

This article was downloaded by: [University of Iowa Libraries]

On: 01 April 2014, At: 10:03

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

Cognitive Neuropsychology

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/pcgn20>

Models of errors of omission in aphasic naming

Gary S. Dell^a, Elisa N. Lawler^b, Harlan D. Harris^c & Jean K. Gordon^d

^a University of Illinois at Urbana-Champaign, USA

^b Georgia Institute of Technology, Atlanta, USA

^c University of Illinois at Urbana-Champaign, USA

^d University of Iowa, Iowa, USA

Published online: 28 Nov 2010.

To cite this article: Gary S. Dell, Elisa N. Lawler, Harlan D. Harris & Jean K. Gordon (2004) Models of errors of omission in aphasic naming, *Cognitive Neuropsychology*, 21:2-4, 125-145

To link to this article: <http://dx.doi.org/10.1080/02643290342000320>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

MODELS OF ERRORS OF OMISSION IN APHASIC NAMING

Gary S. Dell

University of Illinois at Urbana-Champaign, USA

Elisa N. Lawler

Georgia Institute of Technology, Atlanta, USA

Harlan D. Harris

University of Illinois at Urbana-Champaign, USA

Jean K. Gordon

University of Iowa, Iowa, USA

Five computational models of lexical access during production are tested for their ability to account for the distribution of aphasic picture-naming errors. The naming profiles ($N = 14$) were chosen from the literature to represent patients who make a relatively large number of omission errors. The most successful models combined the damage assumptions of the semantic-phonological model of lexical access (Foygel & Dell, 2000) with a treatment of omission errors as largely independent from overt errors (Ruml, Caramazza, Shelton, & Chialant, 2000). An explanation for the occurrence of omission errors was provided by the addition of a lexical-threshold parameter (Laine, Tikkala, & Juhola, 1998) to the model. Suggestions for further testing of these models are introduced, as is a new website that allows other researchers to make use of the models.

INTRODUCTION

Impaired word retrieval in aphasia leads to a variety of speech errors, all of which provide insight into the mechanisms of lexical access during speaking. Most research on aphasic errors focuses on *paraphasias*, errors of commission as opposed to errors of omission. Often, paraphasic utterances bear a semantic relation (e.g., “banana” instead of *strawberry*) or phonological relation (e.g., “treen” for *train*) to the intended word. These errors are

studied, in part, because they provide specific information about hypothesised components of the lexical system (e.g., Buckingham, 1987; Caramazza & Hillis, 1990; Gagnon, Schwartz, Martin, Dell, & Saffran, 1997; Howard & Orchard-Lisle, 1984; Kay & Ellis, 1987; N. Martin & Saffran, 1997; Mitchum, Ritgert, Sandson, & Berndt, 1990; Nickels & Howard, 1995; Rapp & Goldrick, 2000).

In addition to overt errors, aphasic speakers also make errors of omission. These errors, commonly

Correspondence should be addressed to Gary S. Dell, Beckman Institute, University of Illinois, 405 N. Mathews Ave., Urbana IL 61801, USA (Email: gdell@s.psych.uiuc.edu).

This research was supported by DC-00191 (NIH), SBR 98-73450 (NSF), and a post-doctoral award from the National Sciences and Research Council of Canada to Jean Gordon. The authors wish to thank Myrna Schwartz and Adelyn Brecher for providing the breakdowns of the response categories for the profiles from Schwartz and Brecher (2000), and Eleanor Saffran, Nadine Martin, Zenzi Griffin, Prahlad Gupta, Karin Humphreys, Marlene Behrmann, Max Coltheart, and an anonymous reviewer for helpful comments on this work, and Judy Allen for work on the manuscript.

referred to as word-finding blocks, occur in all production tasks, and can be characterised by silence or by evidence of an unsuccessful word search, such as saying "I know what it is, but I can't say it," producing aborted attempts, or describing semantic features of the sought-after word, often called circumlocution (e.g., "It's some kind of animal"). These retrieval blocks have rarely been studied, and for good reason: Without an overt attempt to produce the target, it is difficult to come to any conclusions about the mental processes that are occurring. Nevertheless, clinical observations of word-searching behaviour have allowed some tentative hypotheses to be put forth. For example, comments such as "It's a, no, that's not it..." suggest that speakers sometimes inhibit incorrect responses that come to mind. Descriptions indicate that some of the target's semantic information is available, but not its phonological form. Unfortunately, word-finding difficulties frequently occur without such clues as to their origin.

In picture naming studies, trials on which subjects do not produce a naming attempt are typically excluded from analysis. Yet, such events are common. For example, in Mitchum et al.'s (1990) study of 28 unselected aphasic patients, no response occurred on 11% of naming trials, and for 7 of the patients, no response was the most common error. Mitchum et al.'s no-response category included silence, statements of no response (e.g., "I don't know"), and empty comments (e.g., "I have one of those"). In the studies providing the data for this article, the no-response category was expanded into a general non-naming-response category by including, in addition to silence, all responses in which the individual spoke but was not making an attempt to give the single-word name of the pictured object. For example, semantic descriptions were included.

This study seeks to understand errors of omission by developing computational models of naming and relating those models to aphasic data. Specifically, there are two goals. The first is to expand the data to which existing models can be applied. Current models make precise predictions only for errors of commission. By extending their application to non-naming responses, we can

better discriminate among the models. The second goal is to begin to investigate the omission process itself, by formalising specific hypotheses. To realise these goals, we begin with two existing models that attribute errors of commission to pathological spreading activation within a lexical network: the weight-decay model of Dell, Schwartz, Martin, Saffran, and Gagnon (1997) and the semantic-phonological model of Foygel and Dell (2000). Both have been applied, with varying success, to aphasic naming and repetition errors (Caramazza, Papagno, & Rumel, 2000; Croot, Patterson, & Hodges, 1998; Cuetos, Aguado, & Caramazza, 2000; Dell, Schwartz, Martin, Saffran, & Gagnon, 2000; Gordon, 2002; Hanley, Kay, & Edwards, 2002; Hillis, Boatman, Hart, & Gordon, 1999; Rumel & Caramazza, 2000; Rumel, Caramazza, Shelton, & Chialant, 2000; Schwartz & Brecher, 2000; Schwartz & Hodgson, 2002). However, both models lack an account of non-naming responses, and have consequently only been tested using data sets in which such errors are uncommon.

To address this limitation, we consider three accounts of omission errors gleaned from the literature, the independence model (Rumel et al., 2000), the lexical-editor model (Baars, Motley, & MacKay, 1975), and the lexical-threshold model (Laine, Tikkala, & Juhola, 1998). Each of these is grafted onto the existing naming models to create compound models that can, in principle, explain both errors of commission and omission. The adequacy of these compound models is assessed by matching their performance to the naming profiles of aphasic individuals who make many non-naming responses.

The contrast between errors of commission and errors of omission applies to nonaphasic as well as aphasic speech. Overt speech errors in normal speakers are studied in analyses of natural error corpora (e.g., Garrett, 1975) and experimental tasks that create slips (e.g., Ferreira & Humphreys, 2001). Errors of omission in normal speakers are studied through "tip-of-the-tongue" states that often accompany such word-finding blocks (e.g., Burke, MacKay, Worthley, & Wade, 1991; Harley & Bown, 1998; Meyer & Bock, 1992; Miozzo &

Caramazza, 1997; Vigliocco, Antonini, & Garrett, 1997). Theories of normal production provide informal accounts of both omission and commission errors, but there has been no attempt to give a unified computational account of both (see, however, Levelt, Roelofs, & Meyer, 1999; MacKay, 1987, for work in this direction).

Models of non-naming responses

Independence model. The simplest treatment of errors of omission is to assume that they are independent of the processes that generate other responses (Ruml et al., 2000). The independence model is statistical rather than mechanistic; it specifies how errors distribute when non-naming responses occur, but does not attribute those responses to any particular processing system. To illustrate, consider a sample pattern without non-naming responses: 50% correct responses, 10% each of semantic, formally related, and unrelated word errors, and 20% nonword errors. Next, assume that an unspecified process converts a fraction of the overt responses to non-naming responses. If the independence model is correct, this conversion would leave the relative proportions of each of the overt response categories unchanged. For example, assuming 40% non-naming responses, the remaining categories would each be reduced by that percentage, to 30% correct, 6% semantic, 6% formal, 6% unrelated, and 12% nonwords. In essence, the independent non-naming process “steals” from the overt categories in proportion to their frequency.

Lexical-editor model. Speakers appear to monitor their planned output, suppressing it if it is linguistically deviant (Garnsey & Dell, 1984; Hartsuiker & Kolk, 2001; Levelt, 1983; Postma, 2000). The most commonly hypothesised process of this sort is the lexical editor. Baars et al. (1975) demonstrated that phonological exchange errors are less likely when they create nonwords, suggesting that potential nonword slips are caught and either suppressed or corrected before they are uttered (see also Dell, 1986; Dell & Reich, 1981; Humphreys & Swenson, 2002). This kind of editorial process may be

responsible for many of the non-naming responses that aphasic patients produce (Mitchum et al., 1990). For example, in Schwartz and Brecher's (2000) study of recovery in naming, patient AK had a relatively low rate of nonword errors given the patient's low correctness, along with many no responses, suggesting that potential nonlexical output was converted to non-naming responses. Instead of converting responses from all overt categories like the independence model, the lexical-editor model converts only potential nonword responses.

Lexical-threshold model. The alternative to the claim that omission errors are caused by an editor suppressing deviant responses is that omissions are the consequence of a failure to retrieve any lexical item. This idea is implicit in accounts of the tip-of-the-tongue state (e.g., Burke et al., 1991) and was made concrete in a connectionist production model developed by Laine et al. (1998). The crucial component of this model for our purposes is the assumption that, to be spoken, words must exceed a minimum level of activation. Word units are retrieved by spreading activation, but if no unit's activation exceeds a specified threshold value, the model generates no response.

Interactive two-step models of lexical access in production

The weight-decay and semantic-phonological models of aphasic naming are derived from the interactive two-step model of lexical access (Dell & O'Seaghdha, 1991; N. Martin, Dell, Saffran, & Schwartz, 1994). Similar models in this domain include the restricted interaction model of Rapp and Goldrick (2000), the node-structure theory of MacKay (1987), and interactive activation models developed by Berg and Schade (1992; also Schade & Eikmeyer, 1998), Harley (1984, 1993), and Stemmer (1985). These are all associated with a bidirectional or interactive flow of activation, distinguishing them from discrete-stage (e.g., Levelt et al., 1999) and cascaded-stage theories (e.g., Caramazza, 1997; Cutting & Ferreira, 1999). First, we present the interactive two-step model's

treatment of unimpaired processing, followed by the aphasia models derived from it.

Normal lexical access. Lexical access occurs through spreading activation in a network such as that shown in Figure 1. The network contains units for semantic features, words, and phonemes. Top-down excitatory connections link semantic features to words, and words to phonemes. Bottom-up excitatory connections do the reverse, thus providing interactive feedback from later to earlier levels. There are no inhibitory connections.

The “two-step” aspect of the model refers to distinct word-access and phonological access steps, a feature of many production models (e.g., Bock & Levelt, 1994; Griffin & Bock, 1998; Levelt et al., 1999; Rapp & Goldrick, 2000). Word access begins with a jolt of 10 units of activation to each of the target’s semantic features. Activation then spreads through the network for eight time steps, according to a noisy linear activation rule:

$$A_{j,t} = A_{j,t-1}(1 - \text{decay}) + \sum w_{ij}A_{i,t-1} + \text{noise}$$

where $A_{j,t}$ is the activation of unit j at time step t , *decay* is the decay rate, and w_{ij} is the connection weight from the source unit i to the receiving unit j . During each time step, each unit’s activation level is perturbed by normally distributed noise, specifically the sum of two components, intrinsic noise ($SD = .01$) and activation noise ($SD = .16A_{j,t}$). After eight time steps, the most highly activated word unit of the appropriate syntactic category is selected, completing the word access stage.

During word access, the target (e.g., CAT) will normally obtain the most activation. However, a consequence of spreading activation is that other words, their phonemes, and their semantic features become active as well. Semantic neighbours (e.g., DOG) obtain activation from shared semantic features. Furthermore, because activation spreads from words to phonemes during word access, and feeds back from phonemes to words, phonologically related (or formal) neighbours such as MAT are also activated. Mixed semantic-phonological neighbours such as RAT are especially activated,

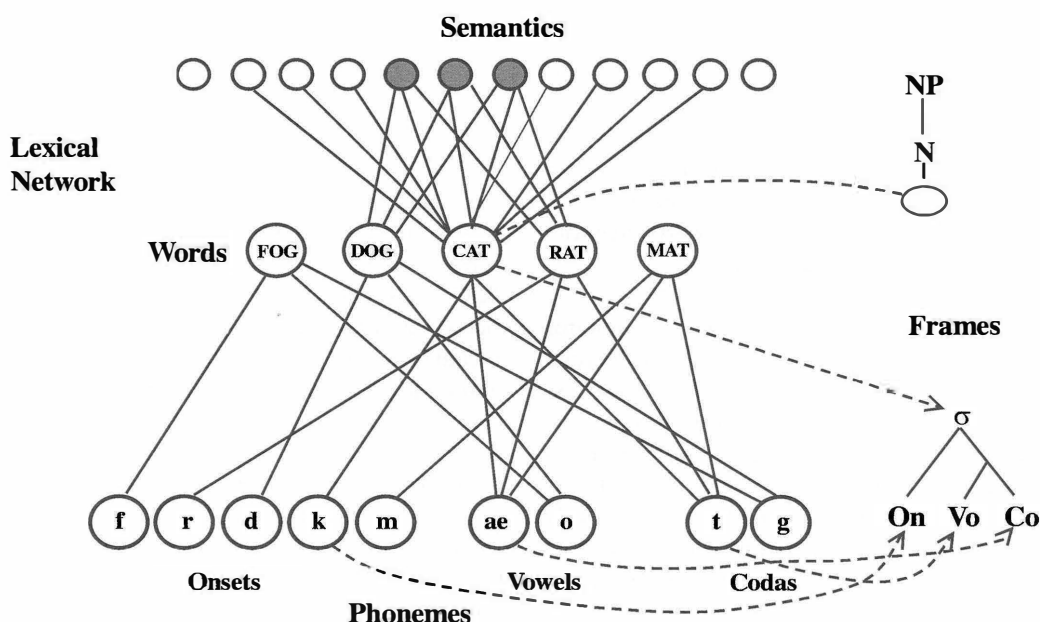


Figure 1. The two-step interactive activation approach to lexical access in spoken word production, adapted from Dell et al. (1997).

gaining activation from both top-down and bottom-up influences, leading to the “mixed-error effect,” (e.g., N. Martin, Gagnon, Schwartz, Dell, & Saffran, 1996). Due to noise, any of these related words may be selected instead of the target. In extreme cases, an unrelated (LOG) word could also be incorrectly chosen.

Phonological access begins with a 100-unit jolt of activation to the selected word unit, and activation again spreads throughout the network, both in a top-down and bottom-up fashion, for eight more time steps. Finally, the most activated phonemes are selected and linked to slots in a phonological frame. The implemented model encodes only single-syllable CVC words. The possible phonemes are organised into sets of onsets, vowels, and codas, and the most active of each set is chosen and linked to the appropriate slot in a CVC frame. If one or more of these phonemes are not in the selected word, a phonological error—either real-word or nonword—occurs.

For the purposes of model evaluation, six categories were considered: *correct responses* (CAT); *semantic errors* (DOG); *formal errors* (MAT); *mixed errors* (RAT); *unrelated word errors* (LOG); and *nonword errors* (LAT). The model was initially set up to simulate the probability of these six categories in normal speakers’ picture naming, based on the 175-item Philadelphia Naming Test (PNT; Roach, Schwartz, Martin, Grewal, & Brecher, 1996). This required specifying the model’s lexical neighbourhoods and its connection weight and decay parameters (see Dell et al., 1997, for details).¹

Aphasic lexical access. Dell et al. (1997) hypothesised that the error patterns produced by aphasic subjects fall on a continuum between the normal error pattern and a “random pattern.” The random pattern is the proportion of each response category that would be observed if there were no lexical

influence on word-form retrieval, but the retrieved forms nonetheless respected the general phonological characteristics of (English) words. Dell et al. estimated the random pattern for English from a number of sources in order to reflect the opportunities afforded by the lexicon, such as the relative numbers of semantic and formal neighbours of words, and the proportion of phonologically legal strings that are words. The model’s neighbourhoods were then constructed to simulate this estimate.

