Chapter 6

Extending the task domain: Effects of abstractness

The final aspect of the H&S model that we investigate is the definition of the task of reading via meaning. Defining a task for a network involves choosing a set of input-output pairs to be presented to the network, as well as specifying how these are represented as patterns of activity over groups of units. Formulating a reasonable task definition for the purposes of modeling human behavior involves a trade-off between being as faithful as possible to what is known about the nature of representations from empirical work, while remaining within the often severe constraints imposed by the available computational resources. As Einstein put it, "Everything should be made as simple as possible, but no more so."

First and foremost, the task that the network performs must adequately approximate the task faced by subjects, or the network's behavior, however interesting in its own right, will have little relevance to understanding human behavior. However, exactly what constitutes "adequate" is very much a matter of debate. In essence, the decisions that are made in creating a simplified version of the task for the network constitute empirical claims about what aspects of the information available to subjects is crucial for understanding their behavior. While our empirical understanding of the nature of how different types of information are represented provides useful constraints, it remains insufficiently detailed to specify the precise representations of each input-output pair as patterns of activity over groups of units. This is where computational considerations of what types of representation networks find easy or difficult to use come into play.

The main computational limitations in specifying a task stem from the fact that the time to train a network increases with the size of the network and the number of examples it is trained on. Thus there is strong pressure to use as few units as possible to represent the input and output, and to keep the size of the training set within reasonable limits. For tasks that require capturing the statistical structure among examples (e.g. mapping orthography to phonology), it may be necessary to use a large number of training cases in order to guarantee good performance on novel inputs. For tasks involving unrelated associations (e.g. mapping orthography to semantics) it may be sufficient to use a small number of examples. However, a drawback of using a small training set is that it becomes difficult to include all of the types of variations among examples that are empirically relevant. The fact that the H&S model was trained on only 40 words is a serious limitation not so much because the nature of the mapping from orthography to semantics would be fundamentally different if more words were involved, but that only the most general semantic distinction, category membership, could be investigated. The influences of many other variables known to affect patients' reading behavior were not investigated.

In particular, a distinction among words known to have a significant effect on the reading behavior of deep dyslexics is their imageability or concreteness. This issue could not be addressed using the original H&S word set because it contains only concrete nouns. The purpose of this chapter is to demonstrate that the approach taken by H&S can be extended to account for additional detailed characteristics of deep dyslexic reading behavior, relating to the effects of the abstractness/concreteness of stimuli and responses, and interactions with visual influences in errors.[1]

6.1 Effects of abstractness in deep dyslexia

The effect of the abstractness of the stimulus on deep dyslexic reading has been investigated in a number of ways. The most basic is its effect on the probability that a word will be read correctly. Coltheart et al. (1987a) claim that all patients who make semantic errors find concrete words easier to read than abstract ones. In many patients a very large difference is observed: 73% vs. 14% for K.F. (Shallice & Warrington, 1980), 67% vs. 13% for P.W. and 70% vs. 10% for D.E. (Patterson & Marcel, 1977).

A more subtle effect is the way that the concreteness of a word can affect the probability of the occurrence of visual errors. Shallice & Warrington (1975) noted in their patient KF that the responses tended to be more concrete than the stimuli when visual errors were made. This has since also been observed in patients B.L. (Nolan & Caramazza, 1982) and G.R. (Barry & Richardson, 1988); patient P.S. (Shallice & Coughlan, 1980) showed a strong trend ($p < .06$) in the same direction. The same effect is also apparent in the corpus of errors made by P.W. and D.E.(see Appendix 2 of Coltheart et al., 1980). The relative concreteness of the stimuli on which different types of responses occur has been investigated in three patients. In two, P.D. (Coltheart, 1980b) and F.M. (Gordon et al., 1987), visual errors occurred on less concrete words than did semantic errors, while in G.R. (Barry & Richardson, 1988) there was no significant difference. Finally, in two patients visual errors occurred significantly more often for stimuli less than a certain level of concreteness by comparison with more concrete stimuli (K.F. (Shallice & Warrington, 1980)

[1] The research described in this chapter was done in collaboration with Tim Shallice. A more condensed description of the major results can be found in Plaut & Shallice (1991b).

$C < 6$ vs. $C > 6$; P.S. (Shallice & Coughlan, 1980) $C < 4.6$ vs. $C > 4.6$). Thus a semantic variable—concreteness—clearly influences the nature of *visual* errors.

There is a single known exception to the advantage for concrete words shown by deep dyslexics: patient C.A.V. with "concrete word dyslexia" (Warrington, 1981). C.A.V. failed to read concrete words like MILK and TREE but succeeded at highly abstract words such as APPLAUSE, EVIDENCE, and INFERIOR. Overall, abstract words were more likely to be correctly read than concrete (55% vs. 36%). In complementary fashion, 63% of his visual error responses were more abstract than the stimulus. However, the incidence of visual errors was approximately equal for words above and below the median in concreteness. While C.A.V. made no more semantic errors than might be expected by chance (see Ellis & Marshall, 1978), he appeared to be relying at least in part on the semantic route because his performance improved when given a word's semantic category. C.A.V. is clearly a very unusual patient, but any account of the relation between visual errors and concreteness can hardly ignore him.

6.2 A semantic representation for concrete and abstract words

The type of semantic feature representation used by H&S is quite similar to that frequently employed in psychological theorizing on semantic memory (e.g. Smith et al., 1974; Smith & Medin, 1981). More complex representations, such as frames (Minsky, 1975), can be implemented using this approach if units can represent a conjunction of a role and a property of what fills it (Derthick, 1988; Hinton, 1981a). More critically for the present purpose, there is a natural extension to the problem of the effect of imageability. Jones (1985) has argued that words vary greatly in the ease with which predicates about them can be generated, and that this measure reflects a psychologically important property of semantic representation. For example, more predicates can be generated for basic-level words than for subordinate or superordinate words (Rosch et al., 1976). Jones showed that there is a very high correlation (0.88) between a measure of ease-of-predication and imageability, and that the relative difficulty of parts-of-speech in deep dyslexia maps perfectly onto their ordered mean ease-of-predication scores. He argued that the effects of both imageability and part-of-speech in deep dyslexia can be accounted for by assuming that the semantic route is sensitive to ease-of-predication. Within the present framework, the natural way to realize this distinction is by representing the semantics of concrete and abstract words in terms of differing numbers of features.

To examine the effect of concreteness on visual errors, a set of 20 abstract and 20 concrete words were chosen such that each pair of words differed by a single letter (see Table 6.1). We represented the semantics of each of these words in terms of 98 semantic features, as shown in Table 6.2. Sixty-seven of these are based on the H&S semantic features for concrete words (e.g. *main-shape-3d, found-woods, living*) with minor changes being made to accommodate the different

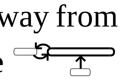
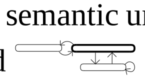
TART	TACT	GRIN	GAIN	FLAN	PLAN	REED	NEED
TENT	RENT	LOCK	LACK	HIND	HINT	LOON	LOAN
FACE	FACT	ROPE	ROLE	WAVE	WAGE	CASE	EASE
DEER	DEED	HARE	HIRE	FLEA	PLEA	FLAG	FLAW
COAT	COST	LASS	LOSS	STAR	STAY	POST	PAST

Table 6.1: The 40 words used in the simulation, consisting of 20 concrete-abstract pairs of words differing by a single letter.

range of meanings in this word set. The 31 additional features (e.g. *has-duration, relates-location, quality-difficulty*) are required to make distinctions among abstract words, but occasionally apply to concrete words as well. Figure 6.1 displays the assignment of semantic features to words. Concrete and abstract words differ systematically in their semantic representations: concrete words have an average of 18.2 features while abstract words have an average of only 4.7 features. The similarity matrix among semantic representations, shown in Figure 6.2, clearly illustrates how there is a range of similarities among concrete words and among abstract words, but very little similarity between these two groups of words. We do not claim that this representation adequately captures the richness and subtlety of the true meanings of any of these words. Rather, we claim that it captures important qualitative distinctions about the relationships *between* word meanings—namely, that similar words (e.g. LACK and LOSS) have similar representations, and that there is a systematic difference between the semantics of concrete and abstract words that reflects their relative ease of predication.

A network that maps from orthography to phonology via semantics will be developed incrementally, as for the networks described in Chapter 4. An "input" network, analogous to the H&S model, will be trained to map from orthography to semantics. A similarly structured "output" network will be trained separately to map from semantics to phonology. These two networks will then be combined into the complete network, shown in Figure 6.3.

6.3 Mapping orthography to semantics

The task of the input network is to generate the semantics of each word from its orthography. Orthography is represented using the same eight feature distributed code used previously (see Figure 4.1, p. 75). The architecture of the input network, shown in the bottom half of Figure 6.3, is broadly similar to the H&S network except that it has (a) full rather than partial (25%) connectivity density, (b) fewer intermediate units (10 vs. 40) and clean-up units (10 vs. 60), (c) no interconnections among semantic units, and (d) a feedback pathway from the semantic units to the intermediate units. In this sense it is something of a hybrid of the  and  networks. The general

Semantic features					
1	max-size-less-foot	35	found-in-transport	68	positive
2	max-size-foot-to-two-yards	36	found-in-factories	69	negative
3	max-size-greater-two-yards	37	surface-of-body	70	no-magnitude
4	main-shape-1D	38	above-waist	71	small
5	main-shape-2D	39	natural	72	large
6	main-shape-3D	40	mammal	73	measurement
7	cross-section-rectangular	41	bird	74	superordinate
8	cross-section-circular	42	wild	75	true
9	cross-section-other	43	does-fly	76	fiction
10	has-legs	44	does-swim	77	information
11	has-arms	45	does-run	78	action
12	has-neck-or-collar	46	living	79	state
13	white	47	carnivore	80	has-duration
14	brown	48	plant	81	unchanging
15	color-other-strong	49	made-of-metal	82	involves-change
16	varied-colors	50	made-of-liquid	83	temporary
17	dark	51	made-of-other-nonliving	84	time-before
18	hard	52	got-from-plants	85	future-potential
19	soft	53	got-from-animals	86	relates-event
20	sweet	54	pleasant	87	relates-location
21	moves	55	unpleasant	88	relates-money
22	indoors	56	dangerous	89	relates-possession
23	in-kitchen	57	man-made	90	relates-work
24	on-ground	58	container	91	relates-power
25	on-surface	59	for-eating-drinking	92	relates-reciprocation
26	otherwise-supported	60	for-wearing	93	relates-request
27	outdoors-in-city	61	for-other	94	relates-interpersonal
28	in-country	62	for-lunch-dinner	95	quality-difficulty
29	found-woods	63	particularly-assoc-child	96	quality-organized
30	found-near-sea	64	particularly-assoc-adult	97	quality-bravery
31	found-near-streams	65	used-for-games-or-recreation	98	quality-sensitivity
32	found-mountains	66	human		
33	found-on-farms	67	female		
34	found-in-public-buildings				

Table 6.2: The 98 semantic features and their assignment to the concrete and abstract words. Features 1–67 are based on the semantic features used by H&S. Features 68–98 are additional features required to make distinctions among abstract words. The ordering of the features, and in particular, the separation of concrete and abstract features, is irrelevant to the operation of the network.

