.7% correct, n.s.). There was a slight trend for shorter words (< 6 phonemes) to be produced better than longer words (70% versus 54% correct, respectively; n.s.) but a robust frequency effect – high-frequency words were produced better than low-frequency words (20/20 correct versus 34/50 correct, respectively; $X^2 = 8.3$, $p<.01$).

An analysis of the distribution of phonological errors was carried out to more fully characterize the nature of D.M.’s naming deficit. For this purpose we only considered errors that resulted in non-word responses. We did not include word responses (there were only 5 such errors) because they may have been influenced by lexical factors. Errors were classified into one of the following categories: substitution, omission, addition or transposition of phonemes. Since each response might contain more than a single error, we distinguished between responses that contained single errors and responses with multiple errors of the same type (i.e. a substitution and an omission). Results are shown in Table 2. It is apparent that the vast majority of errors involved substitutions of phonemes.

TABLE 2

We also assessed the effect of serial position on the distribution of phonological errors, following the procedure of Wing and Baddeley (1980). We collapsed performance for stimuli of different lengths into a single, arbitrary stimulus length. In this procedure, each stimulus is divided into five “phoneme” positions (stimuli with fewer than 5 phonemes were excluded from this analysis). Each position contains one or more phonemes, depending on the number of phonemes that, for each stimulus, exceed 5 or multiples of 5. The phonemes in excess are distributed across the 5 positions so as to maintain a symmetrical structure in the arbitrarily reconstructed stimulus. For example, the word “banana” can be divided into the following positions: B, A, NA, N, A. The word “giraffa” can be divided as G, IR, A, FF, A. Using this procedure, we were able to include in the analysis 71 responses out of the 95 nonword errors produced by the patient.

Results are shown in Table 3a. It is clear that there is a serial position effect – errors are more frequent in the middle than beginnings or ends of words[4]. However, better performance at the end than in the middle of words may simply reflect the fact that Italian words almost always end in a vowel – most frequently an /a/, /o/, or /e/for the singular nouns used in the naming task (or an /a/, /o/, /e/, or /i/ more generally). Thus, the better performance at the end of words may reflect the effects of guessing rather than a true serial position effect. To assess this possibility, the data were re-analyzed, omitting the last phoneme. For example, the word “forchetta” (fork) would become FORCHETT and the resulting 5 positions would be: FO, R, CH, E, TT. Following this procedure fifty nonwords were included in the analysis. The results of this analysis are shown in Table 3b, from which it is clear that the better performance at the ends of words is not due simply to the fact that in Italian words the last position is occupied by a vowel.

TABLE 3

[4] Although there were too few word responses to analyze separately, for these responses, too, there were more errors in the middle than the beginning or end of words.

In summary, the pattern of performance produced by D.M. in speech production tasks is most naturally accounted for by assuming damage at some level of phonological processing, with spared semantic and grammatical processing. Our results do not allow a more precise characterization of the nature of damage within the speech production system, but it clearly implicates processes involved in selecting the segmental content of correctly specified lexical entries. It is astronomically unlikely that D.M.'s pattern of phonological substitutions could have been obtained if lexical entries had been selected incorrectly. The absence of errors of lexical selection implies the absence of damage to lexical and semantic representations and processes.

A test of the globality assumption.

The pattern of results obtained with D.M. clearly replicates the performance profile reported for R.L. by Caplan et al. (1986): Phonological errors in naming can occur in the complete absence of semantic errors. The absence of semantic errors in the case of D.M. is unlikely to reflect merely the insufficient sample size of the naming test. Our naming test had 1.5 times as many stimuli as the Philadelphia Naming Test (Roach, Schwartz, Martin, Grewal, & Brecher, 1996) that was used by Dell et al. (1997) and 5 times as many stimuli as the test used by Caplan et al. and yet there was not a single semantic paraphasia among D.M.'s 115 naming errors. The crucial issue becomes, then, whether or not such a profile can be fit by an interactive activation model that incorporates the globality assumption proposed by Dell et al. (1997). Given the demonstrated ability of Dell et al.'s model to produce a variety of error patterns, it is important to quantitatively test the model rather than relying on intuitions alone to assess its capabilities. We will consider the space of possible error profiles that can be generated when Dell's interactive activation model is lesioned "globally," using the criteria proposed by Dell et al. We can then determine whether D.M.'s error profile falls within the space of possible error profiles.

Since Dell et al.'s model was designed to account for the performance of English speakers, we first had to construct an analogous model for Italian speakers. Following the methodology of Dell et al., we estimated the random error opportunities for Italian, constructed a lexical network that exhibited those opportunities, and then determined parameter values that would allow the model to match the picture-naming performance of normal control participants. Details of the procedure, including our automated method for constructing a lexicon that exhibits desired error opportunities, are given in Caramazza et al. (in preparation). The resulting model for Italian speakers matched the desired error opportunities well (maximum frequency discrepancy in any category was 1.2%) and fully accounted for the performance of normal subjects ($p > .99$ even after more than 2,600 trials from the control participants and 10,000 trials from the model).

Given this Italian version of Dell et al.'s model, we can test its ability to account for D.M.'s performance. Using the assumption of global damage, we systematically simulated the model's response under several thousand possible global lesions. Figure 3 shows several plots of the error patterns that the model is capable of making. As in Figure 2, each panel represents two response categories, and each dot represents a particular error pattern that the model generated. The distribution of dots shows that the model is capable of generating only certain limited combinations of response frequencies. The labeled circle shows the naming profile of D.M.. In almost every panel, D.M.'s particular combination of response frequencies lies outside the space of error profiles that are representable by the model. Using the parameter optimization methods of Ruml and Caramazza (in press), we searched for the closest pattern to D.M.'s that the model could generate. Table 4 shows the best match found, and includes two quantitative measures of fit:

the root mean square difference (RMSD) and a X^2 test (assuming 5 degrees of freedom, which is generous since two parameters were estimated from the data). Clearly, the model cannot account for D.M.'s performance.