The continuum of possible response patterns from normal to random, referred to as *the continuity thesis*, constrains the error patterns that can occur in aphasia. With six response categories in the model, a particular pattern represents one point in a six-dimensional space, with the normal pattern at one end of the space and the random pattern at the other. For the continuity thesis to hold, the set of possible patterns should correspond to a simple continuous region connecting these two points. The weight-decay (Dell et al., 1997) and semantic-phonological (Foygel & Dell, 2000) models are methods of generating this region. Each model specifies how the normal model can be “lesioned,” and this lesioning defines the region between the normal and random points. The semantic-phonological model is lesioned by reducing one or both of two parameters: s , the weight of the top-down and bottom-up connections between semantic features and words, and p , the weight of the corresponding connections between words and phonemes. The weight-decay model also has two lesionable parameters, a single weight parameter, w , which applies to both semantic and phonological connections, and *decay*.²

The two models were compared in a study of 21 patients (Dell et al., 1997; Foygel & Dell, 2000) and in another with 9 patients (Ruml et al., 2000). In these studies, both models performed fairly well in fitting the variety of patterns across the patients.

¹ Since the model’s development, a small change has been made to the lexical neighbourhoods in response to observations that mixed errors were being under-predicted (Foygel & Dell, 2000; Rapp & Goldrick, 2000). In the current version, the chance opportunity for mixed errors has been increased from .004 to .008.

² Decay is set to a constant 0.6 for the semantic-phonological model.

The semantic-phonological model may be theoretically more valid, however, because the ability to create separate semantic and phonological lesions can better explain reported case studies of subjects demonstrating almost exclusively one or the other type of error (Rapp & Goldrick, 2000; Ruml et al., 2000). Moreover, this distinction allows the semantic-phonological model to make more accurate predictions about word repetition in aphasia. Repetition is assumed to depend more on lexical-phonological than lexical-semantic weights, since retrieval from semantics is not required (Foygel & Dell, 2000).

In the current study, we have an opportunity to test the effectiveness of the weight-decay and semantic-phonological models in explaining patterns of errors that previously could not be accounted for by either model; those of aphasic subjects who make many non-naming responses. In our study, we define naming models that account for non-naming responses in different ways, by combining the weight-decay and semantic-phonological models with the independence, lexical-editor, and threshold accounts, and assessing the fit of each to the error profiles of selected aphasic patients from the literature (Dell et al., 1997; Gordon, 2002; Ruml et al., 2000; Schwartz & Brecher, 2000).

MODEL COMPARISON STUDY

Methods

Data sets. Fourteen profiles were chosen from published studies of aphasic naming that used the PNT and its scoring standards. A profile consists of the proportions of correct, semantic, formal, mixed, unrelated, nonword, and non-naming responses to the 175 pictures of the PNT for a single patient. Included in the present study were all profiles from the four articles mentioned above that contained at least 18% non-naming responses,

or in which the non-naming responses comprised at least 80% of all erroneous responses.³ The profiles were those of subjects GB and VP from Dell et al. (1997); IOC, JR1, EA, and JR2 from Ruml et al. (2000); S10, S16, S21, and S31 from Gordon (2002); and WR5, AS1, AK1, and AK2 from Schwartz and Brecher (2000). (Note that JR1 and JR2 are two different individuals, but AK1 and AK2 are profiles collected from the same person at two different times.) Information on these subjects is given in Table 1 and their naming profiles are presented in italic type in Table 2. Other background information can be obtained from the original publications.

The non-naming category comprises all responses in which the subject fails to provide a valid naming attempt. This includes null responses, in which the subject says nothing or otherwise indicates that he or she cannot name the picture, as well as overt responses that do not constitute a single-word attempt to name the target, such as

Table 1. Characteristics of subjects contributing naming profiles

Subject	Age (yrs)	Months post-onset	Education (yrs)
GB ^a	83	4.1	12
VP ^a	64	60.0	12+
IOC ^b	55	18.0	14
JR1 ^b	50	36.0	12
EA ^b	65	216.0	15
JR2 ^b	43	60.0	16
S10 ^c	81	48.0	12
S16 ^c	65	10.0	9
S21 ^c	67	62.0	9
S31 ^c	53	129.0	6
WR5 ^d	62	14.5	12
AS1 ^d	61	9.0	N/A
AK1 ^d	77	14.0	N/A
AK2 ^d	77	17.0	N/A

^a From Dell et al. (1997).

^b From Ruml et al. (2000).

^c From Gordon (2002).

^d From Schwartz and Brecher (2000).

³ We originally required 20% non-naming responses, but then relaxed the criteria to include more subjects.

Table 2. Comparison of five models

Patient	Model	Correct	Semantic	Formal	Mixed	Unrelated	Nonword	Non-naming	Combined	RMSD ^a	χ^2	s / w	p / decay
GB		.39	.07	.09	.01	.03	.08	.32	.40				
	WD-ind	.39	.07	.07	.03	.03	.10	.32		.011	2.31	.044	.695
	SP-ind	.40	.07	.06	.02	.04	.10	.32		.014	3.40	.017	.021
	WD-editor	.36	.08	.11	.02	.08			.35	.032	8.53	.004	.502
	SP-editor	.41	.05	.09	.02	.03			.39	.012	1.84	.021	.012
	SP-thresh ($t = .019$)	.40	.07	.07	.02	.06	.08	.31		.015	4.37	.013	.023
VP		.28	.07	.11	.05	.17	.04	.28	.32				
	WD-ind	.22	.08	.13	.04	.06	.19	.28		.073	56.55	.085	.870
	SP-ind	.28	.11	.14	.02	.14	.04	.28		.024	10.71	.008	.028
	WD-editor	.26	.10	.16	.04	.10			.34	.039	13.93	.055	.764
	SP-editor	.27	.10	.15	.02	.14			.32	.026	9.61	.009	.015
	SP-thresh ($t = .014$)	.27	.10	.14	.02	.14	.04	.29		.022	9.25	.007	.029
IOC		.17	.03	.00	.01	.00	.00	.79	.79				
	WD-ind	.16	.02	.00	.01	.00	.01	.79		.007	4.79	.096	.830
	SP-ind	.16	.02	.01	.01	.01	.00	.79		.008	6.36	.019	.035
	WD-editor	.10	.06	.10	.01	.09			.63	.091	50.41	.002	.533
	SP-editor	.16	.03	.08	.02	.03			.68	.059	24.36	.091	.001
	SP-thresh ($t = .030$)	.18	.02	.01	.01	.01	.00	.78		.012	7.23	.011	.041
JR1		.45	.03	.01	.01	.00	.00	.50	.50				
	WD-ind	.45	.03	.00	.01	.00	.01	.50		.004	1.50	.056	.688
	SP-ind	.46	.02	.01	.01	.00	.00	.50		.005	1.93	.025	.044
	WD-editor	.36	.08	.11	.02	.08			.35	.090	50.57	.004	.502
	SP-editor	.45	.03	.04	.03	.01			.45	.025	9.09	.099	.002
	SP-thresh ($t = .032$)	.45	.02	.02	.01	.01	.00	.49		.007	3.78	.015	.055
EA		.34	.22	.03	.08	.11	.01	.21	.22				
	WD-ind	.34	.10	.12	.06	.05	.14	.21		.078	74.23	.090	.871
	SP-ind	.37	.12	.14	.03	.14	.01	.21		.062	52.89	.009	.036
	WD-editor	.35	.12	.18	.07	.07			.21	.073	38.54	.097	.904
	SP-editor	.36	.13	.16	.03	.16			.17	.075	53.43	.009	.021
	SP-thresh ($t = .013$)	.35	.12	.16	.03	.15	.01	.19		.066	52.18	.008	.039

(continued overleaf)

Table 2. *Continued*

Patient	Model	Correct	Semantic	Formal	Mixed	Unrelated	Nonword	Non-naming	Combined	RMSD ^a	χ^2	s / w	p / decay
JR2		.67	.07	.01	.03	.00	.03	.19	.22				
	WD-ind	.67	.06	.01	.04	.00	.02	.19		.008	2.86	.075	.761
	SP-ind	.68	.05	.03	.01	.01	.02	.19		.014	8.86	.024	.029
	WD-editor	.63	.08	.07	.02	.04			.17	.040	21.69	.005	.504
	SP-editor	.74	.03	.01	.02	.00			.20	.034	10.05	.100	.004
	SP-thresh ($t = .025$)	.68	.05	.03	.02	.02	.03	.19		.014	9.07	.020	.030
S31		.04	.09	.05	.03	.07	.03	.68	.71				
	WD-ind	.08	.03	.06	.02	.03	.09	.68		.039	39.68	.095	.914
	SP-ind	.06	.05	.08	.01	.09	.03	.68		.023	16.13	.001	.025
	WD-editor	.06	.05	.10	.02	.09			.69	.030	14.22	.071	.890
	SP-editor	.06	.05	.09	.01	.10			.69	.029	15.89	.003	.005
	SP-thresh ($t = .017$)	.06	.06	.09	.01	.10	.04	.64		.028	15.72	.001	.025
S10		.55	.07	.02	.02	.02	.02	.30	.32				
	WD-ind	.54	.06	.03	.02	.01	.04	.30		.016	7.73	.040	.660
	SP-ind	.57	.06	.03	.02	.02	.01	.30		.010	2.68	.021	.032
	WD-editor	.51	.08	.09	.02	.06			.24	.048	18.75	.005	.506
	SP-editor	.55	.06	.07	.02	.03			.27	.030	9.80	.022	.015
	SP-thresh ($t = .024$)	.57	.06	.03	.02	.02	.02	.28		.014	2.89	.017	.032
S16		.66	.05	.02	.04	.02	.02	.19	.21				
	WD-ind	.63	.07	.03	.03	.01	.03	.19		.016	6.61	.056	.711
	SP-ind	.66	.06	.03	.02	.02	.01	.19		.012	8.66	.021	.032
	WD-editor	.64	.07	.07	.02	.04			.16	.033	17.38	.006	.508
	SP-editor	.66	.05	.06	.02	.01			.19	.020	10.69	.025	.017
	SP-thresh ($t = .024$)	.67	.06	.03	.02	.02	.02	.19		.011	7.23	.019	.033
S21		.61	.05	.06	.02	.00	.08	.18	.26				
	WD-ind	.60	.06	.04	.02	.02	.07	.18		.012	5.19	.028	.624
	SP-ind	.61	.05	.04	.01	.01	.09	.18		.013	6.00	.024	.021
	WD-editor	.59	.08	.08	.02	.05			.20	.035	13.76	.005	.500
	SP-editor	.60	.04	.06	.02	.01			.27	.007	2.29	.027	.014
	SP-thresh ($t = .022$)	.60	.06	.04	.02	.02	.08	.17		.013	6.10	.019	.024

Table 2. Continued

Patient	Model	Correct	Semantic	Formal	Mixed	Unrelated	Nonword	Non-naming	Combined	RMSD ^a	χ^2	s / w	p / decay
AK1		.42	.05	.02	.01	.01	.01	.48	.49				
	WD-ind	.41	.04	.02	.01	.01	.03	.48		.010	3.99	.041	.662
	SP-ind	.43	.04	.02	.01	.01	.00	.48		.003	0.30	.021	.035
	WD-editor	.35	.08	.11	.02	.08			.36	.076	36.02	.003	.501
	SP-editor	.43	.03	.04	.03	.01			.47	.015	6.66	.099	.002
	SP-thresh ($t = .027$)	.43	.04	.03	.01	.02	.01	.47		.006	0.79	.015	.037
AS1		.15	.05	.11	.02	.21	.19	.27	.46				
	WD-ind	.13	.06	.13	.03	.08	.29	.27		.064	44.94	.086	.895
	SP-ind	.15	.09	.14	.02	.15	.19	.27		.031	8.76	.005	.018
	WD-editor	.14	.07	.14	.03	.11			.50	.049	20.34	.051	.773
	SP-editor	.13	.10	.16	.02	.16			.44	.035	9.35	.003	.012
	SP-thresh ($t = .011$)	.13	.09	.15	.02	.16	.18	.27		.030	8.33	.003	.018
WR5		.52	.07	.05	.03	.06	.05	.23	.28				
	WD-ind	.49	.07	.06	.02	.03	.10	.23		.026	11.98	.035	.656
	SP-ind	.49	.08	.07	.02	.06	.05	.23		.017	3.64	.016	.027
	WD-editor	.50	.08	.09	.02	.06			.25	.024	5.78	.004	.503
	SP-editor	.52	.07	.08	.02	.04			.26	.018	5.41	.020	.016
	SP-thresh ($t = .019$)	.52	.08	.06	.02	.05	.05	.22		.010	2.22	.015	.027
AK2		.85	.01	.00	.01	.00	.00	.13	.13				
	WD-ind	.85	.01	.00	.01	.00	.00	.13		.001	0.05	.067	.583
	SP-ind	.84	.01	.00	.01	.00	.00	.13		.002	0.11	.064	.074
	WD-editor	.78	.06	.04	.01	.02			.09	.042	21.64	.007	.500
	SP-editor	.83	.02	.01	.01	.00			.12	.008	2.67	.069	.009
	SP-thresh ($t = .207$)	.87	.01	.00	.01	.00	.00	.12		.009	0.98	.059	.055

^a These are raw *RMSDs*, with the averaging based on the number of proportions being fit: 6 for editor models and 7 for the others.

semantic descriptions and miscellaneous responses reflecting visual confusions or naming part of the picture.⁴ Our rationale for lumping these together is both practical and principled. On the practical side, three of the experimental studies used an "other" category that closely corresponds to our definition (Dell et al., 1997, Rumel et al., 2000, and Gordon, 2002). Schwartz and Brecher (2000) had a non-naming category, defined as null responses and descriptions, and a miscellaneous category that, for consistency, were combined. The principled reason is that the non-naming responses are a natural class, insofar as the subject fails to provide a label, right or wrong, for the picture. Our treatment of this category is not without drawbacks, of course. There are important differences between silence and semantic description, notably with regard to the subjects' knowledge of the target, their awareness of their disability, and their strategies for adapting to it. However, these issues are beyond the scope of the current investigation.

Models. The weight-decay (WD) and semantic-phonological (SP) models were combined with the independence, lexical-editor, and threshold approaches to create the WD-independence, SP-independence, WD-editor, SP-editor, and SP-threshold models. (For reasons described later in footnote 5, we are not reporting a test of a WD-threshold model.) Each of the five models was fitted to the 14 profiles by finding parameter values that make the model's response proportions as similar as possible to the profiles, specifically by minimising the χ^2 goodness-of-fit value. Each model uses, for each profile, three free parameters to fit seven proportions constrained to add to 1.0. The weight-decay and semantic-phonological models each have two parameters, w and $decay$, and s and p , respectively. Dealing with the non-naming responses adds another free parameter. In the case of the independence models, the additional parameter is the proportion of non-naming responses.

For the lexical-editor models, it specifies the proportion of nonword responses that are converted to non-naming responses. For the SP-threshold model, the third parameter is the threshold, t , the minimum level of activation that the selected word unit must possess for the naming process to continue.

Non-naming response parameters were incorporated into the compound models in different ways depending on the approach taken to non-naming responses. For the independence and editor versions of the models, it was possible to reduce the number of explicitly fitted parameters from three to two, by adjusting the profiles. This adjustment was easy to make in the independence models. Because non-naming responses are assumed to be independent of the processes that generate overt responses, the non-naming proportion is simply removed from the profile, and the remaining proportions are normalised, that is, recalculated as a proportion of the modified total (Rumel et al., 2000). The normalised profile is fitted to determine the remaining two parameters as described below. The lexical-editor versions of the models were also fitted by making a simple adjustment. Because non-naming responses are assumed to correspond to suppressed nonwords, the proportion of non-naming responses is added to that for nonwords, and the adjusted profile is then fitted for the two remaining parameters. The SP-threshold model required an explicit search for all three parameters simultaneously.