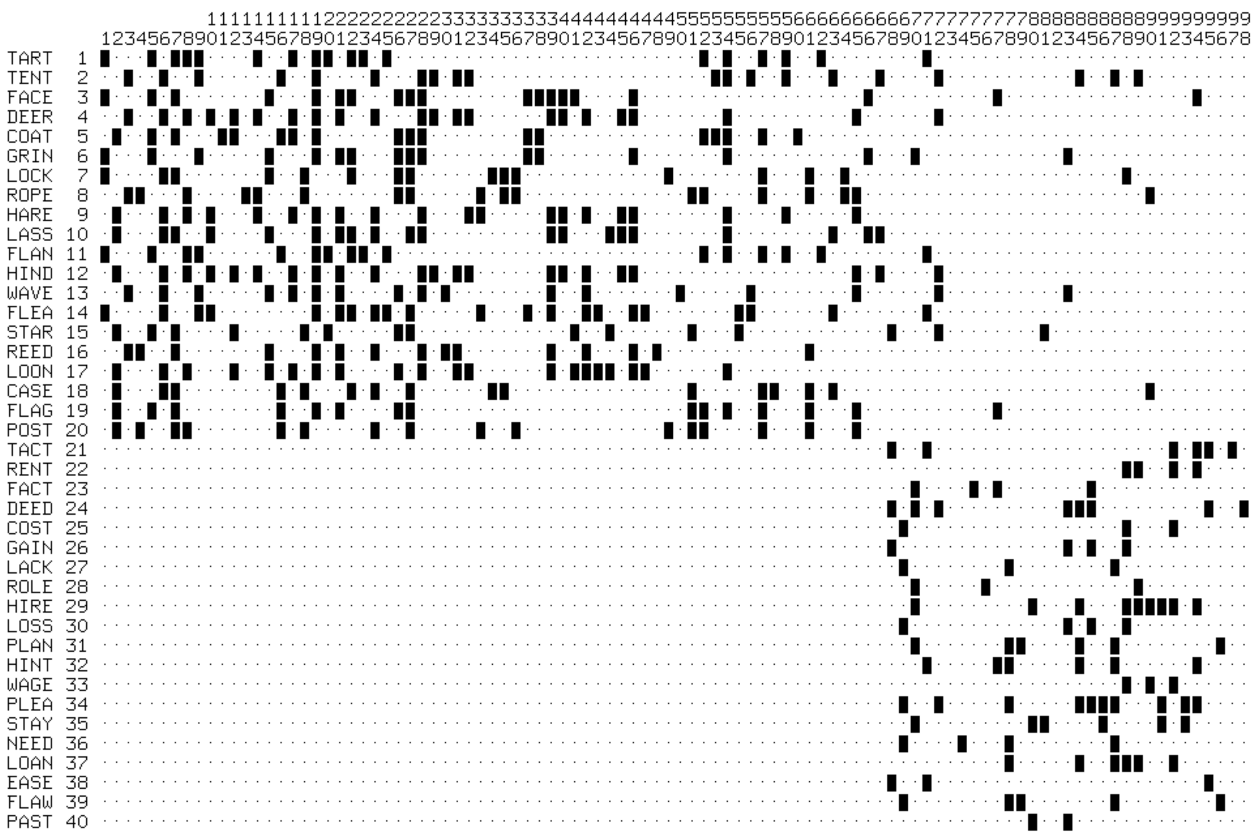


Figure 6.1: The assignment of semantic features to words.

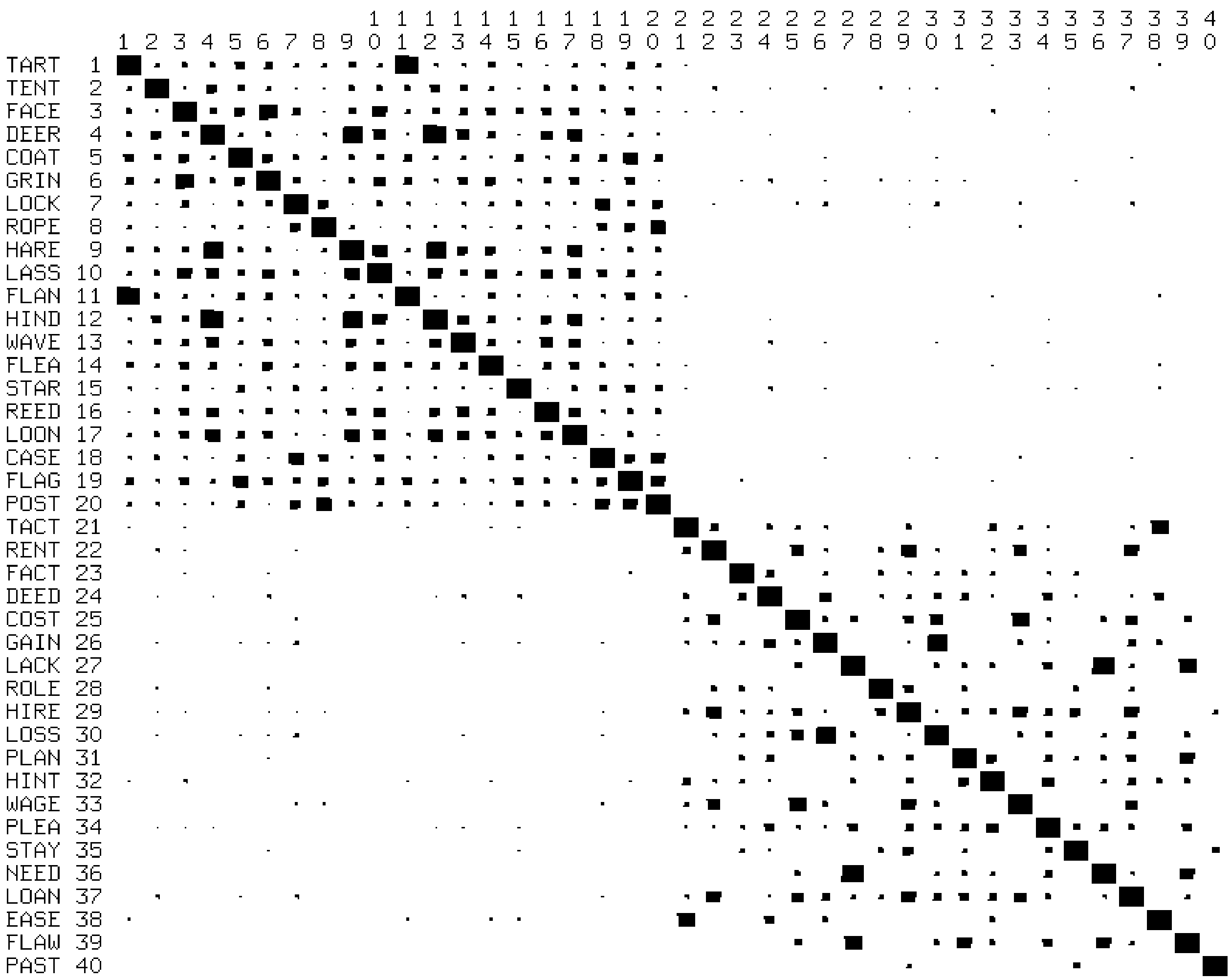


Figure 6.2: The similarity matrix for the semantic representations of words.

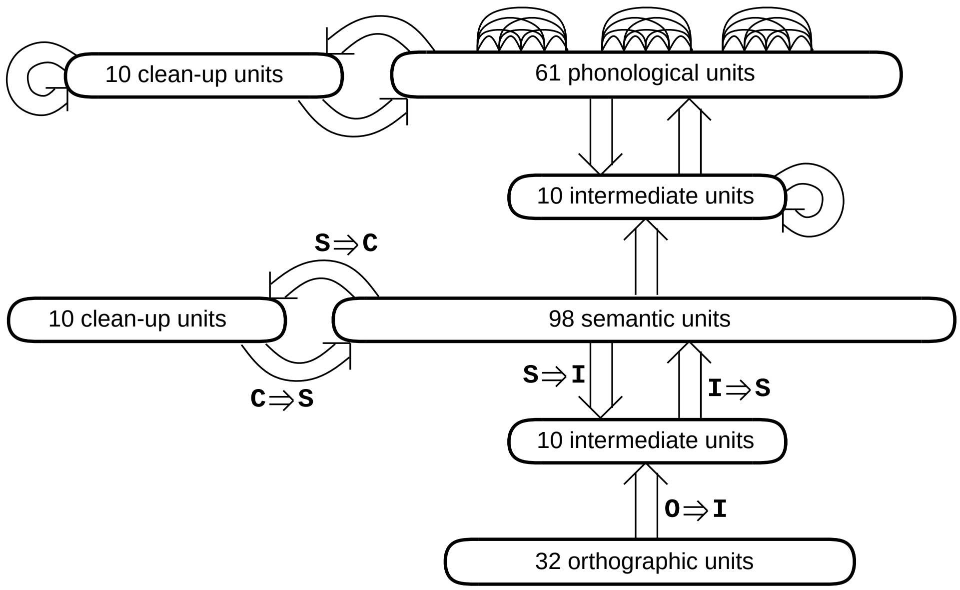


Figure 6.3: The network for mapping orthography to phonology via semantics. The additional recurrent connections at the intermediate and clean-up layers in the output network were intended to facilitate the development of strong phonological attractors.

motivation for these changes was to encourage the network to develop stronger semantic attractors while keeping the number of connections reasonable.

The input network was trained with back-propagation to activate the appropriate semantic units for a word when presented with the word's orthography corrupted by independent gaussian noise with mean 0.0 and standard deviation 0.1. After 4700 sweeps through the training set, the state of each semantic unit was accurate to within 0.1 over the last three of eight iterations for each word.

6.4 Mapping semantics to phonology

The introduction to Chapter 3 presents a number of reasons why is is important to develop an output network to replace the H&S response criteria. The central concern in that chapter was on demonstrating the validity of the criteria as approximations to the behavior of an actual output network. An even more pressing issue for the present purposes is that the criteria are insensitive to the relative semantic and phonological discriminability of words. Any differences found in performance on concrete and abstract words might simply be due to an inherent bias of the response criteria. For this reason, it is essential that we develop a phonological output network that is equally effective for concrete and abstract words under normal operation. We are then guaranteed that systematic differences observed under damage are due to properties of the network rather than properties of an external procedure for interpreting the output.

The word set requires a somewhat more complicated phonological representation than the one used for the H&S word set. Phonology is represented in terms of seven sets of position-specific, mutually-exclusive phoneme units. These groups consist of three slots for phonemes from the initial (onset) consonant cluster, one slot for the vowel, and three slots for phonemes from the final (coda) consonant cluster. Table 6.3 shows the allowable phonemes for each slot, and the resulting phonological representation for each word. Each of the six consonant slots includes a unit for the "null" phoneme in order to explicitly represent the absence of any phoneme at that slot in the pronunciation of a word. As a result, the representation of every word has exactly one active unit in each slot. A total of 66 phoneme units are required to represent the pronunciations of all 40 words.