FIGURE 3 and TABLE 4

The poor fit we have obtained between the model and our data is not merely due to peculiar properties of our Italian version of Dell et al.'s model. If we plot D.M.'s performance against the English version of Dell et al.'s model, we obtain equally poor fits. Figure 4 shows the error patterns that the English version of Dell.'s model is able to generate. Here, too, it is clear that D.M.'s performance lies outside the space of error patterns that the model can generate. The similar shapes of the model's ranges across the two languages suggest that it is not the details of the simulated lexicon that are causing the model's poor performance, but rather its fundamental assumptions.

FIGURE 4

Further tests of Dell et al.'s model of naming

Our tests of the globality assumption show that the globally lesioned model proposed by Dell et al. (1997) cannot account for the performance of fluent aphasics such as D.M.. The results undermine the assumptions implemented in the model as a whole. That is, they show that at least one of the assumptions in the model is unlikely to be true. However, they do not allow more precise attribution of relative blame for the model's failure; they do not, on their own, allow us to conclude that it is specifically the globality assumption that is at fault for the model's poor showing.

There are two approaches we could use to hone in on which assumption in the model may be responsible for the observed failure in a specific patient. One approach is to obtain independent evidence about the locus of damage to the patient's naming system. For example, it might be possible to establish empirically that the globality assumption cannot be applied to the performance of our patient because there is independent evidence that the semantic (or lexical or phonological) level of processing is intact. The other approach is to implement various assumptions of damage and assess the model's ability to account for the patient's patterns of naming performance. The latter approach allows us to determine whether it is possible to find a type of damage to the implemented model that would account for the patient's performance. In the measure to which we fail in this effort, we reduce our confidence in the core assumptions of the model.

To evaluate the impact of different assumptions regarding the functional extent of brain damage, we tried the modeling approach proposed by Foygel and Dell (in press). They proposed damage in the transmission of activation between levels of representation, with different degrees of damage between different levels. Using our adaptation of their model into Italian, we again simulated the results of several thousand possible combinations of parameter values. The results are shown in Figure 5. **The localized effects of the simulated damage causes our systematic sampling to appear in many of the panels as an array of streaks.** Although the assumption of local damage allows the model to approach certain aspects of D.M.'s performance, such as low levels of semantic and unrelated errors, there are still several response categories in which the model cannot begin to capture his error profile. (Recall that the panels reflect each pair of response categories independently, and that even if D.M.'s performance were inside the model's range in every panel, this would not imply that the model could account for his particular response frequencies in combination.)

The combination of few formal or mixed errors with a moderate number of nonwords seems problematic. Close examination of the streaks in the center panel shows that damage to the connections between the semantic and lexical levels increases the frequency of formal errors without introducing nonword responses, while damage to the connections between the lexical and the phonological levels always yields both formal and nonword errors. At high levels of damage, performance converges to complete randomness, but moderate levels of nonword error production such as D.M.'s, the interactivity of the model necessitates a prediction of at least 10% formal errors.

Table 5 shows the best match to D.M. that we obtained after fine-tuning the model's parameters. Although the semantic and unrelated categories are better modeled than with the assumption of global damage, the model still predicts too many formal errors and too few correct and nonword responses. We are led to conclude, therefore, that not only the globality assumption but also other aspects of Dell et al.'s (1997) and Foygel and Dell's (in press) models are undermined by D.M.'s performance. Among these is the interactivity assumption.

FIGURE 5 and TABLE 5

<u>Discussion</u>

D.M.'s performance is striking in at least one respect: his naming errors consist exclusively of phonological distortions of the target response. D.M. produced nearly 100 phonological errors in naming but not a single semantic error. Thus, D.M. replicates the pattern of performance of patient R.L. described by Caplan et al. (see also Hillis et al., 1998), but with a much larger number of stimuli and using the scoring procedure proposed by Dell et al. D.M.'s performance falls well outside the range of performance predicted by Dell et al.'s model of lexical access, using a lexicon reflecting either English (Figure 4) or Italian (Figure 3), and using either global (Figures 3 and 4) or localized damage assumptions (Figure 5).

Dell et al. (1997) interpreted the putative good fits of their model to their patients' naming data as providing support not only for the globality assumption but more generally for the core assumptions of the model in which the globality assumption was implemented. They especially emphasize the assumption of interactivity between levels of representation. Thus, they argued that "The good fits between the patient data and the model … extends support for the interactive two-step approach to naming … [and] … the hypothesis that variation in patient error patterns can be associated with global lesions in activation transmission, representational integrity, or both." (pg. 820). But what happens when the model can't fit the data, as in the case of D.M.'s performance? Do such data undermine both the globality and the interactivity assumptions? We are certainly justified in concluding that D.M.'s performance supports <u>neither</u> assumption. But, can we draw the stronger conclusion that a specific assumption (from the set of assumptions implemented in the model) is to blame for the model's failure? The analyses we have reported do not allow a more precise attribution of relative blame for the model's failure. All that the results allow is that the model failed because one of its assumptions is false: either the globality assumption, the interactivity assumption, both, or some other assumption in the implemented model. Replacing the globality assumption with the more localized damage proposal of Foygel and Dell didn't allow the model to account for D.M.'s performance, but such an outcome doesn't let us determine whether both damage assumptions are faulty or if some other fundamental assumption is to blame. To establish that any one specific assumption is responsible for the failure we would have to obtain much stronger evidence than the results of a single task – evidence that could be used to independently establish which assumption is at fault for the failure (see Ruml & Caramazza, in press; see also Caramazza, 2000).

