Brain and Language **68,** 241–248 (1999)
Article ID brln.1999.2074, available online at http://www.idealibrary.com on IDEAL®

What Do Category-Specific Semantic Deficits Tell Us about the Representation of Lexical Concepts?

Roberto G. de Almeida

Department of Linguistics, University of Alberta, Edmonton, Alberta, Canada

A reassessment of category-specific semantic deficits in light of their contribution to a theory of the representation of lexical concepts is proposed. Two theories are examined: one, held by the majority of researchers in the field, claims that concepts are represented by sets of features; another, in contrast, claims that concepts are atomic representations. An analysis of category-specific semantic deficits in terms of inferential relations (of the meaning-postulates type) between atomic concepts is elaborated. It is argued that this theory can better account for the pattern of performance exhibited by patients with semantic deficits. © 1999 Academic Press

Key Words: category-specific semantic deficits; lexical concepts; atomism; meaning-postulates; semantic features.

CONCEPTS AS SETS OF FEATURES

In the past decade or so, there has been a surge of reports on cases of category-specific semantic deficits caused by diverse types of brain injuries. Dissociations arising from such deficits (e.g., impaired ANIMAL with spared ARTIFACT or the other way around; see Hillis & Caramazza, 1991) are usually claimed as evidence for a prototype or a definitional theory of focal representation of concepts in the brain (see, e.g., Smith, 1995). Common to both theories is the idea that concepts—and, of particular concern, *lexical concepts* (i.e., concepts labeled by monomorphemic items)—are complex mental representations.[1]

Preparation of this paper was supported by a grant from the Social Sciences and Humanities Research Council of Canada. I am grateful to Gary Libben, Shannon McEwen, and two anonymous reviewers for helpful suggestions and comments on an earlier draft of this article. In addition, I thank Jerry Fodor for discussions on some of the ideas developed here.

Address correspondence and reprint requests to Roberto G. de Almeida, Department of Linguistics, University of Alberta, Edmonton, AB, Canada T6G 2E7. Fax: (780) 492-0806. E-mail: roberto@ualberta.ca.

[1] For the present purposes I am not going to distinguish between concepts as definitions (necessary and sufficient features) from concepts as prototypes (fuzzy sets of features). I simply refer to them as sets of features.

0093-934X/99 $30.00

Although there have been many discussions regarding the proper characterization of category-specific semantic deficits, most researchers seem to be in agreement concerning the very *nature* of concepts: They are represented by *sets of features* (see, e.g., Rapp & Caramazza, 1991; Gonnerman, Andersen, Devlin, Kempler, & Seidenberg, 1997; De Renzi & Lucchelli, 1994; McCarthy & Warrington, 1994). Rapp and Caramazza (1991), for instance, assume that the representation of *dog* is composed of features such as ANIMAL, MAMMAL, DOMESTIC, FOUR LEGS, BARKS, and so on. The concept CAT presumably is represented by a similar set except that MEOWS is part of the set instead of BARKS. The features postulate can, thus, be simply stated in the following way: For any concept *X,* there is a set of features *A* ($A = \{a, b, c, d,$ etc.$\}$); and for any concept *Y* there is a set of features *B* such that *X* and *Y* are related if $A \cap B$. The assumption here is that category-specific deficits arise because the set of features shared by *X* and *Y* is affected, and if any member of the intersecting set is affected so is the cluster of concepts that are dependent upon that set.

The features theory, however, has many problems. For instance, it is not clear what should count as a feature (i.e., which aspects or properties go into the set) and whether a feature in one set contributes the same property to different concepts (e.g., FOUR LEGS in the sets of CHAIR and DOG). But notably its main problem is its apparent failure to account for compositionality. A well-known example is PET FISH (see Fodor, 1998): the prototypical PET FISH (goldfish) is neither a prototypical pet (which is, say, dog) nor a prototypical fish (e.g., trout). That is, the lexical concepts PET and FISH do not contribute their content (the statistically most salient sets of features that give rise to the prototypical exemplars dog and trout) to the phrasal concept PET FISH. And if concepts as bundles of features (such as prototypes) are not compositional, they fail to account for the productivity and systematicity of mental representations and processes.