The task of the output network is to generate the phonological representation of each word from its semantic representation. The architecture of this network, shown in the top half of Figure 6.3, was designed to facilitate the development of strong phonological attractors. Each major pathway shown has full connectivity density, and phoneme units in the same consonant (or vowel) cluster are fully interconnected. This connectivity allows units within a slot to develop a "winner-take-all" strategy while still cooperating with units in other slots within the same cluster. Coordination and competition between clusters can only be accomplished via the clean-up units.

In order to minimize the number of blends produced under damage, the output network was trained in a way that maximizes the strength of the attractors it develops—no attempt was made to simulate the development or mode of operation of the human speech production system. Specifically, the "direct" pathway (from semantics to phonology) was trained to produce the correct phonemes of each word during the last two of five iterations when presented with its semantics corrupted by gaussian noise with standard deviation 0.1. After about 3000 sweeps through the training set, the activity of each phoneme unit was accurate to within 0.2 of its correct value for each word. At this point, intra-phoneme connections and the clean-up pathway were added and the amount of input noise was increased to 0.2. In this way the clean-up pathway learned to compensate for the limitations of the direct pathway when pressed by severely corrupted input.[2] The network was trained to produce the correct phonemes over the last three of eight iterations to within 0.1 of their correct values. The amount of noise prevented the network from achieving this criterion consistently, and after 18,000 training sweeps performance had ceased to improve. However, the network easily satisfied the criterion for every word given uncorrupted input.

The output network was then combined with the input network to produce a network that maps from orthography to phonology via semantics. In order to ensure that the output network would operate appropriately with its input generated by the input network, the complete network was given additional training at generating the correct phonology of each word over the last three of 14 iterations when given the uncorrupted orthography of the word. The weights of the input network

[2]This procedure is slightly different than the one used to train the phonological output networks for the original H&S stimuli (see Section 3.2.2), in which the direct and clean-up pathways were trained separately and then combined.

Phonemes allowed in each position
`s -`
`b ch d dy f g h k m n p sh t v z -`
`l r w y -`
`a ai air ar aw e ee eer ew i ie ire o oa ow u uu`
`l m n s -`
`b d j f g k p sh t v z -`
`s t z -`

Phonological representation of each word							
TART	`/-t-`	`ar`	`-t-/`	TACT	`/-t-`	`a`	`-kt/`
TENT	`/-t-`	`e`	`nt-/`	RENT	`/--r`	`e`	`nt-/`
FACE	`/-f-`	`ai`	`s--/`	FACT	`/-f-`	`a`	`-kt/`
DEER	`/-d-`	`eer`	`---/`	DEED	`/-d-`	`ee`	`-d-/`
COAT	`/-k-`	`oa`	`-t-/`	COST	`/-k-`	`o`	`st-/`
GRIN	`/-gr`	`i`	`n--/`	GAIN	`/-g-`	`ai`	`n--/`
LOCK	`/--l`	`o`	`-k-/`	LACK	`/--l`	`a`	`-k-/`
ROPE	`/--r`	`oa`	`-p-/`	ROLE	`/--r`	`oa`	`l--/`
HARE	`/-h-`	`air`	`---/`	HIRE	`/-h-`	`ire`	`---/`
LASS	`/--l`	`a`	`s--/`	LOSS	`/--l`	`o`	`s--/`
FLAN	`/-fl`	`a`	`n--/`	PLAN	`/-pl`	`a`	`n--/`
HIND	`/-h-`	`ie`	`nd-/`	HINT	`/-h-`	`i`	`nt-/`
WAVE	`/--w`	`ai`	`-v-/`	WAGE	`/--w`	`ai`	`-j-/`
FLEA	`/-fl`	`ee`	`---/`	PLEA	`/-pl`	`ee`	`---/`
STAR	`/st-`	`ar`	`---/`	STAY	`/st-`	`ai`	`---/`
REED	`/--r`	`ee`	`-d-/`	NEED	`/-n-`	`ee`	`-d-/`
LOON	`/--l`	`ew`	`n--/`	LOAN	`/--l`	`oa`	`n--/`
CASE	`/-k-`	`ai`	`s--/`	EASE	`/---`	`ee`	`z--/`
FLAG	`/-fl`	`a`	`-g-/`	FLAW	`/-fl`	`aw`	`---/`
POST	`/-p-`	`oa`	`st-/`	PAST	`/-p-`	`a`	`st-/`

Table 6.3: The phonemes allowed in each position, and their assignment to words. The definitions are based on British rather than American pronunciations. In the top table, each of the seven rows constitutes a set of mutually-exclusive phonemes, and each of the three blocks represents a consonant (or vowel) cluster. The letter(s) used to represent phonemes are not from a standard phonemic alphabet but rather are intended to have more intuitive pronunciations. A “`-`” stands for the “null” phoneme.

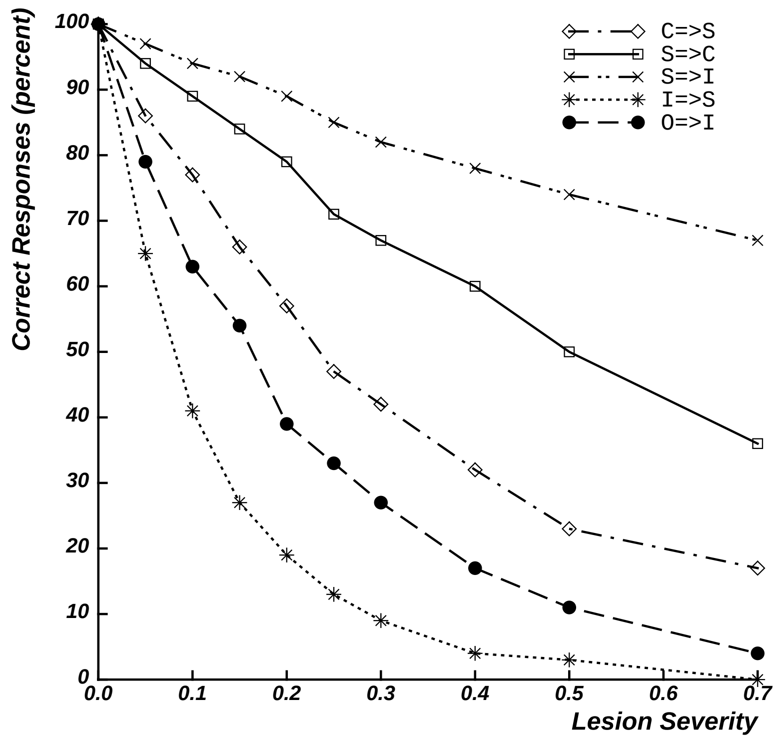


Figure 6.4: Overall rates of correct response for lesions of increasing severity to each of the five main sets of connections in the input network.

were not allowed to change during training to ensure that it continued to generate the correct semantics of each word. This final training required less than 100 sweeps through the words.

6.5 The effects of lesions

After training, the complete network successfully derives the semantics and phonology of each word when presented with its orthography. Each of the five main sets of connections in the input network was subjected to lesions of a wide range of severity, in which a proportion of the connections were chosen at random and removed. Fifty instances of each location and severity of lesion were carried out, and correct, omission, and error responses were accumulated using a criterion of 0.6 for the minimum phoneme response probability, as described in Section 3.1.4. Figure 6.4 shows the overall correct performance of the network as a function of lesion severity. In general, damage to the direct pathway (O⇒I and I⇒S) is more debilitating than damage to the clean-up pathway (S⇒C and C⇒S). Figure 6.5 shows the same data separately for concrete and abstract words. Comparing the two, clean-up lesions impairs performance on concrete words more than abstract words, while the opposite is true for lesions to the direct pathway. In fact, abstract words appear to be particularly sensitive to I⇒S lesions, showing quite severe impairment even with slight amounts of damage.

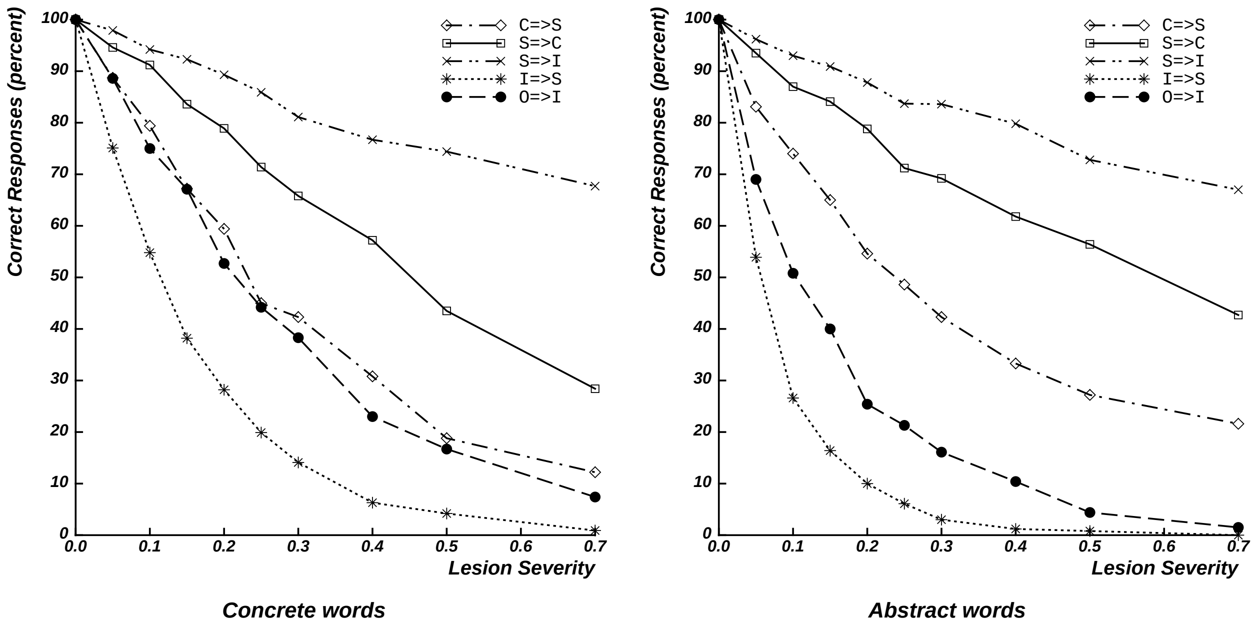


Figure 6.5: Correct performance on concrete (left) and abstract (right) words for each lesion location as a function of lesion severity.

