Systematicity and Specialization in Semantics: A Computational Account of Optic Aphasia

Sean McGuire (sean@roughy.psy.cmu.edu)
Department of Psychology
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213–3890

David C. Plaut (plaut@cmu.edu)
Center for the Neural Basis of Cognition
Mellon Institute 115
4400 Forbes Avenue
Pittsburgh, PA 15213–2683

Abstract

Optic aphasic patients are selectively impaired at naming visually presented objects but demonstrate relative intact comprehension of those objects (e.g., by gesturing or categorization) and are able to name them when presented in other modalities (e.g., via tactile input). This and other modality-specific naming deficits have been taken as evidence that semantics is organized into distinct modality-specific subsystems. We adopt an alternative view in which semantics is a set of learned, internal representations within a parallel distributed processing system that maps between multiple input and output modalities. We account for the critical aspects of optic aphasia in terms of the effects of damage to such a system, despite its lack of modality-specific specialization. We show that the robustness of a task in such a system depends critically on its systematicity, and that modality-specific naming deficits can arise because naming is an unsystematic task.

Introduction

The lexical semantic system can be thought of as a set of representations which mediates between multiple input and output modalities. Perhaps the most immediately intuitive model of semantics is what has been termed the *unitary semantics* model (e.g., Caramazza, Hillis, Rapp, & Romani, 1990; Hillis, Rapp, Romani, & Caramazza, 1990). A generic version of this model is shown in Figure 1a. In such a model, semantics takes input from any of several different modalities, and generates output in one or more other modalities.

Shallice (1987) claimed that certain aspects of neuropsychological data pose a serious challenge to unitary semantics approaches. One problem comes from modality-specific naming deficits, such as optic aphasia. Optic aphasia is a relatively rare neuropsychological disorder, typically caused by damage to the left medial occipital lobe (i.e., visual cortex and the underlying white matter), in which patients exhibit a selective impairment in naming visually presented objects (see Iorio, Falanga, Fragassi, & Grossi, 1992; Endo, Makishita, & Sugishita, 1996, for reviews of cases). For example, patient JB (Riddoch & Humphreys, 1987) was substantially impaired at naming visually presented objects, providing correct answers on only 46% of test trials. However, he was 75% correct on miming the uses of visually presented objects, suggesting that his naming deficit could not be explained in terms of a more fundamental visual recognition impairment. Furthermore, he was 75% correct on naming objects from tactile presentation, ruling out an explanation in term of a more

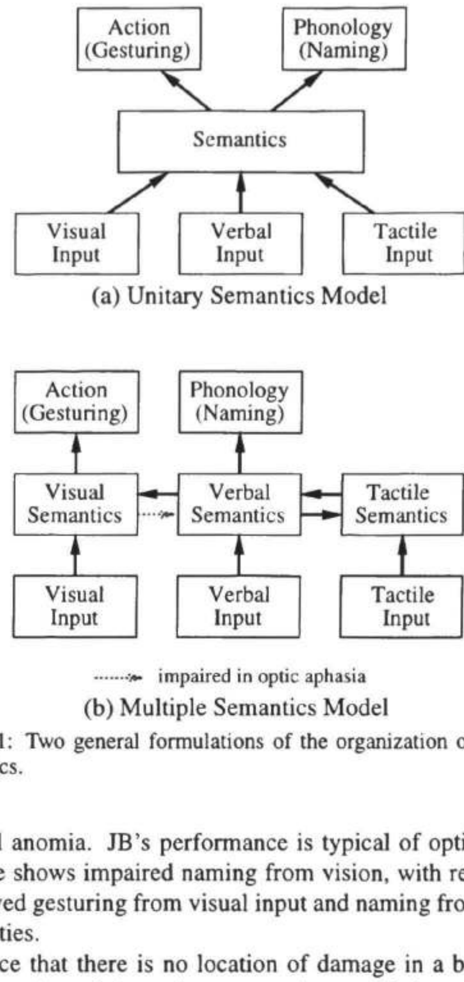

Figure 1: Two general formulations of the organization of lexical semantics.

general anomia. JB's performance is typical of optic aphasics; he shows impaired naming from vision, with relatively preserved gesturing from visual input and naming from other modalities.

