



Bringing Computational Models of Word Naming down to the Item Level

Author(s): Daniel H. Spieler and David A. Balota

Source: *Psychological Science*, Vol. 8, No. 6 (Nov., 1997), pp. 411-416

Published by: Sage Publications, Inc. on behalf of the Association for Psychological Science

Stable URL: <http://www.jstor.org/stable/40063226>

Accessed: 03-01-2017 15:51 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at

<http://about.jstor.org/terms>



Sage Publications, Inc., Association for Psychological Science are collaborating with JSTOR to digitize, preserve and extend access to *Psychological Science*

Research Article

BRINGING COMPUTATIONAL MODELS OF WORD NAMING DOWN TO THE ITEM LEVEL

Daniel H. Spieler and David A. Balota

Washington University

Abstract—*Early noncomputational models of word recognition have typically attempted to account for effects of categorical factors such as word frequency (high vs. low) and spelling-to-sound regularity (regular vs. irregular). More recent computational models that adhere to general connectionist principles hold the promise of being sensitive to underlying item differences that are only approximated by these categorical factors. In contrast to earlier models, these connectionist models provide predictions of performance for individual items. In the present study, we used the item-level estimates from two connectionist models (Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989) to predict naming latencies on the individual items on which the models were trained. The results indicate that the models capture, at best, slightly more variance than simple log frequency and substantially less than the combined predictive power of log frequency, neighborhood density, and orthographic length. The discussion focuses on the importance of examining the item-level performance of word-naming models and possible approaches that may improve the models' sensitivity to such item differences.*

Research on visual word recognition has traditionally involved factorial designs in which item variables (e.g., word frequency, spelling-to-sound regularity, neighborhood density) are “manipulated,” and the effect of these variables on the speed and accuracy of naming or making lexical decisions is measured. In the vast majority of such studies, a mean is calculated for each subject across items (or for each item across subjects) within a condition and entered into an analysis of variance, and the effects of “factors” are measured. A reliable influence of a factor such as neighborhood density is typically interpreted as being consistent or inconsistent with a given model. An enormous amount of work has taken this approach and demonstrated if, and under what circumstances, targeted factors influence performance in word-processing tasks. Moreover, there has been considerable development of models of word recognition that do a reasonable job of accounting for the main effects of and interactions among such factors. Most of the models in this area have been what we refer to as first-wave cognitive models. First-wave models are metaphorical, noncomputational models of word recognition performance. For example, consider the effect of word frequency in some of the first-wave cognitive models. According to Forster (1976), word frequency influences the position of a particular item within a frequency-ordered, orthographically defined bin, whereas according to Morton's (1969) model, frequency modulates the thresholds of word recognition devices called logogens. Of course, these models become complex rela-

tively quickly when one considers the plethora of factors beyond word frequency that influence performance in word recognition tasks.

More recently, there has been a second wave of word recognition models. These are computational models that adhere to general connectionist principles. Here we emphasize the computational model of word-naming performance developed by Seidenberg and McClelland (1989; referred to as SM89 hereafter) and the more recent implementation by Plaut, McClelland, Seidenberg, and Patterson (1996, Simulation 3; referred to as PMSP hereafter). Although it is beyond the scope of the present work to provide a full description of these models, it is sufficient to indicate here that both follow the basic architectural assumptions consistent with connectionist modeling. For example, both models include an input level that codes orthographic information across a set of simple processing units and another set of units that code the output of the model in the form of phonological information. The input and output units are connected to hidden units in the network (in the PMSP model, the output units are also connected to other output units). The connections between units are modified during training to capture the regularity of the mapping of spelling to sound.

The second-wave models take a very different approach to modeling word recognition than the first-wave models. First, these models learn particular spelling-to-sound patterns via training. Effects of factors are not built in, as in the first-wave models, but rather arise through the models' exposure to patterns in the language. Second, although both the SM89 and the PMSP models recognize the importance of additional inputs to the naming processes, the focus of these models has been on the computation of phonological information based on orthographic input, rather than on other inputs, such as semantics. We return to a discussion of the contribution of more abstract lexical-semantic factors to the naming process in the General Discussion (also see Plaut et al., 1996).