Fitting routines for independence and editor models. The goal of the fitting procedure is to find the two parameters that generate the closest fits overall to the adjusted profiles. Because we required each model profile to be calculated over 10,000 runs, it is very time-consuming to search for the best fit. Consequently, we generated and stored model profiles in advance. In some areas of the space defined by the two parameters, small changes in

⁴ The vast majority of non-naming responses consisted of null responses and semantic descriptions. To illustrate, the breakdown for the four profiles selected from Schwartz and Brecher (2000), the only study for which these numbers are available, is: 65% no responses, 29% semantic descriptions, 6% miscellaneous.

the parameters result in large changes in the model profiles; in other areas, small parameter changes result in insignificant profile changes. Rather than generate a uniformly dense array of model profiles to handle the areas in which small changes make a difference, a variable-resolution map was used, as in Foygel and Dell (2000).

We used a new method for determining which model profiles were stored—that is, which points on the variable-resolution map were occupied. Briefly, quasi-random points were generated iteratively, but the associated model profile was only generated and stored if the point's nearest neighbours (in parameter space) were sufficiently dissimilar (in model profile space). The result is a map consisting of model profiles that have very small differences from their neighbours, but that together define the space of possible error patterns as densely as necessary.

More specifically, points were generated using a Sobel' sequence, which is a way of subrandomly generating points that maximises coverage and minimises "lumpiness" (Press, Teukolsky, Vetterling, & Flannery, 1992). Five hundred initial points were generated, and the following procedure was then used to determine which subsequent points to keep. The average model profile of the eight nearest neighbours of a potential point was compared to each of those neighbours. If a χ^2 comparison between the average value and all eight of the neighbours was less than an arbitrary threshold (here, 8, using $N = 175$ for the PNT), then the point was determined to be superfluous and was skipped. Otherwise, the model profile for that point was generated and stored, and the process continued. The map was considered complete once 500 points in a row were skipped, giving 3782 stored model profiles for the semantic-phonological model, and 7481 stored model profiles for the weight-decay model. These fitting routines and the underlying model can be used at the following web address:

<http://langprod.cogsci.uiuc.edu/cgi-bin/webfit.cgi>

Fitting routine for SP-threshold model. A third free parameter, representing the lexical threshold t , was added to the lexical-semantic and lexical-

phonological weight parameters, s and p , in the SP model. If the activation of no word unit is greater than t at the moment that the model chooses the most activated word unit, the trial is assigned to the non-naming-response category. Otherwise, the selected word undergoes phonological access and an overt response is generated.

Because three free parameters have to be specified in the SP-threshold model, it is not possible to use the variable-resolution two-dimensional map described above. Instead, we used a hill-climbing technique. First, we started with values of s and p generated from the SP-independence model, and made an educated guess about the best value of t , based on our experience from exploring the parameter space of the SP-threshold model. Then we evaluated the 26 locations adjacent to this starting location that are generated by increasing or decreasing at least one parameter by a step of .01. The location that led to the smallest value of χ^2 was then selected. The procedure was then repeated with the selected location replacing the starting location. This exploration continued until none of the adjacent locations led to a smaller χ^2 value than the current location. At this point, the step size was decreased to .001, and the process begun again. The final location was selected as the fitted parameter values. Note that this procedure does not guarantee finding the best fit for a profile, or even that the best fit is near the final location. We repeated the fitting process for each profile, starting from a randomly determined region of the parameter space (bounded by 0 and .1). In all cases, the fits starting from the initial locations described above were either better than or identical to those obtained from the random starting locations. Nonetheless, it must be recognised that the addition of a third parameter whose value must be discovered makes the fits of the SP-threshold model more likely to be suboptimal than the fits generated from the variable-resolution map.

Results and discussion

Table 2 presents the proportions of responses in each of the seven categories for each of the 14 profiles, along with the proportions and parameters

generated from each of the five models. Two measures of goodness of fit are given for each profile: χ^2 , the statistic used to discover the fit, and *root mean squared deviation (RMSD)*. RMSD is the more intuitive measure of fit as it summarises the degree of deviation of the obtained proportions from the models' predictions. (An RMSD of .030 means that the model and patient proportions are on average .030 apart.) How close should these proportions be? It is hard to say, because there are no established benchmarks of quantitative modelling in this area. Until that is the case, measures of fit are best interpreted comparatively, rather than absolutely. Our discussion, therefore, focuses on which models provide the best fit.

There were clear differences among the models in overall quality of fit, with the most obvious being the superiority of the semantic-phonological models over the weight-decay models. The total χ^2 values for the SP-independence (130.4), SP-threshold (130.1), and SP-editor (171.1) models were much smaller than those for the WD-independence (262.4) and WD-editor (331.6) models.⁵ The mean RMSD measures tell a similar story: SP-independence (.017), SP-threshold (.018), SP-editor (.028), WD-independence (.026), and WD-editor (.050). As Table 2 shows, there is no profile for which the two WD models fit better on average than the three SP models, and several where the SP models are much better (e.g., AS1, VP). In particular, the WD models have difficulty predicting the unrelated and nonword categories. The difficulty stems from the use of global parameters. Parameter settings that promote unrelated word errors necessarily also

create many nonword errors. However, AS1's and VP's nonword errors are not particularly numerous relative to word errors. The ability of the SP models to separate influences on word errors (determined largely by the s parameter) from nonword errors (determined largely by p) gives them the edge in fitting these profiles.

A second result is that the independence and threshold models provide better fits than the lexical-editor models. Considering all five models, the SP-independence and SP-threshold models are the clear winners. The SP-independence model has significantly smaller RMSD's than both the SP-editor model, $t(13) = 2.52$,⁶ and the WD-independence model, $t(13) = 2.20$. The SP-threshold model is also clearly superior to the SP-editor model, $t(13) = 2.73$, and (marginally) better than the WD-independence model $t(13) = 2.00$. The mean RMSD's for the SP-independence (.017) and SP-threshold (.018) models indicate a close agreement between model and profile proportions. Moreover, these two models account for 96.3% and 95.9% respectively of the variance in profile proportions.⁷ The 96.3% value for the SP-independence model includes the variance within the non-naming category, which this model fits trivially by setting its third free parameter to the proportion of non-naming responses. Even if the non-naming category is left out, the model still accounts for 93.0% of the variance in the remaining categories.

Both the SP-independence and SP-threshold models fit the individual profiles fairly well, with one exception (EA, to be discussed later). There are no profiles for which the SP-editor model

⁵ We explored the WD-threshold model as well, but were not able to find parameters that led to fits comparable to the WD-independence model. Given that the SP-threshold and SP-independence models had similar fits, more stringent parameter searches might lead to similar fits for the WD-threshold and WD-independence models. However, these would still be inferior to those for the SP-threshold and SP-independence models.

⁶ The critical value for $t(13)$ is 2.16. The RMSDs for the two independence models used in these t -tests were adjusted upward so that these models were not credited for matching the non-naming response proportion (which they do perfectly because their third parameter is just this proportion). Specifically, the adjustment averages the total squared deviations over six rather than seven proportions for these models. The adjustment is conservative given our conclusion that the independence models are favoured over the editor models.

⁷ This measure compares the squared deviations between the model and profile proportions to the squared deviations between each profile proportion and the mean proportion for that category across all profiles.

provides a close fit (say, $\text{RMSD} < .020$) and the SP-independence and SP-threshold models do not, while the reverse is clearly true. For example, the SP-independence and SP-threshold models are considerably better for IOC, JR1, JR2, S10, and S16. IOC's profile is a case in point. Under the SP-editor model, the many non-naming responses (.79) are assumed to be suppressed nonwords. To generate such a large proportion of nonwords, the model must set a very low value for p (.001). However, the processes that generate nonword errors create formal errors as well. With the parameters assigned to IOC by the SP-editor model, the

expected proportion of formals is .08, a significant contrast to the obtained proportion of zero.

This over-prediction of formal errors is a general problem with the SP-editor model when compared to the other SP models. Figure 2 illustrates how the predicted proportions for each model deviate from the obtained proportions in each response category, separately for each profile. (EA was left out because this profile was not fit well by any model; possible reasons for this will be discussed shortly.) A positive number indicates that the profile proportions exceeded the model ones; a negative number means that the profile

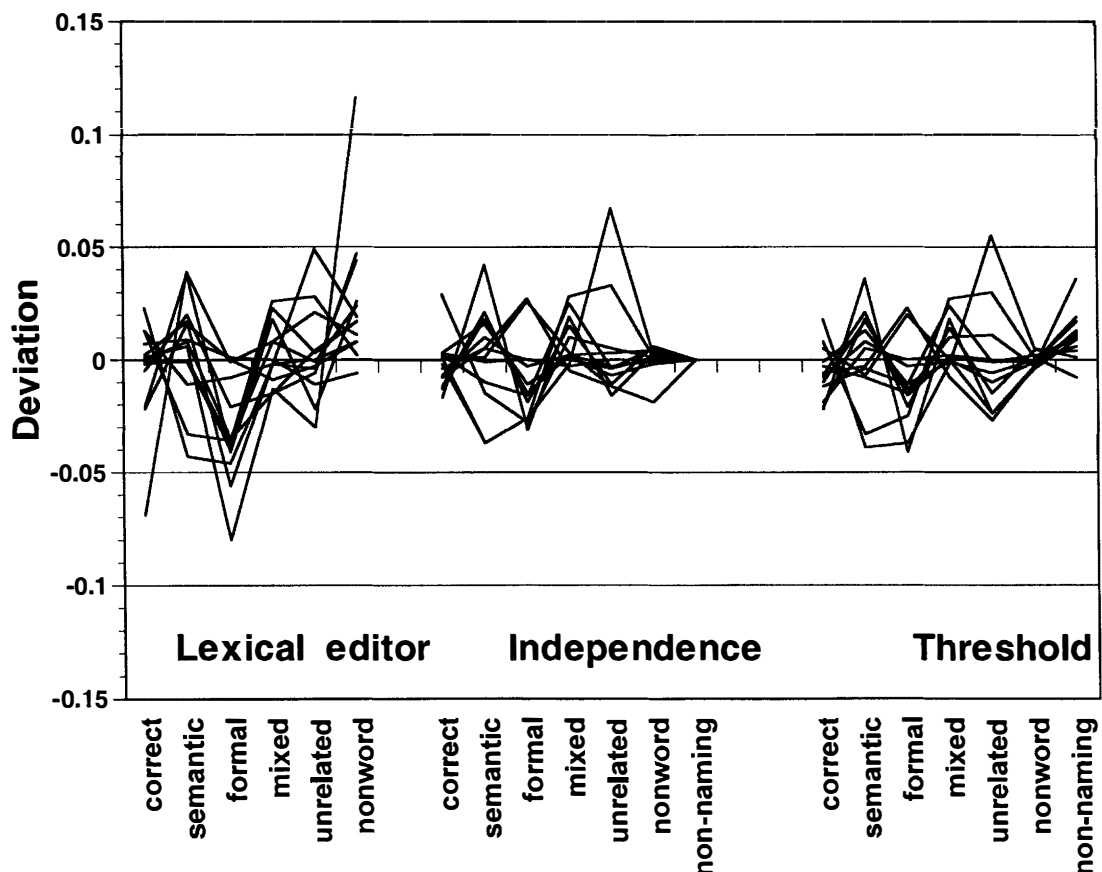


Figure 2. Deviations between obtained response proportions and model response proportions, as a function of response category for the SP-editor model (left), the SP-independence model (middle), and the lexical threshold model (right). Each line represents a single profile (EA is not included).

proportions are too small. The deviations for the SP-independence and SP-threshold models are small, and do not tend to one side or the other of zero. The SP-editor model not only has larger deviations, but they appear to be systematic. For example, the nonword category is almost always over-predicted, while the proportion of formals is underestimated. Although these deviations are largest for IOC (the most deviant line in the figure) they are present on most of the other profiles.

A third important finding is that none of the models fits EA's profile very well. The large number of semantic (.22), mixed (.08), and unrelated (.11) errors, along with a smaller proportion of formal errors (.03), challenges all of the models, even the SP models, which can assign lexical-semantic lesions without creating errors during phonological access. These models' neighbourhoods, however, do not support the generation of such high percentages of semantic and mixed errors. Moreover, when the SP models create many mixed and unrelated word errors, they must also generate a fair number of formal errors. Mixed errors suggest strong phonological feedback, which also activates formal neighbours of the target; unrelated errors signal that the target's activation is sufficiently degraded that these formal neighbours will often be selected instead of the target. Thus, EA's low proportion of formal errors presents a problem.

A clue to EA's naming pattern is the unusual finding that many of his semantic and mixed errors were verbs rather than nouns (Ruml et al., 2000). The naming models, however, assume that the processes ensuring selection of a noun operate without error, even for patients. Perhaps EA's profile reflects a failure of the syntactic system to constrain lexical access. Gordon and Dell (2003) implemented a model of such failure. Lesioning the weights from syntactic/sequential states to word units led to "agrammatic" production, and

specifically to naming deficits in which words were deleted or replaced by other words often not in the target syntactic category. We note that EA's sentence production was described as nonfluent, but lacking in articulatory problems (Ruml et al., 2000), and "agrammatic" (Berndt, Mitchum, Haendiges, & Sandson, 1997), which is consistent with the hypothesised lesion in syntactic/sequential states.⁸

The dominance of semantic and mixed errors along with the absence of phonological errors make EA's profile similar to three other cases in the literature that have challenged the semantic-phonological model: PW (Rapp & Caramazza, 1998; Rapp & Goldrick, 2000), JF (Foygel & Dell, 2000), and DP (Cuetos et al., 2000). Such cases might be consistent with the model if the naming difficulty occurs secondary to a problem with semantic representations. Recall that the models assume intact semantic input, and that semantic and other word errors occur in the process of mapping that input onto lexical units. The assumption of intact semantic input is probably false for EA, at least according to Shelton and Weinrich (1997), whose test of EA's single-word comprehension indicated that he had a "mild deficit to semantic knowledge" (p. 122). This explanation, however, does not hold for PW and DP, who are reported to have preserved semantic-level representations (Cuetos et al., 2000; Rapp & Goldrick, 2000). Instead, Rapp and Goldrick were able to account for PW's error pattern (which is similar to DP's) by positing that lexical-phonological feedback is restricted. That is, the bottom-up weights are weaker than the top-down ones. This restriction allowed Rapp and Goldrick's model to generate semantic errors without leading to formal errors at the word access stage, since the latter are promoted by strong feedback.

Restricted feedback, though, cannot by itself account for EA's profile. Ruml et al. (2000) explored restricted feedback versions of these

⁸ Berndt et al. (1997) found that most examples of the uncommon verb-for-noun substitution were made by agrammatic aphasics. Those patients, however, were generally more impaired in accessing verbs, and more often replaced them with nouns than the other way around.

models and still were not able to generate high enough proportions of semantic, mixed, and unrelated errors. Consequently, we suggest that the failure to fit EA stems from the model's assumptions of intact semantic and syntactic inputs, either or both of which may have been violated in this case. In any event, this error pattern points to a clear limitation in the model as a general account of aphasic naming.

EA aside, the SP-independence and SP-threshold models provide a good account of the patient profiles in our sample, certainly better than the other three models examined. Consequently, the damage assumptions of the semantic-phonological model and the treatment of non-naming responses proposed in the independence and threshold models are supported over the weight-decay and lexical-editor models. This is not to say that the independence or threshold models apply universally or that the suppression of nonword responses—as assumed by the lexical-editor model—does not occur. Rather, the evidence from our sample is that the independence and threshold models are, on balance, more powerful in describing patient performance.

Moreover, they allow us to extend the observations that can be brought to bear on models of aphasic lexical access.

Thus far, we have treated the SP-independence and SP-threshold models together. On the basis of fit quality, they are indistinguishable, though clearly superior to the other models. This is true for the individual profiles as well as overall measures of fit. Only EA is poorly fit, both models having trouble matching the proportions in the formal, mixed, and semantic categories. For the remaining profiles, the two models fit reasonably well and offer similar characterisations. For example, for S10 the predicted profiles for SP-threshold and SP-independence differ only for the nonword (.02 and .01) and non-naming (.28 and .30) categories. The models assign similar values of s (.021 and .017) and p (.032 for both), characterising this patient as having more of a lexical-semantic than lexical-phonological deficit.