The more likely candidate for the model's failure is the globality assumption (Foygel & Dell, in press). This assumption is implausible on its face. It is unreasonable to assume that the vagaries of brain damage would result in uniform damage across all levels of the lexical access system. It is much more plausible that damage affects different parts of the system to different degrees, and that in some rare cases it results in highly selective damage to a single level or component of the system. Our patient D.M. and patients R.L. (Caplan et al., 1986) and J.B.N. (Hillis et al., 1999) would seem to be examples of such highly localized deficit to the speech production system. In the case of D.M., a plausible candidate is damage at the level where the phonological segments of a word are selected for production. This hypothesis is consistent with D.M.'s overall lexical processing performance, which is characterized by the production of phonological errors in all production tasks and essentially normal performance in semantic and syntactic processing tasks.

The clear dissociation between phonological and semantic errors found for D.M. and for R.L. (Caplan et al., 1986) complements the pattern of performance seen in patients such as R.G.B. (Caramazza & Hillis, 1990), K.E. (Hillis et al, 1990), D.P. (Cuetos, Aguado, & Caramazza, submitted), and P.W. (Rapp, Benzing, & Caramazza, 1997; Rapp & Caramazza, 1998; Rapp & Roldrick, in press). The latter patients produce many semantic errors but no phonological errors in naming tasks[5]. This performance profile falls well outside the range of performance predicted by Dell et al.'s model (see Figure 2), providing further evidence against the model (see also analysis of patient P.W. in Rapp & Goldrick, in press). Furthermore, for these patients it is possible to independently establish that the globality assumption is false. For example, patient R.G.B. (Caramazza & Hillis, 1990) produced semantic errors only in oral production tasks (naming and reading); he did not make

---

[5] See also case RCM (Hillis, Rapp, & Caramazza, 1999) who produced semantic errors in writing but not in speaking.

semantic errors in written production and performed flawlessly in all word comprehension tasks. This pattern of performance across tasks allows the inference that the semantic level is intact and that damage must be localized at the level where phonological lexical nodes are selected for production. Similarly, the pattern of performance of patient K.E. (Hillis et al., 1990), who made semantic but not phonological errors in all lexical comprehension and production tasks, allows the inference that the damage is restricted to the semantic level. Thus, D.M.'s performance dovetails with other fairly compelling, independent reasons for rejecting the globality assumption.

What are the implications of our results for the interactivity assumption? Here the issue is far more complex than the case of the globality assumption. As already noted, the evidence from naming does not support the specific interactive model implemented by Dell et al. (1997) and Foygel and Dell (in press). But we do not have independent evidence that the model's failure is due to its interactive nature and, therefore, we must remain agnostic on this issue. In any case, the evidence and arguments for the interactivity assumption are much richer and better motivated than those for the globality assumption (on this issue see Rapp & Goldrick, in press). Still, if one assumes that the fundamental shape of the model's possible error spaces is intrinsic to the model's core assumptions, then D.M.'s performance is of deep concern. If, as we have seen in the center and lower center panels of every data figure, a two-step interactive spreading-activation model must predict an association between increased nonword errors and increased formal and mixed errors, then cases in which a patient of moderate severity makes phonological substitutions resulting in many nonwords and few formal or mixed word errors seem to challenge those basic theoretical foundations.

Finally, there is a methodological moral that emerges from this study. To begin with, and to reiterate, the data from D.M. (but also R.L., J.B.N., R.G.B., K.E., D.P., P.W., and other well-studied cases) do not support the globality assumption. Note that this does not mean, however, that the globality assumption might <u>not</u> be true of some patients. It could very well be the case that some or perhaps even most fluent aphasics have sustained uniform damage (unlikely as this may seem) to all levels of the lexical access system. But from this we cannot draw the conclusion that the cause of naming deficits in <u>fluent aphasia</u> is a global lesion to the lexical access system. This is like saying that the cause of Broca's aphasia is damage to the syntactic mechanisms involved in processing transformationally moved elements (Grodzinsky, Pinango, Zurif and Drai, 1999)[6], or that the cause of acquired dyslexia is uniform damage to all levels of the reading system. A major problem with such pronouncements is that the categories over which the theoretical claims are made – fluent aphasia, Broca's aphasia, acquired dyslexia – are not homogeneous over the domain of interest. That is, there is nothing in the classification procedure of a patient as a fluent aphasic or a Broca's aphasic or an acquired dyslexic that guarantees that the cognitive or neural mechanisms that are damaged in a patient of a given clinical type are the same as those in another patient of the same clinical type. And indeed, as shown here, it is not hard to find patients with the same classification but whose underlying deficits are radically different one from another.

Is there any use for the globality assumption? Dell, Schwartz, Martin, Saffran, and Gagnon (in press) continue to suggest that globality may be useful as a modeling approximation even if it is not strictly true for all patients. First, in light of the methodological discussion above, we must ask what a model incorporating the globality assumption is supposed to represent. Since it cannot be said to represent all fluent aphasics, which patients is it intended to account for? If we restrict the scope of the model to fluent aphasics who are known to have suffered global damage, this begs the question of what theory and model we can use to diagnose this distinction

[6] See Caramazza, Capitani, Rey, & Berndt (in press) for a detailed critique of this claim.

in patients. Presumably we would need a way of classifying those patients whose behavior is incompatible with any combination of different degrees of damage to different levels of lexical access---a difficult task. If we cannot even recognize those patients to whom a model is intended to apply, the prospects for validating it are slim. Second, as we have seen, the globality assumption is counterproductive as a pragmatic modeling approximation. It makes accounting for patients more difficult without decreasing the complexity of the model. The effects of such a fundamental oversimplification are dramatic, as we saw in Figures 3 and 5. The diversity of possible error patterns are decreased enormously by insisting on global damage, preventing the model from matching any patient who makes many errors yet produces few nonword responses. And as Foygel and Dell (in press) and Ruml et al. (in press) have shown, modeling localized damage need not increase the number of free parameters in the model.