THE ALTERNATIVE: ATOMISM AND MEANING POSTULATES

The theory I argue for has two main characteristics. (1) Lexical concepts are atomic representations, and complex concepts have lexical concepts as their constituents and (2) conceptual relations are not obtained via feature-sharing but can be captured by the sets of inferential relations drawn from elementary and complex concepts. Although these characteristics are controversial in philosophy and cognitive science, I suggest nonetheless that a theory that combines both characteristics (together with a few other assumptions that I will mention below) can better explain the pattern of category-specific semantic deficits.

Sketch of the Theory

The thesis of conceptual atomism (see, e.g., Fodor, 1990), in short, holds that there are nomic relations between objects, words, and events and their

mental representations and that those mental representations are independent of others one might possess. Thus, for someone to have the mental representation DOG does not depend on someone also having the mental representations PET, FURRY, FOUR-LEGGED, CAT, ANIMAL, and so on. In sum, the mental representation of *dog* is the ''Mentalese'' translation (i.e., not the English word) DOG and, by assumption, the mental representation DOG does not enter into *constitutive relations* to any other concept.

Atomism radically contrasts with the features theory of conceptual representation for which possessing the concept *X* depends upon possessing the concepts (or features) {*A, B, C,* etc.} that are constitutive of *X.* But perhaps the major divide between the two theories, for our present purposes, resides on conceptual relations and the way those relations capture the content of the representation of linguistic expressions, objects, and events. Feature theories hold that conceptual relations are the product of feature-sharing. Thus, for concepts *X* and *Y* to be related, some of the features that constitute the content of *X* and some of those that constitute the content of *Y* have to be the same. For the atomic theory there are no such constitutive relations, so conceptual relations have to be captured by a mechanism that does not break the strongest of the atomistic assumptions—that symbols (like *X*) do not belong to symbol systems (see Fodor, 1990).

The theory I propose to account for category-specific deficits relies heavily on the notion that semantic representation and, by assumption, conceptual relatedness can be captured by sets of meaning postulates (MPs) derived from atomic concepts. Certainly, MPs are not new in semantics and psycholinguistics (e.g., see Carnap, 1956; Fodor, Fodor, & Garrett, 1975; de Almeida, 1998), but they have not been considered as a tool for the analysis of the patterns of semantic deficits. What follows is an extension of the standard view of MPs as applied to effects of category-relatedness and, thus, category-specific deficits.

I begin by assuming that for each concept *X* there is a set of MPs that constitutes the inferential domain of *X* (a lexical or a phrasal concept). I take the *inferential domain* of *X* to be the set $A(A = \{a, b, c, \ldots n\})$ of inferences that are *caused* by *X* and also the subset *B* of inferences that are caused by *Y,* but of whose entailments *X* takes part. The notion of inferential domain can be exemplified by the MPs that are related to DOG in (1), of which the last expression symbolizes that there are indefinitely many properties that can be inferred from *X* (that is, *caused* by *X*).[2]

[2] For simplicity, I refrain from employing modal operators. Also, I avoid discussing the analytic/syntectic distinction. I assume that all inferences that *X* causes are valid (*latu sensu*) inferences. But it is plausible that for some cases (such as the causative and inchoative representations of verbs; see de Almeida, 1998) certain inferences may play a more ''prominent'' role (that is, may be considered ''logical'' entailments).

$$
\begin{aligned}
dog = \ & \text{DOG} \\
& [\text{DOG}(x)] \rightarrow [\text{ANIMAL}(x)] \\
& [\text{ANIMAL}(x)] \rightarrow [\text{LIVING}(x)] \\
& [\text{ANIMAL}(x)] \rightarrow [\text{ANIMATE}(x)] \\
& (\forall x\ [P(x)] \rightarrow [Q(x)])_n
\end{aligned}
\tag{1}
$$