What is particularly promising about the second-wave models is that they do not merely predict the influence of a categorical factor on performance, but rather have the potential to capture performance at an item level. This is an important aspect of these models because they have the ability to reflect the more continuous nature of relevant factors (e.g., frequency, regularity) in addition to the categorical manipulations reflected in the designs of typical word recognition studies. Moreover, because of the complexity of the factors that have been identified in the word recognition literature, it is difficult to categorize each item in an experiment on the multiple dimensions necessary for a factorial crossing of relevant variables. For example, the role of spelling-to-sound correspondence will depend on a number of factors such as the frequency of a given target word, the number and frequency of neighbors with similar spelling-to-sound correspondences (friends), the number and frequency of neighbors with different spelling-to-sound correspondences (enemies), and probably many other variables. A factorial experimental design might require one to identify a sufficient number of items that are high frequency, irregular, from low-density neighborhoods, and so on. Such categorization can

Address correspondence to Daniel H. Spieler, Department of Psychology, State University of New York at Binghamton, Binghamton, NY 13902; e-mail: spieler@binghamton.edu, or to David A. Balota, Department of Psychology, Washington University, St. Louis, MO 63130; e-mail: dbalota@artsci.wustl.edu.

be difficult and also potentially misleading. For example, consider a word such as *pint*. This word might be categorized as an irregular word although, as noted by Plaut et al. (1996), it is, in fact, regular in three of the four letters. Plaut et al. pointed out that the "connectionist approach, by contrast, avoids the need to impose such unfortunate divisions and leaves a mechanism that exhibits sensitivity to all of these partially regular aspects of so-called exception words" (p. 102). The notion is that the effects of the factors in factorially crossed word recognition studies should naturally fall from a network that "learns" the correct spelling-to-sound correspondences in a frequency-based, item-sensitive fashion.

In the present study, we begin to address how well these second-wave models capture performance at the item level (see Besner, in press; also see Besner, Twilley, McCann, & Seergobin, 1990; Seidenberg & McClelland, 1990). In pursuit of this goal, we obtained naming latencies for each of 2,870 words that were used to train the PMSP and SM89 models and compared the models' output with the mean naming performance for humans at the item level.¹ We should emphasize that accounting for performance at the item level was not the focus of the PMSP and SM89 models. However, because these models make specific predictions for specific items in terms of error scores (SM89) or settling times (PMSP), we can consider how well the models account for performance on individual items. This is a particularly seductive aspect of these models and is very much in the spirit of the connectionist modeling approach, which aims to capture the continuous graded nature of factors affecting naming performance. In fact, one might argue that a potential gold standard of these models is to account for variability across items, instead of accounting for influences of categorical factors.

METHODS

Participants

Thirty-one individuals were recruited from the undergraduate student population at Washington University. Individuals were paid \$20 for their participation. The individuals had a mean age of 22.6 years ($SD = 5.0$) and a mean of 14.8 years of education ($SD = 2.1$).

Stimuli

The words consisted of 2,870 single-syllable words appearing in the training corpora of the PMSP and SM89 models. These words range in frequency from 68,246 to 0 counts per million according to Kucera and Francis (1967). The words range from two to eight letters (only *straight* and *strength* have eight letters).

Apparatus

An IBM-compatible CompuDyne 486 computer was used to control the display of stimuli and to collect response latencies to the nearest millisecond. The stimuli were displayed on a NEC4G 14-in. color VGA monitor in 40-column mode in white on a blue background. The naming latency for each word was measured using a Gerbrands Model G1341T voice-operated relay interfaced with the computer.

1. Item-naming latencies for the 2,870 words used in this study can be obtained from the authors upon request.

Procedure

Each individual participated in two experimental sessions, naming 1,435 words in each session. Words were presented in a different random order for each participant. At the beginning of each session, the individual was seated in front of the computer and given the instructions for the experiment. Participants were told that they would be shown single words at the center of the computer screen and that their task was to name the words aloud as quickly and as accurately as possible. Each trial consisted of the following sequence of events: (a) A fixation consisting of three plus signs ("+++") appeared in the center of the computer screen for 400 ms, (b) the screen went blank for 200 ms, and (c) the word appeared at the position of the fixation and remained on the screen until 200 ms after the initial triggering of the voice key. After each naming response, the participant pressed a button on a mouse to go on to the next word. The participant was told to press the right button on the mouse if there was an error or an extraneous sound triggered the voice key and to press the left button on the mouse if everything appeared to have worked properly on that trial. Pressing the mouse button initiated a 1,200-ms intertrial interval.