Dell et al.'s (1997) claim that the naming deficits in fluent aphasics are caused by global lesions to the lexical access system harks back to an approach to the study of cognitive disorders that privileged claims about clinical categories over an approach designed to provide strong tests of cognitive theory. Why this emphasis on clinical categories? For example, why the emphasis on the fact that most fluent aphasics produce various mixtures of semantic, formal, and phonological errors in naming tasks? A possible reason is that strongly interactive theories naturally predict associations of error types no matter where the lesion is placed in a system (e.g., Hinton & Shallice, 1991; Plaut & Shallice, 1993). Perhaps unsurprisingly, then, proponents of such theories may be more inclined to consider the performance of patients who show the expected associations (i.e., a syndrome). By contrast, many patterns of association of error types do not provide suitable data for constraining non-interactive theories of language and cognition. This is because these patterns of errors may simply reflect damage to multiple levels of a processing system and therefore the resulting combinations of error types may be too complex and ambiguous to play a determining role in deciding among theoretical alternatives. Note, however, that while a clear pattern of dissociation of error types can falsify a strong interactive model, a pattern of association of error types, or even very many such cases, is (are) not necessarily problematic for non-interactive or weakly interactive theories. Thus, the relative importance of these two types of evidence for testing claims about the structure of cognitive systems is not equal. To distinguish between theories we must seek the evidence that provides an adequate test of the theory. In the case of the globality assumption the evidence seems to be quite clear.

References

Benton, A.L., de Hamsher, K.D., Varney, N.R. & Spreen, O. (1983). Contributions to neuropsychological assessment: A clinical manual. Oxford University Press, Oxford.

Brazzelli, M., Capitani, E., Della sala, S., Spinnler, H. & Zuffi, M. (1994). A neuropsychological instrument adding to the description of patients with suspected cortical dementia: The Milan Overall Dementia Assessment. Journal of Neurology, Neurosurgery and Psychiatry, 57, 1510-1517.

Burani, C., Laudanna, A., Capasso, R. & Miceli, G. (unpublished). Batteria per la valutazione dei deficit di scrittura.

Burani, C., Laudanna, A., Miceli, G. (unpublished). Batteria per la valutazione dei disturbi di lettura.

Capasso, R. & Miceli, G. (unpublished). Batteria di denominazione per categorie-semantiche.

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

Caramazza, A. (2000). Aspects of lexical access: Evidence from aphasia. In Y. Grodzinsky, L. Shapiro, & D. Swinney (Eds.), Language and the Brain: Representation and Processing. San Diego: Academic Press.

Caramazza, A., Capitani, E., Rey, A., & Berndt, R.S. (in press). Agrammatic Broca's aphasia is not associated with a single pattern of comprehension performance. Brain and Language.

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

Caramazza, A., Ruml, W., Capasso, R., & Miceli, G. (in preparation). Patterns of naming errors in Italian fluent aphasics: A test of interactive activation models of naming.

Cuetos, F., Aguado, G., & Caramazza, A. (submitted). Dissociation of semantic and phonological errors in naming.

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

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

Dell, G. Schwartz, M. F., Martin, N., Saffran, E.M., & Gagnon, D.A. (in press). The role of computational models in neuropsychological investigations of language: Reply to Ruml and Caramazza. Psychological Review.

De Renzi, E. & Faglioni, P. (1978). Normative data and screening power of a shortened version of the Token Test. Cortex, 14, 41-49.

Foygel, D., & Dell, G.S. (in press). Models of impaired lexical access in speech production. Journal of Memory and Language.

Grodzinsky, Y., Pinango, M. M., Zurif, E., & Drai, D. 1999. The critical role of group studies in neuropsychology: Comprehension regularities in Broca's aphasia. Brain and Language, 67, 134-147.

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., Rapp, B., & Caramazza, A. (1999). When a rose is a rose in speech but a tulip in writing. Cortex, 35, 337-356.

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

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

Novelli, G., Papagno, C., Capitani, E., Laiacona, M., Cappa, S.F. & Vallar, G. (1986). Tre test clinici di memoria a lungo termine verbale: taratura su soggetti normali. Archivio di Neurologia, Psichiatria e Psicologia, 47, 278-295.

Orsini, A., Grossi, D., Capitani, E., Laiacona, M., Papagno, C., & Vallar, G. (1987). Verbal and spatial immediate memory span: Normative data from 1355 adults and 1112 children. The Italian Journal of Neurological Sciences, 8, 539-548.

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

Rapp, B., Benzing, L. & Caramazza, A. (1997). The autonomy of lexical orthography. Cognitive Neuropsychology, 14, 71-104.

Rapp, B., & Caramazza, A. (1998). A case of selective difficulty in writing verbs. Neurocase, 4, 127-140.

Rapp, B. & Goldrick, M. (in press). Discreteness and interactivity in spoken word production. Psychological Review.

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.

Ruml, W., & Caramazza, A. (in press). An evaluation of a computational model of lexical access: Comments on Dell et al. (1997). Psychological Review.

Ruml, W., Caramazza, A., Shelton, J.R., & Chialant, D. (in press). Testing assumptions in computational theories of aphasia. Journal of Memory and Language.

Saffran, E.M., Dell, G., & Schwartz, M.F. (2000). Computational modeling of language disorders. In M.S. Gazzaniga (Ed.), The new cognitive neurosciences, 2nd Edition. Cambridge: MIT Press.

Schuell, H. & Jenkins, J.J. (1959). The nature of language deficit in aphasia. Psychological Review, 66, 45-67.

Snodgrass, J., & Vanderwart, M. (1980). A standardised set of 260 pictures: Norms for name agreement, familiarity, and visual complexity. Journal of Experimental Psychology: Human Learning and Memory, 6, 174-215.

Wilshire, C.E. & McCarthy, R.A. (1996). Experiemental investigations of an impairment in phonological encoding. Cognitive Neuropsychology, 13, 1059-1096.