Clearly, the SP-threshold model behaves like the SP-independence model. Figure 3 shows why. The figure shows what types of responses a threshold tends to inhibit using a sample model with typical aphasic parameters. A model with $s = .02$

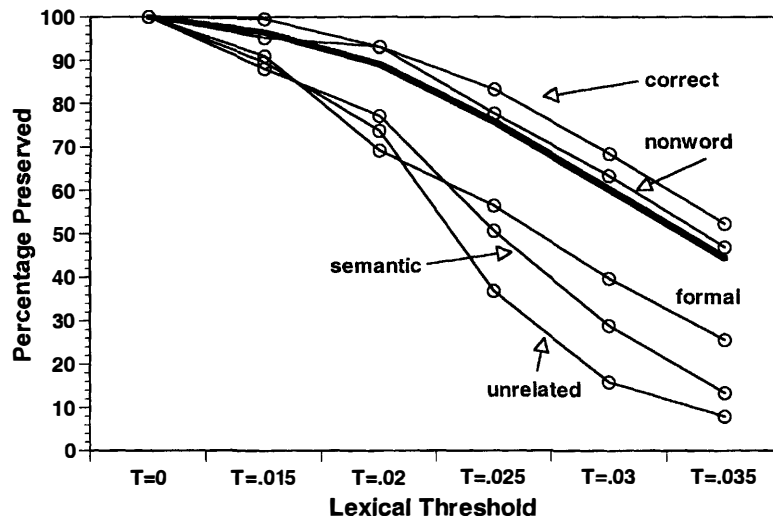


Figure 3. The effect of threshold (T) on the proportions of each response category that are preserved. The bold line represents the average percentage of responses, across categories, preserved at each threshold value.

and $p = .02$ and no threshold is compared to the same model with a variety of threshold values. The ordinate presents the percentage of responses exceeding the threshold. With t equal to zero (i.e., no threshold), all responses will emerge as overt responses. As the threshold increases, the proportions of supra-threshold responses are reduced. The narrow lines illustrate that the ability of the threshold to inhibit responses varies somewhat across response categories. Nevertheless, *all* of the categories decrease dramatically as t increases; the threshold robs from all response types and turns them into null responses. The single thick line indicates the mean percentage of responses preserved across all categories. In essence, this represents the independence model, in which non-naming responses are stolen from each category in proportion to its frequency of occurrence. The proximity of the narrow lines to the thick one thus demonstrates the degree of similarity between the SP-threshold and SP-independence models.

Despite their similarity, however, the two models are not identical. For example, the threshold is more likely to suppress a potential unrelated-word response than a correct response. A situation in which the most activated word unit corresponds to an unrelated word is most likely to be one in which all of the activations are very low, and hence subthreshold. Nevertheless, the differences between the two models are subtle with respect to their ability to account for the naming profiles. The overt response categories that are most common in the profiles are the correct and nonword categories. Figure 3 shows that the threshold takes away from these categories at about the same rate that it would take away if independence held. Consequently, the two models are difficult to distinguish.

Are there any reasons, then, to prefer either the SP-independence or the SP-threshold model? The SP-independence model has the advantage of being easy to use and can, therefore, be taken as the default assumption for treating non-naming responses when the focus of the analysis is on the overt error categories (e.g., Rumel et al., 2000). Its main drawback is that it does not specify a mechanism for non-naming responses. Although it succeeds in characterising error distributions

statistically, it does not explain them. The SP-threshold model may provide the missing explanation. Thresholds are common in biological systems, and particularly so in neural network models of such systems. Thus, an activation-level threshold governing whether or not a lexical unit proceeds to phonological access is a plausible mechanism. In the general discussion, we further consider the contrast between these models as well as other ramifications of our findings.

GENERAL DISCUSSION

The principal findings of the current model comparison were as follows:

1. The semantic-phonological model (Foygel & Dell, 2000) is better suited to matching naming profiles with a high percentage of non-naming responses than is the weight-decay model (Dell et al., 1997).
2. The distribution of non-naming responses is better accounted for by models incorporating an independence assumption (Rumel et al., 2000) or threshold mechanism (Laine et al., 1998) than a lexical-editor mechanism (Baars et al., 1975).
3. The SP-threshold model is largely indistinguishable from the SP-independence model in terms of its predictive power, but offers a more theoretically motivated mechanism for explaining the occurrence of non-naming responses.
4. The profile of EA, with its relatively high percentage of semantic and mixed errors, and low percentage of formal errors, is not fit well by any of the models, a finding that may call into question the models' assumptions of intact syntactic and semantic input.

It is important to put these conclusions in perspective. The superiority of the SP-independence and SP-threshold models does not mean that the suppression of deviant responses does not occur. On the contrary, one can construe a threshold as a strategically determined value that serves to prevent potentially erroneous responses. Moreover, the relatively poor performance of the lexical-editor models does not suggest that patients never

preferentially suppress nonword errors. Rather, a pure lexical-editor account is less favoured by the current evidence. Given that any single account is unlikely to apply to all patients, it is worth noting that several of our profiles were well fit by the SP-editor model, for example, WR5, AK1, and AK2. Each of these profiles is associated with few nonword errors and a surfeit of non-naming responses. In one or more of these cases, the lexical-editor model might be the correct explanation. Nevertheless, we emphasise that the SP-independence and SP-threshold models also accounted for these profiles, as well as for other profiles that the SP-editor model was less successful in characterising.

Finally, we recognise that one can construct alternative editor models that may give more accurate results. For example, one can hypothesise that all semantically unrelated responses (nonwords, formals, and unrelated errors) are subject to suppression. Models based on these broader editors, however, will be difficult to distinguish from the independence assumption because they, like the independence model, convert potential responses from multiple categories into non-naming responses. We have focused on a *lexical* editor because not only is it supported in the literature, but also it is maximally distinct from models that behave like the independence model. Model testing is best served by examining motivated models whose predictions stand in sharp contrast.

Ultimately, the decision about the best account of non-naming responses, both in general and for particular patients, cannot be made on the basis of naming profiles alone. In addition, the competing models should be used to generate and test predictions. Here we give two examples of this strategy: predicting word repetition performance from word naming performance, and predicting variations in naming profiles as a function of semantic neighbourhood.

Predicting repetition from naming

The semantic-phonological and weight-decay models assign parameters to profiles based on naming performance. Dell et al. (1997) and Foygel

and Dell (2000) used these parameters to predict auditory word repetition by identifying repetition performance with the phonological access stage of the naming process. This assumes that repetition entails access of the appropriate word unit through an intact perceptual system, followed by the mapping of that word unit onto its phonological representation (see R. C. Martin, Lesch, & Bartha, 1999, for evidence of separate input and output lexical-phonological mappings). Foygel and Dell found that naming parameters from the SP model predicted repetition performance better than did parameters from the WD model. This is because word repetition depends more on lexical-phonological mappings than on lexical-semantic mappings, and the SP model is able to characterise these differentially. The WD model's parameters do not isolate relatively impaired or preserved lexical-phonological abilities; hence, its repetition predictions are less accurate.

Predicting repetition can also discriminate between the independence/threshold and editor accounts of the non-naming responses. Note in Table 2 that the independence and threshold versions of the SP model assign higher values of p than s for 13 of 14 profiles, associating these patients with more severe lexical-semantic than lexical-phonological lesions. In contrast, the SP-editor model assigns a lower p than s in 10 of 14 cases, because treating non-naming responses as suppressed nonwords increases the evidence for a phonological deficit, by the model's logic. Because good repetition depends more on the lexical-phonological weight, the SP-independence and SP-threshold models predict that repetition should be relatively good for these patients in contrast to their naming, whereas the SP-editor model predicts poorer repetition.

Ruml et al. (2000) reported word repetition data for some of the subjects in their study, two of whom (EA and JR2) are represented in this study. Both EA and JR2 repeated the items of the PNT without error. Which of the models, if any, can account for this good performance? It turns out that both the SP-independence and SP-threshold models can. Using the selected parameters for these two patients to carry out the

repetition task, the two models perform at 97% and 98% correct respectively for EA, and 96% and 95% for JR2. The SP-editor model, on the other hand, predicts accuracy levels of only 76% for EA and 80% for JR2. The superiority of the SP-independence and SP-threshold models in predicting the repetition of EA and JR2 arises because they assigned larger values of p than did the SP-editor model, such that repetition was better preserved.

Semantic neighbourhood effects

In a study of naming errors by an aphasic individual (MW), Blanken, Dittmann, and Wallesch (2002) found that items differed in their propensity to create errors of either commission or omission. Target items that had close semantic competitors (as determined by independently gathered ratings) tended to create semantic errors, whereas those with few competitors often led, instead, to failures to respond. These two error types traded off such that the sum of the two categories differed very little as a function of the competitiveness of target's semantic neighbourhood. As semantic errors and no responses were the dominant error categories for MW, this trade-off implies that the proportion of correct responses was similar for competitive and less competitive neighbourhoods. Blanken et al.'s finding demonstrates, not surprisingly, that semantic-error probability increases with the opportunities for semantic errors. More importantly, though, it suggests that omissions are promoted more by the failure to strongly activate any lexical items, than by the activation of many competing items.

With respect to the current article, the trade-off between semantic errors and null responses is a clear prediction of the lexical-threshold model, provided that these are common error types for a particular patient. Consider an example using the SP-threshold model with $s = .02$, $p = .03$, and

$t = .03$. If the model neighbourhood with one semantic neighbour of the target (e.g., DOG for the target CAT) is tested, the model is correct 57% of the time, fails to respond 35% of time, and makes 3% semantic errors. If the neighbourhood is made more competitive by including DOG and RAT as semantic neighbours of CAT, these percentages become 60% correct, 28% no responses, and 9% semantic errors.⁹ The increase in competition creates more semantic errors *at the expense of null responses*, just as was found by Blanken et al. (2002). The model behaves this way because competition increases the activation of both the target and its competitors. Activation reverberates between words and the shared properties of these words. The more competitors there are, and the more properties they share, the more reverberation occurs. The end result is that the most activated word unit, whether it is correct or not, has a greater chance of being above threshold. Hence, an overt response is more likely to occur. The presence of more competition, of course, leads to more semantic errors, but this is almost entirely compensated for by the decreased likelihood of an omission.

The lexical-editor approach does not make the same prediction. Consider a set of parameters (e.g., s and p both at .01) that creates a large number of nonwords, the responses that the editor can convert into omissions. Using the less competitive neighbourhood, the model generates 18% correct responses, 7% semantic errors, 52% nonwords, and 23% other errors (formals and unrelateds). When the more competitive neighbourhood is tested, the results are 18% correct, 15% semantic errors, 52% nonwords, and 15% other errors. Greater competition naturally leads to more semantic errors, but not at the expense of the nonword category, which remained at 52%. Rather it is at the expense of other errors. Consequently, the increase in semantic errors would not be associated with fewer omissions, on the assumption that these are suppressed nonwords.

⁹ Note that RAT is also a phonological neighbour of CAT. This only tends to make it more of a competitor because of the model's interactive assumption, the point being to demonstrate that additional competition increases lexical substitution errors, while decreasing failures to respond.

It is difficult to say what the independence model would predict about Blanken et al.'s (2002) trade-off. Because that model lacks a true mechanism for generating no responses, it cannot create different proportions of omissions as a function of neighbourhood competitiveness without effectively adding parameters. So, unlike the lexical threshold model, which makes the correct prediction, and the lexical-editor model, which makes the wrong prediction, the independence model makes no prediction at all.

The repetition and semantic-neighbourhood predictions show how the models can be tested using data outside of the naming profiles. Predictions for repetition supported the SP-independence and SP-threshold models over the SP-editor model, and the semantic-neighbourhood prediction was consistent only with the SP-threshold model. These tests, though, are more illustrative than definitive, as they concerned only two cases for the repetition prediction, and only one for the semantic-neighbourhood prediction.

CONCLUSION

The current study demonstrates that it is possible to develop and test models of aphasic naming using errors of omission, as well as errors of commission. Our analysis favoured two models that were largely indistinguishable, the SP-independence model and the SP-threshold model. For the most part, these models provided close fits to the naming profiles of individuals who tend to make many non-naming responses. The analysis helps to constrain the interactive two-step theory of lexical access that is the basis of those models, and specifically supports the semantic-phonological approach to lesioning, as well as the independence and lexical-threshold accounts of non-naming responses. The weight-decay account of aphasic lesions suffered from its assumption that pathological parameter values are global, that is, they cannot differentially characterise lexical-semantic and lexical-phonological abilities (see Rapp & Goldrick, 2000; Rumel & Caramazza, 2000). The lexical-editor account of non-naming responses

forced the profiles to treat too many responses as nonwords. Although the lexical-editor account, in combination with the semantic-phonological model, fit several profiles quite well, the other approaches also worked well with these profiles.

By providing a method of analysing non-naming responses, we hope to expand the sources of evidence that can inform models of normal and aphasic speech production. Although the hypotheses presented here were able to explain a number of observations of aphasic behaviour in language production tasks, they must be considered preliminary. Further evidence from both normal and aphasic subjects, and from a variety of tasks, is needed to strengthen these findings and generate new predictions. We encourage other researchers interested in deficits in spoken word production to explore the models presented here using the website introduced in the Methods section of the study.

REFERENCES

- Baars, B., Motley, M., & MacKay, D. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, 14, 382–391.
- Berg, T., & Schade, U. (1992). The role of inhibition in a spreading activation model of language production, Part 1: The psycholinguistic perspective. *Journal of Psycholinguistic Research*, 22, 405–434.
- Berndt, R. S., Mitchum, C. C., Haendiges, A. H., & Sandson, J. (1997). Verb retrieval in aphasia. *Brain and Language*, 56, 68–106.
- Blanken, G., Dittmann, J., & Wallesch, C.-W. (2002). Parallel or serial activation of word forms in speech production? Neurolinguistic evidence from an aphasic patient. *Neuroscience Letters*, 325, 72–74.
- Bock, J. K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). San Diego, CA: Academic Press.
- Buckingham, H. W. (1987). Phonemic paraphasias and psycholinguistic production models for neologistic jargon. *Aphasiology*, 1, 381–400.
- Burke, D. M., Mackay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language*, 30, 542–579.

- Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, 14, 177–208.
- Caramazza, A., & Hillis, A. (1990). Where do semantic errors come from? *Cortex*, 26, 95–122.
- Caramazza, A., Papagno, C., & Rumel, W. (2000). The selective impairment of phonological processing in speech production. *Brain and Language*, 75, 428–450.
- Croot, K., Patterson, K., & Hodges, J. R. (1998). Single word production in nonfluent progressive aphasia. *Brain and Language*, 61, 226–273.
- Cuetos, F., Aguado, G., & Caramazza, A. (2000). Dissociation of semantic and phonological errors in naming. *Brain and Language*, 75, 451–460.
- Cutting, J. C., & 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 language production. *Psychological Review*, 93, 283–321.
- Dell, G. S., & O'Seaghdha, P. G. (1991). Mediated vs. convergent lexical priming in language production: Comment on Levelt et al. *Psychological Review*, 98, 604–614.
- Dell, G. S., & Reich, P. A. (1981). Stages of sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior*, 20, 611–629.
- Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, 104, 801–838.
- Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (2000). The role of computational models in the cognitive neuropsychology of language: Reply to Rumel and Caramazza. *Psychological Review*, 107, 635–645.
- Ferreira, V. S., & Humphreys, K. R. (2001). Syntactic influence on lexical and morphological processing in language production. *Journal of Memory and Language*, 44, 52–80.
- Foygel, D., & Dell, G. S. (2000). Models of impaired lexical access in speech production. *Journal of Memory and Language*, 43, 182–216.
- Gagnon, D. A., Schwartz, M. F., Martin, N., Dell, G. S., & Saffran, E. M. (1997). The origins of form-related paraphasias in aphasic naming. *Brain and Language*, 59, 450–472.
- Garnsey, S. M., & Dell, G. S. (1984). Some neuro-linguistic implications of prearticulatory editing in production. *Brain and Language*, 23, 64–73.
- Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 133–175). San Diego, CA: Academic Press.
- Gordon, J. K. (2002). Phonological neighborhood effects in aphasic speech errors: Spontaneous and structured contexts. *Brain and Language*, 82, 113–145.
- Gordon, J. K., & Dell, G. S. (2003). Learning to divide the labor: An account of deficits in light and heavy verb production. *Cognitive Science*, 27, 1–40.
- Griffin, Z. M., & Bock, J. K. (1998). Constraint, word frequency, and the relationship between lexical processing levels in spoken word production. *Journal of Memory and Language*, 38, 313–338.
- Hanley, J. R., Kay, J., & Edwards, M. (2002). Imageability effects, phonological errors, and the relationship between auditory repetition and picture naming: Implications for models of auditory repetition. *Cognitive Neuropsychology*, 19, 193–206.
- 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. (1993). Phonological activation of semantic competitors during lexical access in speech production. *Language and Cognitive Processes*, 8, 291–310.
- Harley, T. A., & Bown, H. E. (1998). What causes a tip-of-the-tongue state? Evidence for lexical neighbourhood effects in speech production. *British Journal of Psychology*, 89, 151–174.
- Hartsuiker, R. J., & Kolk, H. H. J. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology*, 42, 113–157.
- 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.
- 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.
- Humphreys, K. R., & Swendsen, A. (2002). *Asymmetric lexical bias in speech errors*. Presented at the 15th CUNY Conference in Human Sentence Processing, New York.
- Kay, J., & Ellis, A. W. (1987). A cognitive neuropsychological case study of anomia: Implications for psychological models of word retrieval. *Brain*, 110, 613–629.

