To appear in D. Heinke, G. W. Humphreys, and A. Olson (Eds.), *Connectionist models in cognitive neuroscience: Proceedings of the Fifth Annual Neural Computation and Psychology Workshop.* New York: Springer-Verlag.

Systematicity and Specialization in Semantics

David C. Plaut
Carnegie Mellon University and the Center for the Neural Basis of Cognition
Pittsburgh PA, USA

Abstract

Most models of cognitive processing assume a considerable amount of built-in structure in the cognitive system, such as modality-specific semantic systems. Such structural distinctions are thought to be necessary to account for various neuropsychological dissociations, including the selective impairment of visual object naming with spared tactile naming and visual gesturing, known as optic aphasia. A distributed connectionist simulation is presented in which optic aphasia occurs following damage as a result of more graded representational specialization. This specialization emerges through learning under the combined influence of two general factors: the relative systematicity of mappings between input and output modalities, and a topographic bias favoring short connections. The results raise the possibility that much of the structure of the lexical/semantic system can be derived from general learning principles and need not assumed a priori.

1 Introduction

A central issue in the study of language and cognition concerns the organization of semantic representations for words, objects and their associated actions. A natural perspective on this issue is what has been termed the *unitary semantics* account [1, 2, 3, 4]: that the meanings of objects are stored in a central, amodal semantic system that can be accessed from various input modalities (e.g., vision, touch, spoken language) and can be used to direct behavior in various output modalities (e.g., physical action, writing, speaking). Such an organization is parsimonious in that it allows knowledge derived from one modality to generalize automatically to others.

There are, however, a number of empirical findings that seem problematic for a unitary semantics account. These findings come largely from the study of patterns of impaired and preserved cognitive abilities of certain types of brain-damaged patients. In particular, the current work focuses on the modality-specific aphasias, in which patients have naming deficits specific to a particular input modality. For example, optic aphasic patients (see, e.g., [5]) fail to name visually presented objects even though they can demonstrate that they recognize the objects (e.g., by gesturing appropriately) and can name them from definition or when presented in another input modality (e.g., tactile or auditory). Interesting, the visual naming deficit appears to be less severe when generating the names of actions compared with objects [6]. Table 1 lists the correct performance of three optic aphasic patients on naming and gesturing from visual input and naming from tactile input, and the performance of one patient on generating the name of the action associated with a visually presented object (e.g.,

Study	Correct Performance			
	Visual Naming	Visual Gesturing	Tactile Naming	Action Naming
Lhermitte & Beauvois [5]	73%	100%	91%	
Manning & Campbell [6]	27%	75%	90%	67%
Coslett & Saffran [10]	0%	50%	92%	

Table 1: Correct performance of three optic aphasic patients on various tasks.

"sleep" for *bed*). Analogous selective naming deficits have been documented in the auditory modality [7] and in the tactile modality [8]. The pattern of impaired and preserved performance in optic aphasia is difficult to reconcile with standard forms of the unitary semantics account, under the assumption that naming must proceed via semantics (although see [9] for an alternative view). Modality-specific damage prior to semantics would impair comprehension; damage between semantics and output phonology would impair naming from other modalities; damage to semantics itself would impair both comprehension and naming across modalities.

Note, however, that modality-specific aphasias can be explained easily if the semantic system is subdivided into modality-specific subsystems (e.g., visual, tactile, verbal) [11, 12]; optic aphasia, for example, would arise from a disconnection between visual semantics (sufficient for recognition) and verbal semantics (required for naming) [5, 13]. Such strict subdivisions within semantics, however, are considered by many researchers to be unparsimonious [4] if not theoretically incoherent [1]. Moreover, detailed empirical testing [14] suggests that visual object recognition and comprehension in optic aphasia may not be fully intact, raising the possibility that partial impairment of a modality-specific input pathway to a unitary semantic system may be sufficient to account for the pattern of performance [2, 14, 15]. Specifically, comprehension and gesturing might be preserved relative to naming following input damage due to the "privileged access" of visual structural descriptions to certain semantic features [2, 15] and/or to action representations [14].