Notice that there is no location of damage in a box-and-arrow version of the unitary semantics model that will give rise to this pattern of performance. Damage between vision and semantics would result in visual agnosia, wherein a visually presented object would not be recognized and so its use could not be gestured. Damage between semantics and phonology would lead to a modality-independent naming deficit. Finally, damage to semantics itself would degrade performance on tasks in all modalities.

Shallice (1987, also see Beauvois, 1982) solved this problem by dividing semantics into multiple, modality-specific subsystems (see Figure 1b). On this view, optic aphasia results from a disconnection of visual and verbal semantics.

Thus, access to the verbal semantic representations needed for naming is prevented only for visual input, and all other tasks are unimpaired.

While it might seem that this model provides an adequate account of the data, there are major problems with it. First and foremost, it seems distinctly unparsimonious. The need to develop several relatively independent modules in which to store semantic information significantly increases the difficulty of learning about objects. Instead of forming a single, amodal representation for a particular object, the brain would have to learn a number of different representations. The processes for maintaining consistency of these representations, and other implementational factors, are unclear.

Moreover, while the multiple-semantics model provides a convenient explanation for the major aspects of optic aphasia, it does a poor job at accounting for other relevant factors. In particular, performance on miming and on non-visual naming are typically also impaired, although not to the same extent as naming from vision. It is difficult to see how a single lesion to the multiple-semantics model would result in minor impairment of miming and tactile naming, and substantial impairment of visual naming. In fact, this would seem to require three separate lesions. Furthermore, optic aphasics' ability to discriminate between visually similar objects in a semantic categorization task may also be mildly impaired (Riddoch & Humphreys, 1987), suggesting difficulty in accessing semantics from vision.

In light of these problems, Caramazza and colleagues (Caramazza et al., 1990; Hillis & Caramazza, 1995; Hillis et al., 1990) argue that the multiple-semantics model does not provide an adequate explanation of the actual patient data. Moreover, they argue that the unitary-semantics approach can, in fact, account for optic aphasia if certain predictive properties of the environment are taken into account. They postulate that visual properties tend to be highly predictive of functional properties. This is similar to Gibson's (1979) notion of *affordances*—the fact that a cup has a handle and a concave shape is highly (although not perfectly) predictive of its function; it affords holding liquid, and the particular physical manipulations involved in drinking. However, these same visual features provide no systematic information about the object's name. Thus, there are many objects which we could call "cup" but only a few which afford drinking.

The predictiveness of a relationship can be recast in terms of the *systematicity* of a mapping. A mapping is systematic to the extent that it preserves similarity; that is, similar inputs map to similar outputs. Thus, an identity mapping is completely systematic in that it preserves similarity exactly, whereas a random mapping is completely unsystematic in that input similarity is entirely unrelated to output similarity. Another way to characterize systematicity is in terms of how many input features are needed to predict each output feature. In an identity mapping, each output feature is perfectly predicted by a single (corresponding) input feature; in a random mapping, each output feature can be predicted only by knowing the entire input. A highly predictive relationship, such as that between vision and action, corresponds to a highly systematic mapping, whereas a relationship with little predictive value, such as that between vision and naming, can be approximated by a random mapping.

If visual information is systematically related to action/function, then it may be possible to determine functional properties (and, hence, gesture accurately) from partially degraded information. On the other hand, such information may be inadequate for supporting accurate naming given that small differences in input must produce completely different outputs. Thus, partial damage to the mapping between vision and semantics in a unitary-semantics model might be expected to give rise to the overall pattern of performance in optic aphasia (also see Riddoch & Humphreys, 1987).

This idea can be tested by implementing a system which performs tasks similar to those performed by the semantic system, and then examining the performance of that system when it is damaged. We chose to implement such a system using a parallel distributed processing (PDP) framework, for a number of reasons. Primary among them is the fact that the type of computations performed by a PDP system, although not perfectly faithful to those performed by neurons, nonetheless share certain fundamental properties with them. As a result, it is natural to damage a PDP system to varying degrees. Moreover, such systems have been shown to be sensitive to relative degrees of systematicity within a single task, both in terms of rate of acquisition and in terms of the effects of damage (Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989). In this paper, we explore whether optic aphasia can be accounted for by the effects of damage to a PDP network in which multiple input-output mappings of varying systematicity are mediated by the same internal (semantic) representations.

