Available online at www.sciencedirect.com
ScienceDirect
Brain and Language 100 (2007) 95–100
Brain and Language
www.elsevier.com/locate/b&l

Commentary

Evaluating computational models in cognitive neuropsychology: The case from the consonant/vowel distinction

Mark Knobel, Alfonso Caramazza *

Department of Psychology, Cognitive Neuropsychology Laboratory, Harvard University, Cambridge, MA 02138, USA

Accepted 24 June 2006
Available online 1 August 2006

Abstract

Caramazza et al. [Caramazza, A., Chialant, D., Capasso, R., & Miceli, G. (2000). Separable processing of consonants and vowels. *Nature*, *403*(6768), 428–430.] report two patients who exhibit a double dissociation between consonants and vowels in speech production. The patterning of this double dissociation cannot be explained by appealing to sub-phonemic distinctions, such as sonority level or damage to specific phonological features. They argue that consonant/vowel status is an autonomous level of representation. Monaghan and Shillcock [Monaghan, P., & Shillcock, R. (2003). Connectionist modelling of the separable processing of consonants and vowels. *Brain and Language*, *86*(1), 83–98.] present computational models which supposedly exhibit a similar double dissociation. They contend that these models can explain the patient data, without appeal to such supra-phonemic distinctions as consonant/vowel status. Here we argue that their claim fails to meet two necessary criteria. Their models do not fit the pattern of the patient data, either quantitatively or qualitatively. Furthermore, the motivation for these models is unclear beyond just being an attempt to explain this specific phenomenon. We conclude that these models, in their current form, do not provide an alternative explanation to the representation of consonants and vowels.

Keywords: Aphasia; Computational modeling; Dissociations; Phonology

1. Introduction

Do consonants and vowels have a categorical reality in the mental representations used for speech, or are they only convenient labels for some underlying process that does not make such a distinction? The patterns of deficits in producing consonants and vowels has been used to support the idea that there is indeed a psychological reality to these labels (Caramazza, Chialant, Capasso, & Miceli, 2000). However, a recent computational modeling study by Monaghan and Shillcock (2003) questions this conclusion. The disagreement is not just over consonants and vowels, but cuts deeply into the question of how we evaluate arguments for or against representations in general. An evaluation of the hypotheses must be made on two levels: (1) how well do the hypotheses capture the data, and (2) how well is each hypothesis motivated. We intend to show that Monaghan and Shillcock's models do not support their conclusions and cannot be used as an alternate explanation for Caramazza et al.'s results.

Caramazza et al. present the cases of patients AS and IFA. In a word repetition task, AS made nearly three times as many errors on vowels as on consonants, whereas IFA made nearly five times as many errors on consonants as on vowels. These complementary patterns of performance disallowed an explanation based on either consonants or vowels being more difficult to produce. Though this double dissociation suggested that consonant and vowel status was represented,[1] Caramazza et al. recognized that there was another possibility. These patients may have been affected

---

[1] Though we and others use many terms, such as consonants and vowels "having a categorical reality", "having autonomous status", and being "categorically distinct objects", these refer to this same idea of there being a representation of consonant/vowel status somewhere in the speech system.

* Corresponding author. Fax: +1 617 496 6262.
E-mail address: caram@wjh.harvard.edu (A. Caramazza).

0093-934X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.bandl.2006.06.008

by a deficit in processing lower-level phonological information such as sonority or specific phonological features.

All phonemes vary along a spectrum of sonority, with vowels being more sonorous than consonants. It is possible that this underlying sonority spectrum could have produced the double dissociation observed. Since sonority is a continuously varying property, this hypothesis predicts that a patient who has trouble with vowels (such as AS) should perform worse on the more sonorous consonants compared to the less sonorous consonants, and vice versa for a patient with consonant impairment (such as IFA). Caramazza et al. tested this possibility and found no correlation, in either patient, between a consonant's sonority rating and error rate. We must put special emphasis on this lack of a sonority effect in the patient data, as this will become important later on in the discussion.