Recently, McGuire and Plaut [16] developed a connectionist implementation consistent with this proposal. They trained a network to map a visual or tactile representation of an object onto its phonological and action representations via a common hidden representation corresponding to a unitary semantics system. Visually similar objects were assumed to have similar associated actions [17, 18] but unrelated names. The sensitivity of learning to the systematicity between vision and action provided the basis for the "privileged access" assumed in other accounts. McGuire and Plaut demonstrated that, due to differences in task systematicity, mild damage between vision and semantics did, in fact, impair naming more than gesturing (and other tests of comprehension). The magnitude of the effects in the simulation were, however, relatively small compared with those observed in some patients (e.g., [10]). Moreover, an account based solely on systematicity (or another form of privileged access) would not seem to generalize to the analogue of optic aphasia in the auditory domain [7], given the relative lack of systematicity between object sounds and actions.

The current work extends the approach of McGuire and Plaut to include an additional pressure for functional specialization in semantics: a topographic bias on learning favoring short connections [19]. The basic idea is that brain organization must permit sufficient connectivity among neurons to carry out the necessary information processing, but the total axon volume must fit within the confines of the skull. This constraint is severe: If the brain's 10^{11} neurons were placed on a sphere and fully interconnected with axons 0.1 μm in radius, accommodating the axon volume would require a sphere over 12.5 miles in diameter [20]. Clearly, connectivity must be as local as possible. Within a connectionist approach, this bias can be instantiated by assigning functional locations to units and by reducing the effectiveness of weight changes for each connection as a function of its length (i.e., distance between the connected units). As a result, during learning the network will use short connections (i.e., nearby hidden units) as much as possible, and will develop significant weights on longer connections only if necessary. Jacobs and Jordan [19] demonstrated that a bias favoring short connections can induce varying degrees of functional specialization among hidden units in a network trained to derive both the identity and position of a visual object.

In the current context, hidden units form semantic representations that mediate among multiple input and output modalities. Under a topographic bias, the degree to which hidden units participate in a particular input-output mapping will depend on their proximity to the relevant modalities. Hidden (semantic) regions that are equidistant from multiple modalities may learn to function in a relatively amodal way, whereas regions near a particular modality may serve more modality-specific functions. Within such a system, the degree of modality specificity is entirely graded and subject to the demands of the relevant tasks—in this way, it constitutes a middle ground between the extreme views of semantics as a unitary, amodal system on the one hand [1, 4] and as multiple, modality-specific systems on the other [11, 13]. On this account (and as demonstrated below), optic aphasia arises from damage to the connections from high-level visual representations to a region of semantics that is partially specialized for naming.

The notion of graded functional specialization within semantics derives from a perspective first articulated by Allport [21] and Warrington and McCarthy [22] and later elaborated by Shallice [23, pp. 302–304]:

> It may be useful to think of it (i.e., the semantic system) as a giant distributed net in which regions tend to be more specialised for different types of process.... The basis on which differentiation between processing regions within semantics would develop would include the most favoured modality of input for the process. Modality-specific pre-semantic classification subsystems would, thus, come to be more closely linked with some of the processing regions within the overall semantic system. So "visual semantic" and "verbal semantic" could be thought of as partially specialized subregions.... However, for explanations of this sort to be more than a speculation, a simulation of the hypothetical semantic system would be required.

This paper presents just such a simulation.

2 Simulation

2.1 Method

A continuous recurrent attractor network was trained to map either visual or tactile input to action and/or phonological output (see [24] for additional details and results). The architecture of the network is shown in Figure 1. It has two input groups (Vision and Touch) and two output groups (Action and Phonology), each of which contains 20 units. These groups are connected with 225 Semantic (hidden) units, organized in a 15×15 grid. The Semantic units receive inputs from both Vision and Touch and are bidirectionally connected with both Action and Phonology (which are each fully interconnected). The network also has two Task units which project to Semantics, Action, and Phonology and whose function is described below. Each non-input unit was also given a bias connection (from a unit whose state is fixed to 1) that determined the unit's "resting" activation in the absense of other input. In total, the network had 28,555 connections. Units were assigned functional positions in 2 dimensions as depicted in Figure 1 with connection length defined as euclidean distance.

The activations of units in the network change continuously in time as a function of their summed input from other units. To simulate on a digital computer, this continuous process is approximated by a finite difference equation (with a discretization of $\tau = 0.2$), such that the new activation of a unit is a sigmoid function of a weighted average of its old summed input and the new input it is currently receiving (where τ is the weighting factor).

No attempt was made to model the detailed structure of any of the input or output modalities. Rather, sets of more abstract representations were defined in such a way that the similarity structure both within and between modalities approximated the central theoretical claims about the relevant tasks—namely, that there is considerable systematicity among visual, tactile, and action representation, but no systematicity between any of these and phonology.