In the following analyses we include data only from lesions producing overall correct performance between 15–85%. We used a slightly wider range of correct performance for including lesions than in previous experiments (20–80%) because some of the phenomena we are interested in arise specifically in cases of severe impairment. Considering correct responses to concrete and abstract words separately, there is a significant advantage for concrete words (52.7% correct) over abstract words (45.0% correct, $F(1,2598) = 62.4$, $p < .001$). For a given lesion location and severity, we define the relative difference in correct performance between concrete and abstract words to be $(C - A)/(C + A)$, where C and A are the number of correct responses to concrete and abstract words, respectively. This measure can range from ± 1—positive values reflect superior performance on concrete words relative to abstract words. Figure 6.6 displays the relative difference in correct performance between these two sets of words as a function of the overall level of incorrect performance produced by each lesion location and severity. Two main results are apparent from the figure. The first is that the advantage for concrete over abstract words overall arises almost entirely from lesions to the direct pathway, where the majority (82.7%) of errors are produced. The second, unexpected result is that severe lesions of the clean-up pathway, producing the lowest levels of overall correct performance, result in the reverse advantage—abstract words are responded to more accurately than concrete words ($F(1,49) > 22$, $p < .001$ for each of S⇒C(0.5,0.7) and C⇒S(0.5,0.7)). This type of lesion and pattern of performance are consistent with what is known about the concrete word dyslexic, C.A.V. (Warrington, 1981). His reading disorder was quite severe initially, and he also showed an advantage for abstract words in picture-word matching with auditory presentation, suggesting modality-independent damage at the level of the

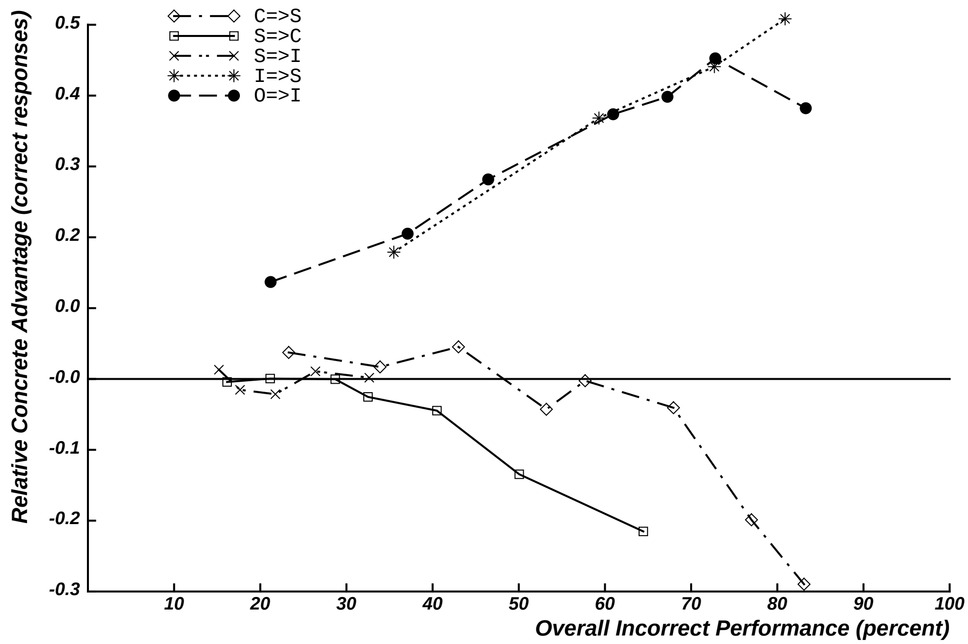


Figure 6.6: Relative difference in correct performance between concrete and abstract words as a function of overall incorrect performance, for lesion locations and severities producing overall correct performance between 15–85%. The data are plotted in terms of incorrect rather than correct performance to be consistent with data plotted as a function of lesion severity.

semantic system.

Figure 6.7 presents the correct performance on individual words after lesions to the direct pathway, or after severe lesions (i.e. 0.5 and 0.7) to the clean-up pathway. Lesions to the direct pathway affect words fairly evenly, with abstract words being consistently worse than concrete words. Severe lesions to the clean-up pathway produce a wider range of performance across words. Performance on most concrete words is quite impaired although a few (e.g. WAVE and REED) are much better than the rest. Many abstract words are also impaired under these conditions, but a larger number of them retain a reasonable level of performance than for concrete words. In fact, 35% of the abstract words (7/20) account for over 60% of the total correct responses. Apparently, the advantage for abstract words after severe clean-up lesions is due to a fairly uniform impairment of concrete words combined with the selective preservation of a relatively small subset of abstract words.

Analyzing error responses, we tested whether responses tend to be more concrete than stimuli by counting how often a stimulus and response were of the opposite type. Overall, abstract words are over twice as likely to produce a concrete response than *vice versa* (33.4% vs. 15.6% of total errors, $F(1, 2598) = 53.9$, $p < .001$). *Post hoc* analyses for each lesion location and severity showed a similar pattern as for correct performance: a tendency for responses to be more concrete for all lesions within the direct pathway, but the opposite tendency for severe lesions within the

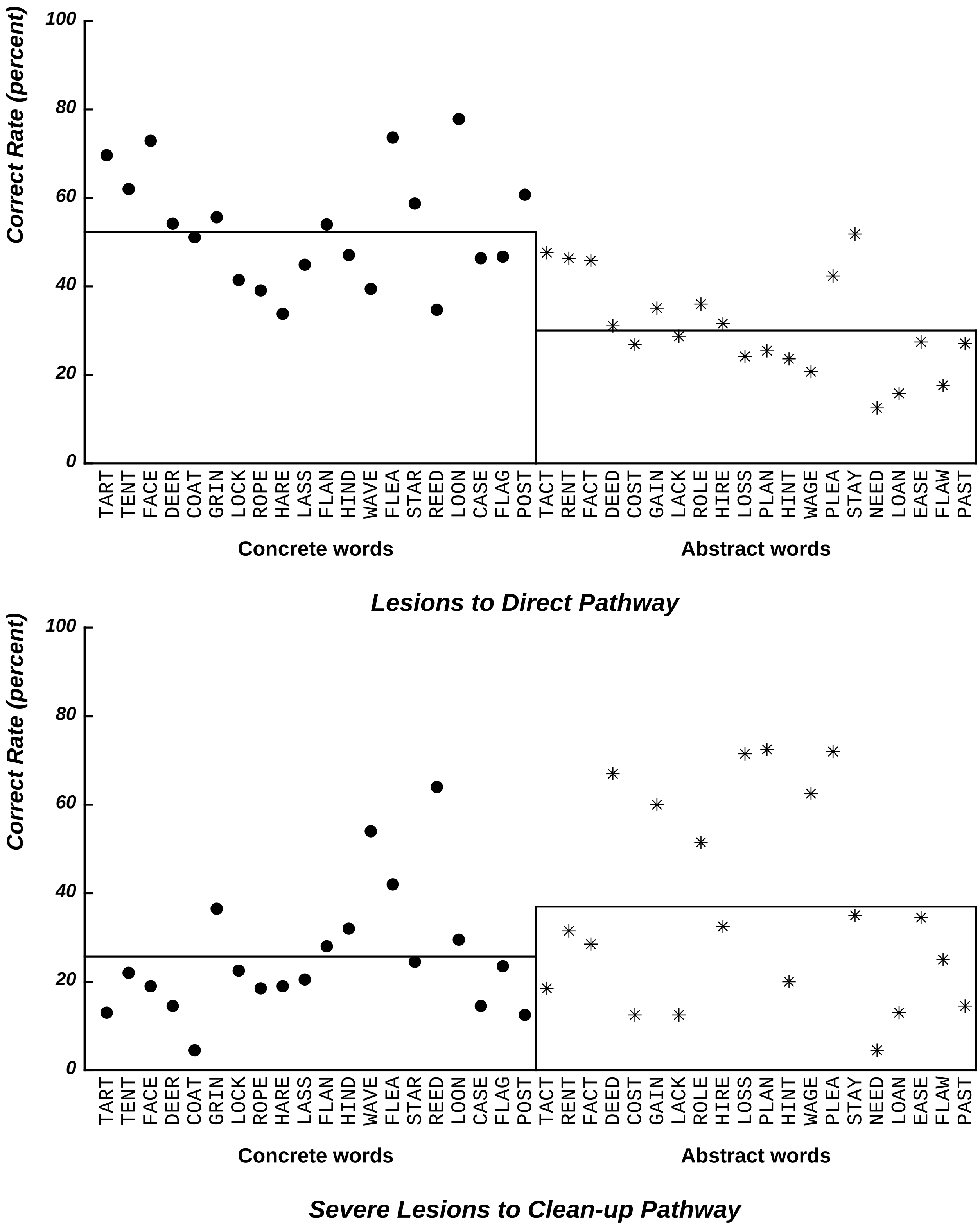


Figure 6.7: Correct performance for individual words after lesions to the direct pathway (top) and after severe lesions of the clean-up pathway.

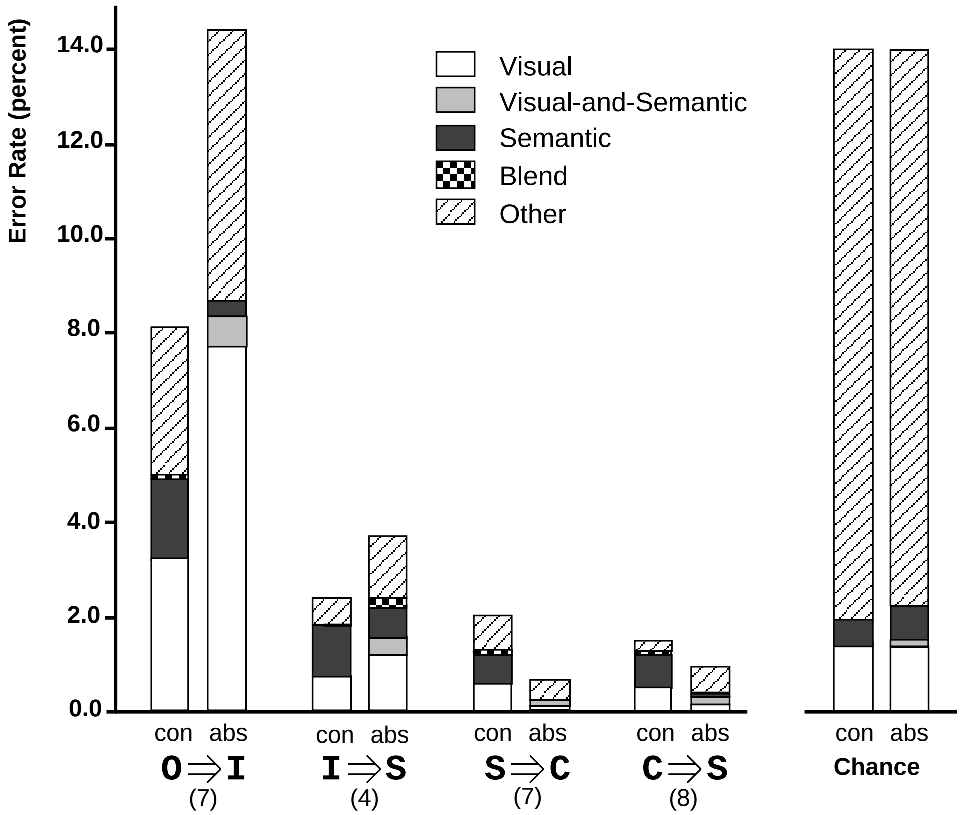


Figure 6.8: Overall rates of each error type for concrete (con) and abstract (abs) words for each lesion location (except S⇒I lesions which produce virtually no explicit errors).

semantic clean-up pathway.