Wing, A.M. & Baddeley, A.D. (1980). Spelling errors in handwriting: A corpus and a distribution analysis. In U. Frith (Ed.), Cognitive processes in spelling. London: Academic Press.

Acknowledgments

The research reported here was supported in part by NIH grant NS 22201 and by the National Science Foundation under grants CDA-94-01024 and IRI-9618848. We thank D.M. for his patience and good cheer during testing.

Figure Captions

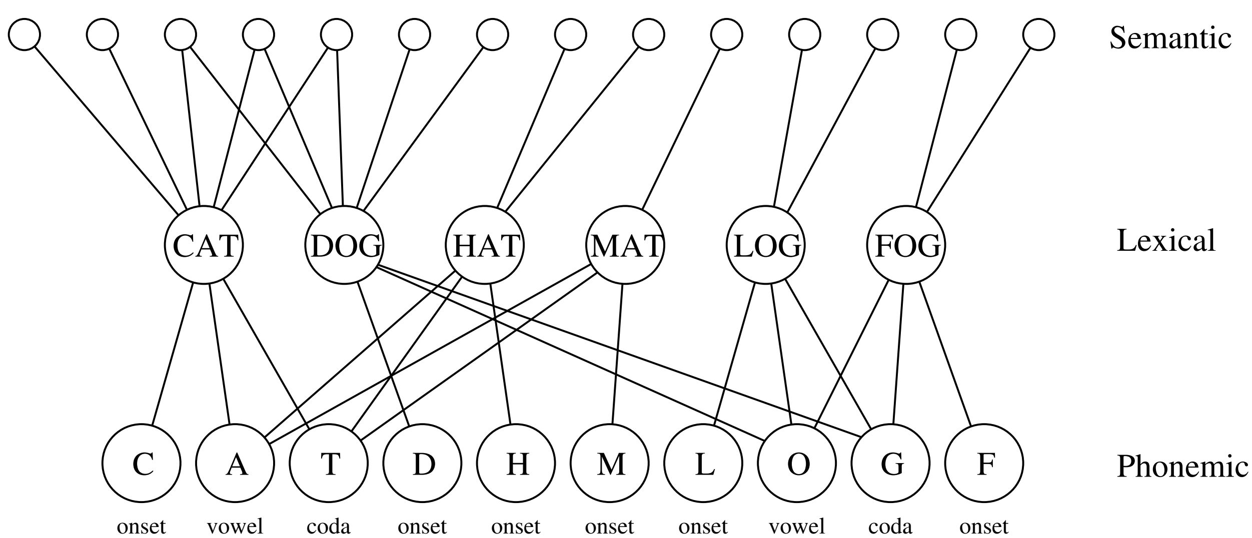
1. The structure of part of Dell et al.'s (1997) model of lexical access.

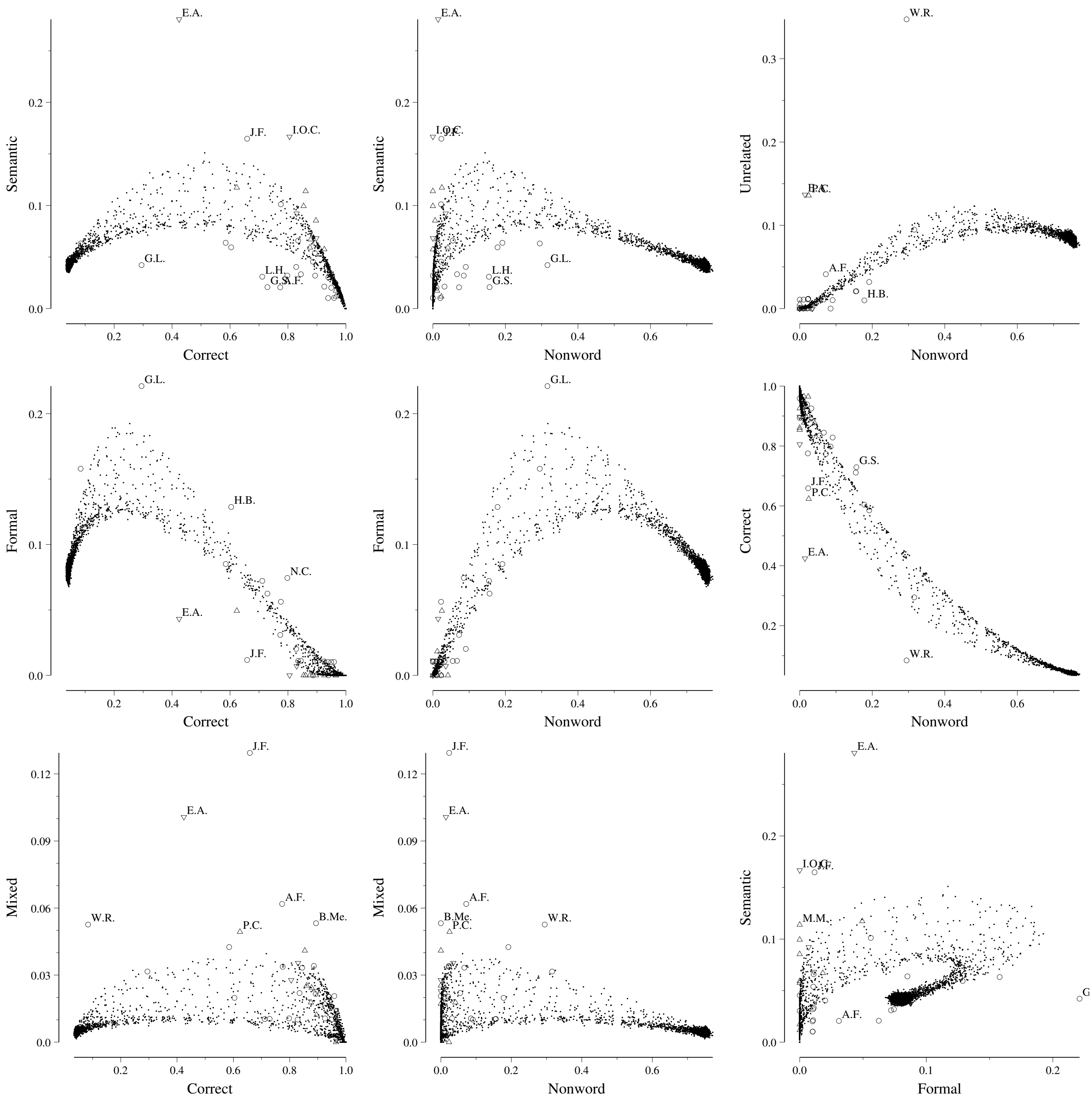
2. Views of error distributions. Dots represent possible patterns achievable by Dell et al.'s model of English under the assumption of global damage. Circles represent patients reported by Dell et al. and triangles represent patients reported by Ruml et al. (in press).