Within each modality other than Phonology, 100 representations were generated to form 5 categories of 20 objects each. This was done by first generating 5 random prototype patterns (with 10 of 20 features equal to 1) and then generating exemplars by changing each feature with probability 0.1. Partial systematicity among Vision, Touch, and Action was enforced by assigning representations to objects in such a way that, across all pairs of objects in the same Vision or Touch category, 80% of them were also in the same Action category.

Phonological representations were generated to form CVC strings over three slots (with 7, 6, and 7 units, respectively). Each phoneme was coded by 2 active features in a slot, yielding 16 possible consonants, 5 possible vowels, and $16 \times 5 \times 16 = 1280$ possible CVC strings. Of these, 200 were chosen randomly an assigned as the names of objects (100) and as the names of the actions associated with the objects (100).

During training, the network was presented with either the visual or tactile representation of an object (with the other modality set to all zeros) and trained to perform either of two tasks: 1) generate the phonological representation of the name of the object (object naming) with no targets for action, or 2) generate both the action representation (gesturing) and the phonological representation of the action associated with

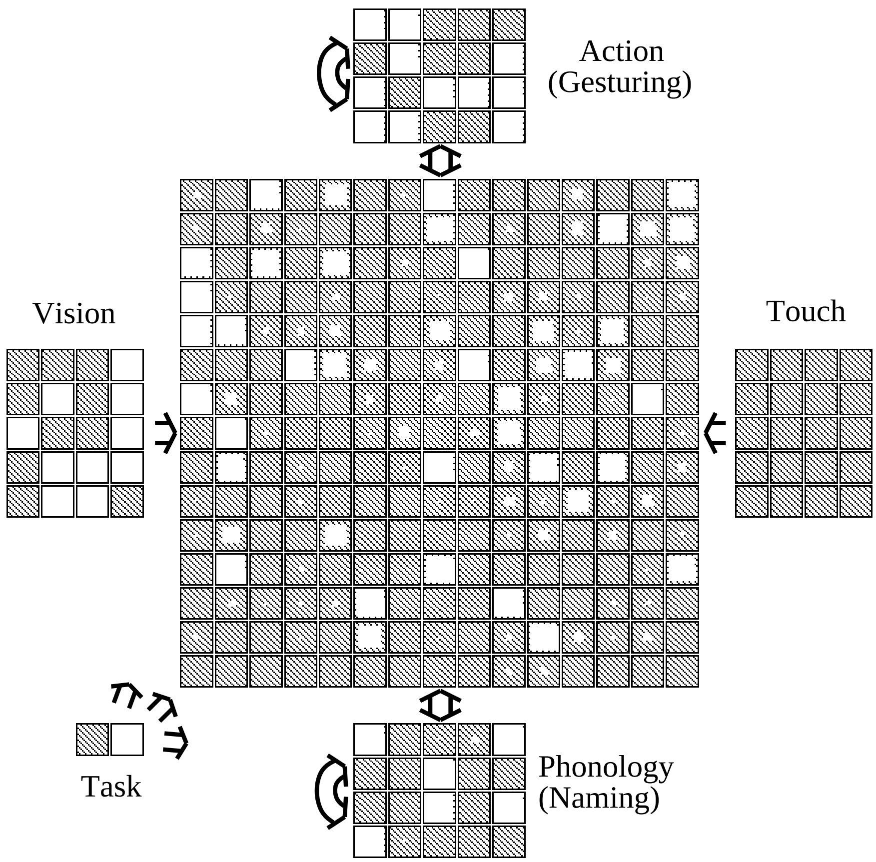

Figure 1: The architecture of the network. Each grey square constitutes a unit whose activity value is indicated by the size of the white region within the square. The activations shown are those generated by the fully trained network when presented with a visual object and instructed (with the Task units) to generate the action representation and name of the action associated with the object. Bold arrows indicate fully connectivity between the indicated unit groups; bidirectional arrows indicate two separate projections. The Task units are connected to all non-input groups. The positions of units in the figure correspond to their functional positions with respect to calculating connection lengths (as euclidean distance).

the object (action naming). The particular task that the network was to perform (object or action) was indicated by the activity of one of two Task units (see Figure 1).

