In: Graham Katz, Sabine Reinhard, and Philip Reuter, eds. (2002), *Sinn & Bedeutung VI,*
Proceedings of the Sixth Annual Meeting of the Gesellschaft für Semantik, University of Osnabrück

INDIVIDUALS AND PREDICATION - A NEUROSEMANTIC PERSPECTIVE*

Michael Klein[1,2] and Hans Kamp[1]
[1] Institute for Natural Language Processing, Univ. of Stuttgart
[2] Section Experimental MR of the CNS, Univ. of Tuebingen

Abstract

In this paper we reflect on the constraints that current knowledge and assumptions in the neurosciences impose on theories of concepts and of linguistic meaning. We focus on the operation of predication. A central contention of this paper is that at the neural level there is no fundamental difference between predicate and predicable - representations of individuals and representations of categories are both feature complexes with basically the same structure. We discuss the structure of such representations and the ways in which they interact in certain simple situations of perception and verbal communication.

1 Introduction

Knowledge about information processing in the brain can be gathered in many ways. The most important are: (i) correlating brain damage with behavioral changes, (ii) measuring neurophysiological changes reflecting certain cognitive processes, (iii) tracing the pathways of information flow in the brain, and (iv) investigating the types and computational properties of neurons in certain brain structures. In addition, (v) hypotheses about information processing in the brain can be tested by simulations (computational neuroscience). In computational and experimental neuroscience these and other methods have meanwhile led to theories and experimental results about the nature of neural representations and computations as well as about the regions of the brain where they are located. In this way they have created a platform on which it now seems possible to develop accounts of how the human brain understands and produces language.

There are areas of experimental brain research, such as functional magnetic resonance imaging, where much of what has been learned about the representation and processing of language by the methods of linguistics has thus far been ignored. This has been true in particular of the semantic aspects of these representations and processes. Neuroscience has succeeded in establishing correlations between some aspects of language processing and local changes in neural activity, or detected lesions that cause the impairment of certain linguistic capabilities etc., but these results are still very distant from a neural explanation of language processes as the linguist understands and describes them.

Both disciplines, neuroscience and linguistics alike, carry a responsibility for trying to close this gap. For their part, linguists, with their insights into the structure and use of language, ought to

*We would like to thank Michael A. Arbib, Valentino Braitenberg, Christoph v.d. Malsburg, Julie C. Martin-Malivel, and Almut Schuez for discussions.

give guidance to neuroscientific research to a greater extent than has been the case hitherto. Such guidance is needed especially in connection with questions of linguistic meaning. Semanticists have, in spite of their growing knowledge of how meaning is expressed in language, hardly played a part in efforts to account for meaning-related phenomena in neural terms.

We believe that it is time for linguistic semantics to take a more active part in neurolinguistics. But we hasten to add that the theories which linguistic semantics has thus far developed do not seem a useful basis for the interpretation of neurophysiological data or the design of new neuroscientific experiments. The inadequacies are particularly salient when we turn to very basic semantic relations and operations which the working semanticist usually takes for granted as he grapples with problems that count as the real challenges within his community. In the light of results that have been emerging within the neurosciences about the neural representation of information generally, semanticists will, if they are to make a useful contribution to neurolinguistics, have to rethink their theories, and establish a new conceptual foundation for them which is compatible with these results.

A rethinking of the foundations of semantics will have to take into account that *in the brain* representations and computations manifest themselves in states of physically real entities (i.e. by neurons and synapses) and it is the properties of those entities and their interaction that determines what kind of computations or representations are possible. Our current understanding of neural information processing is based on experimental investigations of these entities (both on the microscopic and on the macroscopic level). Two major discoveries of such experiments are (i) certain neurons in certain areas that respond to certain features in the environment and (ii) synapses between neurons in many areas that increase their synaptic strength when the connected neurons fire at approximately the same time. Based on these and other findings, there is a growing consensus that basic concepts are neurally represented by comparatively strongly connected neurons, which thereby form so called *cell assemblies*; coordinated firing activity in such an assembly signifies activation of the represented concept.

This understanding is crucial to a reanalysis of the semantic concept of *predication*. This concept is arguably the most important building block of the entire edifice of linguistic semantics. As most linguists and philosophers see it, predication is an operation in which one entity, the *predicate*, is applied to another, the *predicable*; the result of the application is either positive, when the predicate is true of the predicable, or negative, when it is not. An important aspect of this conception of predication is the view usually associated with Frege (1879) that predicate and predicable are entities of fundamentally different sorts. The predicable has a reality all of its own (it can, so to speak, stand on its own feet), but the predicate has only a kind of virtual or potential reality - it yields something of genuine reality only through predication, i.e. when it is combined with a predicable and the result is either a *truth* or its opposite.

When we try to relate this apsychologistic (or, as some would see it, antipsychologistic) conception of predication to the mental act of predication - to what it is people do when they attribute a property to a thing and to what goes on in their heads when they do so - we run into all sorts of problems. In fact, as soon as one starts to think about the relation, one can't help noticing that there isn't just one type of mental act that corresponds to the linguists' and philosophers' notion of predication, but a number of them, and that it is crucial not to confuse these. Mental acts of predication involve particular ways of representing, and sometimes recognizing, both the predicate and the individual involved, and the recognition and representation of these two components of the predication may vary significantly from one instance to the next.