- Laine, M., Tikkala, A., & Juhola, M. (1998). Modelling anomia by the discrete two-stage word production architecture. *Journal of Neurolinguistics*, 11, 275–294.
- Levelt, W. (1983). Monitoring and self-repair in speech. *Cognition*, 14, 41–104.
- Levelt, W., Roelofs, A., & Meyer, A. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–75.
- MacKay, D. G. (1987). *The organization of perception and action: A theory for language and other cognitive skills*. New York: Springer-Verlag.
- Martin, N., & Saffran, E. M. (1997). Language and auditory-verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology*, 14, 641–682.
- Martin, N., Dell, G. S., Saffran, E., & Schwartz, M. F. (1994). Origins of paraphasias in deep dysphasia: Testing the consequence of a decay impairment of 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, R. C., Lesch, M. F., & Bartha, M. C. (1999). Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language*, 41, 3–29.
- Meyer, A. S., & Bock, K. (1992). The tip-of-the-tongue phenomenon: Blocking or partial activation? *Memory and Cognition*, 20, 715–726.
- Miozzo, M., & Caramazza, A. (1997). The retrieval of lexical-syntactic features in tip-of-the-tongue states. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 1410–1423.
- Mitchum, C. C., Ritger, B. A., Sandson, J., & Berndt, R. (1990). The use of response analysis in confrontation naming. *Aphasiology*, 4, 261–280.
- Nickels, L., & Howard, D. (1995). Aphasic naming: What matters? *Neuropsychologia*, 33, 1281–1303.
- Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition*, 77, 97–131.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes in C: The art of scientific computing* (2nd ed.). Cambridge: Cambridge University Press.
- Rapp, B., & Caramazza, A. (1998). A case of selective difficulty in writing verbs. *Neurocase: Case Studies in Neuropsychology, Neuropsychiatry, and Behavioural Neurology*, 4(2), 127–139.
- Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, 107, 460–499.
- Roach, A., Schwartz, M. F., Martin, N., Grewal, R. A., & Brecher, A. (1996). The Philadelphia Naming Test: Scoring and rationale. *Clinical Aphasiology*, 24, 121–133.
- Ruml, W., & Caramazza, A. (2000). An evaluation of a computational model of lexical access: Comments on Dell et al. (1997). *Psychological Review*, 107, 609–634.
- 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.
- Schade, U., & Eikmeyer, H.-J. (1998). Modeling the production of object specifications. In J. Grainger & A. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 257–282). Mahwah, NJ: Lawrence Erlbaum Associates Inc.
- Schwartz, M. F., & Brecher, A. (2000). A model-driven analysis of severity, response characteristics, and partial recovery in aphasic picture naming. *Brain and Language*, 73, 62–91.
- Schwartz, M. F., & Hodgson, C. (2002). A new multi-word naming deficit: Evidence and interpretation. *Cognitive Neuropsychology*, 19, 263–288.
- Shelton, J. R., & Weinrich, M. (1997). Further evidence of a dissociation between output phonological and orthographic lexicons: A case study. *Cognitive Neuropsychology*, 14, 105–129.
- Stemberger, J. P. (1985). *The lexicon in a model of language production*. New York: Garland.
- Vigliocco, G., Antonini, T., & Garrett, M. F. (1997). Grammatical gender is on the tip of Italian tongues. *Psychological Science*, 8, 314–319.

- title: MODELS OF ERRORS OF OMISSION IN APHASIC NAMING
- authors: Gary S. Dell and Elisa N. Lawler and Harlan D. Harris and Jean K. Gordon
- year: 2004
- journal: COGNITIVE NEUROPSYCHOLOGY,
- volume: 21 (2/3/4)
- pages:125-145

失語症の命名課題における無回答エラーモデル

要約

産出時の語彙アクセスに関する 5 つの計算モデルについて、失語症の絵画命名検査時のエラー分布を説明する能力を検証した。命名プロファイル (N=14) は、比較的多くの無回答エラーを犯す患者を代表する文献から選ばれた。最も成功したモデルは、語彙アクセスの意味-音韻モデル (Foygel & Dell, 2000) の損傷仮定と、無回答エラーを回答エラーとはほぼ独立したものとして扱う方法 (Rumel, Caramazza, Shelton, & Chialant, 2000) を組み合わせたものであった。無回答エラーの発生を説明するために、語彙-しきい値パラメータ (Laine, Tikkala, & Juhola, 1998) をモデルに加えた。これらのモデルをさらに検証するための提案や、他の研究者がモデルを利用できるようにするための新しいウェブサイトも紹介している。

1. 導入

失語症では、言葉の検索がうまくいかないため、さまざまな発話の誤りが生じるが、これらの誤りはすべて、会話中の語彙アクセスのメカニズムを説明するものである。失語症の誤りに関するほとんどの研究は、パラフェシア (省略の誤りとは対照的な遂行の誤り) に焦点を当てている。多くの場合、パラフェシアの発話には、意図した単語との意味的な関係 (例えば、strawberry の代わりに banana) や音韻的な関係 (例えば、train の代わりに treen) がある。これらの誤りは、語彙システムの仮説的な構成要素に関する具体的な情報を提供するという理由で、研究されている (例えば、Buckingham, 1987; Caramazza & Hillis, 1990; Gagnon, Schwartz, Martin, Dell, & Saffran, 1997; Howard & Orchard-Lisle, 1984; Kay & Ellis, 1987; N. Martin & Saffran, 1997; Mitchum, Ritger, Sandson, & Berndt, 1990; Nickels & Howard, 1995; Rapp & Goldrick, 2000)。

失語症の話者は、明示的な言い誤りエラーに加えて、無回答エラーも犯す。これらのエラーは、一般的に単語検索ブロックと呼ばれ、すべての産出課題で発生する。その特徴は、沈黙や、「何かは知っているけど言えない」と言ったり、産出を中断したり、探している単語の意味上の特徴を説明したり (例: 動物の一種) するなど、単語検索がうまくいかなかったことを示す証拠がある。これらの検索ブロックはほとんど研究されたことがないが、それには理由がある。ターゲットを明らかに生み出そうとしない限り、発生している精神的処理過程について結論を出すことは困難だからである。とはいえ、単語検索行動の臨床観察によって、いくつかの暫定的な仮説を立てることができた。例えば "It's a, no, that's not it ..." のようなコメントは、話し手が間違えた言葉を抑制することがあることを示唆している。この記述はターゲットの意味情報の一部は利用可能だが音韻は利用できないことを示している。残念なことに、単語検索の困難さは、そのような手がかりなしに起こることが多い。

絵画命名研究では、被験者が命名しなかった試行は、通常、分析から除外される。しかし、このような事象はよくある。例えば、Mitchum ら (1990) の 28 人の失語症患者を対象とした研究では、無反応は命名試行の 11 % で発生しており、7 人の患者では無反応が最も一般的なエラーであった。Mitchum らの無反応のカテゴリーには、沈黙、無反応発言 (例: 「わからない」)、空のコメント (例: 「こんなの持ってるよ」) が含まれていた。この論文のデータを提供してくれた研究では、無回答カテゴリーは、沈黙に加えて、話してはいるが絵の具で描かれた物の単語名を言おうとしていないすべての回答を含む、一般的な無回答カテゴリーに拡大された。例えば、意味的な記述も含まれる。

本研究では、命名課題の計算モデルを開発し、そのモデルを失語症のデータと関連付けることで、無回答エラーを理解を意図した。具体的には、2 つの目標がある。1 つは、既存のモデルを適用できるデータを増やすことである。現在のモデルは、回答エラーに対してのみ正確な予測を行う。無回答反応に適用を拡大することで、モデル間の識別をより正確に行うことができる。2 つ目の目標は、特定の仮説を公式化することで、無回答自体の調査を開始することである。これらの目標を実現するために、我々は、語彙ネットワーク内の病的な広がりのある活性化に委ねられたエラーを原因とする 2 つの既存モデルから始める。Dell, Schwartz, Martin, Saffran, and Gagnon (1997) の重み-崩壊 weight-decay モデルと、Foygel and Dell (2000) の意味-音韻 semantic-phonological モデルである。両者とも、様々な成功を収めながら、失語症の命名・反復エラーに適用されている (Caramazza, Papagno, & Rumel, 2000; Croot, Patterson, & Hodges, 1998; Cuetos, Aguado, & Caramazza, 2000; Dell, Schwartz, Martin, Saffran, & Gagnon, 2000; Gordon, 2002; Hanley, Kay, & Edwards, 2002; Hillis, Boatman, Hart, & Gordon, 1999; Rumel & Caramazza, 2000; Rumel, Caramazza, Shelton, & Chialant, 2000; Schwartz & Brecher, 2000; Schwartz & Hodgson, 2002)。しかし、どちらのモデルも無回答を考慮していないため、結果的にはそのようなエラーが少ないデータセットでしか検証されていない。

この問題を解決するために、我々は文献から得られた 3 つの省略エラーの説明、すなわち、独立モデル Independence model (Rumel et al., 2000)、語彙-編集モデル lexical-editor model (Baars, Motley, & MacKay, 1975)、語彙-しきい値モデル lexical-threshold model (Laine, Tikkala, & Juhola, 1998) を検討した。これらのモデルを既存の命名モデルに接ぎ木 (grafted) して複合モデルを作り、原理的にはエラーと省略の両方を説明することができる。これらの複合モデルの妥当性は、非命名反応を多くする失語症患者の命名プロファイルとのマッチングによって評価される。

回答エラーと無回答エラーの対比は、失語症だけでなく非失語症にも当てはまる。健常話者の顕在的な発話エラーは、自然なエラーコーパス分析 (Garrett, 1975 など) や、スリップを生じさせる実験課題 (Ferreira & Humphreys, 2001 など) で研究されている。健常話者における無回答エラーは、そのような単語発見ブロックをしばしば伴う「tip-of-the-tongue」の状態を通して研究されている (例えば、Burke, MacKay, Worthley, & Wade, 1991; Harley & Bown, 1998; Meyer & Bock, 1992; Miozzo & Caramazza, 1997; Vigliocco, Antonini, & Garrett, 1997)。健常者の産出に関する理論では、無回答エラーと回答エラーの両方を非公式に説明しているが、両者を統一的に計算で説明する試みはない (ただし、この方向での研究については、Levett, Roelofs, & Meyer, 1999; MacKay, 1987 を参照)。

1.1 無回答反応のモデル

独立モデル

無回答エラーの最も単純な扱いは、他の反応を生成する処理過程から独立していると仮定することである (Rumel et al., 2000)。独立モデルは機械的なものではなく統計的なもので、無呼称反応が起きたときにエラーがどのように分布するかを指定するが、それらの反応を特定の処理システムに帰属させるものではない。例として、無回答ではないサンプルパターンを考えてみよう。50 %の正解、10 %ずつの意味エラー、形式エラー、および無関連エラー、20 %の非単語エラー。次に、ある不特定多数の処理過程が、正答の一部を非名称回答に変換すると仮定する。独立モデルが正しいければ、この変換によって、各明白な回答カテゴリの相対的な割合は変わらない。例えば、40 %の非名称回答を仮定すると、残りのカテゴリはそれぞれその割合で減少し、30 %の正解、6 %の意味的誤答、6 %の形式的誤答、6 %の無関係誤答、12 %の非名称誤答となる。要するに、独立した非名称化処理は、その頻度に比例して、明白なカテゴリから「盗む」のである。

語彙編集 Lexical-editor モデル

話者は自分の計画した出力を監視し、言語的に逸脱している場合はそれを抑制しているように見える (Garney & Dell, 1984; Hartsuiker & Kolk, 2001; Levelt, 1983; Postma, 2000)。この種の処理過程として最も一般的に仮説が立てられているのは、語彙編集 lexical editor である。Baars (1975) は、非単語を作るときに音韻交換の誤りが少なくなることを実証しており、潜在的な非単語のスリップが発話される前に捕獲され、抑制されるか修正されることを示唆している (Dell, 1986; Dell & Reich, 1981; Humphreys & Swendsen, 2002 も参照)。このような編集処理過程が、失語症患者が発する名前のない反応の多くに関与している可能性がある (Mitchum et al., 1990)。例えば、Schwartz & Brecher (2000) の命名の回復に関する研究では、患者 AK の正解率が低いことを考えると、非単語エラーの割合は比較的低く、無回答も多かったことから、潜在的な非語彙的出力が非命名反応に変換されていることが示唆されている。独立モデルのようにすべての顕在化したカテゴリからの応答を変換するのではなく、語彙編集 lexical editor モデルは潜在的な非単語の応答のみを変換する。

語彙-しきい値モデル

無回答エラーは編集者が逸脱した反応を抑制するために起こるという主張の代わりに、無回答はどの語彙項目も取り出せないことの結果であるという主張がある。この考え方は、tip-of-the-tongue 状態 (Burke et al. 1991 など) の説明に含まれており、Laine ら (1998) によって開発されたコネクションリスト産出モデルで具体化された。このモデルの重要な要素は、単語が話されるためには、最低レベルの活性化を超えなければならないという仮定である。単語ユニットは活性化の広がりによって検索されるが、どのユニットの活性化も指定されたしきい値を超えない場合、モデルは何の反応も起こさない。

1.2 発話における語彙アクセスの相互 2 段階モデル

失語症の命名課題での重み-崩壊 weight-decay モデルと意味-音韻 semantic-phonological モデルは、語彙アクセスの相互 2 段階モデルから派生したものである (Dell & O'Seaghdha, 1991; N. Martin, Dell, Saffran, & Schwartz, 1994)。この領域の類似モデルとしては、Rapp and Goldrick (2000) の制限付き相互作用モデル、MacKay (1987) のノード構造理論、Berg and Schade (1992; Schade & Eikmeyer, 1998 も)、Harley (1984, 1993)、Sternberger (1985) によって開発された相互活性化モデルがある。これらモデルはすべて、活性化の双方向または相互の流れに関連しており、離散的段階説 (例えば、Levelt et al. 1999) やカスケード段階理論 (例えば、Caramazza, 1997; Cutting & Ferreira, 1999) とは異なる。まず、相互 2 段階モデルによる損傷なしの処理の扱いを紹介し、続いて、そこから派生した失語症モデルを紹介する。

健常者の語彙アクセス 語彙アクセスは、図 1 に示したようなネットワークの活性化が広がることで起こる。このネットワークには、意味特徴、単語、音素のユニットが含まれている。トップダウンの興奮性接続は、意味特徴を単語に、そして単語を音素に結びつける。ボトムアップの興奮性接続はその逆で、後のレベルから前のレベルへの相互的フィードバックを提供する。抑制性の接続はない。