Participants were given breaks after every 150 trials. Two buffer trials consisting of filler words not appearing in the training corpora were inserted at the beginning of each block of trials. In addition, at the beginning of each session, subjects were given 20 practice trials to familiarize them with the task. Each experimental session lasted approximately 60 min.

RESULTS

Response latencies for trials that participants marked as errors and response latencies faster than 200 ms or slower than 1,500 ms were excluded from all analyses, as were response latencies that fell more than 2.5 standard deviations beyond each subject's mean response latency. These criteria eliminated 4.8% of the observations. Mean latency was then computed for each item across subjects. The present analyses included only those words for which we had both error scores from the SM89 model and settling times for the PMSP model. This restriction eliminated 37 words, leaving a total of 2,833 words. In addition, the 12 heterophonic homographs (e.g., *wound*, *house*) were eliminated from all analyses, as was a word that inadvertently appeared twice in the list of stimuli, leaving 2,820 words.

In the analyses reported, we first examined the predictive power of factors such as log frequency in accounting for the variance in naming latency. By using variables such as log frequency, neighborhood density, and length in letters to account for variance in naming latency, we attempted to provide some baseline for evaluating the performance of the models. We then turned to the primary question of how much of the variance in human naming latencies is accounted for by each model.

Simple Predictors of Item Variance

A considerable amount of research has centered on the crucial role that frequency plays in naming performance (e.g., Huey, 1908/1968). Thus, the most obvious place to start looking at predictors of naming latency was frequency (or more accurately, log frequency). When log frequency (Kucera & Francis, 1967) was entered into a regression equation as the sole predictor of naming latency, it accounted for 7.3% of the variance in our data (variances were unadjusted; see Table 1 for detailed results). The two additional predictors of naming latency that we used were the length of the word in letters and a measure of the

Table 1. Regression analyses with mean item-naming latency as the criterion variable and each variable reported as the sole predictor

Statistic	Standard predictor variable		
	Log frequency	Length	Coltheart's <i>N</i>
Regression coefficient (β)	-0.2705	0.3793	-0.3576
$t(2818)$	14.92***	21.76***	20.33***
r^2	.0732	.1439	.1279

*** $p < .001$.

neighborhood density known as Coltheart's *N* (Coltheart, Davelaar, Jonasson, & Besner, 1977). Coltheart's *N* is defined as the number of words that can be formed by changing one letter in the targeted word. The results of regression analyses when each of these was entered as a sole predictor are reported in Table 1. When all three of these predictors were added simultaneously to the regression equation, they collectively accounted for 21.7% of the variance (see Table 2).² We now turn to the performance of the models.

PMSP Model

It is important to note that Plaut et al. (1996) presented four simulations, each with a slight variation in the underlying architecture. For present purposes, we used the settling times obtained from Simulation 3, which appears to be the central focus in the article. (We discuss Simulation 4 in the General Discussion.) The settling times from the PMSP model are intended to be directly analogous to naming latency. When the settling times from this model for all of the 2,820 items were entered into a linear regression equation as the sole predictor of naming latency, they accounted for approximately 3.3% of the variance (see Table 3). Note that this figure contrasts with 7.3% for simple log frequency and 21.7% for the three standard predictors. To see what the model captures beyond word frequency, we added the settling times after partialing out the effect of log frequency. The model's settling times then accounted for a reliable although attenuated 1.6% of the variance.

Thus, the PMSP model accounts for a reliable amount of the variance, although the actual percentage was clearly less than expected. Indeed, the settling times accounted for only about half as much variance as was attributable to simple log frequency. This result is surprising given that the degree to which weights were modified during training was a function of word frequency. Thus, frequency was directly involved in the training of the model.

2. There are a number of the words with very low frequency (i.e., 0 and 1 occurrences per million) that might have been quite unfamiliar to participants in this study, and it is possible that performance on these items does not reflect normal lexical processing. Therefore, we repeated analyses eliminating all words with frequencies of 0 or 1, a total of 618 words. The pattern of results was very similar for this reduced set of words.