Another possible explanation of the double dissociation is damage to specific phonological features, namely either those that distinguish between vowels (in the case of AS) or those that distinguish between consonants (in the case of IFA). This hypothesis predicts that a patient who has difficulties with vowels would also have difficulty with consonants that only vary on those features that are used to distinguish vowels (since those would be the ones that are damaged). Caramazza et al. looked specifically at the /l/ and /r/ phonemes, which only vary on features that are important in distinguishing vowels. An account based on phonological features would predict that these consonants should be selectively more damaged than other consonants in patient AS. They found that there was no difference between the error rates on these consonants and the error rates for the other consonants, in *either* patient.

These latter data are also a direct test between the consonant/vowel hypothesis and the sonority hypothesis. The consonants /r/ and /l/ are closer to vowels with regard to sonority than to the majority of the other consonants. If sonority is truly the underlying variable then we would expect /r/ and /l/ to behave more like vowels and be spared for IFA while damaged for AS. One the other hand, if a true consonant/vowel distinction underlies these phenomena, then these consonants should behave like all the other consonants and be damaged for IFA, and spared for AS. The data patterned just like the consonant/vowel representation account would predict.

In arguing for the autonomous status of consonants and vowels, Caramazza et al. were careful to rule out likely alternative hypotheses. Having shown that the two subphonemic explanations above were incompatible with the data, they concluded for the categorical representation of consonant and vowel status. They wrote: "the fact that error performance depends neither on the sonority value of individual phonemes nor on their feature properties suggest[s] that consonants and vowels are categorically distinct objects at some level of representation even though they are not categorically distinguishable at the phonetic level" (p. 430). Their claim is specific to content, but indefinite to location. The consonant or vowel status of a phoneme must be represented somewhere in the speech system. However, they do not specify, nor is it possible based on the empirical data, where in the system this representation is stored. A probable location for this representation is at the level of phonological processing in the perception and/or production systems.

The cases reported by Caramazza and colleagues are not unique. Several studies have produced similar evidence of disproportionate damage to both consonants (e.g., Beland, Caplan, & Nespoulous, 1990; Boatman, Hall, Goldstein, Lesser, & Gordon, 1997) and vowels (e.g., Romani, Grana, & Semenza, 1996). Furthermore, a close parallel to this work can also be found in the domain of spelling, where much evidence has been presented in support of the representation of *graphemic* consonants and vowels. One type of evidence that has been used is the pattern of consonants substituting for consonants and vowels for vowels much more often than chance in graphemic buffer patients (e.g., Caramazza & Miceli, 1990; Jonsdottir, Shallice, & Wise, 1996; McCloskey, Badecker, Goodman-Schulman, & Aliminosa, 1994). The other, more striking evidence, also from graphemic buffer patients, has been the relative impairment of vowels (Cotelli, Abutalebi, Zorzi, & Cappa, 2003; Cubelli, 1991) or consonants (Kay & Hanley, 1994; Miceli, Capasso, Benvegnu, & Caramazza, 2004).

Monaghan and Shillcock (2003) present a challenge to Caramazza et al.'s (2000) interpretation of their patient data. Monaghan and Shillcock reject the idea that there is a categorical representation of consonants and vowels anywhere in the speech system. Instead, they claim that a connectionist model can account for the observed double dissociations without postulating a distinction between consonants and vowels. They write: "Distinguishing a higher, phonemic level above the level of phonological features may be descriptively useful and convenient, but we will show in terms of cognitive processing that modularization at the lower level of representation can result in distinctions at the higher level without requiring that higher level as an explanatory concept" (p. 2).

The idea behind their modeling is that if there are two resources used in processing phonemes (i.e., both consonants and vowels), and processing in one of these resources was more important to vowels than to consonants and vice versa in the other, then it would be possible to see a double dissociation between vowels and consonants by damaging each of these resources in turn. However, the same logic needs to be applied to the models as was applied to the patients. Specifically, if the system functions without representing the consonant/vowel status of phonemes, then it should display a signature of the underlying distinction that allows the resources to differentially specialize, such as an effect of sonority or of another phonological feature.