Simulation 1: Basic Effects

As a first step to illustrate the basic effects of systematicity in PDP systems, we trained a simple three-layer feedforward network on either a systematic task or an unsystematic task, and compared its performance on these tasks over the course of learning and following damage.

Method

The network had 20 input units, 40 hidden units, and 20 output units. Each hidden unit received a connection from each input unit and sent one to each output unit. Weights on these connections were initialized to random values uniformly distributed between -0.5 and 0.5 and were unconstrained during learning. In addition, hidden and output units had *bias* connections whose weights were initialized between -0.5 and 0 and were constrained to remain nonpositive during learning. All units used the standard logistic activation function with activations ranging from 0 to 1.

The input to the network consisted of 100 random patterns over the input units, such that each unit had a probability of 0.5 of being active in each pattern. For the systematic task,

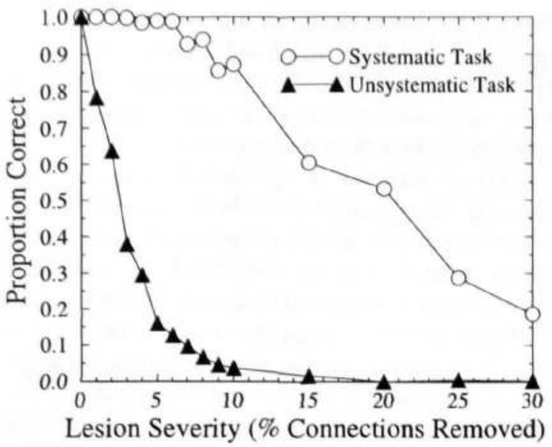

Figure 2: Correct performance on the systematic and unsystematic tasks as a function of the percent of input-to-hidden connections removed in Simulation 1.

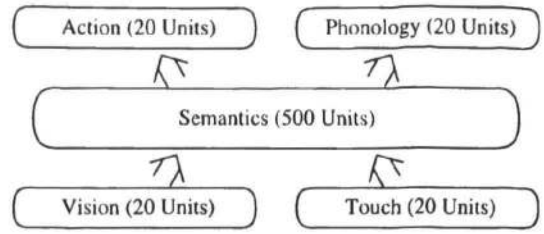

Figure 3: The architecture of the network trained to map between multiple input and output modalities in Simulation 2.

the network was trained to regenerate the identical pattern over the output units (i.e., an identity mapping). For the unsystematic task, a new set of 100 random patterns were generated and paired randomly with the input patterns (i.e., an arbitrary mapping). Note that these mappings were not intended to correspond in any direct way to the actual mappings involved in visual naming and gesturing, but simply to capture the basic distinction between a systematic versus unsystematic task.

The network was trained with back-propagation (Rumelhart, Hinton, & Williams, 1986) using the cross-entropy error function (Hinton, 1989), a learning rate of 0.1 and no weight decay or momentum. If an output unit was within 0.1 of its target, then it was counted as correct and no error was generated for that unit. Training was halted when, for each input presented, all output units were within 0.1 of their targets.

After training, each version of the network was lesioned by randomly selecting and removing a proportion of input-to-hidden connections (ranging from 1% to 30%). At each level of severity, 10 repetitions were run, wherein a new randomly chosen set of connections was removed and the model's performance on all 20 patterns was determined (where, in this context, an output was considered correct if all of the output units had activations on the correct sides of 0.5).

Results and Discussion

Task systematicity had a dramatic effect both on rate of acquisition and on robustness to damage. The systematic task was mastered after only 50 epochs of training. By contrast, the unsystematic task was at floor until 100 epochs. It reached 50% correct at epoch 277 and only achieved perfect performance at epoch 392.

Similarly, performance on the systematic task was far more robust to damage (see Figure 2). Removal of only 1% of input-to-hidden connections left the systematic task unaffected but reduced correct performance on the unsystematic task to 78%. With a 10% lesion, performance on the systematic task remained at 87% correct while the unsystematic task was near floor at 3.7% correct. Even with 20% of input-to-hidden connections removed, correct performance on the systematic task was better than 50%.