Table 2. Multiple regression analysis with mean item-naming latency as the criterion variable

Statistic	Standard predictor variable		
	Log frequency	Length	Coltheart's <i>N</i>
Regression coefficient (β)	-0.2300	0.2491	-0.1723
$t(2816)$	13.72***	11.47***	7.92***
Partial r^2	.0627	.0447	.0218

$R^2 = .2169, F(3, 2816) = 262$

*** $p < .001$.

SM89 Model

We turn next to the SM89 model. It is important to note that, unlike the PMSP model, which provided settling times, the SM89 model had no direct analogue for the empirical naming latencies. Instead, Seidenberg and McClelland (1989) evaluated the phonological output of the model by computing a phonological error score, which was a measure of how close the model's output was to the desired output. They then used the phonological error score to map the model's performance onto human naming latencies. The assumption was that a low error score corresponded to a strong signal to the articulatory stages of processing, whereas a higher error score resulted in a weaker signal that would take longer to assemble into the appropriate articulatory response. Because Seidenberg and McClelland suggested that the error score obtained from the model is linearly related to naming latency, it is appropriate to ask how well the model's error scores do in accounting for variance in naming latencies for these items.

When the error scores were entered as the sole predictor of naming latency, the model accounted for 10.1% of the variance (see Table 3). As for the analysis of the PMSP model, the error scores were also entered into the analysis after log frequency to determine how much variance the model captured beyond simple log frequency. After we entered log frequency, error scores accounted for 5.2% of the variance. Thus, the SM89 model does quite a bit better than the PMSP model both as the sole predictor and above and beyond log frequency. In fact, the SM89 model captures slightly more variance than simple

Table 3. Regression analyses with mean item-naming latency as the criterion variable and output from each model entered as the sole predictor

Statistic	Model	
	SM89	PMSP
Regression coefficient (β)	0.3173	0.1824
$t(2818)$	17.76***	9.85***
r^2	.1007	.0333

*** $p < .001$.

Item Variance

log frequency. Of course, it is noteworthy that neither model approached the 21.7% of variance accounted for by the combined influence of log frequency, neighborhood density (Coltheart's N), and word length.³

Analyses With Onset-Phoneme Coding

The SM89 and the PMSP models are intended to account for those factors that affect the computation of orthography to phonology. These models are not intended to account for perceptual factors such as the effect of degradation on naming performance. Similarly, these models do not attempt to account for factors that affect the articulatory-motor component of naming performance. It is possible that at least a portion of the effects due to variables such as frequency may be at the level of articulation. If so, then some of the variance picked up by log frequency might be considered outside the models. One way of assessing the contribution of articulatory processes to naming performance is to use a delayed naming paradigm, in which the participant has sufficient time to recognize the word before pronouncing the word aloud (e.g., 2 s). To separate out the articulatory contribution to naming performance, one might partial out delayed naming performance. However, this approach is not without complications (see Balota & Chumbley, 1990; Monsell, Doyle, & Haggard, 1989, for a discussion). It is unlikely that delayed naming is a pure measure of the articulatory component of naming. Given the cascaded nature of processing (McClelland, 1979), factors that affect an early process can also affect processes downstream (Balota & Abrams, 1995). Hence, partialing out delayed naming performance may remove more variance than is actually appropriate. Related to this complication is a problem of multicollinearity. Normal speeded naming and delayed naming may be correlated sufficiently to make it difficult to test effects of a factor after partialing out delayed naming performance.⁴

The approach that we chose was intended to control for some of the variance due to the articulation component and also to control for differences between words in their ability to trigger the voice key used to measure naming latency. Our approach (modeled after Balota & Chumbley, 1985; also Treiman, Mullennix, Bijeljac-Babic, & Rich-