Once an input was clamped on a particular input modality, units in the network updated their states over 5.0 units of time with $\tau = 0.2$, corresponding to 5.0/0.2 = 25 updates. Cross-entropy error was injected only over the last unit of time, based on the task being performed. Error derivatives were then calculated using a version of back-propagation through time adapted for continuous units [25]. The topographic bias on learning was implemented by scaling the magnitude of the derivative on each connection by a Gaussian function (*SD* = 10) of its length.[1] This produces a scaling of near 1.0 for the shortest connections, and near 0.1 for the longest. The weights were then updated using a learning rate of 0.01, momentum of 0.9, and weight decay of 0.00005.

The network was trained on a total of 110,000 object presentations, corresponding to 275 presentations per condition (object $\times$ modality $\times$ task). At this point, all output activations generated by the network in all conditions were on the correct side of 0.5.

On the hypothesis that optic aphasia arises from impaired semantic access from vision [14], lesions were administered to connections from Vision to Semantics. Each lesion had a specified "center" in Semantics. The probability of removing a given Vision-to-Semantics connection was a 2D Gaussian function of the position of the receiving Semantic unit relative to the center of the lesion. Thus, connections to the unit at the center of the lesion were lesioned with probability 1.0; connections to progressively more distant Semantic units were less and less likely to be lesioned. The severity of the lesion was controlled by the standard deviation of the Gaussian.

2.2 Results and Discussion

Figure 2 shows the correct performance of the model on various tasks after Vision-to-Semantics lesions centered in Semantics very close to Phonology (7th unit in the bottom row; see Figure 1), as a function of the severity of the lesion. Results are averaged over 40 instances of lesion at each level of severity. The tasks are a) visual object naming (Vision input, Phonology output, Task: object); b) tactile object naming (Touch input, Phonology output, Task: object); c) visual gesturing (Vision input, Action output, Task: action); and d) visual action naming (Vision input, Phonology output, Task: action). For comparison, the figure also shows (with vertical lines and asterisks) the levels of performance of the three patients in Table 1 on visual naming (bottom), visual gesturing (top), and for one, action naming (middle).

As the figure shows, the lesioned network is far more impaired at visual naming than at either visual gesturing or tactile naming, thus exhibiting the hallmark characteristics of optic aphasia. In fact, when compared with the levels of performance of the three patients, the network does a reasonable job at matching the magnitudes of the dissociation between visual naming versus gesturing across a range of severity. Moreover, like at least one patient, the network is far better at naming the action associated

[1] For simplicity, this scaling was not applied to the connections from Task units as none of the current theoretical issues relate to topographic influences on executive control. Also note that this procedure differs from the one used by Jacobs and Jordan [19] who scaled the magnitude of weight decay rather than error derivatives.

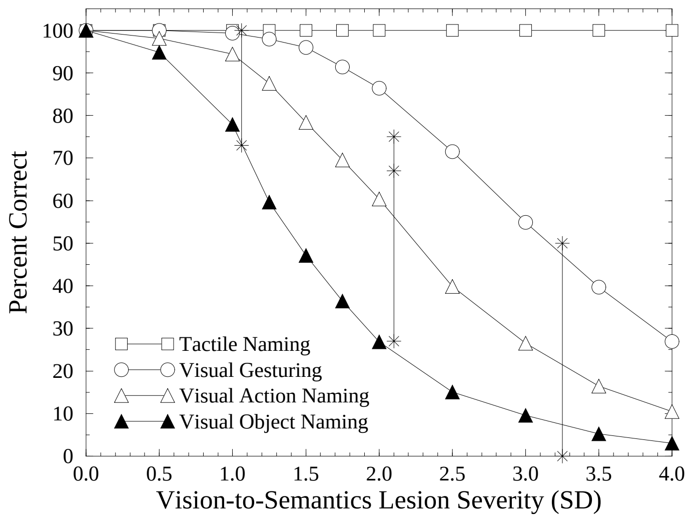

Figure 2: Correct performance of the network on various tasks as a function of lesion severity, and corresponding levels of performance of three optic aphasic patients (see Table 1).

with a visually presented object than at naming the object itself.

Two factors underly the relatively selective impairment of visual naming relative to visual gesturing and tactile naming exhibited by the network. The first is the greater systematicity in the relationship between visual objects and their associated actions than between the objects and their names. In connectionist networks, systematic mappings are acquired more quickly and are more robust to damage than are unsystematic mappings [16]. (Although tactile naming is also unsystematic, it is unaffected by Vision-to-Semantics lesions.) The second factor is the graded modality-specific functional specialization that develops as a result of the topographic bias on learning favoring short connections. The region in Semantics near Phonology learns to become somewhat specialized for generating names compared with actions. Damage to connections from Vision to this region thus impairs naming more than gesturing.