I assume, in short, that any conceptual relations can be represented in terms of entailments that are obtained in a system of derivations between concepts and other conceptual expressions. Thus, if MPs constitute the mechanism of conceptual inferences, and if, in many cases, MPs allow for two-way entailments, we take that, all else being equal, expressions such as (2a) and (2b) are both valid and also that they generally typify the kinds of inferential relations obtained between concepts. That is, they materialize many cases of relations such as hyponymy, synonymy, meronymy, and opposition, (see Cruse, 1986).

$$
\begin{aligned}
&\text{a. } \forall x, [\text{VEGETABLE}(x) \text{ or } \text{ANIMAL}(x)] \rightarrow \text{LIVING}(x) \\
&\text{b. } \forall x, \text{LIVING}(x) \rightarrow [\text{VEGETABLE}(x) \text{ or } \text{ANIMAL}(x)]
\end{aligned}
\tag{2}
$$

There are three related observations to be made regarding the nature of these inferences. First, since I assume that the elements (or the ''nonlogical'' elements, to borrow Carnap's term) of the MPs are atomic representations, this theory distances itself from inferential-role theories (see, e.g., Block, 1986) for which the *contents* of expressions are determined by the inferences in which the expressions and their constituents enter. Second, it is important to distinguish the present theory from the features theory: here the set *A* (or *B* or *C*) is a set of *inferences* or MPs, not a set of concepts or feature-concepts. This is not a simple case of notational difference because of what follows. Third, I assume that the inferences in *A* are causally connected to *X* but are not *X-content* constitutive. That is, although *X* causes *A*, the inferences constitutive of *A* are not where *X* gets its content from.[3] What *A* does is to determine the epistemic conditions by virtue of which *X* and *Y* are related—where *X* causes *A*, *Y* causes *B*, and *A* and *B* are said to have some of the same MPs (i.e., $A \cap B$).

An Analysis of Category-Specific Semantic Deficits

In the framework presented above, we can postulate that category-specific semantic deficits arise from functional damage to particular concepts or sets of concepts. In addition, we can postulate that inferential domains are affected, thus producing category-specific effects. The assumption is that if a

[3] Where *X* gets its content from is not a matter to be solved here. I rather side with Fodor (1990) and assume that '' '*X*' means X.'' The point is that, given a lexical item (or an object or an event) its token corresponds to a Mentalese symbol *X*.

concept *X* is affected and if *X* is in the inferential domain of *Y, Y* can also be affected. For ''basic-level'' concepts, this assumption is straightforward: If DOG is damaged, patients cannot say (recognize/define) ''dog,'' and by virtue of the fact that DOG is in the inferential domain of CAT, COW, and many other concepts, patients cannot name (recognize/define) them either.

That is a plausible assumption. However, what the data appear to reveal is that the deficits are not *concept*-specific but *category*-specific. This means that what seem to be affected are the properties (features) that hold a class of concepts together. Thus, for instance, if patients have difficulty with DOG, CAT, and COW, what is affected is the superordinate concept LIVING. My suggestion is that this categorization effect can be treated in a way similar to the *concept*-specificity assumption: If the pattern of performance reveals that there is a deficit related to LIVING, the inferential domain of LIVING might be disrupted. A functional damage to the concept LIVING, then, might affect the concept ANIMAL and its respective inferential domain, as shown in (3), including all or most exemplars—and as many as the lexical concepts that introduce LIVING among their MPs. (Impaired inferential relations are represented by an asterisk.)

$$\forall x,\ \text{ANIMAL}(x)^* \rightarrow \text{LIVING}(x) \tag{3}$$

This can be seen further in (4), which follows from the entailments of the representation for DOG, such as in (1) above.