3. The dual-route cascaded (DRC) model of Coltheart, Curtis, Atkins, and Haller (1993) has been proposed to account for aspects of naming performance similar to those the two models we have reviewed attempt to account for. The DRC appears to be a hybrid of the first- and second-wave models. It uses two computational routes for computing phonology. One route applies learned grapheme-to-phoneme correspondence rules to the word; the other route contains localist representations of the visual and phonological features of the word and operates in a similar fashion to the interactive activation model of McClelland and Rumelhart (1981). The most complete published specification of the model describes in detail only the assembled route. Nonetheless, we examined the model's item-level predictions in the same manner as for the PMSP and SM89 models, with the DRC naming time added to the regression analysis as the sole predictor of naming latency. However, we had item-level predictions from the DRC model for only 1,803 words, so the analysis was done on these items only. For these items, the DRC model accounts for 6.9% of the total variance. In comparison, log frequency accounts for 10.3% of the variance. For this same set of items, the PMSP model accounts for 3.1% and the SM89 model accounts for 10.7% of the variance.

4. We should note that Besner (in press) took the approach of partialing out delayed naming performance and obtained results generally consistent with those presented here, although the analysis was on only 300 words, compared with the 2,820 words in the present analysis.

Table 4. Summary of hierarchical regression analysis with onset-phoneme variables entered in the first step and other predictors entered in separate second steps

Variable	β weight	t value	R^2
Step 1			
Affricative	-0.2843	2.52*	.2986
Alveolar	1.0614	3.77**	
Bilabial	0.9068	3.60**	
Dental	0.3381	3.95**	
Fricative	-0.5996	2.11*	
Glottal	0.2096	1.65	
Labiodental	0.5899	3.70**	
Liquid	-0.7474	3.49**	
Nasal	-0.6254	4.12**	
Palatal	0.5106	2.64**	
Stop	-0.9460	3.23**	
Velar	0.9526	4.68**	
Voiced	0.0918	4.08**	
Step 2			
SM89	0.2795	18.48**	.0760
Step 2			
PMSP	0.1621	10.37**	.0259
Step 2			
Log frequency	-0.2704	17.82**	.0712
Step 2			
Log frequency	-0.2431	16.74**	.1326
Coltheart's N	-0.1411	7.22**	
Length	0.1492	7.43**	

* $p < .05$. ** $p < .01$.

mond-Welty, 1995) was to code the onset of each word. We used dummy variable coding for place of articulation, manner of articulation, and voicing. Each word was coded by a 1 or 0 representing the presence or absence, respectively, of each feature. For words with vowel onsets, we coded 0 for both manner and place of articulation and 1 for voicing. In these analyses, the phonemic features were entered into the regression analysis first. The results of this analysis are reported in Table 4. The phonemic features collectively accounted for 29.9% of the variance. The first analysis examined the variance accounted for by the error scores from the SM89 model. The error scores accounted for 7.6% of additional variance. An identical analysis using settling times from the PMSP model showed that settling times accounted for 2.6% of additional variance. A separate analysis using log frequency, length, and Coltheart's N revealed that these three predictors captured an additional 13.3% of the variance after partialing out variance due to the onset phoneme. The total amount of variance accounted for by using onset-phoneme features, log frequency, length, and Coltheart's N was 43.1%. These results clearly indicate that a considerable amount of systematic variance in our naming data can be accounted for by these variables.⁵

5. These results have been replicated on 1,856 words from this same set with another group of 20 subjects. The results showed the same pattern as that reported here for the standard predictors, onset coding, and model performance. Also, the results reported for the SM89 model appear to be consistent with analyses reported by Seidenberg and McClelland (1990) on the Seidenberg and Waters (1989) word-naming corpus.

GENERAL DISCUSSION

The present analyses have revealed two interesting aspects of the second-wave computational models of naming performance. First, it appears that the close correspondence between the performance of the models and the empirical data at the factor level does not hold at the level of individual items. Second, although the PMSP model has clearly addressed some problems with the SM89 model (see Plaut et al., 1996, for discussion), this may have been accomplished at the cost of its ability to reflect performance at the item level.

There were at least two problems with the original SM89 model that resulted in the modifications contained in the implementation of the PMSP model. An early criticism of the SM89 model focused on its difficulty in naming pronounceable nonwords (Besner et al., 1990; but see also Seidenberg & McClelland, 1990; Seidenberg, Petersen, MacDonald, & Plaut, 1996) and the need to map the model's error scores onto response latency. These difficulties were addressed in the PMSP model, and this model might be viewed as the successor to the SM89 model. Unfortunately, the PMSP model, while alleviating some of the problems with the SM89 model, appears to have traded off some of its ability to fit the data at the item level.