2. The models

In the first simulation, Monaghan and Shillcock implement a three-layer (input, output, and hidden) connectionist model to clean up "noisy" phonemes. The input to the model is an 11-dimensional vector representing the phonological features of a phoneme. Each of the feature values in this vector has random noise added to it, and the model's job is to "clean-up" the input. The hidden layer is split into two parts, each of which consists of six units and receives slightly different input from all cells in the input layer.[2] The output of these two parts then combines into one output layer. The final output of the model is also a phoneme represented by an 11-dimensional vector of phonological features.

To simulate brain damage, Monaghan and Shillcock lesion their model by severing, in turn, the connections between each of the hidden layer parts and the output layer. What they find is indeed a double dissociation between consonants and vowels. When one of the hidden layer parts is lesioned (henceforth the vowel layer), the performance on vowels is worse than on consonants (31% correct for vowels vs. 48% correct for consonants), whereas when the other hidden layer part (henceforth the consonant layer[3]) is lesioned the performance on consonants is worse than vowels (66% correct for vowels vs. 35% correct for consonants).

In their second simulation, Monaghan and Shillcock's model is similar in structure to their first one. This time, however, the two parts of the hidden layer receive coarse and fine input. The coarse layer receives as input the current phoneme and its two temporal neighbors, whereas the fine-layer only receives the current phoneme (as was the case for the hidden layer in the first model). The rest of the model remains the same as in the first simulation. Monaghan and Shillcock base this separation of processing in their model on coarse and fine grained processing in the visual system and from possibly similar processes in the auditory system.

Unfortunately, Monaghan and Shillcock's second model fails to deliver the proper dissociation between consonants and vowels. With a 100% lesion to either the coarse- or fine-layer (i.e., cutting all the connections from that layer), the performance on consonants and vowels does not vary between coarse and fine layers (65.63% vs. 66.91% for vowels, and 56.25% vs. 54.01% for consonants). However, with a 50% lesion, vowels remain mostly unaffected, but now there is worse performance on consonants when the fine-layer is lesioned, but this is still not a double dissociation. Monaghan and Shillcock claim that using a more stringent criterion (mean square error) one can see that vowels are more damaged with a lesion to the coarse layer (1.53 *MSE* vs. 1.31 *MSE*). This may be the case, but using this measure does not seem to produce much difference in performance on consonants between the course and fine layers (1.99 *MSE* vs. 1.90 *MSE*), and thus the second model still fails to give us a double dissociation.

3. Do the models capture the data?

It is intuitively clear that goodness of fit must be evaluated in claims for or against a representation. But how does one decide whether a computational model is any good at modeling the data? The model must fit the relevant data, with relevance being dictated by those aspects of the data that actually distinguish between theories. In this case, goodness of fit applies not only to qualitative aspects of the data, but also to the quantitative aspects.

The qualitative aspects of the data are both the double dissociation *and* the lack of sonority or feature effects. It is not enough for a model to show a double dissociation. It must also show no effect of sonority. If sonority effects occur along with a double dissociation then the model is simply an instantiation of the sonority hypothesis that Caramazza et al. tested and ruled out.

The quantitative aspect of the data that must be explained is the magnitude of the double dissociation observed between patients AS and IFA. Alternative explanations of double dissociations based on connectionist architectures have been around for some time (e.g., Plaut, 1995; Plaut & Shallice, 1993). Showing that a model can produce a double dissociation qualitatively is not sufficient in this case. A theory that predicts that consonants and vowels can be damaged independently predicts arbitrarily large dissociations between damage to vowels and damage to consonants. A model of these phenomena must also be able to account for this quantitative aspect of the data (see Ruml & Caramazza, 2000 for a more in depth discussion of the importance of quantitative modeling).

Monaghan and Shillcock's two models do not fare well with regard to both the qualitative and quantitative elements of goodness of fit. From the previous section, we saw that the second model was already having difficulty reproducing a double dissociation. On the other hand, the first model has no problem producing the double dissociation. Qualitatively, at least the first model captures this aspect of the patient data. Indeed, Monaghan and Shillcock use this fact to argue against Caramazza et al.'s conclusion that there is a categorical distinction between consonants and vowels. However, their argument seems to ignore the fact that both of their models display a strong sonority effect, a key aspect that clearly distinguishes between theories.