$$\begin{aligned} &\text{a. } \forall x,\ \text{DOG}(x) \rightarrow \text{ANIMAL}(x) \\ &\text{b. } \forall x,\ \text{ANIMAL}(x)^* \rightarrow \text{LIVING}(x) \\ &\text{c. } \forall x,\ \text{DOG}(x)^* \rightarrow \text{LIVING}(x) \end{aligned} \tag{4}$$

Notice that, in principle, if expressions like (4a) are not impaired, we can predict that subjects should be able to produce semantic paraphasias (*dog* → ''cat''), and in fact they are (see Hillis & Caramazza, 1991). But if what is affected is the concept LIVING and its inferential domain (arguably, all LIVING things), why can't the subjects say ''dog''? Why can't they define concepts within the affected category? Certainly, answers to these questions are essential for the success of the present analysis.

The first question is a challenge for atomism: after all, if there is a nomic relation between *dog* and DOG, and if the assumption is that the concept DOG is spared but some of its MPs are impaired (e.g., those involving LIVING), why do patients perform paraphasias? Why can't they say *dog* when they see a picture of a dog or are presented with the word *dog*? There are two hypotheses to consider. The first says that the problem is in the sets of inferences unleashed by the tokening of DOG: since they are disrupted by the broken entailments that involve LIVING [as in (4c) above], patients have trouble selecting the appropriate lexical items. The assumption is that the tokening of DOG (if it occurs at all; see below) causes the computation

of its inferential domain, which intersects with the inferential domain of many other concepts, thus giving rise to paraphasias. The second—and perhaps stronger—hypothesis places the problem in the selection of the appropriate concept, given a certain stimulus. That is, it is possible that the specificity of the problem is in the *causal link* between the proximal stimulus *dog* and the concept DOG. Thus, in this sense, when *dog* is presented, DOG may be accessed but the inferences it unleashes (e.g., to ANIMAL) may lead the patient to consider other alternatives (CAT, COW) that are in the inferential domain of ANIMAL.

There is some evidence in favor of the ''causal link'' hypothesis. In the picture/word matching task, the patients studied by Hillis and Caramazza (1991), for instance, never rejected a correct match. Similar results were obtained by Laiacona, Barbarotto, and Capitani (1993), with participants producing 75 and 80% correct responses (against a norm of 93%) for stimuli in the ''impaired'' category. In Gonnerman et al.'s (1997) study, the picture/word task was the only task that did not produce significant differences between impaired and spared categories. If there is a deficit in the concept itself or in a cluster of concepts (or their defining features or in the bundle of features that is central to a given category), patients should not be able to respond correctly even in the presence of both stimuli. In the picture/word matching task, however, the patients have two sources of evidence on which to verify the *hypothesis* that, e.g., *dog* causes DOG (and not CAT or COW) and, thus, they rarely commit errors.

Further evidence for the causal link hypothesis comes from the types of errors more often committed in oral naming: semantic paraphasias. This pattern of errors suggests two additional hypotheses concerning the nature of semantic deficits. First, the fact that patients are able to produce items that are within the category (or, possibly, inferential domain) of the target item suggests that they are able to determine the nature of the stimuli (i.e., by assumption, *dog* is probably causing [DOG or CAT or COW]) but are unable to produce ''dog'' and in most trials they produce an incorrect response. In the cases studied by Hillis and Caramazza (1991), semantic paraphasias accounted for 65 and 100% of the errors of naming in the affected categories. Second, the fact that they are able to produce semantic paraphasias may be an indication that some of the MPs within the inferential domain of the target item are intact. Thus, the patient is able to produce items that represent concepts ''related'' to the target—related in the sense that both elicit similar MPs.