There are at least two differences between these models that might have resulted in this decrease in performance at the item level. First, when the PMSP model settles into a stable pattern representing its phonological output, the stable pattern is not a perfect representation of the correct phonological output but rather an approximation. Thus, the model settles into a representation that is close to but still some distance from the correct representation. However, only the time that the model takes to settle into some response is counted as its naming latency, as long as the final response is within some criterion distance of the correct output. Thus, the distance to the final representation does not matter in measuring response latency. In contrast, the distance of the SM89 model's output to the correct output was completely contained in the error score, perhaps providing a more accurate account for individual item performance.⁶

The second difference between the two models is that they were trained using two variants on the back-propagation algorithm, one for a feed-forward and one for a recurrent network. The weights in the SM89 model were modified to minimize the sum of squared error between the obtained and the target output. The PMSP model used a modification of the back-propagation algorithm that attempted to minimize cross entropy rather than the sum of squared error. Although it is possible that this difference in training might contribute to differences in the models' performance, it is unclear exactly what impact it might have in accounting for variance at the item level.

One factor contributing to the PMSP and SM89 models' difficulty accounting for item-level variability may have been the requirement that the models provide "correct" pronunciations for virtually all of the words in the training corpus. As noted by Plaut et al. (1996), this requirement entailed some compression of the frequency effect because training on the items with very low frequency had to be sufficient to produce correct pronunciations. It would be interesting to know if loosening this restriction on correct performance might increase the models' ability to capture item-level performance. In other words, there may be a nonmonotonic function relating the training and the models' ability to capture item performance. Of course, reducing

the degree of training will increase the error rate. The unexplored question is whether the current level of training in these models represents the optimal trade-off between accuracy and variance accounted for at the item level.

One of the most intriguing aspects of both the SM89 model and the PMSP (Simulation 3) model is that they do not have implemented lexical (whole word form) or semantic (meaning level) representations contributing to speeded naming performance. Thus, the focus of these models has been to capture the systematicity in mapping spelling patterns to speech patterns in English orthography. The surprising and exciting result of this research endeavor is that both of these models have achieved considerable success at capturing extant factor-level results in the word-naming literature without implementing a lexicon or meaning-level network. However, the present analyses indicate that although the models are clearly successful at the factor level, they do have some difficulty accounting for item-level effects. Possibly, this item-level difficulty should be taken as a signal for the importance of implementing a second, lexical-semantic processing route that may contribute to naming performance (Balota, Ferraro, & Connor, 1991; Besner, 1990).

Consider, for example, the role of word frequency. The present results suggest that both models appear to underestimate the rather large influence of simple word frequency in accounting for naming latencies, most likely because of the compression of frequency that was necessary to have the models reach acceptable levels of accuracy. If one considers the possibility that the frequency of experiencing a word might modulate the speed of accessing an additional processing route at the lexical-semantic level, then the strong influence of frequency shown in the results we report here may suggest that an additional route is necessary to fully capture speeded naming performance. In this light, it is important to note that Plaut et al. (1996) found it necessary to implement a type of frequency-sensitive lexical-semantic influence in their Simulation 4 to account for the performance of surface dyslexics. Plaut et al. acknowledged that this implemented pathway could only be viewed as an approximation of a lexical-semantic pathway, because its function was simply to provide an additional "correct" input to the phonological units. Although it appears that further work is needed to characterize the contribution of a lexical-semantic pathway, we believe that the consideration of additional pathways that may contribute to lexical processing is an important step in this modeling endeavor.