Caramazza et al. stress that there is no sonority effect in the patient data and this finding is crucial to their conclusion. Looking closer at the first model we see that when the vowel layer is lesioned, not only is there worse performance on vowels, but there is also worse performance on the more sonorous consonants (i.e., the nasals, laterals, and approximants; Table 2, p. 90) relative to the less sonorous ones. Conversely, when the consonant layer is damaged, there is likewise a worse performance on the less sonorous consonants (i.e., the plosives, fricatives, and affricates). The model shows a clear sonority effect and is at odds with the patient data.

---

[2] The input vector has different random noise added to it depending on which of the two hidden layers it is bound for.

[3] Monaghan and Shillcock refer to these parts of the hidden layer as the "vowel" and "non-vowel" layers.

The second model fares no better. Even if we grant that it can produce a double dissociation, it still shows an effect of sonority like the first model. Performance on nasals, laterals, and approximants (i.e., the more sonorous consonant groups) is not affected with a 50% lesion to either hidden layer, which matches a similar lack of damage to vowels (Table 4, p. 94). However, when the fine-layer is damaged, it leads to worse performance not only on consonants, but specifically on the plosives, fricatives, and affricates (i.e., the less sonorous consonant groups). Thus this second model is also inconsistent with the patient data.

To be fair, in the case of the first model, Monaghan and Shillcock also report that they get a similar double dissociation from models whose input lacked the sonority and/or consonantal features of the original input. It may be argued that these modified models would not display a sonority effect. However, Monaghan and Shillcock do not provide a breakdown by consonant types for these modified models so that this type of analysis is impossible to perform. Nevertheless, we would expect these models to still show an effect of sonority, since the voice and nasal features that remain are sufficient to distinguish between the more and less sonorant consonants.

We have shown how these two models do not qualitatively reproduce important characteristics of the data. However, it is also as important for models to quantitatively capture the data. To quantify how well the models can accommodate the size of the dissociations produced by the patients, we compare AS and IFA to Monaghan and Shillcock's first model by analyzing a hypothetical run that has an equivalent magnitude of difference. Given the 10 published runs of the model,[4] what is the chance that the model will produce a run that has as great a magnitude of difference between vowels and consonants as observed in patients AS and IFA?[5] Hypothetical runs that produce a dissociation as large as shown by the patients are *at least* 3.1 and 3.6 standard deviations away from the mean of the runs for the first model (for AS and IFA respectively). These *minimized* values are clearly out of the range of the standard model runs.

It is understandable that Monaghan and Shillcock do not consider their models as an explanation of the patient's deficits,[6] so is it fair to criticize them on aspects of the data that they never intended to model? In some cases such a criticism would not be warranted, since simple extensions to the model would be able to better fit the data. However, in this case, it is not obvious how it is possible to bridge the gap between the models and the patient data, since the results are based on maximally lesioning the models. Given the theoretical importance of this quantitative aspect of the data (see above), the burden lies with Monaghan and Shillcock to explain how their models can be modified to account for such a large dissociation in the patient data.

4. Motivation for the modeling

We have provided a critical look at the goodness of fit of Monaghan and Shillcock's models. It is equally important, if not more so, to evaluate the motivation for these models. There is always the possibility that these models can be modified to overcome our criticisms, so then the only point of comparison between the two competing hypotheses would be at the level of their motivation. Monaghan and Shillcock's work follows a strong tradition of questioning the psychological reality of representations through connectionist modeling (e.g., Farah & McClelland, 1991; Plaut, McClelland, Seidenberg, & Patterson, 1996). We are not questioning this underlying framework, only pointing out motivational deficiencies for specific choices in the modeling not covered by the broad connectionist framework.