In summary, although this first simulation is highly simplified it serves to illustrate the powerful effect that task systematicity has on the performance of PDP networks, both in learning and following damage.

Simulation 2: Multiple Modalities

Simulation 1 was simplified both in the extreme form of systematicity it employed and in the fact that the systematic and unsystematic mappings were learned separately. In the second simulation, we trained a network to map from multiple input modalities (vision and touch) to multiple output modalities (action and phonology), employing a more realistic formulation of systematicity for the vision-to-action mapping.

Method

The network, depicted in Figure 3, consisted of five groups of units: two input groups of 20 units each, named "Vision" and "Touch"; one hidden layer of 500 units, named "Semantics"; and two output groups of 20 units each, named "Action" and "Phonology." The large number of units in Semantics is useful for helping the model learn multiple arbitrary mappings in a reasonable amount of time. (Qualitatively similar results obtain with fewer hidden units, e.g., 100.). Semantics received connections from both of the input groups, and both of the output groups received connections from Semantics. Weights and biases were initialized and constrained as in Simulation 1,

The training environment consisted of 100 objects, divided into 10 categories of 10 objects each. Each object consisted of patterns for Vision, Touch, Action, and Phonology.

Vision input patterns were made to cluster into categories using the following procedure. We first generated 10 random prototype patterns, such that each of 20 features had a probability of 0.5 of being present and all prototypes differed from each other by at least 5 features. For each prototype, we then generated 10 exemplars by choosing two features of the prototype and reversing them. We constrained all exemplars to differ from each other by at least two features. Each exemplar was used as a Vision input pattern.

Action output patterns were generated in the same way as were Vision inputs, although different prototypes were used.

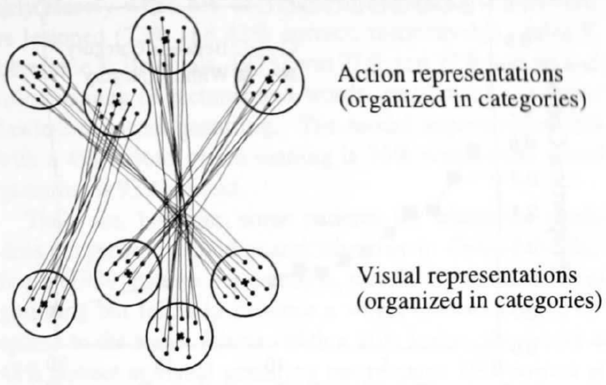

Figure 4: A depiction of the relationship between visual inputs and action outputs. Corresponding patterns are represented as points with a line connecting them.

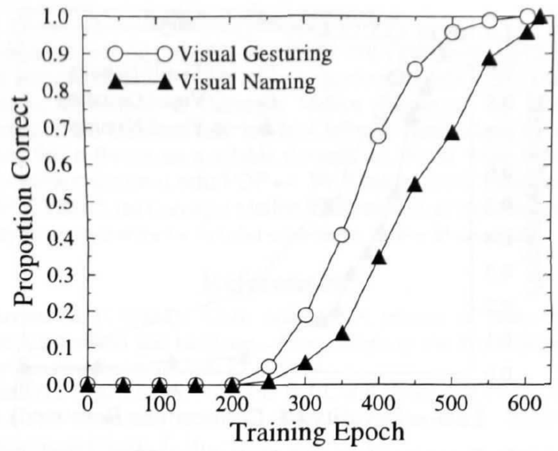

Figure 5: The proportion of Action versus Phonology outputs correct given Visual input, as a function of training epoch.

Exemplars generated from a single prototype can be considered to form a category. In order to create a systematic mapping, we assigned Vision inputs to Action outputs such that, if two objects had input patterns from the same Vision category, then they would have output patterns from the same Action category. Thus, membership in a visual category was predictive of membership in an action category, but individual visual features were not perfectly predictive of individual action features (see Figure 4).

Although these representations are by no means faithful to actual visual and functional representations, they do capture some basic aspects of their structure and relationship. People categorize objects at least partially on the basis of visual features (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976), and there is evidence that our representations of actions have a categorical structure (Klatzky, Pellegrino, McCloskey, & Lederman, 1993). Thus, the use of an environment with visual and functional categories provides a sufficient setting in which to test the implications of relative task systematicity.