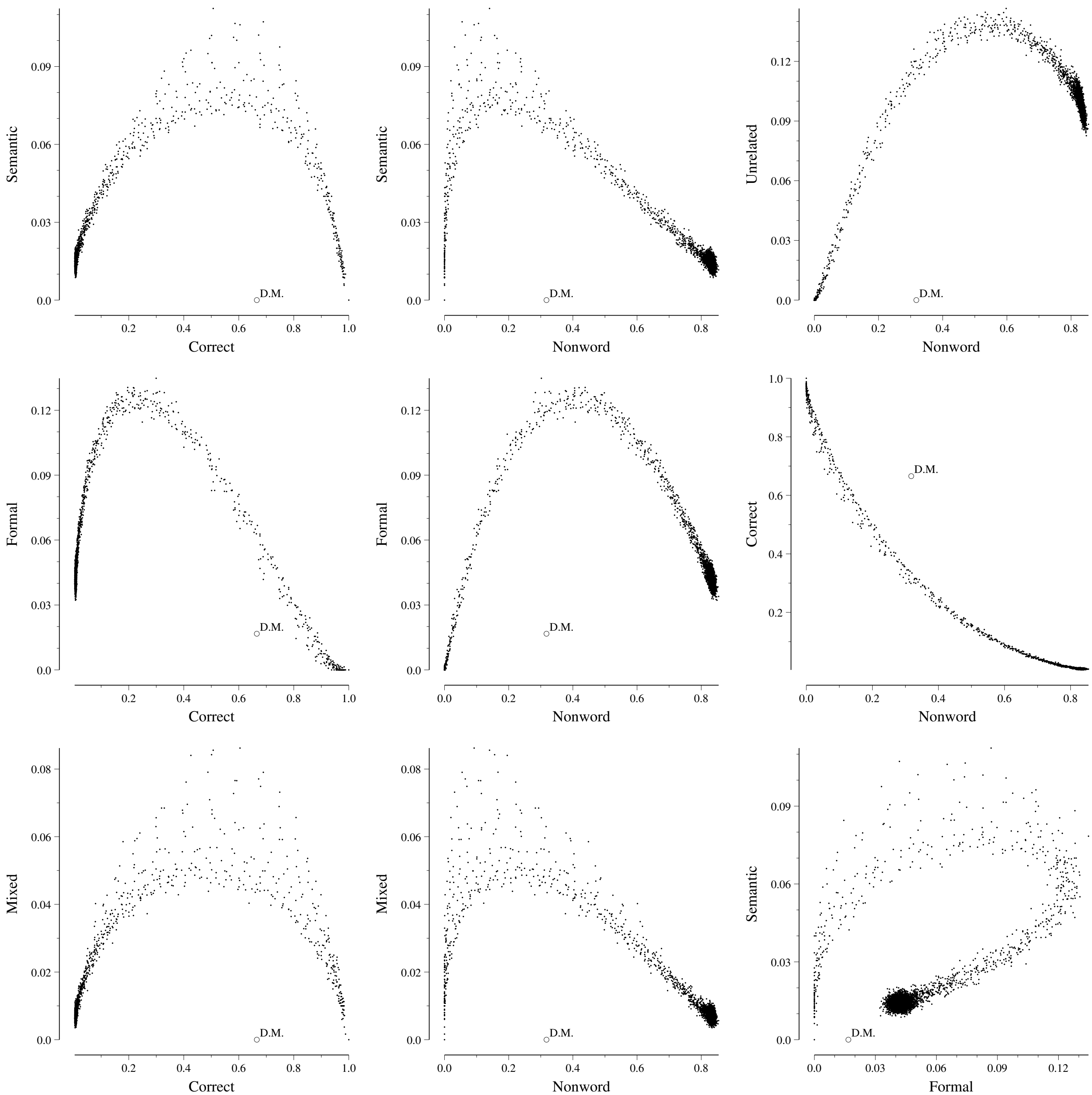
3. Error patterns that can be achieved using our Italian analogue of Dell et al.'s (1997) model under the assumption of global damage. The performance of patient D.M. is shown for comparison.