A final question concerns what empirical constraints researchers should use for evaluating models. The factorial studies have taken the field quite far in understanding basic processes in word recognition. This research forms the primary empirical data set for developing models of word recognition. However, we believe that with the advent of second-wave computational models, item-level regression techniques can provide an important complement to the traditional factorial-experiment approach in model evaluation. Obviously, the ultimate goal is to elucidate the factors that influence language processing and to understand the mechanisms underlying these effects. Factorial experiments are crucial in this work because they drive the model building and aid in the communication between researchers about how particular factors affect performance. At the same time, knowledge about the underlying nature of these influences on naming performance, as revealed in regression analyses, is necessary if researchers are to avoid losing sight of the large influences (e.g., word frequency) while pursuing accounts for the smaller ones. In fact, we would argue

6. We thank Gary Dell for pointing this out.

Item Variance

that these models have been, at least in part, motivated by the desire to capture the more complex aspects of the computation of phonology from orthography, even under conditions in which such complex aspects may account for a relatively small amount of variance. Thus, these aspects of naming performance may have been overemphasized at the cost of the larger influences on naming performance, such as simple word frequency and length. Perhaps future implementations can move toward redressing the balance in emphasis. At the very least, we hope that item-level analyses and traditional factorial approaches are both used in future model construction.

Acknowledgments—We gratefully acknowledge Mark Seidenberg, David Plaut, and Max Coltheart for providing their respective models' item-level predictions that were used in this article. In addition, we thank Derek Besner, Max Coltheart, Gary Dell, Ken Forster, David Plaut, and Mark Seidenberg for helpful comments on an earlier version of this manuscript. This work represents a totally collaborative effort and was supported by National Institute on Aging Grant PO1 AG03991 and Training Grant AG00030.

REFERENCES

- Balota, D.A., & Abrams, R.A. (1995). Mental chronometry: Beyond onset latencies in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1289–1302.
- Balota, D.A., & Chumbley, J.I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory and Language*, 24, 89–106.
- Balota, D.A., & Chumbley, J.I. (1990). Where are the effects of frequency in visual word recognition tasks? Right where we said they were! Comments on Monsell, Doyle, and Haggard (1989). *Journal of Experimental Psychology: General*, 109, 231–237.
- Balota, D.A., Ferraro, F.R., & Connor, L.T. (1991). On the early influence of meaning in word recognition: A review of the literature. In I.P. Schwanenflugel (Ed.), *The psychology of word meaning* (pp. 187–221). Hillsdale, NJ: Erlbaum.
- Besner, D. (1990). Does the reading system need a lexicon? In D.A. Balota, G.B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 73–99). Hillsdale, NJ: Erlbaum.
- Besner, D. (in press). Basic processes in reading: Multiple routines in localist and connectionist models. In P.A. McMullen & R.M. Klein (Eds.), *Converging methods for understanding reading and dyslexia*. Cambridge, MA: MIT Press.
- Besner, D., Twilley, L., McCann, R., & Seergobin, K. (1990). On the connection between connectionism and data: Are a few words necessary? *Psychological Review*, 97, 432–446.
- Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, 100, 589–608.
- Coltheart, M., Davelaar, E., Jonasson, J.T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance* (Vol. 6, pp. 535–555). Hillsdale, NJ: Erlbaum.
- Forster, K.I. (1976). Accessing the mental lexicon. In E. Walker & R.J. Wales (Eds.), *New approaches to language mechanisms* (pp. 257–287). Amsterdam: North-Holland.
- Huey, E.B. (1968). *The psychology and pedagogy of reading*. Cambridge, MA: MIT Press. (Original work published 1908)
- Kucera, H., & Francis, W.N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- McClelland, J.L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, 86, 287–330.
- McClelland, J.L., & Rumelhart, D.E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, 88, 375–407.
- Monsell, S., Doyle, M.C., & Haggard, P.N. (1989). The effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General*, 118, 43–71.
- Morton, J. (1969). The interaction of information in word recognition. *Psychological Review*, 76, 165–178.
- Plaut, D.C., McClelland, J.L., Seidenberg, M.S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103, 56–115.
- Seidenberg, M.S., & McClelland, J.L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523–568.
- Seidenberg, M.S., & McClelland, J.L. (1990). More words but still no lexicon: Reply to Besner et al. (1990). *Psychological Review*, 97, 447–452.
- Seidenberg, M.S., Petersen, A., MacDonald, M.C., & Plaut, D.C. (1996). Pseudohomophone effects and models of word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 48–62.
- Seidenberg, M.S., & Waters, G.S. (1989). Word recognition and naming: A mega study [Abstract 30]. *Bulletin of the Psychonomic Society*, 27, 489.
- Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E.D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General*, 124, 107–136.

(RECEIVED 8/2/96; ACCEPTED 11/6/96)