Why, then, is generating action names relatively preserved compared with generating object names? The reason is that the network learns to rely on support from the Action representation when generating an action name. It does this because a given visual input by itself is ambiguous with respect to the correct phonological output—it could be either the object name or the action name. Because the network is trained to generate the Action representation in conjunction with generating the action name over Phonology, it is natural for it to use the derived Action representation to resolve the ambiguity and override the object name. The damaged network generates the Action representation relatively successfully from visual input (as evidenced by the good visual gesturing performance), and thus this information is available to support relatively good (although far from perfect) naming of actions in the face of impaired object naming.

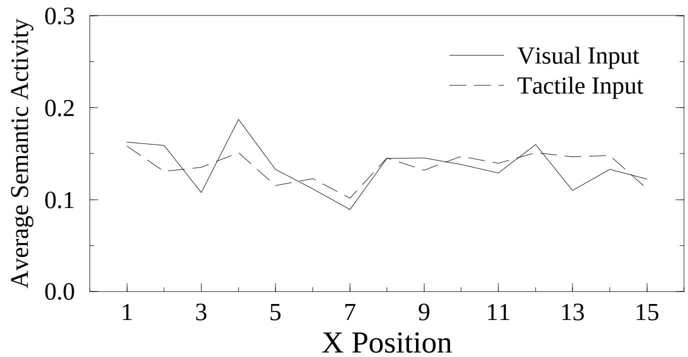

Figure 3: Average activations of Semantic units as a function of their horizontal position, for visual and tactile input.

One final issue to consider is whether the topographic bias has induced the system to develop modality-specific semantic representations for visual versus tactile input. Two analyses suggest this is not the case. Figure 3 shows the average activations of Semantic units for visual and tactile input, as a function of the horizontal position of the unit (recall that Vision is on the left of Semantics and Touch is on the right). There is no differential involvement of Semantic units in representing visual versus tactile input (paired $t_{14} < 1$)—the entire semantic system is involved in representing both types of information.

The second analysis involved comparing the similarity among semantic representations generated within versus between modalities. The mean correlation of the semantic representations generated by the same object presented in different modalities (.625, *SD* = .084) was much greater than that for pairs of objects from the same category presented in the same modality (.357, *SD* = .185) or in different modalities (.217, *SD* = .161), each of which was much greater than for pairs from different categories in the same modality (.026, *SD* = .121). Thus, representational similarity in Semantics was governed far more by identity and semantic relatedness than by modality of presentation.

3 Conclusion

In the current work, a distributed connectionist model of semantics is presented in which semantic representations develop under the pressure of learning to mediate among multiple input and output modalities. The system has a topographic bias on learning favoring short connections, which leads to a graded degree of modality-specific functional specialization within semantics. As a result, damage to connections from vision to regions of semantics near phonology impairs visual object naming far more than visual gesturing or tactile naming, as observed in optic aphasia. Moreover, as in patients, the system is better at generating the name of an action associated with an object than at generating the name of the object itself, because action naming is supported by action representations.

Although not reported in this paper, system also exhibits modality-specific impairments for grammatical categories (nouns vs. verbs; [26, 27, 28]) following lesions to or from regions of semantics which are partially specialized for objects (nouns) or actions (verbs; see [24] for details). In this way, it avoids the rather unpalatable alternative of proposing an anatomic separation of lexical representations by grammatical category within each of the input and output modalities [26].

It is important to note that the topographic bias during learning did not lead to the development of modality-specific subsystems within semantics—rather, the learned functional specialization is of a far more graded nature. Thus, the current account of optic aphasia in particular, and of semantic organization more generally, benefits not only by being computationally explicit, but also by constituting an intermediate theoretical perspective between strictly amodal [1, 4] and strictly modality-specific [11, 13] accounts of semantics.

Acknowledgments

This work was supported by an NIH/NIMH FIRST award (Grant MH55628). Correspondence may be sent to David Plaut, Mellon Institute 115-CNBC, Carnegie Mellon University, 4400 Fifth Avenue, Pittsburgh PA, 15213-2683, USA (email plaut@cmu.edu).

References

[1] A. Caramazza, A. E. Hillis, B. C. Rapp, and C. Romani. The multiple semantics hypothesis: Multiple confusions? *Cognitive Neuropsychology*, 7:161–189, 1990.