Touch input patterns and Phonology output patterns were generated by setting each feature of each pattern to 1 with a probability of 0.5, with the additional constraint that no two Touch patterns or Phonology patterns could be identical. This design results in a systematic relationship between Vision and Action, and a random relationship between Vision and Phonology and between Touch and both of the output modalities. Note that, in addition to being unstructured, the Touch modality had an entirely unsystematic relationship with both Action and Vision. Of course, in actuality, the domain of touch has a high degree of structure that is closely related to the structure among visual and functional representations. However, we chose not to implement this structure nor the relevant relationships because we were primarily interested in the effects of the systematic relationship of vision and action. In fact, by making all of the other relationships random, we ensured that the network can take advantage only of those regularities in the mapping from Vision to Action.

The model was trained using the same learning procedure, parameters, and stopping criteria as in Simulation 1.

Results

Acquisition. Figure 5 shows, as a function of training epoch, the proportion of correct outputs in each modality for Visual inputs. An output pattern was considered correct in this context if all of the output units were within 0.1 of their targets. As expected, the model learned the Vision-to-Action mapping more quickly than the Vision-to-Phonology mapping, $F(1,99)=112.0$, $p<.001$, due to its greater systematicity.

Effects of Damage: Visual Naming versus Gesturing. The trained network was lesioned by removing randomly selected connections from the Vision layer to the Semantics layer. Levels of lesion severity ranged from 1% to 30% of connections removed. At each level, ten repetitions were run, wherein a new randomly chosen set of connections was removed. The model's performance on all mappings was then determined. An output in a particular modality was considered correct, for this task, if all of the output units were on the correct side of 0.5. Average performance at each level of severity is shown in Figure 6.

The model's ability to map from Vision to both Action and Phonology was impaired by the lesions, and, as expected, performance decreased as lesion severity increased. However, at low and intermediate severities, the model performed much better at visual gesturing than at visual naming. Overall, the advantage for the Vision-to-Action mapping was significant, $F(1,99)=963.2$, $p<.001$, as was the interaction of output modality and lesion severity, $F(13,1287)=64.76$, $p<.001$. Note that the model's performance on mapping from Touch to either of the output modalities remained unimpaired. Since the model was feedforward, it is unsurprising that the removal of connections from Vision to Semantics had no effect on the model's performance on Touch mappings.

Effects of Damage: Semantic Categorization. One source of evidence in support of the claim that optic aphasic patients have an impairment in accessing semantics from

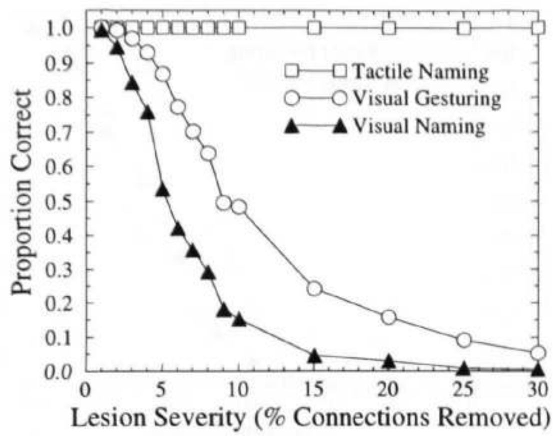

Figure 6: Correct performance on tactile naming, visual gesturing, and visual naming as a function of the percent of Vision-to-Semantics connections removed.

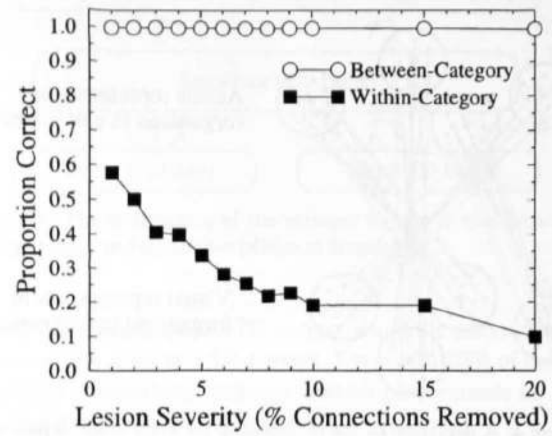

Figure 7: Correct performance at choosing which of three objects is least semantically related to the other two when that object is in a different category ("Between-Category") versus when all three objects are from the same category ("Within-Category"), as a function of the percent of Vision-to-Semantics connections removed.