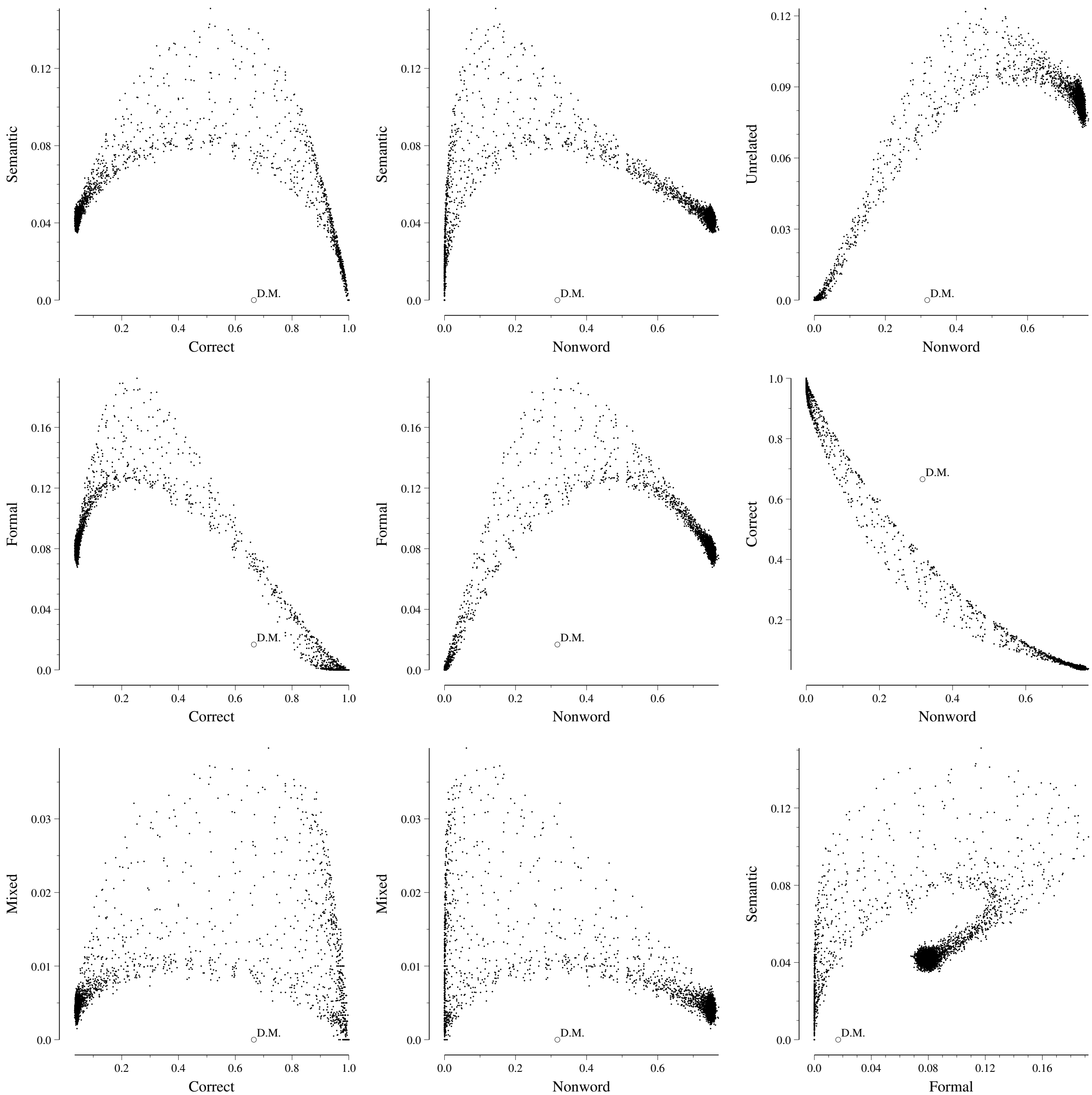
4. The error patterns of Dell et al.'s (1997) English model, with patient D.M. shown for comparison.

5. Error patterns achievable by our Italian model using Foygel and Dell's (in press) assumption of local damage. Patient D.M. is shown for comparison.


Semantic
Lexical
Phonemic
CAT
DOG
HAT
MAT
LOG
FOG
C
A
T
D
H
M
L
O
G
F
onset
vowel
coda
onset
onset
onset
onset
vowel
coda
onset


Semantic
Correct
Semantic
Nonword
Unrelated
Nonword
Formal
Correct
Formal
Nonword
Correct
Nonword
Mixed
Correct
Mixed
Nonword
Semantic
Formal
E.A.
J.F.
I.O.C.
G.L.
L.H.
G.S.
A.F.
W.R.
P.C.
H.B.
N.C.
B.Me.
M.M.


Semantic
Correct
D.M.
Semantic
Nonword
D.M.
Unrelated
Nonword
D.M.
Formal
Correct
D.M.
Formal
Nonword
D.M.
Correct
Nonword
D.M.
Mixed
Correct
D.M.
Mixed
Nonword
D.M.
Semantic
Formal
D.M.


Semantic
Correct
D.M.
Semantic
Nonword
D.M.
Unrelated
Nonword
D.M.
Formal
Correct
D.M.
Formal
Nonword
D.M.
Correct
Nonword
D.M.
Mixed
Correct
D.M.
Mixed
Nonword
D.M.
Semantic
Formal
D.M.