[2] A. E. Hillis and A. Caramazza. Cognitive and neural mechanisms underlying visual and semantic processing: Implications from "optic aphasia". *Journal of Cognitive Neuroscience*, 7(4):457–478, 1995.

[3] A. E. Hillis, B. Rapp, C. Romani, and A. Caramazza. Selective impairments of semantics in lexical processing. *Cognitive Neuropsychology*, 7:191–243, 1990.

[4] M. J. Riddoch, G. W. Humphreys, M. Coltheart, and E. Funnell. Semantic systems or system? Neuropsychological evidence re-examined. *Cognitive Neuropsychology*, 5(1):3–25, 1988.

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

[6] L. Manning and R. Campbell. Optic aphasia with spared action naming: A description and possible loci of impairment [note]. *Neuropsychology*, 30(6):587–592, 1992.

[7] G. Denes and C. Semenza. Auditory modality-specific anomia: Evidence from a case of pure word deafness. *Cortex*, 11:401–411, 1975.

[8] M.-F. Beauvois, B. Saillant, V. Meininger, and F. Lhermitte. Bilateral tactile aphasia: A tacto-verbal dysfunction. *Brain*, 101:381–401, 1978.

[9] G. Ratcliff and F. A. Newcombe. Object recognition: Some deductions from the clinical evidence. In A. W. Ellis, editor, *Normality and Pathology in Cognitive Functions*, pages 147–171. Academic Press, New York, 1982.

[10] H. B. Coslett and E. M. Saffran. Preserved object recognition and reading comprehension in optic aphasia. *Brain*, 112:1091–1110, 1989.

[11] E. K. Warrington. The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology*, 27:635–657, 1975.

[12] T. Shallice. Impairments of semantic processing: Multiple dissociations. In M. Coltheart, G. Sartori, and R. Job, editors, *The Cognitive Neuropsychology of Language*, pages 111–128. Erlbaum, Hillsdale, NJ, 1987.

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

[14] M. J. Riddoch and G. W. Humphreys. Visual object processing in optic aphasia: A case of semantic access agnosia. *Cognitive Neuropsychology*, 4(2):131–185, 1987.

[15] D. C. Plaut and T. Shallice. Perseverative and semantic influences on visual object naming errors in optic aphasia: A connectionist account. *Journal of Cognitive Neuroscience*, 5(1):89–117, 1993.

[16] S. McGuire and D. C. Plaut. Systematicity and specialization in semantics: A computational account of optic aphasia. In *Proceedings of the 19th Annual Conference of the Cognitive Science Society*, pages 502–507, Hillsdale, NJ, August 1997. Erlbaum.

[17] J. J. Gibson. *The Ecological Approach to Visual Perception.* Houghton-Mifflin, Boston, 1979.

[18] E. Rosch, C. B. Mervis, W. Gray, D. Johnson, and P. Boyes-Braem. Basic objects in natural categories. *Cognitive Psychology*, 8:382–439, 1976.

[19] R. A. Jacobs and M. I. Jordan. Computational consequences of a bias toward short connections. *Journal of Cognitive Neuroscience*, 4(4):323–336, 1992.

[20] M. E. Nelson and J. M. Bower. Brain maps and parallel computers. *Trends in Neurosciences*, 13:403–408, 1990.

[21] D. A. Allport. Distributed memory, modular systems and dysphasia. In S. K. Newman and R. Epstein, editors, *Current Perspectives in Dysphasia*. Churchill Livingstone, Edinburgh, 1985.

[22] E. K. Warrington and R. McCarthy. Categories of knowledge: Further fractionation and an attempted integration. *Brain*, 110:1273–1296, 1987.

[23] T. Shallice. *From Neuropsychology to Mental Structure*. Cambridge University Press, Cambridge, 1988.

[24] D. C. Plaut and S. McGuire. Systematicity and specialization in semantics: A computational account of optic aphasia. Manuscript in preparation.

[25] B. A. Pearlmutter. Learning state space trajectories in recurrent neural networks. *Neural Computation*, 1(2):263–269, 1989.

[26] A. Caramazza and A. E. Hillis. Lexical organization of nouns and verbs in the brain. *Nature*, 349:788–790, 1991.

[27] A. E. Hillis and A. Caramazza. The representation of grammatical categories in the brain. *Journal of Cognitive Neuroscience*, 7(3):457–478, 1995.

[28] B. Rapp and A. Caramazza. The modality-specific organization of grammatical categories: Evidence from impaired spoken and written sentence production. *Brain and Language*, 56(2):248–286, 1997.