Error responses were categorized in terms of their visual and semantic similarity to the stimulus. Words were considered visually similar if they overlapped in two or more letters, and semantically similar if their semantic representations overlapped by at least 84% for concrete words and 95% for abstract words. The definition of semantic similarity is more complicated because of the systematic differences between concrete and abstract semantics and because the semantic representations are not organized into categories as in the H&S simulations. Note that two typical unrelated words have roughly 67% overlap if both are concrete and 91% if both are abstract. Thus the values of the semantic relatedness criteria for concrete and abstract words are each approximately half way between the corresponding expected value for unrelated word pairs of the same type and 100%.

Figure 6.8 shows the rates of each error type produced by each lesion location, for concrete and abstract words separately. Also included in the figure is the distributions of each error type for "chance" error responses to chosen randomly from the word set in response to concrete or abstract stimuli. Notice that the criteria for visual and semantic relatedness are quite stringent—almost 85% of all possible stimulus-response pairs are unrelated. One consequence of this is that only four of the 190 pairs of abstract words are both visually and semantically related, and *none* of the concrete pairs are. Thus concrete words cannot produce mixed visual-and-semantic errors. Nonetheless, when errors to concrete and abstract words are taken together, the ratios of the rates of each error type with that of "other" errors is at least four times the chance value for every lesion location. In fact, this also holds for each word set separately, except for visual errors to abstract words produced by clean-up lesions, where the ratios are only about twice the chance value, and for `S⇒C` lesions which produced no semantic errors to abstract words. Also, the rates of mixed visual-and-semantic errors among the abstract words for all lesion locations are at least three times the rates expected from the independent rates of visual and semantic errors. Thus, the network replicates (on a different word set) the H&S finding of mixtures of error types for lesions throughout the network, including purely visual errors for lesions entirely within the semantic clean-up system. In addition, as with the networks trained on the original H&S word set, a number of the "other" errors are actually of the visual-then-semantic type found in deep dyslexia (e.g. PLAN ⇒ (flan) ⇒ "tart").

A comparison of error types for concrete and abstract words revealed that the proportion of errors which are visual is higher for abstract words (41.4% vs. 36.4%, $F(1, 1036) = 3.95$, $p < .05$), while the proportion of errors which are semantic is higher for concrete words (32.3% vs. 6.4%, $F(1, 1036) = 155.1$, $p < .001$). This effect is most clearly shown in Figure 6.8 for lesions of the direct pathway. As a measure of the "abstractness" of the errors produced by a lesion, we used the number of errors to abstract words minus the number of errors to concrete words. Applying this measure to visual and semantic errors separately reveals that visual errors are more abstract than semantic errors (means 0.201 vs. -0.161 per lesion, $(F(1, 2598) = 85.0$, $p < .001$). Finally, for each pair of visually similar words of contrasting types (e.g. TART and TACT), we compared how often each word produced the other as an error. Overall, abstract words are more likely to produce the paired visually similar concrete word as an error than *vice versa* (13.1% vs. 6.2% of total errors, Wilcoxon signed-ranks test $n = 520$, $Z = 3.24$, $p < .001$). Considering lesions to the direct and clean-up pathways separately, the effect is quite pronounced for the direct pathway (15.6% abs vs. 3.9% con, $n = 220$, $Z = 6.16$, $p < .001$) while lesions of the clean-up pathway produce the opposite effect (0.0% abs vs. 23.8% con, $n = 300$, $Z = 1.83$, $p < .05$).

To provide a further comparison of the effects of lesions to the direct pathway vs. severe lesions of the clean-up pathway, Figure 6.9 presents the confusion matrix for errors produced after lesions of the former type, while Figure 6.10 presents the same information for the latter type of lesion. Considering direct-pathway lesions first, the advantage in correct performance for concrete over abstract words is clearly reflected in the size of the squares along the main diagonal for these two sets of words. These values are exactly those plotted in the top of Figure 6.7. The tendency for abstract words to produce concrete responses is shown in the greater frequency of errors in the lower left quadrant of the matrix compared with the other quadrants. The most frequent errors among these are along the diagonal band halfway below the main diagonal—these error responses

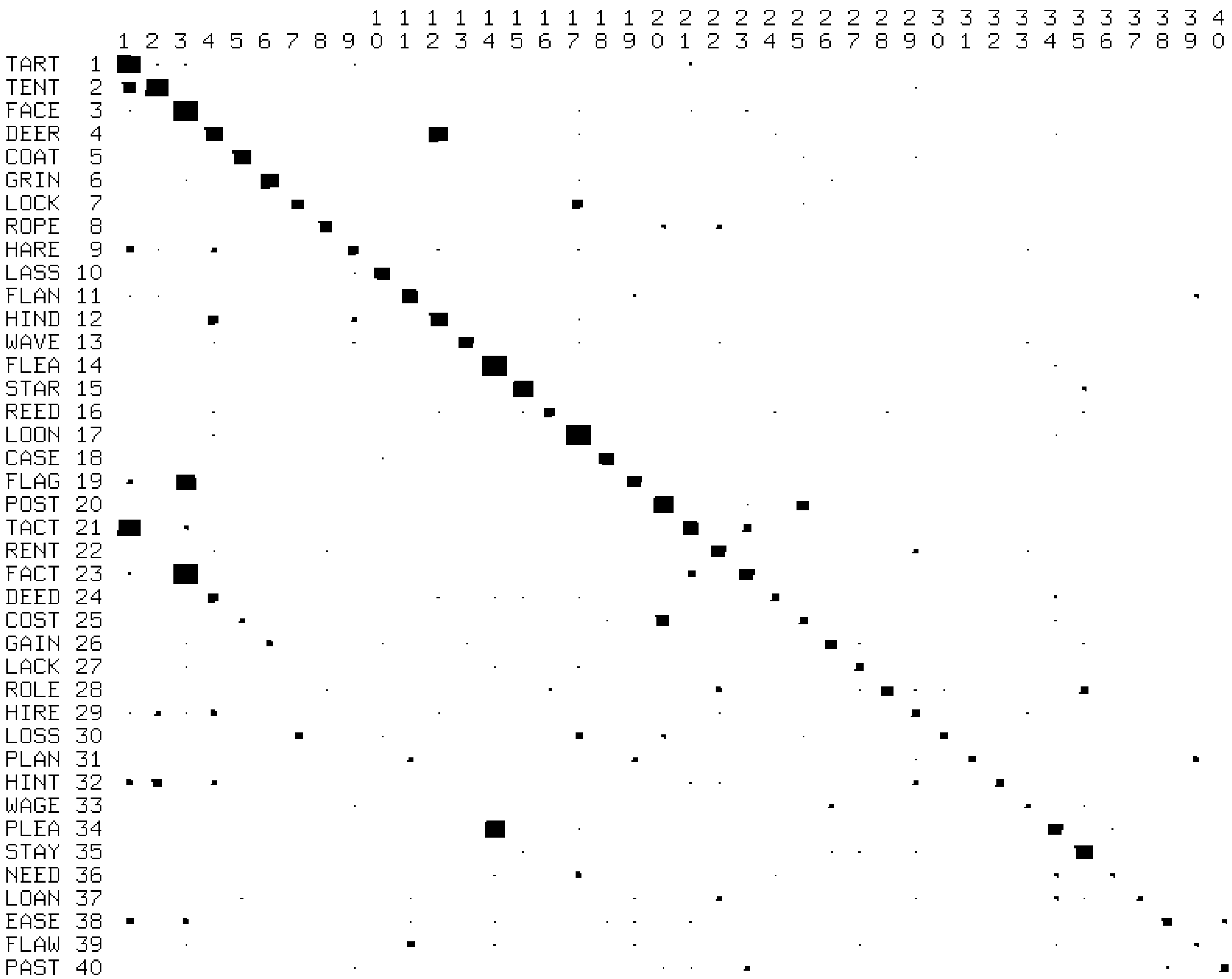


Figure 6.9: The confusion matrix for errors produced by lesions to the direct pathway.

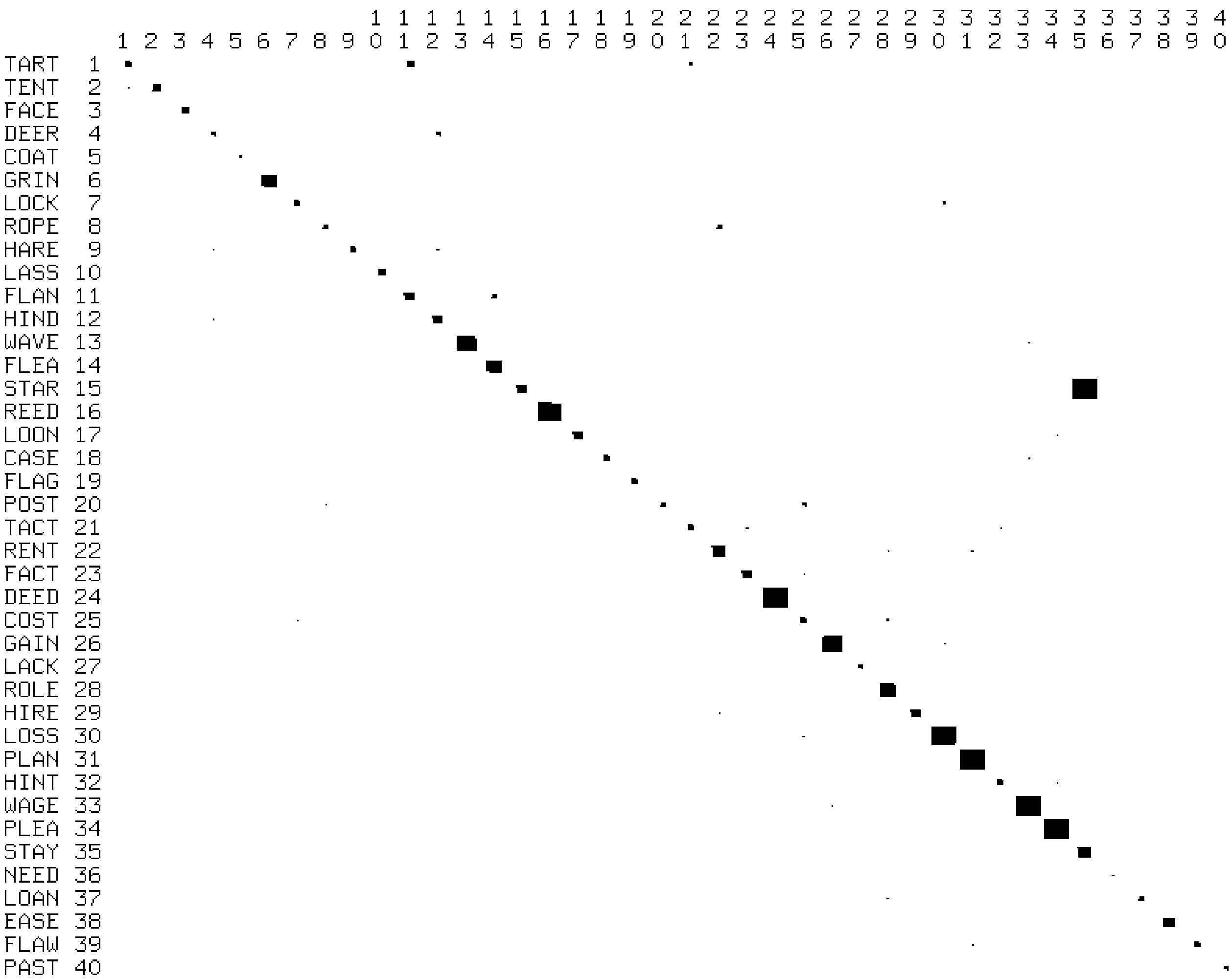


Figure 6.10: The confusion matrix for errors produced by severe lesions to the clean-up pathway.

correspond to the paired visually similar concrete word (e.g. FACT ⇒ "face").