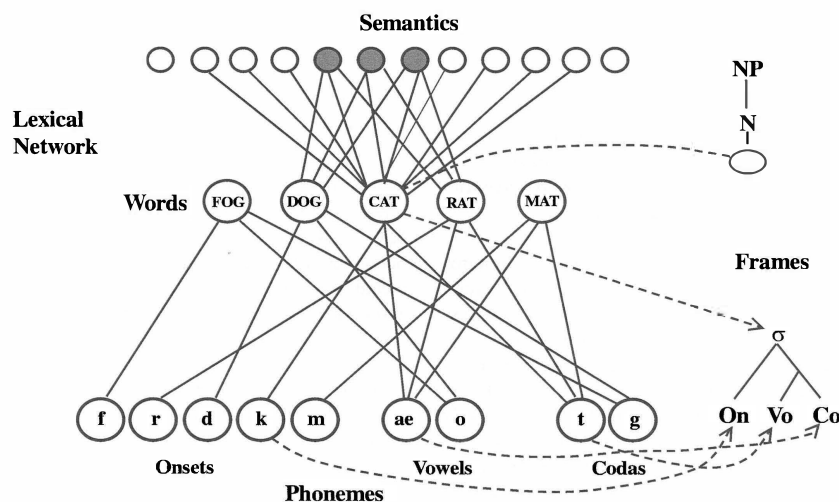


図 1. 音声合成における語彙アクセスに対する 2 段階相互活性化アプローチ。Dell et al. (1997) から引用

このモデルの「2 段階」とは、単語へのアクセスと音韻へのアクセスのステップが異なることを意味しており、多くの産出モデルの特徴となっている (例えば、Bock & Levelt, 1994; Griffin & Bock, 1998; Levelt et al, 1999; Rapp & Goldrick, 2000)。単語へのアクセスは、ターゲットの意味特徴のそれぞれに 10 ユニットの活性化の衝撃を与えることから始まる。活性化は、ノイズ月線形活性化規則に従って、8 時刻ステップでネットワークに拡散する:

$$A_{j,t} = A_{j,t-1}(1 - \text{decay}) + \sum w_{ji} A_{i,t-1} + \text{noise}.$$

ここで $A_{j,t}$ は時刻ステップ t におけるユニット j の活性化、 decay は減衰率、 w_{ij} はソースユニット i から受信ユニット j への結合係数である。各時刻ステップの間、各ユニットの活性化レベルは正規分布ノイズ、具体的には内在性ノイズ ($\text{SD}=0.01$) と活性化ノイズ ($\text{SD}=0.16 A_{j,t}$) の 2 つの成分の合計によって摂動される。8 時刻ステップの後、適切な構文カテゴリの最も高く活性化された単語ユニットが選択され、単語アクセスページが完了する。

単語アクセス時には、通常、ターゲット (例: CAT) が最も活性化される。しかし、活性化が広がると、他の単語やその音素、意味上の特徴も活性化されるようになる。意味上の隣接語 (例: DOG) は、共有された意味上の特徴から活性化を得る。さらに、単語へのアクセス時には、活性化が単語から音素へと広がり、音素から単語へとフィードバックされるため、MAT のような音韻的に関連する (または正式な) 隣接語も活性化される。また、RAT のような意味論的・音韻的に混在した隣接領域は特に活性化され、トップダウンとボトムアップの両方の影響による活性化が見られ、「混合スエーラ効果」(例えば、N. Martin, Gagnon, Schwartz, Dell, & Saffran, 1996) と呼ばれている。ノイズの影響で、ターゲットではなく、これらの関連語のいずれかが選択される可能性がある。極端な場合には、関連性のない (LOG) 単語が誤って選択されることもある。

音韻へのアクセスは、選択された単語ユニットに 100 ユニットの活性化が起こることから始まり、さらに 8 つの時間ステップで、活性化がトップダウンとボトムアップの両方の方法でネットワーク全体に広がる。最後に、最も活性化された音素が選択され、音韻フレームのスロットに連結される。このモデルでは、単音節の CVC 単語のみを符号化する。可能な音素はオフセット、母音、コードのセットに整理され、各セットの中で最も活性化されたものが選ばれ、CVC フレームの適切なスロットにリンクされる。これらの音素のうち 1 つ以上が選択された単語に含まれていない場合、音韻エラー (実在の単語または非単語) が発生する。

モデルの評価では、正解 (CAT)、意味エラー (DOG)、形式的エラー (MAT)、混合エラー (RAT)、無関係単語エラー (LOG)、非単語エラー (LAT) の 6 つのカテゴリを考慮した。このモデルは、175 項目の Philadelphia Naming Test (PNT; Roach, Schwartz, Martin, Grewal, & Brecher, 1996) に基づいて、健常話者の絵画命名におけるこれら 6 カテゴリの確率をシミュレートするように初期設定された。そのためには、モデルの語彙的近隣関係と、接続の重みと減衰のパラメータを指定する必要があった (詳細は Dell et al. 1997 参照)。(脚注 1)

脚注 1. このモデルを開発して以来、混合エラーが過小評価されているという観察結果を受けて、語彙の近傍性に小さな変更を加えた (Foygel & Dell, 2000; Rapp & Goldrick, 2000)。現在のバージョンでは、混合エラーの確率は 0.004 から 0.008 に増加している。

失語症者の語彙アクセス Dell ら (1997) は、失語症者のエラーパターンは、健常者のエラーパターンと“ランダムパターン”の間の連続体であるという仮説を立てた。ランダムパターンとは、単語形式の検索に語彙の影響がなく、検索された形式が (英) 単語の一般的な音韻特性を尊重していた場合に観察される、各応答カテゴリの割合のことである。Dell らは、単語の意味隣接語と形態隣接語の相対的な数や、音韻的に合法的な文字列が単語である割合など、語彙によって得られる機会を反映するために、いくつかの情報源から英語のランダムパターンを推定した。そして、この推定値をシミュレートするために、モデルの近傍領域を構築した。

連続性仮説と呼ばれる、正常からランダムまでの可能な応答パターンの連続性は、失語症で起こりうるエラーパターンを制約する。モデルには 6 つの応答カテゴリがあり、特定のパターンは 6 次元空間の 1 つの点を表し、その空間の一方の端には正常なパターンがあり、もう一方の端にはランダムなパターンがある。連続性仮説が成立するためには、可能なパターンのセットが、これら 2 つの点を結び単純な連続領域に対応する必要がある。重み-崩壊 weight-decay (Dell et al., 1997) と意味-音韻 semantic phonological (Foygel & Dell, 2000) モデルは、この領域を生成する方法である。それぞれのモデルでは、通常モデルをどのように「病変」させるかが規定されており、この病変によって、通常点とランダム点の間の領域が定義される。意味-音韻モデルは、意味特徴と単語の間のトップダウンおよびボトムアップの接続の重みである s と、単語と音素の間の対応する接続の重みである p の 2 つのパラメータの一方または両方を減少させることによって病変させる。また、重み-崩壊モデルでは、病変可能なパラメータとして、意味的接続と音素的接続の両方に適用される単一の重みパラメータ w と decay の 2 つがある。(脚注 2)

脚注 2. Decay は、semantic-phonological モデルでは一定の 0.6 に設定されている。

この 2 つのモデルは、21 人の患者を対象とした研究 (Dell ら 1997; Foygel & Dell, 2000) と、9 人の患者を対象とした研究 (Rumel ら, 2000) で比較された。これらの研究では、どちらのモデルも、患者の多様なパターンにかなりよく適合していた。しかし、理論的には、意味的・音韻的モデルの方がより有効であると考えられる。というのも、意味的・音韻的病変を別々に作ることができれば、ほとんどどちらか一方のタイプの誤りしか示さない被験者のケーススタディが報告されていることをよりよく説明できるからである (Rapp & Goldrick, 2000; Rumel et al., 2000)。さらに、この区別により、意味的・音韻的モデルは、失語症における単語の繰り返しについて、より正確な予測を立てることができる。意味論からの検索が必要ないため、復唱は語彙-意味論的重みよりも語彙-音韻的重みに依存すると想定される (Foygel & Dell, 2000)。

本研究では、これまでどちらのモデルでも説明できなかった、無応答反応を多くする失語症者の誤りのパターンを説明する上で、重み-崩壊モデルと意味-音韻モデルの有効性を検証することができた。本研究では、無応答反応を異なる方法で説明する命名モデルを定義し、重み-崩壊モデルと意味-音韻モデルを独立モデル、語彙-編集モデル、語彙-しきい値モデルの各説明と組み合わせ、文献から選ばれた失語症者のエラープロファイルに対する各モデルの適合性を評価した (Dell et al., 1997; Gordon, 2002; Rumel et al., 2000; Schwartz & Brecher, 2000)。

2. モデル比較研究

2.1 方法

データセット PNT とその採点基準を用いた失語症の命名に関する発表された研究から 14 個のプロファイルを選んだ。1 つのプロファイルは、1 人の患者の PNT の 175 枚の絵に対する正解、意味、形式、混合、無関連、非単語、無反応の割合からなる。本研究では、上述の 4 つの論文から、少なくとも 18 % の非命名応答を含むすべてのプロファイルを対象とした。PNT とその採点基準を用いた失語症の命名に関する研究の中から 14 のプロファイルを選んだ。1 つのプロファイルは、1 人の患者の PNT の 175 枚の絵に対する正解、意味、形式、混合、無関連、非単語、無応答の割合からなる。本研究では、上記の 4 つの論文から、無応答応答が 18 % 以上含まれている、あるいは非命名応答が誤応答全体の 80 % 以

上を占めているプロファイルをすべて対象とした (脚注3)。プロファイルは、Dell ら (1997) の被験者 GB と VP、Ruml ら (2000) の IOC, JR1, EA, JR2, Gordon (2002) の S10, S16, S21, S31, Schwartz and Brecher (2000) の WR5, AS1, AK1, AK2 のものである。(JR1 と JR2 は 2 人の異なる人物であるが、AK1 と AK2 は同一人物から 2 つの異なる時期に採取されたプロファイルであることに注意)。これらの被験者の情報を表 1 に、彼らの命名プロファイルを表 2 にイタリック体で示す。その他の背景情報は、オリジナル出版物から得ることができる。

脚注 3 当初は 20 % の無命名応答を求めているが、より多くの被験者を対象とするために基準を緩和した。

Table 1. Characteristics of subjects contributing naming profiles

Subject	Age (yrs)	Months post-onset	Education (yrs)
GB ^a	83	4.1	12
VP ^a	64	60.0	12+
IOC ^b	55	18.0	14
JR1 ^b	50	36.0	12
EA ^b	65	216.0	15
JR2 ^b	43	60.0	16
S10 ^c	81	48.0	12
S16 ^c	65	10.0	9
S21 ^c	67	62.0	9
S31 ^c	53	129.0	6
WR5 ^d	62	14.5	12
AS1 ^d	61	9.0	N/A
AK1 ^d	77	14.0	N/A
AK2 ^d	77	17.0	N/A

^a From Dell et al. (1997).
^b From Ruml et al. (2000).
^c From Gordon (2002).
^d From Schwartz and Brecher (2000).

表 2

患者	正解応答	意味エラー	形態エラー	混合エラー	無関連エラー	非単語エラー	無応答エラー	前 2 者合算
GB	0.39	0.07	0.09	0.01	0.03	0.08	0.32	0.40
VP	0.28	0.07	0.11	0.05	0.17	0.04	0.28	0.32
IOC	0.17	0.03	0.00	0.01	0.00	0.00	0.79	0.79
JR1	0.45	0.03	0.01	0.01	0.00	0.00	0.50	0.50
EA	0.34	0.22	0.03	0.08	0.11	0.01	0.21	0.22
JR2	0.67	0.07	0.01	0.03	0.00	0.03	0.19	0.22
S31	0.04	0.09	0.05	0.03	0.07	0.03	0.68	0.71
S10	0.55	0.07	0.02	0.02	0.02	0.02	0.30	0.32
S16	0.66	0.05	0.02	0.04	0.02	0.02	0.19	0.21
S21	0.61	0.05	0.06	0.02	0.00	0.08	0.18	0.26
AK1	0.42	0.05	0.02	0.01	0.01	0.01	0.48	0.49
AS1	0.15	0.05	0.11	0.02	0.21	0.19	0.27	0.46
WR5	0.52	0.07	0.05	0.03	0.06	0.05	0.23	0.28
AK2	0.85	0.01	0.00	0.01	0.00	0.00	0.13	0.13

無応答反応カテゴリーは、被験者が有効な命名を試みられなかったすべての回答からなる。このカテゴリーには、被験者が何も言わずに絵画の名前を言えないことを示す無応答や、意味的な説明や視覚的混乱を反映した雑多な反応、あるいは絵の一部を名付けるなど、ターゲット名を一言で言い表すことができない明白な反応が含まれる (脚注 4)。実用面では、3 つの実験的研究が、我々の定義に近い「その他」カテゴリーを使用していた (Dell et al., 1997, Ruml et al., 2000, and Gordon, 2002)。Schwartz and Brecher (2000) は、無応答の記述と定義される無命名カテゴリーと、一貫性を保つために雑多なカテゴリーを統合した。原則的な理由は、無記名回答は、被験者が絵に正解か不正解かのラベルを付けられない限り、自然な分類であるからである。もちろん、このカテゴリーを扱うことに問題がないわけではない。沈黙と意味的記述の間には、特に被験者のターゲットに関する知識、自分の障害に対する認識、障害に適応するための戦略などについて、重要な違いがある。しかし、これらの問題は今回の調査の範囲を超えている。

モデル 重み-崩壊 weight-decay (WD) モデルと 意味-音韻 semantic-phonological (SP) モデルを、 独立 independence, 語彙-編集 lexical-editor, しきい値 threshold の各アプローチと組み合わせて、WD-独立モデル, SP-独立モデル, WD-編集, SP-編集, SP-しきい値, の各モデルを作成した。(後述の脚注 5 の理由により、WD-しきい値モデルの検証は報告していない)。5 つのモデルはそれぞれ、モデルの反応の割合がプロファイルにできるだけ近くなるようなパラメータ値を見つけ、特に χ^2 の適合度の値を最小にすることで、14 のプロファイルに適合させた。各モデルは、プロファイルごとに 3 つの自由パラメータを使用して、1.0 になるように制約された 7 つの割合に適合させている。重み減衰 WD モデルと意味音声 SP モデルには、それぞれ w and $decay$ と s and p という 2 つのパラメータがあった。無応答を扱うと、さらに自由パラメータが追加される。独立モデルの場合、追加のパラメータは、非名称の回答の割合である。語彙編集 lexical-editor モデルの場合は、非単語の回答が非名称の回答に変換される割合を指定する。SP-しきい値モデルの場合、3 番目のパラメータは閾値 t で、選択された単語ユニットが命名プロセスを継続するために保有しなければならない活性化の最小レベルである。

無応答反応のパラメータは、無応答反応に対するアプローチに応じて、異なる方法で複合モデルに組み込まれた。独立モデルと編集バージョンのモデルでは、プロファイルを調整することで、明示的にフィットするパラメータの数を 3 つから 2 つに減らすことができた。この調整は、独立モデルでは簡単に行うことができた。無応答反応は、表出反応を生成する処理過程から独立していると仮定されるため、無応答の割合は単純にプロファイルから削除され、残りの割合は正規化される、つまり修正された合計の割合として再計算される (Rumel et al., 2000)。正規化されたプロファイルは、後述する残りの 2 つのパラメータを決定するためにフィッティングされる。語彙-編集モデルも、簡単な調整を行うことで適合させた。無応答反応は抑制された非単語に対応すると仮定されるため、無応答の割合を非単語の割合に加え、調整されたプロファイルを残りの 2 つのパラメータのために適合させた。SP-しきい値モデルでは、3 つのパラメータを同時に明示的に探索する必要があった。

独立モデルと編集モデルのフィッティングルーチン フィッティング手順の目的は、調整後のプロファイルに対して全体的に最も近い適合度を生み出す 2 つのパラメータを見つけることである。各モデルのプロファイルを 10,000 回の実行で計算する必要があるため、最適な適合を探すのは非常に時間がかかる。そのため、事前にモデルプロファイルを生成して保存した。2 つのパラメータで定義される空間の中には、パラメータを少し変えただけでモデルのプロファイルが大きく変化する領域と、パラメータを少し変えただけでプロファイルがほとんど変化しない領域がある。小さな変化が大きな変化をもたらす領域に対応するために、一様に高密度なモデルプロファイルの配列を生成するのではなく、Foygel and Dell (2000) のように、可変解像度マップを使用した。