vision rather than an impairment in semantics per se comes from Riddoch and Humphreys (1987), who demonstrated that their patient, JB, had difficulty performing categorization tasks when fine-grained visual discrimination was required. Riddoch and Humphreys argued that, because the semantic representation is formed from inaccurate and/or incomplete information, it is generally sufficient to drive tasks which do not require a high level of detail, but inadequate for harder tasks. Accordingly, we examine the extent to which our model exhibits similar behavior.

We implemented a forced-choice task by presenting the model with three objects and determining which two it considered most similar. Similarity was judged by computing the normalized dot product of the model's Semantic representations generated by Visual presentation of two objects. The larger the normalized dot product, the more similar the objects were considered to be.

We examined the model's performance on two forced-choice tasks. In the *between-categories* task, it was presented with two objects from the same category and a third object from a different category. In the *within-category* task, it was presented with three objects from the same category. When testing a damaged model, we defined the correct response for each comparison to be the response generated by the undamaged model. For both tasks, we presented the model with all of the triples of objects relevant to that task. This resulted in 40,500 triples for the between-categories task (10 categories $\times \binom{10}{2}$ pairs in each category $\times$ 90 objects from outside the category) and 3600 triples for the within-categories task (10 categories $\times \binom{10}{3}$ triples in each category).

We acquired performance data for the model at twelve levels of damage, ten ranging from one to ten percent of connections from Vision to Semantics lesioned, one at fifteen percent lesioned, and one at twenty percent lesioned. Only one repetition was performed at each level, largely because of the computational difficulty of performing dot products on thousands of 500 element vectors.

Figure 7 shows the model's performance on each task for each level of severity. At all levels, the model's performance on the between-categories task was nearly perfect. At 20% of connections lesioned, the model still had a correct performance rate of nearly 99.5%. By contrast, the model's performance on the within-categories task was very poor even for extremely mild lesions. At a mere 1% of connections lesioned, the model's correct performance was only about 58%. This dropped to slightly less than 10% correct at 20% of connections lesioned.

Our model, therefore, displays the expected behavior: it performs much more poorly on a task that requires fine-grained visual discriminations. However, JB's error rate on such a task was only about 8%, whereas our model's error rate starts at above 40%. This extremely high error rate is an effect of the structure of our Visual domain. Members of a Vision category do not differ enough to give the model a good chance at performing the within-categories task. With a more realistically structured environment, one would expect the within-category error rates to decrease substantially.

Discussion

The simulation demonstrates that the category-based systematicity of the Vision-to-Action mapping provides an advantage for both learning and performance under damage. As a result, when damaged, the model exhibits the central characteristics of optic aphasia: an impairment in visual naming with relatively spared performance on visual gesturing and on naming from other modalities (e.g., touch). The model also accounts for preserved between-category discrimination with impaired within-category discrimination.

The model even provides a fairly good quantitative match to data from some specific optic aphasic patients. JB (Riddoch & Humphreys, 1987) was 75% correct at miming the use of visually presented objects but only 46% correct at naming them. The graph indicates that the model matches this fairly closely when 6% of Vision-to-Semantics connections are lesioned (77% and 42% correct, respectively). Jules F. (Lhermitte & Beauvois, 1973) was 72% and 77% correct and visual naming of pictures and words, respectively, and was flawless at visual gesturing. The model approximates this with a 4% lesion: visual naming is 76% correct and visual gesturing is 93% correct.

There are, however, some patients for whom the model does not provide a good quantitative match. Coslett and Saffran's (1989) patient, for example, was 50% correct at visual gesturing but failed to produce a single correct naming response to the same objects. With a 10% lesion, the model is 48% correct at visual gesturing but remains 15% correct at visual naming; with a 20% lesion, visual naming is reduced to 3% correct but visual gesturing reaches only 16% correct.