The confusion matrix for errors following severe clean-up lesions looks quite different. The diagonal values, representing correct performance, are also plotted in the bottom of Figure 6.7. Although some abstract words are very poorly read, many others are relatively preserved, whereas most concrete words are severely impaired. A single error appears to predominate—STAR ⇒ "stay"—although a number of other concrete words also produce abstract responses. In contrast, almost no abstract words produce concrete responses under this type of damage.

Overall, the network successfully reproduces the behavior of deep dyslexics after lesions to the direct pathway, showing better correct performance for concrete over abstract words, a tendency for error responses to be more concrete than stimuli, and a higher proportion of visual errors in response to abstract compared with concrete words. In contrast, severe lesions to the clean-up pathway produce the reverse advantage for abstract words, quite similar to a patient with concrete word dyslexia.

6.6 Network analysis

The effects of abstractness on the performance of the network under damage can be understood in the following way. As abstract words have fewer semantic features, they are less effective than concrete words at engaging the semantic clean-up mechanism and must rely more heavily on the direct pathway. Concrete words are read better under lesions to this pathway because of the stronger semantic clean-up they receive. In addition, abstract words are more likely to produce visual errors as the influence of visual similarity is strongest in the direct pathway. Slight or moderate damage to the clean-up pathway impairs what little support abstract words receive from this system, but also impairs concrete words, producing no relative difference. Under severe damage to this pathway, the processing of most concrete words is impaired but many abstract words can be read solely by the direct pathway, producing an advantage of abstract over concrete words in correct performance.

In order to provide more direct evidence for this interpretation, we examined a number of aspects of the operation of the undamaged network. One measure that should be particularly informative is the similarity of concrete and abstract word representations at different times and locations in the network with their final semantic representations. One hypothesis is that, if abstract words rely more heavily on the direct pathway and less on the clean-up pathway, their representations should be more semantically organized than those of concrete words prior to the influence of semantic clean-up.

Figure 6.11 shows the similarity matrices for the intermediate layer representations at iterations 1, 3, and 5, together with their correlations with the matrices for the input (visual) and output (semantic) representations. As was the case for networks trained on the original H&S word set, the initial intermediate representations are more visually than semantically organized. For

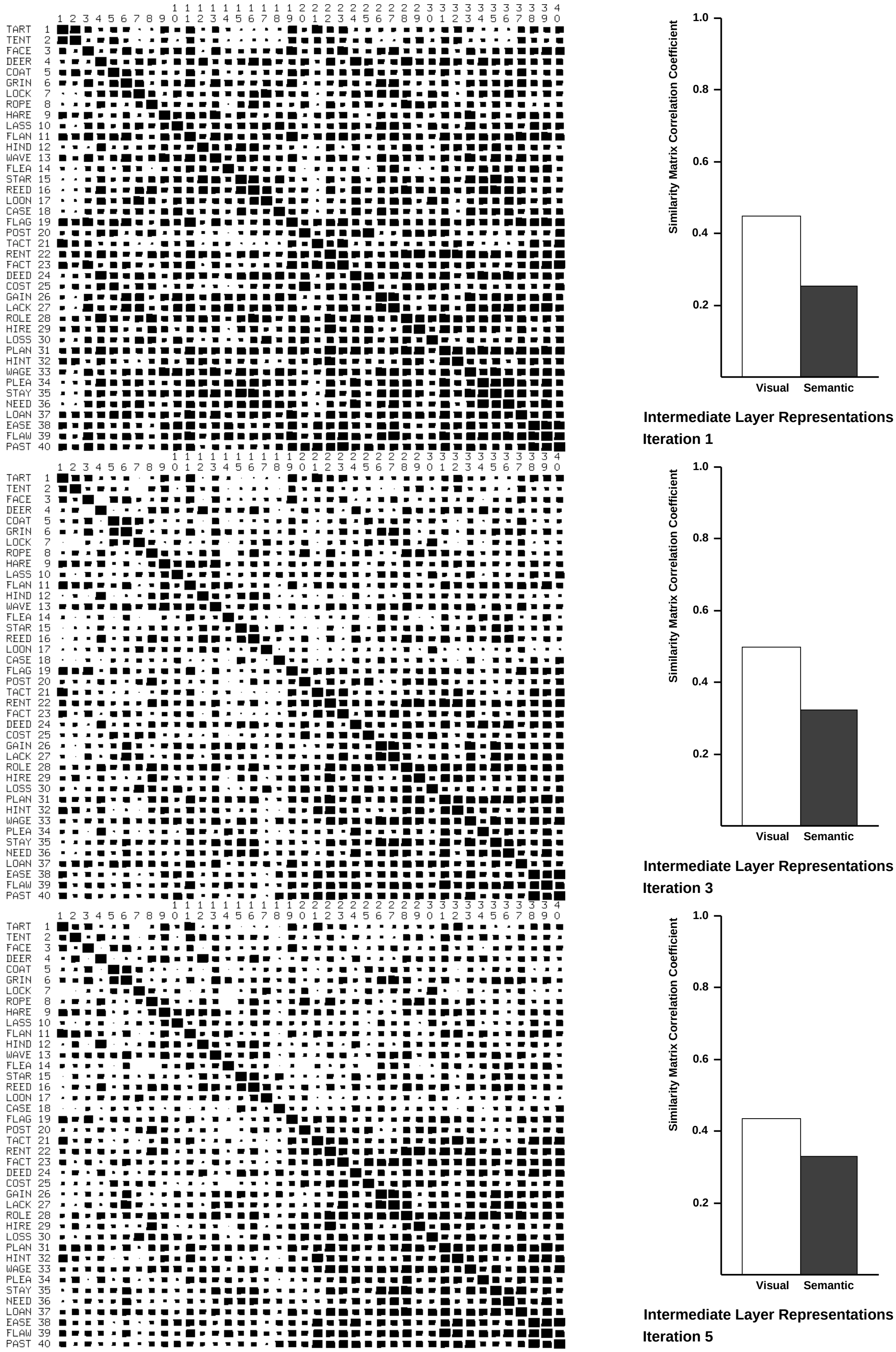


Figure 6.11: Similarity matrices and their correlation coefficients with matrices for visual (orthographic) and semantic similarity, for representations at the intermediate layer at iterations 1, 3, and 5.

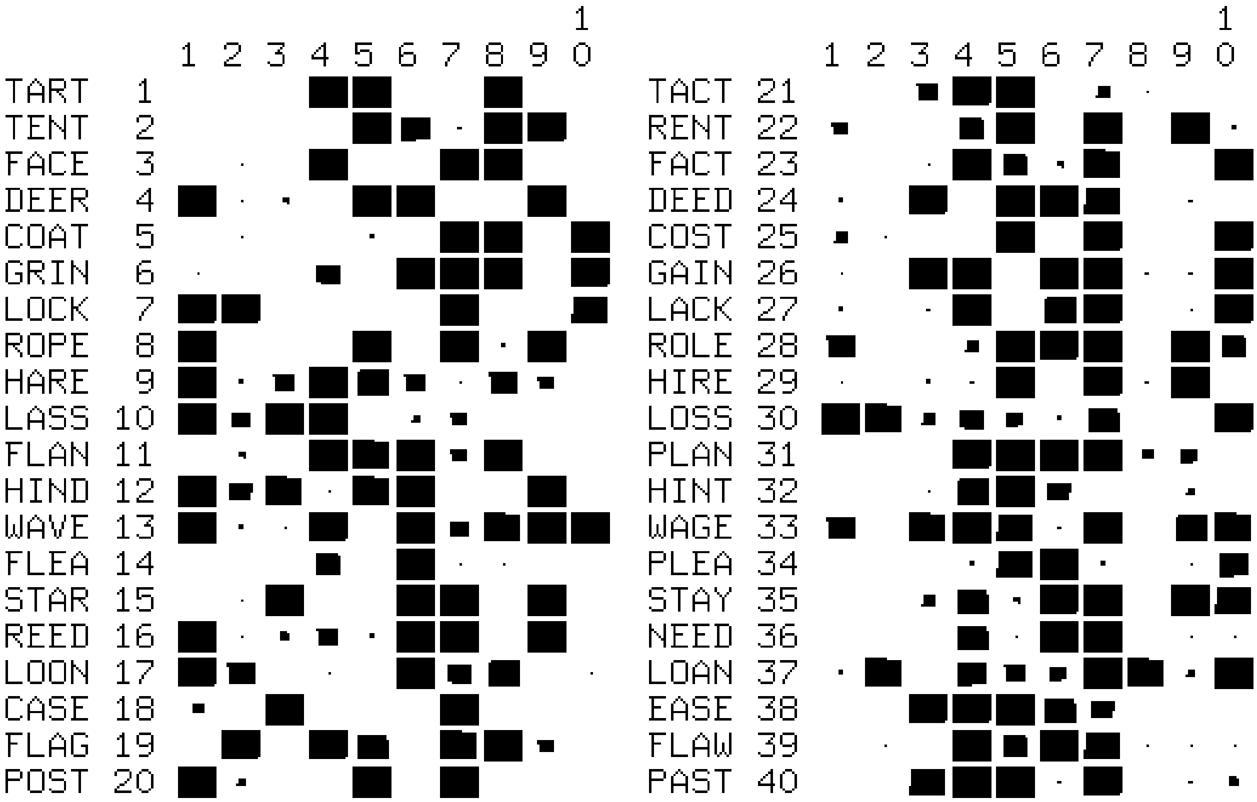


Figure 6.12: The final states of the intermediate units for concrete (left) and abstract words.

example, the visual similarity of the concrete-abstract word pairs is reflected in diagonal bands halfway above and below the main diagonal. (These may be seen more clearly by viewing the matrices along the diagonal at a sharp angle with the page.) However, even though the degree of semantic organization increases somewhat over iterations, even by iteration 5 the word representations remain more visually organized. The similarity matrix at this iteration shows a wide range of similarities among concrete words that begin to approximate their final semantic similarities (Figure 6.2, p. 157). However, the abstract words are almost uniformly similar to each other, with little differentiation among them. This difference can be seen more directly in the actual final intermediate representations for concrete and abstract words, shown in Figure 6.12. The representations for concrete words show greater variety than those for abstract words. The difference between concrete and abstract words can also be seen in the progression of visual and semantic similarity of the intermediate representations for these two types of words across iterations (see the left half of Figure 6.13). The correlations for concrete words are consistently higher than for abstract words, contrary to the hypothesis that the latter would show more semantic organization in the absence of semantic clean-up.