そこで、どのモデルプロファイルを保存するか、つまり可変解像度マップ上のどの点が占有されているかを決定するために、新しい方法を用いた。簡単に言うと、準ランダムな点が繰り返し生成されるが、関連するモデルプロファイルは、その点の (パラメータ空間での) 最近傍が (モデルプロファイル空間での) 十分に異なっている場合にのみ生成され、保存される。その結果、隣人との差が非常に小さいモデルプロファイルで構成されたマップが作成されるが、これらのモデルプロファイルが一緒になって、可能なエラーパターンの空間を必要なだけ高密度に定義する。

具体的には、Sobel 系列を用いて点を生成した。Sobel 系列とは、準ランダムに点を生成する方法で、カバレッジを最大にし、「しこり」を最小にするものである (Press, Teukolsky, Vetterling, & Flannery, 1992)。500 個の初期点が生成され、以下の手順で後続のどの点を残すかを決定した。候補となる点の 8 つの最近傍のモデルプロファイルの平均値を、それらの近傍のそれぞれと比較した。平均値と 8 つの近傍のすべてとの χ^2 比較が任意の閾値 (ここでは 8, PNT の $N=175$ を使用) よりも小さい場合、そのポイントは余計なものと判断され、スキップされた。それ以外の場合は、その点のモデルプロファイルを生成・保存し、処理を続ける。500 個のポイントが連続してスキップされた時点で、マップは完成したとみなされ、SPモデルでは 3782 個、WD モデルでは 7481 個のモデルプロファイルが保存された。これらの適合ルーチンと基礎となるモデルは、次の web アドレスで使うことができる: <http://langprod.cogsci.uiuc.edu/cgi-bin/webfit.cgi>

SP-しきい値モデルのフィッティングルーチン SPモデルの語彙-意味、語彙-音韻の重みパラメータ s と p に、語彙の閾値を表す第 3 の自由パラメータを追加した。モデルが最も活性化された単語ユニットを選択した時点で、どの単語ユニットの活性化も t より大きければ、そのトライアルは非命名反応カテゴリーに割り当てられる。そうでなければ、選択された単語は音韻アクセスを経て、顕在化した反応が生成される。

SP-しきい値モデルでは 3 つの自由パラメータを指定する必要があるため、上述の可変解像度の 2 次元マップを使用することはできない。その代わりに、丘登りの手法を用いた。まず、SP-独立モデルから得られた s と p の値からスタートし、SP-しきい値モデルのパラメータ空間を探索した経験に基づいて、 t の最適値を推測した。そして、このスタート地点に隣接する、少なくとも 1 つのパラメータを 0.01 ステップで増減させた 26 地点を評価した。そして、 χ^2 の値が最も小さくなる位置を選択した。そして、選択された場所を出発地に置き換えて、この手順を繰り返した。この作業は、現在の場所よりも χ^2 の値が小さくなる隣接した場所がなくなるまで続けた。この時点で、ステップサイズを 0.001 に下げて、このプロセスを再び開始した。そして、最終的に得られた位置が、パラメータ値のフィッティングに用いられた。なお、この手順は、プロファイルに最適なフィットを見つけることを保証するものではなく、また、最適なフィットが最終的な位置の近くにあることを保証するものでもない。各プロファイルについて、パラメータ空間のランダムに決定された領域 (0 と 0.1 で囲まれた領域) から始めて、適合処理を繰り返した。すべての場合において、上述の初期位置から始めたフィットは、ランダムに始めた位置から得られたフィットよりも良いか、あるいは同じであった。しかし、値を発見しなければならない第 3 のパラメータを追加したことで、SP-しきい値モデルのフィットは、可変解像度マップから生成されたフィットよりも最適でない可能性が高くなっていることを認識しなければならない。

3. 結果と考察

表 2 は、14 のプロファイルのそれぞれについて、7 つのカテゴリーのそれぞれにおける回答の割合と、5 つのモデルのそれぞれから生成された割合とパラメータを示している。また、各プロファイルについて、適合度を表す 2 つの尺度を示した。適合度を発見するために使用された統計量である χ^2 と、平均平方根 (RMSD) である。RMSD は、得られた比率がモデルの予測値からどの程度乖離しているかを要約したもので、より直感的な適合度の指標である。(RMSD が 0.030 ということは、モデルと患者の比率が平均して 0.030 離れていることを意味する) これらの比率はどの程度まで近づけるべきだろうか? この分野では、定量的なモデリングのベンチマークが確立されていないため、一概には言えない。そのような状況になるまでは、適合度の測定は絶対的なものではなく、比較的に解釈するのが最善である。そのため、ここではどのモデルが最もフィットしているかを中心に議論する。

全体的な適合度にはモデル間で明確な違いがあり、最も顕著なのは、意味-音韻モデルが重み-崩壊モデルよりも優れていることである。SP-独立モデル(130.4)、SP-しきい値モデル(130.1)、SP-編集モデル(171.1) の χ^2 の合計値は、WD-独立モデル(262.4)、WD-編集モデル(331.6) の χ^2 の合計値よりもはるかに小さかった。(脚注 5) 平均 RMSD 測定値も同様のことを物語っている。SP-独立(0.017)、SP-しきい値(0.018)、SP-編集(0.028)、

WD-独立(0.026), WD-編集(.050)である。表2が示すように、2つのWDモデルが3つのSPモデルよりも平均的にフィットしているプロファイルはなく、SPモデルの方がはるかに優れているプロファイルがいくつかある(AS1, VPなど)。特に、WDモデルは、無関係な単語と非単語のカテゴリを予測するのが難しい。この問題は、大域パラメータを使用していることに起因する。無関係単語エラーを促進するパラメータ設定は、必然的に多くの非単語エラーも生み出す。しかし、AS1とVPの非単語の誤りは、単語の誤りに比べて特によくはない。SPモデルは、単語エラー(主に s パラメータで決定される)と非単語エラー(主に p パラメータで決定される)への影響を分離することができるため、これらのプロファイルを適合させる上で優位に立つことができる。

脚注5. WD-しきい値モデルについても検討したが、WD-独立モデルと同等の適合度をもたらすパラメータを見つけることはできなかった。SP-しきい値モデルとSP-独立モデルが似たような適合度であったことを考えると、より厳密なパラメータ探索を行えば、WD-しきい値モデルとWD-独立モデルも似たような適合度になるかもしれない。しかし、それでもSP-しきい値モデルやSP-独立モデルよりは劣っていると思われる。

もう一つの結果は、独立モデルと閾値モデルが、語彙-編集モデルよりも優れた適合性を示すことである。5つのモデルすべてを考慮すると、SP-independenceモデルとSP-thresholdモデルが最も優れていることがわかる。SP-独立モデルの $t(13) = 2.52$ (脚注6)やWD-独立モデルの $t(13) = 2.20$ よりも有意に小さくなっている。また、SP-thresholdモデルは、SP-editorモデルよりも明らかに優れており、 $t(13)=2.73$, WD-independenceモデルよりも(わずかに)優れている。SP-独立モデルの平均RMSD(0.017)とSP-しきい値モデルの平均RMSD(0.018)は、モデルとプロファイルの比率がほぼ一致していることを示している。さらに、これら2つのモデルは、プロファイル比率の分散のそれぞれ96.3%と95.9%を占めている(脚注7)。SP-独立モデルの96.3%の値は、非命名カテゴリ内の分散を含んでおり、このモデルは、3番目の自由パラメータを非命名の回答の割合に設定することで、些細なことだが適合する。非命名カテゴリを除外しても、このモデルは残りのカテゴリの分散の93.0%を占めている。

脚注6. $t(13)$ の臨界値は2.16である。これらの t 検定で使用された2つの独立モデルのRMSDは、これらのモデルが非命名回答の割合と一致したことが評価されないように上方修正された(これらのモデルの第3パラメータはちょうどこの割合なので、完全に一致している)。具体的には、この調整では、これらのモデルの7つの割合ではなく、6つの割合での二乗偏差の合計を平均化しています。この調整は、独立モデルが編集モデルよりも有利であるという我々の結論からすると、保守的なものである。

脚注7. この指標は、モデルとプロファイルの比率の二乗偏差を、各プロファイルの比率と全プロファイルにわたるそのカテゴリの平均比率の二乗偏差と比較する。

SP-独立モデルとSP-しきい値モデルは、1つの例外(EA後述)を除いて、個々のプロファイルにかなりよく適合した。SP-編集モデルが近似的に適合し(例えば $RMSD < 0.020$)、SP-独立モデルとSP-しきい値モデルがフィットしないプロファイルはないが、その逆は明らかである。例えば、IOC, JR1, JR2, SIO, S16では、SP-independenceモデルとSP-thresholdモデルがかなり良い結果を示している。IOCのプロファイルがその一例である。SP-編集モデルでは、多くの無応答回答(0.79)は、抑制された非言語であると想定される。これだけの割合の非語を生成するためには、モデルは p に非常に低い値(0.001)を設定しなければならない。しかし、非単語エラーを生成する処理過程では、形式エラーも生成される。SP-編集モデルがIOCに割り当てたパラメータでは、期待される形式エラーの割合は0.08であり、得られた割合が0であることは対照的である。

このような形態エラーの過剰予測は、他のSPモデルと比較した場合、SP-編集モデルの一般的な問題である。図2は、各プロファイルごとに、各モデルの予測された割合が、各回答カテゴリで得られた割合からどのようにずれているかを示している。(EAはどのモデルでもうまくフィットしなかったので除外した。この理由については後述する)。正の数は、プロファイルの比率がモデルの比率を上回っていることを示し、負の数は、プロファイルの比率が小さすぎることを示す。SP-独立モデルとSP-しきい値モデルの偏差は小さく、ゼロの片側にも片側にも傾かない。SP-編集モデルは偏差値が大きだけでなく、体系的であるように見える。例えば、非単語のカテゴリはほとんど常に過剰に予測され、フォーマルの割合は過小評価される。これらの偏差はIOC(図の中で最も逸脱した線)で最大だが、他のほとんどのプロファイルにも存在する。

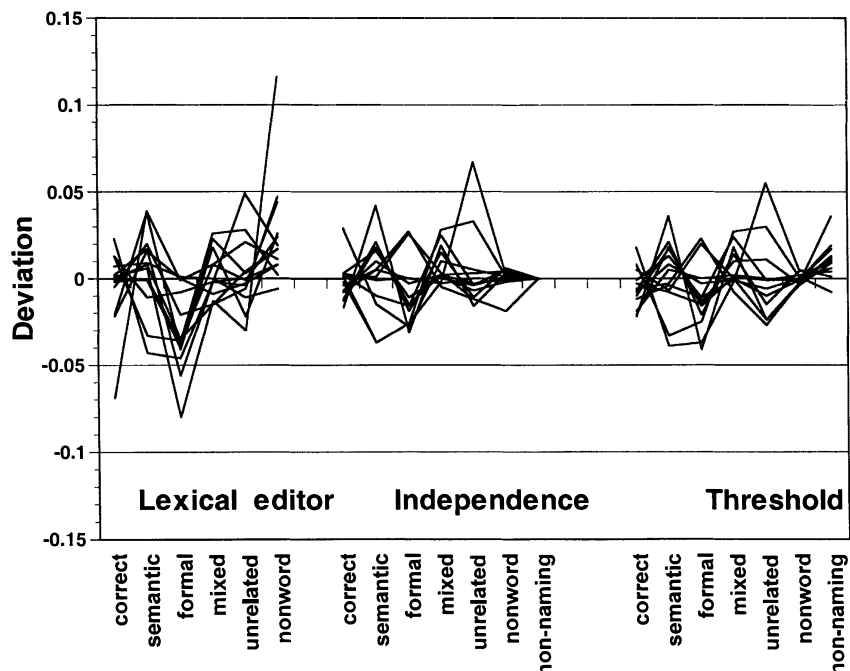


図2. SP-編集モデル(左), SP-独立モデル(中), 語彙-しきい値モデル(右)における、得られた応答比率とモデル応答比率との乖離を、応答カテゴリの関数として表したものの。各線は1つのプロファイルを表している(EAは含まれていない)。

3 つ目の重要な発見は、どのモデルも EA のプロファイルにあまり適合していないということである。意味的誤り(0.22), 混合的誤り(0.08), 無関係な誤り(0.11)が多く、形式的な誤り(0.03)が少ないことから、SP モデルを含めたすべてのモデルが課題となっている。しかし、これらのモデルの近隣領域では、このような高い割合の意味エラーや混合エラーを生成することはできない。さらに、SP モデルが多くの混合エラーや無関係な単語のエラーを生成する場合、かなりの数の形式エラーも生成する必要がある。また、無関係語エラーは、ターゲットの活性化が十分に低下しており、ターゲットの代わりにこれらの形式的隣接語が選択されることが多いことを示している。このように、EA の形態エラーの割合の低さは問題である。

EA の命名パターンを知る手がかりとして、意味エラーや混合エラーの多くが名詞ではなく動詞であるという珍しい発見がある (Rumel et al. 2000)。しかし、命名モデルは、患者であっても名詞の選択を保証する処理過程がエラーなく作動することを前提としている。おそらく EA のプロファイルは、語彙へのアクセスを制限する構文システムの失敗を反映しているのだろう。Gordon and Dell (2003) は、そのような失敗のモデルを実装した。統語的／順序的な状態から単語単位への重みを減少させると、非文法的な産生、特に、単語が削除されたり、目標とする統語的カテゴリに属さない別の単語に置き換えられたりする命名障害が生じた。EA の文生成は、非流暢ではあるが、調音性の問題はなく (Rumel et al., 2000)、非文法的 (Berndt, Mitchum, Haendiges, & Sandson, 1997) と表現されており、これは統語的／順序的な状態の病変の仮説と一致していることに注意したい (脚注 8)。

意味的エラーと混合エラーが多く、音韻エラーがないことから、EA のプロファイルは、意味的音韻モデルに疑問を投げかけた文献にある他の 3 つのケースに似ている: PW (Rapp & Caramazza, 1998; Rapp & Goldrick, 2000), JF (Foygel & Dell, 2000), DP (Cuetos et al., 2000)。このような患者は、命名困難が意味表現の問題から二次的に発生しているのであれば、モデルと矛盾しないかもしれない。これらのモデルは、意味的な入力に損なわれていないことを前提としており、意味的エラーやその他の単語エラーは、その入力を語彙単位にマッピングする過程で起こることに留意せよ。Shelton and Weinrich (1997) は、EA の単語理解力をテストした結果、EA には「意味知識の軽度の欠損」があると指摘している (p.122) が、意味入力に損なわれていないという仮定は、少なくとも EA には当てはまらないだろう。しかし、意味レベルの表現が保持されていると報告されている PW と DP には、この説明は当てはまらない (Cuetos et al., 2000; Rapp & Goldrick, 2000)。その代わりに、Rapp & Goldrick は、語彙的な音韻のフィードバックが制限されていると仮定することで、PW のエラーパターン (これは DP と似ている) を説明することができた。つまり、ボトムアップの重みがトップダウンのものよりも弱いということである。この制限により、Rapp & Goldrick のモデルは、単語アクセスの段階で形式的な誤りを引き起こすことなく、意味的な誤りを発生させることができた (後者は強いフィードバックによって促進されるからである)。

しかし、フィードバックを制限しても、それだけでは EA のプロファイルの説明することはできない。Rumel ら (2000) は、これらのモデルのフィードバックを制限したバージョンを検討したが、それでも意味エラー、混合エラー、無関連エラーの割合を十分に高くすることはできなかった。その結果、EA に適合しないのは、意味的な入力と統語的な入力に損なわれていないというモデルの仮定に起因すると考えられる。いずれにしても、このエラーパターンは、失語症の命名を一般的に説明するモデルとしては明らかに限界があることを示している。

SP-独立モデルとSP-しきい値モデルは、我々のサンプルの患者プロファイルをよく説明しており、他の 3 つのモデルよりも優れていることは確かである。その結果、意味-音韻モデルの損傷の仮定と、独立モデルとしきい値モデルで提案された非命名反応の扱いは、重み-崩壊モデルと語彙-編集モデルよりも支持されている。これは、独立モデルやしきい値モデルが普遍的に適用されるとか、語彙-編集モデルで想定されている非単語反応の抑制が起こらないということではない。むしろ、我々のサンプルから得られた証拠は、独立モデルとしきい値モデルが、患者の成績を説明する上で、バランスよくより強力であるということである。さらに、これらのモデルは、失語症の語彙アクセスのモデルに適用できる観察範囲を広げることができる。