Also note that, because it has a feedforward architecture, the model does not account for cases in which tactile naming is less than perfect. Jules F., for instance, was only 91% correct on naming from touch. In a recurrent version of the current model, interactions across damaged Vision-Semantics connections might lead to some naming errors for stimuli presented to the undamaged modality. It may also be the case that some optic aphasic patients have additional mild damage to the semantic system itself; such damage would be expected to lead to a mild deficit in naming from other modalities and to exacerbate the visual naming deficits.

Despite its limitations, the simulation does provide support for the central claim of the current work, that optic aphasia and other modality-specific naming deficits are not incompatible with a unitary-semantics account if one takes into account the robustness of tasks of differing systematicity.

Conclusions

Semantic knowledge for objects is standardly thought to be represented within a single, amodal system. One challenge to this point of view is that modality-specific naming deficits such as optic aphasia are not easily explained on such an account. In this paper we have shown that a PDP implementation of a unitary semantic system can, in fact, account for central characteristics of optic aphasia under the assumption that input-output relationships vary in their systematicity.

It should be acknowledged, though, that the model does not account for all of the data, including the quantitative magnitude of the difference between visual naming versus gesturing performance in some patients (e.g., Coslett & Saffran, 1989). This discrepancy may simply reflect limitations in the scale of the simulation and in the sophistication of the representations. However, if anything, these simplifications may have amplified the effect in the model. Thus, the current results should be taken as indicating that systematicity is an important contributing factor in understanding these deficits, but may not provide a complete account. We leave it for future research to determine whether it is possible to provide a fully adequate account of optic aphasia and related disorders without at least some graded degree of modality-specific specialization within the semantic system (also see Shallice, 1993).

Acknowledgements

This research was supported by an NIMH FIRST award (MH55628) to the second author. Simulations were performed using the PDP++ simulator, developed at Carnegie Mellon University by Randy O'Reilly, Chad Dawson, and Jay McClelland; instructions for accessing this software are available through the World Wide Web at http://www.cnbc.cmu.edu/PDP++. We thank Marlene Behrmann, Jay McClelland, the Carnegie Mellon PDP research group, and three anonymous reviewers for helpful comments and/or discussions.

References

Beauvois, M.-F. (1982). Optic aphasia: A process of interaction between vision and language. *Proceedings of the Royal Society of London, Series B, 298*, 35–47.

Caramazza, A., Hillis, A. E., Rapp, B. C., & Romani, C. (1990). The multiple semantics hypothesis: Multiple confusions? *Cognitive Neuropsychology, 7*, 161–189.

Coslett, H. B., & Saffran, E. M. (1989). Preserved object recognition and reading comprehension in optic aphasia. *Brain, 112*, 1091–1110.

Endo, K., Makishita, H., & Sugishita, M. (1996). Modality specific naming and gesture disturbances: A case with optic aphasia, bilateral tactile aphasia, optic apraxia and tactile apraxia. *Cortex, 32*, 3–28.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton-Mifflin.

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

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

Hinton, G. E. (1989). Connectionist learning procedures. *Artificial Intelligence, 40*, 185–234.

Iorio, L., Falanga, A., Fragassi, N. A., & Grossi, D. (1992). Visual associative agnosia and optic aphasia: A single case study and review of the syndromes. *Cortex, 28*, 23–37.

Klatzky, R. L., Pellegrino, J., McCloskey, B. P., & Lederman, S. J. (1993). Cognitive representations of functional interactions with objects. *Memory and Cognition, 21*, 294–303.

Lhermitte, F., & Beauvois, M.-F. (1973). A visual-speech disconnexion syndrome: Report of a case with optic aphasia, agnosic alexia and colour agnosia. *Brain, 96*, 695–714.

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.

Riddoch, M. J., & Humphreys, G. W. (1987). Visual object processing in optic aphasia: A case of semantic access agnosia. *Cognitive Neuropsychology, 4*, 131–185.

Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382–439.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*, 533–536.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*, 523–568.

Shallice, T. (1987). Impairments of semantic processing: Multiple dissociations. In M. Coltheart, G. Sartori, & R. Job (Eds.), *The cognitive neuropsychology of language* (pp. 111–128). Hillsdale, NJ: Erlbaum.

Shallice, T. (1993). Multiple semantics: Whose confusions? *Cognitive Neuropsychology, 10*, 251–261.