In motivating the splitting of resources in their models, Monaghan and Shillcock focus on neural plausibility. Their first proposal is that such a split occurs due to both hemispheres being involved. In fact, their second model instantiates the coarse and fine processing styles of the right and left hemispheres, respectively. However, they also suggest the possibility that the division of labor that they are modeling may instead be intra-hemispheric, especially for the first model, but they do not elaborate on this beyond giving the example of magno- and parvocellular processing. While this may be enough to establish the possibility for such a processing split, it does not demonstrate why such a split is likely. More importantly, Monaghan and Shillcock do not explore whether a split in processing resources is even necessary for their task. Although there is some basis for a neural split in resources, a stronger case could be made by demonstrating on a functional level why two groups of resources are better than one (or three or more). It would be fairly straightforward to see if a model with two hidden layers of 6 units each cleans up noisy phonemes better than a model with a single hidden layer of 12 units.

In attempting to provide an existence proof, Monaghan and Shillcock question the need for the representation of consonant/vowel status, but they spend little time arguing for why their theory is superior. It remains unclear why the assumption of a division of neural resources, whether intra- or inter-hemispheric, is preferable to the assumption of a categorical distinction between consonants and vowels. One does not seem to be *a priori* more parsimonious than the other. The two must be empirically compared. One possible way of showing the advantage of a division of neural resources is to compare the current models to a model that actually assumes a categorical distinction between consonants and vowels (see Mirkovic, MacDonald, & Seidenberg, 2005 for a similar comparison in the domain of lexical gender). A simple model could be made where one bank of hidden units is only trained on consonants, while the other

---

[4] The means and standard deviations are reported for 10 runs of the model in Table 2, p. 90.

[5] 2.89 times more vowel errors as consonant errors for AS and 5.32 more vowel errors than consonant errors for IFA.

[6] Caramazza et al. also do not provide an explanation of their patients' deficits.

is only trained on vowels. The lesioning of this type of model would serve as a baseline for comparing models with and without the representation assumption. If Monaghan and Shillcock's models were to perform similarly or better than this type of baseline model, then there would be room to argue not only that the representation is unnecessary, but more importantly, that the representation provides no advantage.

In tying their work to the literature, Monaghan and Shillcock suggest that their modeling work can be the basis for modeling hemispheric processing. However, their hedging on the neural basis for the split in resources in their models, undermines this possibility (see above). No other attempts are made to tie their work into current models of phonological processing, either for perception or production. Without such broader connections, their work risks being construed as a narrow attempt aimed only at explaining Caramazza et al.'s specific patient data.

5. Conclusion

The current debate in phonology mirrors a similar debate in orthography. We can gain valuable insight from the latter debate, however, since it was started much earlier. Caramazza and Miceli (1990) originally argued for the representation of consonants and vowels using substitution data from their graphemic buffer patient (see Section 1 for later evidence). As an alternative, Houghton and colleagues (Houghton, Glasspool, & Shallice, 1994; Shallice, Glasspool, & Houghton, 1995) presented a model of the graphemic buffer that lacked an explicit representation of consonants and vowels. Their model was successful at modeling certain aspects of the data, such as the distribution of error types and relative frequency of errors at different positions in the word. However, the model could not capture the fact that vowels substituted for vowels and consonants for consonants more often than chance.[7] Further evidence of dissociations between consonants and vowels (e.g., Cotelli et al., 2003; Cubelli, 1991; Kay & Hanley, 1994; Miceli et al., 2004) also could not be accommodated by the model.

A model of the graphemic buffer must be able to account for these data from graphemic buffer patients. To resolve this problem with their model, Houghton and colleagues have recently had to introduce an explicit consonant/vowel distinction to their model (Glasspool & Houghton, 2005). In the case of phonology, we would suggest that a similar solution may be necessary to explain Caramazza et al.'s patient data.

Though we have spent the greater part of this paper arguing against Monaghan and Shillcock's conclusions, we wish to stress that we agree with them in the value of their approach. It is necessary to question claims about representation and to propose alternatives. In this case, however, the models presented by Monaghan and Shillcock do not meet necessary criteria for doing so and do not even constitute an existence proof. In their current form, their models are not a compelling challenge to Caramazza et al.'s conclusion that consonants and vowels are represented categorically at some level in the speech production system. In fact, the failure of these models to capture relevant aspects of the data could be taken as *prima facie* evidence for the need for such a representation to explain the patient data.