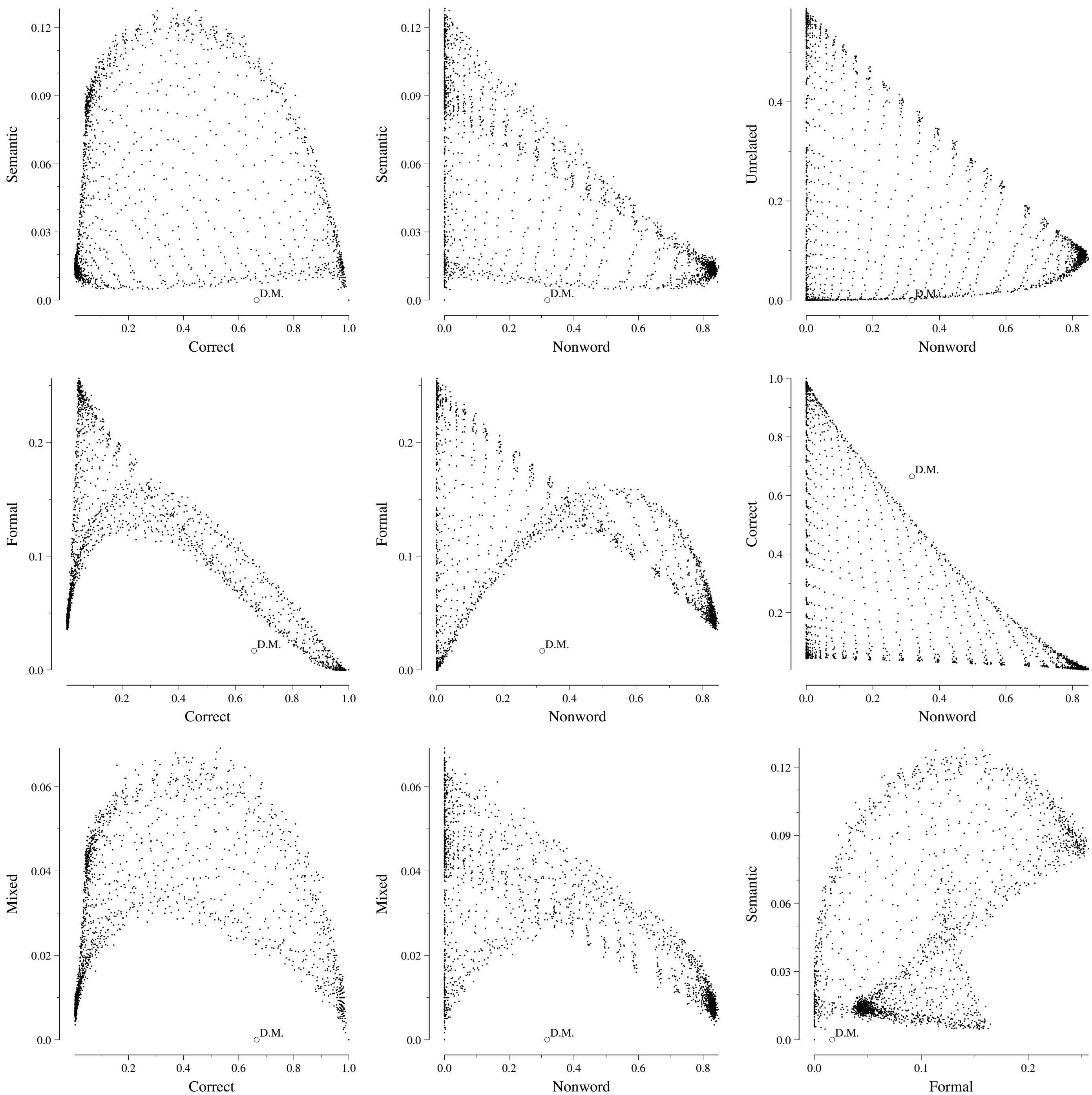

Semantic
Correct
D.M.
Semantic
Nonword
D.M.
Unrelated
Nonword
D.M.
Formal
Correct
D.M.
Formal
Nonword
D.M.
Correct
Nonword
D.M.
Mixed
Correct
D.M.
Mixed
Nonword
D.M.
Semantic
Formal
D.M.

TABLE 1

TABLE 1a. GENERAL NEUROPSYCHOLOGICAL ASSESSMENT

	DM's performance	cut-off
MODA	92.7	>89
Raven Coloured Progressive Matrices	24.5	≥18
Digit span	3	≥3.75
Corsi span	4	≥3.50
Short Story	12	≥8
Spatial learning	9.21	>5
Facial apraxia	20/20	≥17
Ideomotor apraxia	65/72	>53
Constructional apraxia	18/20	>16
Benton Line Orientation Test	23/30	>17

TABLE 1b. LANGUAGE EXAMINATION

	Number of correct responses
Phonological discrimination	60/60
Repetition	
Nonwords	30/36
Words	34/45
Sentences	10/36
Auditory-verbal comprehension	39/40
Written comprehension	40/40
Object Naming	20/30
Action Naming	18/28
Syntactic comprehension	
Auditory	54/60
Written	60/60
Token Test	26/36
Writing to dictation	
Words	101/198
Nonwords	17/80
Reading	
Words	235/272
Nonwords	61/122

TABLE 2 - DISTRIBUTION OF NONWORD ERRORS IN NAMING

Neologisms (unrecognizable target word)	16
Single errors	
Substitution	31
Addition	1
Omission	8
Transposition	5
Multiple errors	
Substitution	17
Omission	4
Mixed errors	
Substitution + Omission	2
Substitution + Omission + Inversion	1

TABLE 3

TABLE 3a – NUMBER OF ERRORS FOR EACH OF THE FIVE POSITIONS IN NAMING (NONWORD RESPONSES ONLY)

Position	A	B	C	D	E
Number of errors	14	31	29	28	4

TABLE 3b – NUMBER OF ERRORS FOR EACH OF THE FIVE POSITIONS IN NAMING (LAST LETTER NOT INCLUDED)

Position	A	B	C	D	E
Number of errors	7	16	31	11	12

TABLE 4: FIT OF GLOBAL DAMAGE MODEL TO D.M.

Source	Corr.	Sem.	Form.	Mix.	Unr.	N W	RMSD	X^2	p
Patient D.M.	0.67	0.00	0.02	0.00	0.00	0.32			
Model*	0.55	0.07	0.09	0.05	0.07	0.18	0.092	111	0.000

*The parameter values used for this model were 0.006 for connection strength and 0.51 for decay.

TABLE 5: FIT OF LOCAL DAMAGE MODEL TO D.M.

Source	Corr.	Sem.	Form.	Mix.	Unr.	N W	RMSD	X^2	p
Patient D.M.	0.67	0.00	0.02	0.00	0.00	0.32			
Model*	0.59	0.01	0.11	0.03	0.00	0.27	0.054	38.9	0.000

*The parameter values used for this model were 0.025 for connection strength of semantic to lexical layer 0.004 for connection strength for lexical to segmental layer.