これまで、SP-独立モデルとSP-閾値モデルと一緒に扱ってきた。適合度の面では、他のモデルよりも明らかに優れている、区別することはできなかった。これは、個々のプロファイルについても、全体の適合度についても同様である。EA だけが適合せず、両モデルとも形式、混合、意味の各カテゴリの比率が一致していない。残りのプロファイルについては、2 つのモデルは適度に適合し、似たような特徴を持っている。例えば、S10 では、SP-しきい値と SP-独立の予測プロファイルは、非単語 (0.02 と 0.01) と無応答 (0.28 と 0.30) のカテゴリでのみ異なる。また、 s (0.021 と 0.017) と p (両方とも 0.032) の値が似ていることから、この患者は語彙・音声障害よりも語彙・意味障害の方が大きいと考えられる。

明らかに、SP-しきい値モデルは SP-独立モデルのように振る舞う。図 3 はその理由を示している。この図では、典型的な失語症のパラメータを持つサンプルモデルを用いて、しきい値がどのような反応を抑制する傾向があるかを示している。 $s=0.02$, $p=0.02$ でしきい値を設定しないモデルと、様々な閾値を設定した同じモデルを比較している。縦軸は、しきい値を超えた回答の割合を示している。 t が 0 に等しい (つまり、閾値がない) 場合、すべての回答が顕在化応答として現れる。しきい値が大きくなるにつれて、しきい値を超える回答の割合は減少する。細い線は、しきい値が反応を抑制する能力が、反応カテゴリによって多少異なることを示している。それにもかかわらず、 t が増加するにつれて、すべてのカテゴリが劇的に減少している。一本の太い線は、すべてのカテゴリで保存された回答の割合の平均値を示している。本質的には、これは独立モデルを表しており、各カテゴリの出現頻度に比例して、無応答が各カテゴリから奪われる。このように、細い線と太い線が近接していることから、SP-しきい値モデルとSP-独立モデルの類似性の度合いがわかる。

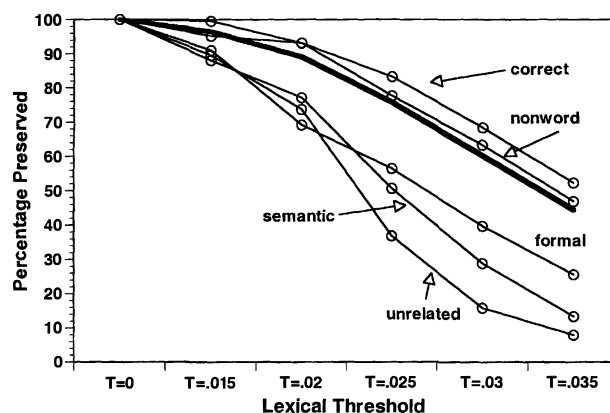


図3. 閾値 (T) が各回答カテゴリの保存される割合に及ぼす影響。太線は、各閾値で保存された回答の平均的な割合をカテゴリごとに示している。

しかし、似ているとはいえ、この2つのモデルは同じではない。例えば、しきい値は、正しい反応よりも、潜在的な無関係な単語の反応を抑制する可能性が高い。最も活性化された単語ユニットが無関係な単語に対応している状況は、すべての活性化が非常に低く、したがってしきい値以下である可能性が最も高い。しかし、これら2つのモデルの違いは、命名プロファイルの説明する能力に関しては微妙である。プロファイルで最も一般的な明白な反応カテゴリは、正解カテゴリと非単語カテゴリである。図3によると、独立性が保たれている場合とほぼ同じ割合で、これらのカテゴリからしきい値が取り除かれている。その結果、2つのモデルを区別することは困難である。

では、SP-独立モデルと SP-しきい値モデルのどちらかを選択する理由はあるのだろうか？ SP-独立モデルは使いやすいという利点があり、したがって、分析の焦点が顕在化したエラーカテゴリにある場合、無応答を扱うためのデフォルトの仮定とすることができる (例えば、Ruml et al., 2000)。その主な欠点は、無応答反応のメカニズムを規定していないことである。このモデルは、エラー分布を統計的に特徴づけることには成功しているが、それを説明することはできない。SP-しきい値モデルは、その説明をしてくれるかもしれない。しきい値は生物学的システムでは一般的であり、特にそのようなシステムのニューラルネットワークモデルでは一般的である。したがって、語彙ユニットが音韻アクセスに進むかどうかを支配する活性化レベルのしきい値は、もっともらしいメカニズムである。総合考察では、これらのモデルの対比や、我々の発見の他の影響についても検討している。

4. 総合考察

今回のモデル比較で得られた主な知見は以下の通りである。

1. 意味-音韻モデル (Foygel & Dell, 2000) は、重み-崩壊モデル (Dell et al., 1997) よりも、無応答の割合が高い命名プロファイルのマッチングに適している。
2. 無応答の分布は、語彙-編集機構 (Baars et al., 1975) よりも独立性機構 (Ruml et al., 2000) や しきい値機構 (Laine et al., 1998) を組み込んだモデルの方がよく説明できる。
3. SP-しきい値モデルは、予測力の点では SP-独立モデルとほとんど変わらないが、無応答反応の発生を説明するためのより理論的な機構を提供している。
4. EA のプロファイルは、意味的・混合的エラーの割合が比較的高く、形態エラーの割合が低いことから、どのモデルでもうまく当てはまらないことがわかった。

これらの結論を視野に入れることは重要である。S-P 独立モデルや S-P しきい値モデルが優れているからといって、逸脱した反応の抑制が起こらないわけではない。逆に、しきい値は、潜在的に誤った反応を防ぐために戦略的に決定された値であると解釈することもできる。さらに、語彙-編集モデルの性能が比較的低いことは、患者が非単語の誤りを優先的に抑制しないことを示唆するものではない。むしろ、純粋な語彙-編集の説明は、現在の証拠ではあまり好ましくない。一つの説明がすべての患者に当てはまるとは考えにくい。WRS, AK1, AK2 など、いくつかのプロファイルが SP-編集モデルによく当てはまったことは注目に値する。これらのプロファイルでは、非単語の誤りが少なく、無応答の回答が多い。これらの患者では、語彙-編集モデルが正しい説明である可能性がある。しかし、SP-独立モデルや SP-しきい値モデルもこれらのプロファイルを説明しており、SP-編集モデルがあまりうまく説明できなかった他のプロファイルも説明していることを強調しておく。

最後に、我々は、より正確な結果を得ることができる別の編集モデルを構築することができることを認識している。例えば、意味的に無関係な回答 (非単語、形態、無関連エラー) はすべて抑制の対象になるという仮説を立てることができる。しかし、このような広範な編集に基づくモデルは、独立モデルと同様に、複数のカテゴリからの潜在的な応答を無応答に変換するため、独立モデルとの区別が難しくなる。我々が語彙編集モデルに注目したのは、文献的に支持されているだけでなく、独立モデルのような振る舞いをするモデルとの区別が最大限につくからである。モデル検証では、予測が対照的な動機付けされたモデルを検証するのが最も効果的である。

最終的には、一般的にも特定の患者についても、無応答反応の最適な説明を決定するためには、命名プロファイルだけでは判断できない。さらに、競合するモデルを用いて予測を行い、検証する必要がある。ここでは、この戦略の2つの例を紹介する。すなわち、単語命名成績から単語復唱成績を予測することと、意味的隣接語の機能として命名プロファイルの変動を予測することである。

4.1 命名から復唱を予測

意味-音韻モデルと重み-崩壊モデルは、パラメータをプロファイルに割り当てる。Dell (2000) は、これらのパラメータを用いて、復唱成績を命名過程の音韻アクセス段階と区別することで、聴覚的単語復唱課題を予測した。これは、復唱には、無傷の知覚系を介して適切な単語ユニットにアクセスし、その後、その単語ユニットを音韻表現にマッピングすることが必要であると仮定している (入力と出力の語彙-音韻マッピングが別々に行われている証拠については、R. C. Martin, Lesch, & Bartha, 1999 を参照)。Foygel と Dell は、重み-崩壊モデルのパラメータよりも意味-音韻モデルの命名パラメータの方が復唱成績を予測することを発見した。これは、単語の繰り返しは、語彙-意味マッピングよりも、語彙-音韻マッピングに依存しており、意味-音韻モデルはこれらを区別して特徴づけることができるからである。重み-崩壊モデルのパラメータは、相対的に損なわれている語彙-音韻の能力を分離することができないため、復唱予測の精度が低くなっている。

また、復唱を予測することで、無応答反応の独立/しきい値モデルと編集モデルの説明を区別することができる。表2によると、意味-音韻モデルの独立モデル版としきい値モデル版では、14例中13例で s よりも高い p の値が割り当てられており、これらの患者は、語彙-音韻よりも語彙-意味の方が重度の病変であると考えられる。対照的に、SP-編集モデルでは、14例中10例で s よりも p の方が低く割り当てられている。これは、このモデルの論理では、無応答を抑制された非単語として扱うことで、音韻障害の証拠が増えるからである。良好な復唱は、語彙-音韻の重みにより大きく依存する。このため、SP-独立モデルと SP-しきい値モデルでは、これらの患者の復唱は、命名とは対照的に比較的良好であると予測されるのに対し、SP-編集モデルでは復唱がより劣ると予測される。

Rumlら (2000) は、彼らの研究に参加した被験者の一部について、単語復唱データを報告しており、本研究ではそのうちの2人 (EA と JR2) を取り上げた。EA も JR2 も、PNT 項目を間違いなく復唱した。この好成績を説明できるモデルがあるとするれば、それはどのモデルだろうか。SP-

独立モデルと SP-しきい値モデルの両方が可能であることがわかった。この 2 人の患者に選択されたパラメータを使って復唱課題を実行すると、EA ではそれぞれ 97% と 98%、JR2 では 96% と 95% の正解率となった。一方、SP-編集モデルでは、EA で 76%、JR2 で 80% の正解率しか予測できなかった。SP-独立モデルと SP-しきい値モデルが EA と JR2 の復唱を予測するのに優れているのは、SP-編集モデルよりも大きな p の値を割り当てたことで、繰り返しがよりよく保存されたためである。

4.2 意味的隣接効果 (Semantic neighbourhood effect)

Blanken, Dittmann, and Wallesch (2002) は、失語症患者 (MW) の命名エラーの研究において、項目によって、非単語または無応答エラーを起こす傾向が異なることを発見した。独立して収集された評価によって、意味的に近い競争相手がいるターゲットアイテムは、意味的エラーを引き起こす傾向があったが、競争相手が少ないターゲットアイテムは、代わりに無応答を引き起こすことが多かった。この 2 つのエラータイプはトレードオフの関係にあり、ターゲットの意味的隣接語の競合の関数として、2 つのカテゴリーの合計はほとんど差がなかった。MW では、意味エラーと無反応が主要なエラーカテゴリーであったため、このトレードオフは、競合する隣接語と競争しない隣接語では、正解の割合が同じであることを意味している。Blanken らの発見は、当然のことながら、意味エラーの確率は、意味エラーの機会に応じて増加することを示している。しかし、より重要なことは、無応答が促進されるのは、競合する多くの項目が活性化されるよりも、どの語彙項目も強く活性化されない場合であることを示唆している。

本論文では、意味エラーと無応答の間のトレードオフは、特定の患者に共通するエラータイプであれば、語彙-しきい値モデルの明確な予測となる。 $s = 0.02, p = 0.03, t = 0.03$ の SP-しきい値モデルを用いた例を考えてみよう。ターゲットの 1 つの意味上の隣接語 (例えば、ターゲット CAT に対して DOG) を持つモデルの隣接語をテストした場合、モデルは 57% の確率で正解し、35% の確率で反応せず、3% の確率で意味エラーを起こす。CAT の意味上の隣接語として DOG と RAT を含めることで隣接語の競争が激しくなると、これらの割合は 60% の正解、28% の無反応、9% の意味エラーとなる (脚注 9)。競合が激しくなると、Blanken ら (2002) が発見したように、無応答を犠牲にして意味エラーが増える。このモデルがこのような挙動を示すのは、競合がターゲットとその競合相手の両方の活性化を増加させるからである。活性化は単語間で反響し、これらの単語の共有された特性に影響を与える。競合語が多ければ多いほど、そしてそれらが共有する特性が多ければ多いほど、反響は大きくなる。その結果、最も活性化された単語ユニットは、それが正しいかどうかに関わらず、閾値を超える確率が高くなる。ゆえに、顕在的な反応が起こりやすくなる。競合単語が多いと、もちろん意味上の誤りも多くなるが、これは無応答の可能性が減ることでほぼ完全に補われる。

脚注9 RATはCATの音韻上の隣人でもあることに注意。これは、モデルが対話的な仮定をしているため、より競争相手になる傾向があるだけで、ポイントは、競争相手が増えると語彙置換エラーが増える一方で、無応答が減ることを実証することである。

語彙-編集アプローチでは、同じ予測はできない。パラメーターセット (例えば、 s と p は共に 0.01) で、編集が省略に変換できる回答である非単語を大量に作成した場合を考えてみよう。競合の小さい隣接語領域を使用した場合、モデルは 18% の正答、7% の意味エラー、52% の非単語、23% のその他のエラー (形態エラーや無関係エラー) を生成する。一方、競争の激しい領域では、18% の正解、15% の意味エラー、52% の非単語エラー、15% のその他のエラーという結果になった。競合が激しくなると、当然ながら意味エラーが多くなるが、非単語カテゴリーが犠牲になることはなく、52% のままであった。むしろ、その他のエラーが犠牲になっている。その結果、意味エラーの増加は、これらが抑制された非単語であるという仮定のもとでは、無応答の減少とは関連しないだろう。

独立モデルは、Blanken ら (2002) のトレードオフについて何を予測するかは難しい。このモデルは、無回答を生成する真のメカニズムを欠いているため、実質的にパラメータを追加しなければ、隣接語の競争力の関数として異なる省略の割合を作り出すことができない。そのため、正しい予測をする語彙しきい値モデルや、間違った予測をする語彙-編集モデルとは異なり、独立モデルは全く予測をしない。

復唱と意味隣接の予測は、命名プロファイル以外のデータを用いてモデルを検証できることを示している。復唱の予測は、SP-編集モデルよりも SP-独立モデルと SP-しきい値モデルを支持し、意味的隣接語による予測は SP-しきい値モデルとのみ一致した。しかし、これらの検証は、反復予測では 2 人の患者、意味的近隣関係予測では 1 人の患者しか関係していないので、決定的なものではなく、例証的なものである。

5. 結論

今回の研究では、誤応答だけでなく、無応答を用いた失語症の命名モデルを開発し、検証することが可能であることを示した。本研究では、SP-独立モデルと SP-しきい値モデルの 2 つのモデルを解析したが、ほとんど区別がつかなかった。これらのモデルは、ほとんどの場合、無応答を多くする傾向のある患者プロフィールに近いものであった。この分析は、これらのモデルの基礎となっている語彙アクセスの相互的 2 段階理論を制約するのに役立ち、特に、無応答反応の独立性と語彙しきい値の説明と同様に、病変の意味論的・音論的アプローチを支持する。失語症病変の重み-崩壊説による説明は、病的なパラメータ値が大域的であり、すなわち、語彙-意味論的能力と語彙-音韻論的能力を区別して特徴づけることができないという仮定に苦しんでいた (Rapp & Goldrick, 2000; Rumel & Caramazza, 2000 参照)。無応答の語彙-編集モデルによる説明では、プロファイルがあまりにも多くの回答を非単語として扱うことを余儀なくされた。語彙-編集モデルの説明と意味-音韻モデルの組み合わせは、いくつかのプロファイルに非常によく適合したが、他のアプローチもこれらのプロファイルによく適合した。

無応答反応を分析する方法を提供することで、正常および失語症の音声生成のモデルに情報を提供できる証拠のソースを増やしたいと考える。今回発表した仮説は、言語産出課題における失語症の行動の多くを説明することができたが、これらは予備的なものと考えなければならない。これらの発見を強化し、新たな予測を生み出すためには、健常者と失語症患者の両方から、またさまざまな課題から得られるさらなる証拠が必要である。話し言葉の産出障害に関心のある他の研究者の方々には、本研究の「方法」の項で紹介したウェブサイトを利用して、今回発表したモデルを検討していただきたい。