The second problem this theory should be able to answer, the problem of definitions, can be rather naturally accounted for. The fact that patients have difficulty producing definitions can be explained by the impairment of some of the MPs. In Hillis and Caramazza's (1991) study, patients produced wrong or ambiguous definitions for 15 to 23% of the target items in the impaired categories (see also Carbonnel, Charnallet, David, & Pellat, 1997). If concept

X is impaired and *Y* is spared and *X* has an entailment relation with *Y*, subjects will not be able to compute some of *Y*'s entailments and, therefore, they will not be able to produce definitions, except for circumlocution, which is often the case.

CONCLUSIONS

The primary concern of this article has been with the nature of the representation of lexical concepts as revealed by category-specific semantic deficits. In general, semantic deficits have been interpreted as supporting the view that lexical concepts are complex mental representations. I have suggested another framework that takes atomic concepts and their inferential relations via MPs to be the basic elements of conceptual representation and processing. There are crucial differences between features and atomic theories: while the former is bound to the infinite regress of conceptual primitives and to the indeterminacy of conceptual representation, only the latter appears to account for compositionality, which is a necessary condition for the postulation of any plausible theory of cognitive representations and processes.

The theoretical framework suggested here appears to account adequately for the pattern of deficits, which are seen as arising from damage to inferential domains of atomic concepts. Research within this framework, which has not been applied to semantic deficits, can lead to new and more fine-grained empirical predictions on the proper characterization of these deficits. A careful exploration of the entailment relations between concepts may produce new evidence concerning the nature of category-specific semantic deficits, thus concerning the very nature of concepts and their representation in the brain.

REFERENCES

Block, N. 1986. Advertisement for a semantics for psychology. *Midwest Studies in Philosophy,* **10,** 615–679.

Carbonnel, S., Charnallet, David, D., & Pellat, J. 1997. One or several semantic system(s)? Maybe none. Evidence from a case study of modality and category-specific ''semantic'' impairment. *Cortex,* **33,** 391–417.

Carnap, R. 1956. *Meaning and necessity* (2nd ed.). Chicago: Chicago Univ. Press.

Cruse, D. A. 1986. *Lexical semantics.* Cambridge, UK: Cambridge Univ. Press.

de Almeida, R. G. 1998. *The representation of lexical concepts: A psycholinguistic inquiry*. Unpublished Ph.D. dissertation, Rutgers University, New Brunswick, NJ.

De Renzi, E., & Lucchelli, F. 1994. Are semantic systems separately represented in the brain? The case of living category impairment. *Cortex,* **30,** 3–25.

Fodor, J. A. 1990. *A theory of content and other essays.* Cambridge, MA: MIT Press.

Fodor, J. A. 1998. *Concepts: Where cognitive science went wrong.* Oxford, UK: Oxford Univ. Press.

Fodor, J. D., Fodor, J. A., & Garrett, M. F. 1975. The psychological unreality of semantic representations. *Linguistic Inquiry,* **6,** 515–531.

Gonnerman, L. M., Andersen, E. S., Devlin, L. T., Kempler, D., & Seidenberg, M. 1997. Double dissociation of semantic categories in Alzheimer's disease. *Brain and Language*, **57,** 254–279.

Hillis, A., & Caramazza, A. 1991. Category-specific naming and comprehension impairment: A double dissociation. *Brain,* **114,** 2081–2094.

Laiacona, M., Barbarotto, R., & Capitani, E. 1993. Perceptual and associative knowledge in category-specific impairment of semantic memory: A study of two cases. *Cortex*, **29,** 727–740.

McCarthy, R. A., & Warrington, E. K. 1994. Disorders of semantic memory. *Philosophical Transaction of the Royal Society, London, B,* **346,** 89–96.

Rapp, B. C., & Caramazza, A. 1991. Lexical deficits. In M. T. Sarno (Ed.), *Acquired aphasias* (2nd ed., pp. 181–222). San Diego, CA: Academic Press.

Smith, E. E. (1995). Concepts and categorization. In E. E. Smith & D. N. Osherson (Eds.), *An invitation to cognitive science: Thinking* (2nd ed., Vol. 3, pp. 3–33). Cambridge, MA: MIT Press.