Figure 6.14 shows the similarity matrices for the semantic layer representations at iterations 2, 3, and 4. The similarities are much sparser overall at this layer than at the intermediate layer

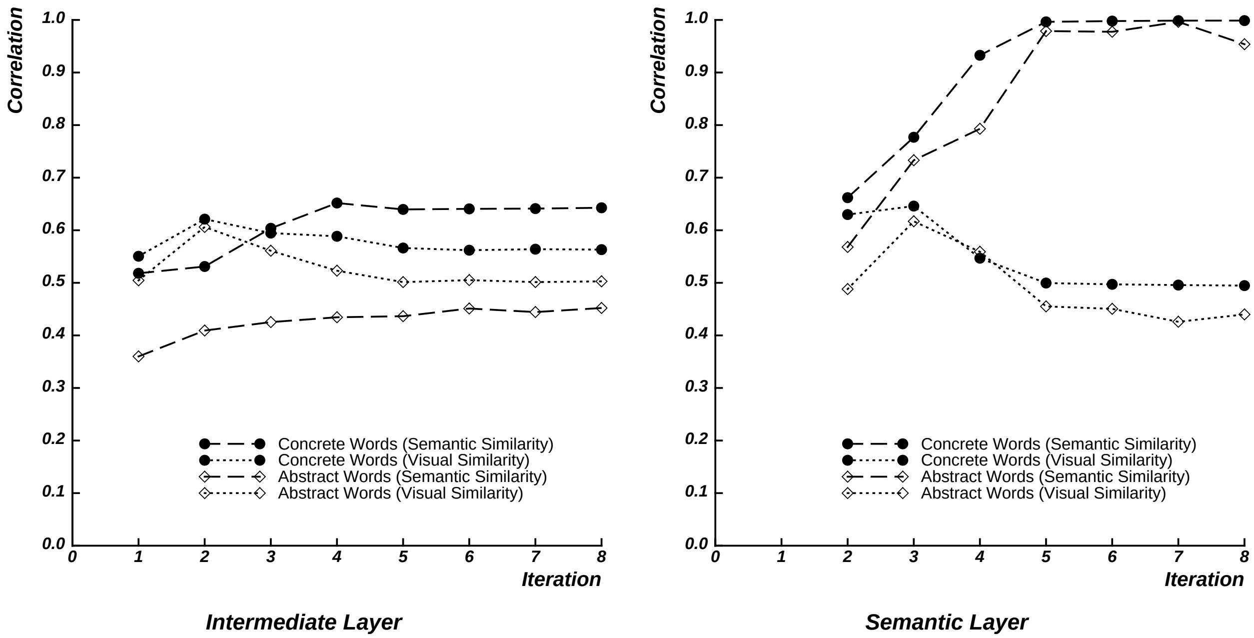


Figure 6.13: The correlations with the visual and semantic similarity matrices for the similarities among concrete and abstract word representations at the intermediate layer (left) and semantic layer (right) of the network for each iteration.

because there are almost ten times as many semantic as intermediate units—in a sense there is more room for vectors to be dissimilar. When input first arrives at iteration 2, the representations are only slightly more semantically than visually organized, but this steadily increases over iterations. Interestingly, visual similarity initially increases as well at iteration 3, but then drops. There is considerable cross-similarity between concrete and abstract words, but this is gradually eliminated. The similarity of word pairs is again evident as diagonals off the main one within the areas of cross-similarity. The right half of Figure 6.13 presents the visual and semantic correlations for the representations of concrete and abstract words separately, at the semantic layer for each iteration. Again, concrete words are consistently more semantically organized than abstract words, except at the final iterations when the representations of each set of words closely approximates their correct semantics. Concrete words appear to be more strongly influenced by the clean-up pathway because at iteration 4, when this influence first arrives, the increase in semantic similarity among concrete words increase much more than among abstract words. However, there is no evidence from these similarities that the *direct* pathway is somehow more effective for abstract words than concrete words.

There is, however, further evidence that the clean-up pathway is particularly important in processing concrete words. Figure 6.15 presents the final clean-up representations of each word, with concrete words on the left and abstract words on the right. The representations for concrete words are far more "binary" than those for abstract words. When processing a concrete word,

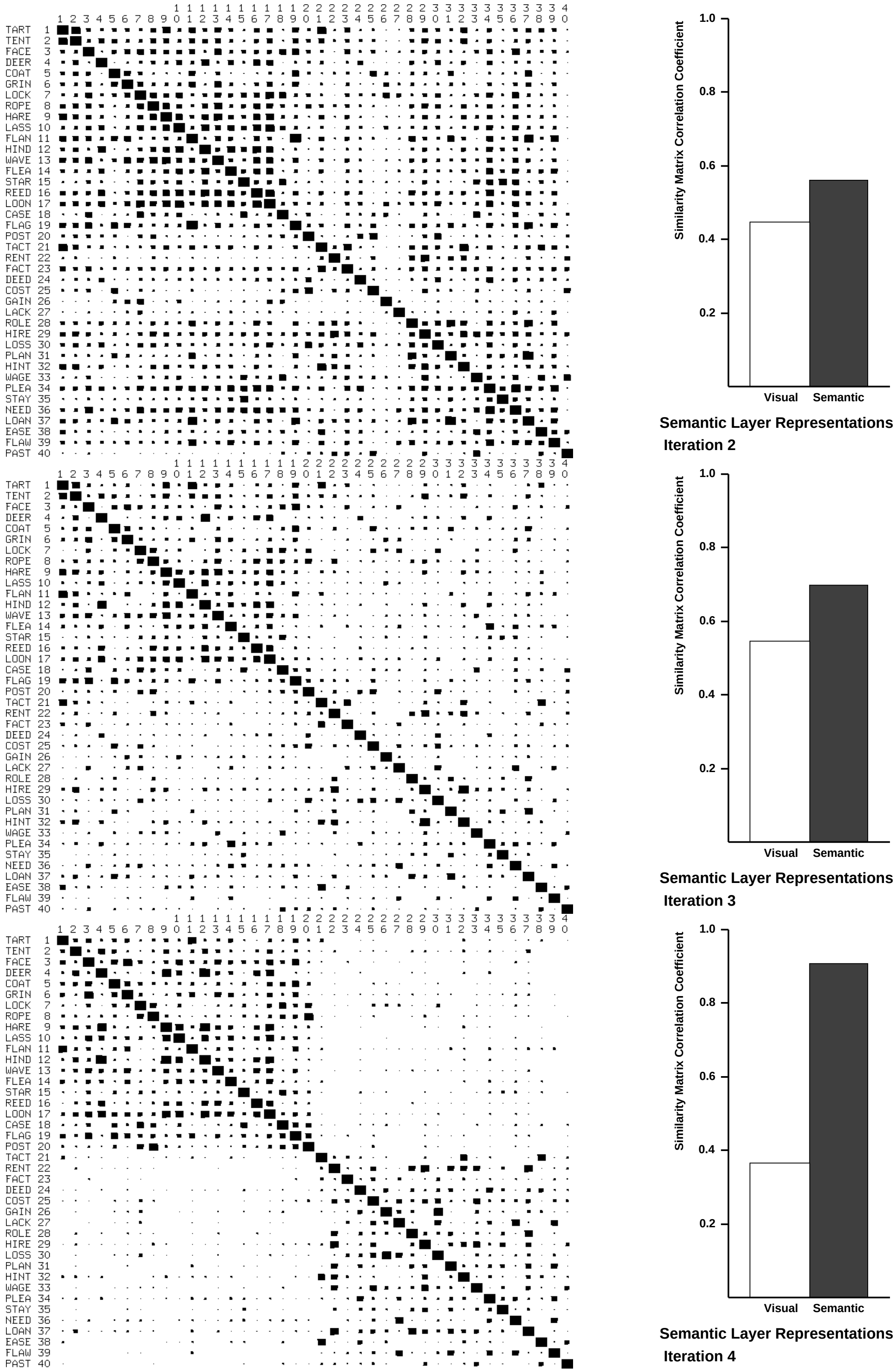


Figure 6.14: Similarity matrices and their correlation coefficients with matrices for visual (orthographic) and semantic similarity, for representations at the semantic layer at iterations 2, 3, and 4.

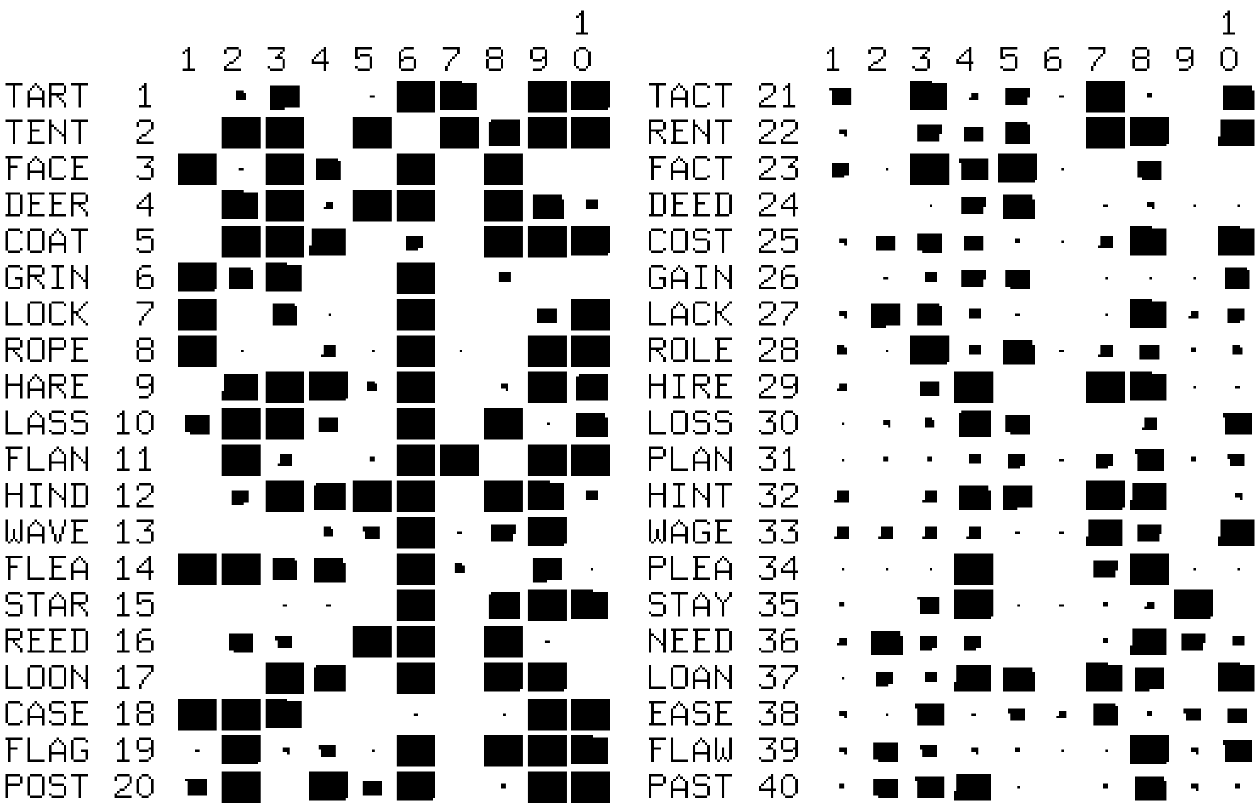


Figure 6.15: The final states of the clean-up units for concrete words (left) and abstract words (right).