References

Beland, R., Caplan, D., & Nespoulous, J.-L. (1990). The role of abstract phonological representations in word production: evidence from phonemic paraphasias. *Journal of Neurolinguistics, 5*(2–3), 125.

Boatman, D., Hall, C., Goldstein, M. H., Lesser, R., & Gordon, B. (1997). Neuroperceptual differences in consonant and vowel discrimination: as revealed by direct cortical electrical interference. *Cortex, 33*(1), 83–98.

Caramazza, A., Chialant, D., Capasso, R., & Miceli, G. (2000). Separable processing of consonants and vowels. *Nature, 403*(6768), 428–430.

Caramazza, A., & Miceli, G. (1990). The structure of graphemic representations. *Cognition, 37*(3), 243–297.

Cotelli, M., Abutalebi, J., Zorzi, M., & Cappa, S. F. (2003). Vowels in the buffer: a case study of acquired dysgraphia with selective vowel substitutions. *Cognitive Neuropsychology, 20*(2), 99–114.

Cubelli, R. (1991). A selective deficit for writing vowels in acquired dysgraphia. *Nature, 353*(6341), 258–260.

Farah, M. J., & McClelland, J. L. (1991). A computational model of semantic memory impairment: modality specificity and emergent category specificity. *Journal of Experimental Psychology: General, 120*(4), 339.

Glasspool, D. W., & Houghton, G. (2005). Serial order and consonant/vowel structure in a graphemic output buffer model. *Brain and Language, 94*(3), 304–330.

Houghton, G., Glasspool, D. W., & Shallice, T. (1994). Spelling and serial recall: insights from a competitive queuing model. In G. D. A. Brown & N. C. Ellis (Eds.), *Handbook of spelling: Theory, process and intervention*. Chichester: John Wiley and Sons.

Jonsdottir, M. K., Shallice, T., & Wise, R. (1996). Phonological mediation and the graphemic buffer disorder in spelling: cross-language differences? *Cognition, 59*(2), 169–197.

Kay, J., & Hanley, R. (1994). Peripheral disorder of spelling: the role of the graphemic buffer. In G. D. A. Brown & N. C. Ellis (Eds.), *Handbook of spelling: Theory, process and intervention*. Chichester: John Wiley and Sons.

McCloskey, M., Badecker, W., Goodman-Schulman, R. A., & Aliminosa, D. (1994). The structure of graphemic representations in spelling: evidence from a case of acquired dysgraphia. *Cognitive Neuropsychology, 11*(3), 341–392.

Miceli, G., Capasso, R., Benvegnu, B., & Caramazza, A. (2004). The categorical distinction of vowel and consonant representations: evidence from dysgraphia. *Neurocase, 10*(2), 109–121.

Mirkovic, J., MacDonald, M. C., & Seidenberg, M. S. (2005). Where does gender come from? Evidence from a complex inflectional system. *Language and Cognitive Processes, 20*(1), 139–167.

Monaghan, P., & Shillcock, R. (2003). Connectionist modelling of the separable processing of consonants and vowels. *Brain and Language, 86*(1), 83–98.

Plaut, D. C. (1995). Double dissociation without modularity: evidence from connectionist neuropsychology. *Journal of Clinical and Experimental Neuropsychology, 17*(2), 291.

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*(1), 56.

---

[7] The model was also not able to reproduce CV-structure effects on error rates, such as the fact that there were significantly more errors on the second consonant relative to the first in a CC consonant cluster.

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: a case study of connectionist neuropsychology. *Cognitive Neuropsychology, 10*(5), 377.

Romani, C., Grana, A., & Semenza, C. (1996). More errors on vowels than consonants: an unusual case of conduction aphasia. *Brain and Language, 55*.

Ruml, W., & Caramazza, A. (2000). An evaluation of a computational model of lexical access: comment on Dell et al. (1997). *Psychological Review, 107*(3), 609–634.

Shallice, T., Glasspool, D. W., & Houghton, G. (1995). Can neuropsychological evidence inform connectionist modelling? Analyses of spelling. *Language and Cognitive Processes, 10*(3), 195–225.