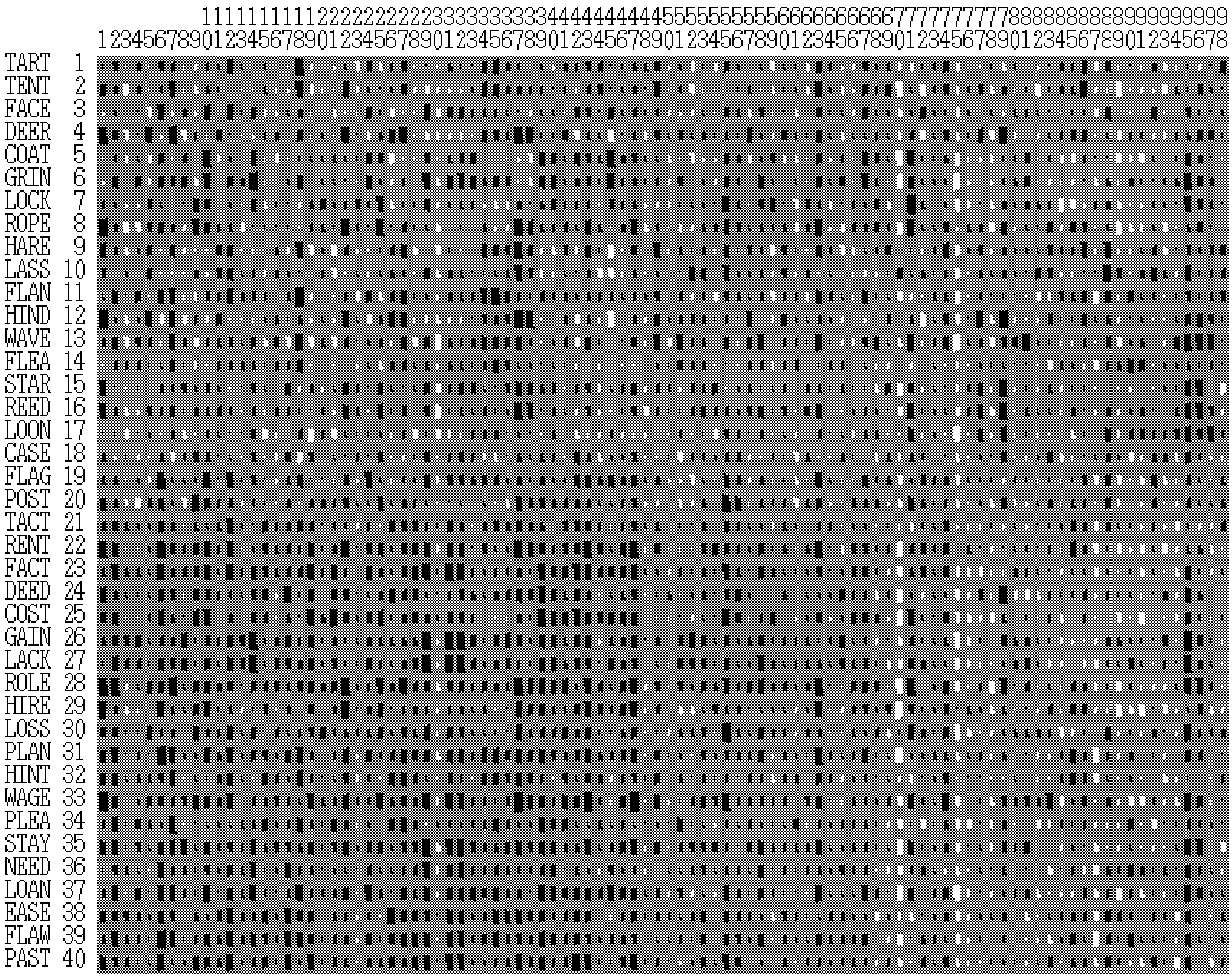


Figure 6.16: The final summed input to each semantic unit from units in the intermediate layer for each word. The largest negative input (WAGE feature 37) has value -16.9.

most clean-up units receive strong input (positive or negative) from semantics and are driven into a state near 0 or 1. In contrast, clean-up units receive relatively weak input from semantics when processing an abstract word, and so tend to remain in a state near 0.5. In this sense, the clean-up units play less of a role in generating the correct semantics of abstract words than they do for concrete words. This difference can be seen more directly by comparing the amount of input that each semantic unit receives from the direct pathway and from the clean-up pathway. The semantics of concrete words consist almost entirely of subsets of the first 67 features, while abstract words only use the last 31 features. Thus differences between the inputs to units representing these two groups of features indicates the influence that the different pathways have on concrete and abstract words, respectively. Figure 6.16 displays the summed input that each semantic unit receives from units in the intermediate layer at the end of processing each word, while Figure 6.17 displays the input from the clean-up layer to each semantic unit. Notice that the direct pathway provides input

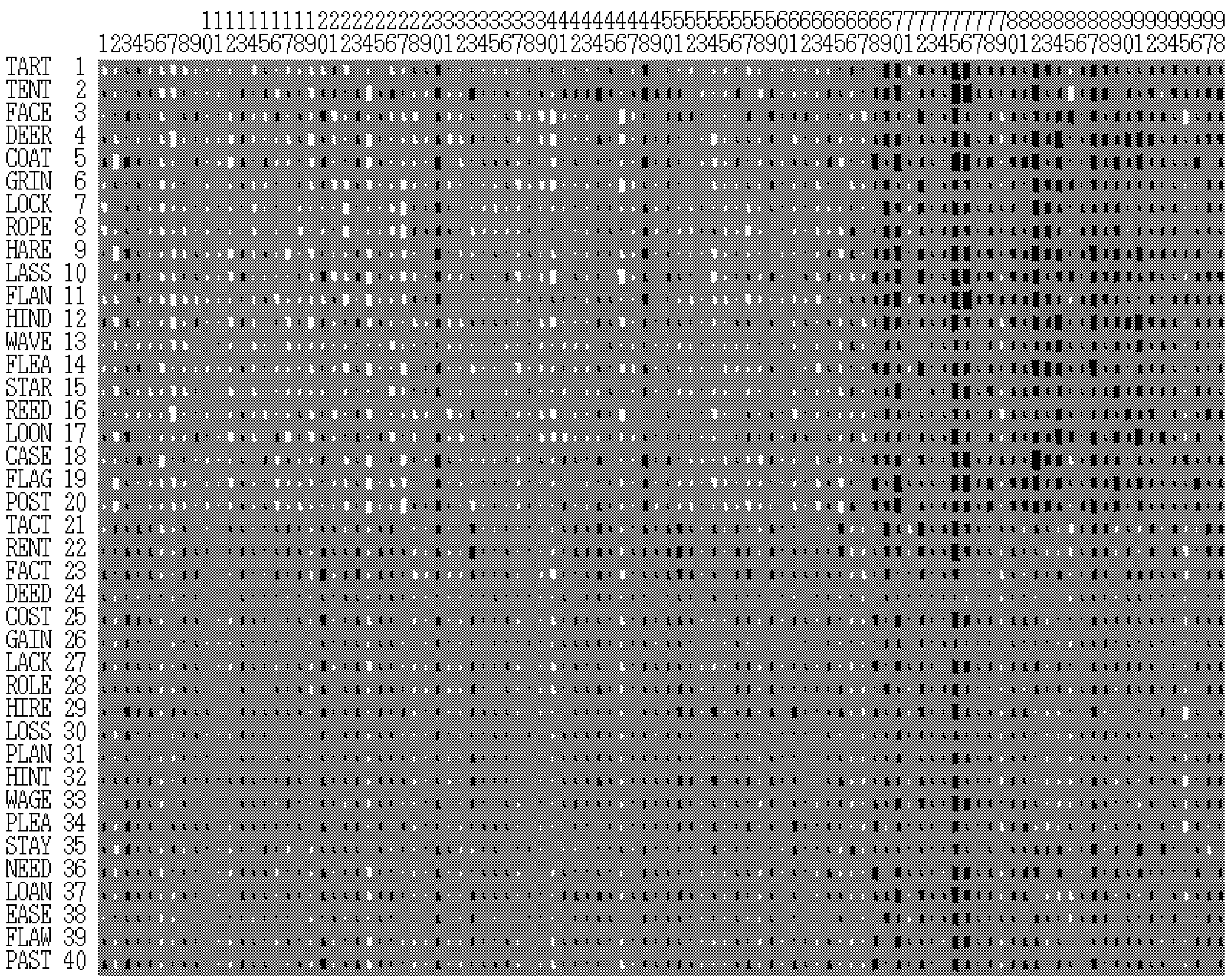


Figure 6.17: The final summed input to each semantic unit from units in the clean-up layer for each word. The largest negative input (TENT feature 75) has value -17.5.

of roughly equal magnitude input to concrete and abstract semantic features when processing both types of word. The input to concrete features for abstract words is largely inhibitory because all of these features must be turned off in the final semantic representation. Some semantic units seem to receive uniformly positive inputs (e.g. 30 and 75) or negative inputs (e.g. 37 and 95) regardless of what word is presented. The network must rely on the clean-up pathway to override these biases for particular words. In fact, as shown in Figure 6.16 the clean-up pathway has the opposite influence on each of these semantic features. Also notice that the clean-up pathway provides a sharp distinction between concrete and abstract features for concrete words. Most concrete features receive support from clean-up units, while abstract features are strongly inhibited. The input to concrete features is significantly reduced when processing abstract words precisely because the states of the clean-up units are less binary for these words. While the clean-up units do make useful distinctions among abstract features, in general their influence on semantics when processing an abstract word is much less than when processing a concrete word. For this reason, severe clean-up lesions produce a selective deficit for concrete words relative to abstract words.

6.7 Summary

The range of empirical phenomena addressed by H&S was quite limited, in part because of limitations of the original model, but also in part because the restricted definition of the task of reading via meaning they used precluded consideration of many aspects of deep dyslexic reading behavior. The simulations in this chapter serve to replicate the original findings of the co-occurrence of error types using a different word set, but more importantly to extend the empirical adequacy of the approach to include the effects of abstractness in deep dyslexia and its interactions with visual influences in errors. Our explanation for these effects hinges on the claim that the semantic representations of abstract words are composed of far fewer features than those of concrete words. This difference causes the direct and clean-up pathways of the network to become differentially important in processing each type of word through the course of learning, and is thus reflected in the behavior of the network under damage. The explanation has some similarities to those previously offered for the interaction between effects of abstractness and visual similarity (e.g. Morton & Patterson, 1980; Shallice & Warrington, 1980) but these were essentially *ad hoc* verbal extrapolations from cascade notions unrelated to other aspects of the syndrome, without even a principled account of the abstract/concrete difference. The present account is supported by a simulation, is linked to explanations of other aspects of the syndrome, and offers the possibility of also addressing concrete word dyslexia.