In the light of such reflections what little current semantic theory has to say about predication may not seem all that helpful; at the very least it is incomplete, and needs to be complemented by an account of the various ways in which the mental representations of predicate and predicable can be combined so as to yield, at the level of consciousness, the sense of a predicational judgment. Below we will formulate some first hypotheses about what the representations involved in these mental predication operations may be like and about some of the ways in which they can interact to produce an act of predication. One aspect of the story will be that the representations of predicates and predicables do not differ in the fundamental way that is generally assumed in linguistics and philosophy.

Before we present these hypotheses, we will first provide a brief summary of the views that are gaining acceptance within the neuroscientific community and which we consider most relevant to the question how the theory of predication should be modified and refined. From this summary it will appear that a theory of mental predication should, at the very least, have something to say about the following mental acts and processes: (i) the acquisition of a categorical concept through perception, (ii) the acquisition of an individual concept through perception, (iii) categorization of a perceived object as belonging to a certain known category, (iv) identification of a perceived object as a known individual, (v) retrieval of an individual concept based on the understanding of a denoting predicate or predicate complex, and (vi) modification of an individual concept through communication.

2 Insights from Neuroscience

By *semantic information* let us understand the kind of information that people get out of words - that is, out of what they read or are told by someone else - as well as the information they put into words. Whenever a person gets such information, it must be somehow represented in his brain. The question how such information is represented cannot be separated from questions concerning the representation of perceptual information - information which people derive from what they see, hear, taste, etc. For one thing, semantic information is often (if not always) perceptually testable: It carries expectations of certain complex perceptions, and when these expectations are not confirmed, that will be evidence that the represented semantic information is false. This correlation between semantic and perceptual information could not exist if representations of the former were not somehow linked to representations of the latter. The simplest and most natural explanation of there being such links is that the representations involve the same components - representations of concepts which can be activated both by the perception of words and by the perception of features connected with the concepts those words denote.

Given these considerations it would make little sense to investigate the neural basis of semantic representations in isolation from the representation of perception, for in all likelihood we are dealing with modes of representation that have a great deal in common. Where the question is concerned how semantic information is represented, there is an additional reason for looking at the representation of perceptual information: As things stand, there is a large quantity of experimental results and theoretical insights about the neural realization of perception processes and their results - very much more than we currently know about the neural aspects of language processing in general and semantic processing in particular. If from the perspective of neural implementation the two kinds of representations are as similar as the above reflections suggest, then what is known about the neural dimension of perception may help us in developing

reasonable hypotheses about how the brain represents what it gets out of verbal input.

As things stand at present, so much is known about neural aspects of perception (even about vision alone!) that anything resembling a survey of the literature would be out of the question. All we can do is mention a few facts and hypotheses which are pertinent to the ideas which we will develop in the following section.

First of all, perception is a hierarchical process. There is convincing evidence that at the lowest level certain sensible features - colors like green and yellow, qualities of taste such as sweet and sour, etc - are *represented* by single cells or small cell clusters (Hubel and Wiesel 1962). The activation of such a cell or cluster is the neural correlate of the feature perception, so that the cell (cluster) can be regarded as a *feature identifier* at this level. In particular, the visual field is represented by a field of identifiers for any one of the basic colors (as well as a number of other visually perceptible features) and the activation of any one of these identifiers signals the instantiation of that feature in the part of the visual field to which the cortical position of this identifier corresponds. In these cases it is the direct neural connection between the identifiers and the corresponding detectors in the sensory organs which, it might be said, gives the identifiers their *meaning*. On the other hand the feature identifiers are linked to more complex cell clusters *higher up* in the processing hierarchy. These represent more complex concepts and derive their meaning from their links to the identifiers *below*. Beyond this second tier of cell clusters there are other tiers; and the farther removed a cluster is from the primary sensory areas, the more abstract the concept it represents.

Furthermore, the geometrical lay-out of the brain is such that identifiers are spatially grouped by perceptual channel. Roughly speaking, visual features are represented in areas of the visual cortex, auditory features in the auditory cortices, taste features in the paralimbic cortex, and so on. These areas are further subdivided according to feature type. Thus, visual features pertaining to form are processed and represented in the so-called ventral stream while other visually detectable characteristics of objects (location, color, motion, orientation) are dealt with elsewhere (Fuster 1995).

These facts are important for our present concerns because they imply that the representations involved in typical cases of predication will have to be distributed. Take the case of some person H who is told that Mary is sick and retains from this the predicational information that the individual named (i.e. Mary) falls under the concept *sick*. We assume that the representation of this information takes the form of a link between separate representations of the individual Mary and the categorical concept *sick*. Let us assume, moreover, that H has a conceptual representation of Mary which enables him to recognize her under normal conditions. This means that his representation of Mary will have links to a number of perceptual features whose coordinated activation will, under normal perceptual conditions, lead to his recognizing Mary (which, in the spirit of what has been said above, we will assume amounts to activation of his individual-representation of Mary). In general, the cell clusters corresponding to the perceptual features that are part of this individual-representation will be distributed over different parts of the brain. (Think for instance of features that pertain to what Mary looks like, features pertaining to the sound of her voice, features that capture the smell of her favoured perfume.) If all these are part of the individual-representation that H has of Mary, then activation of H's individual-representation of her will often give rise to the joint activation of such mutually distant clusters. In particular, if perception activates enough features this will trigger the activation of other properties that are part of this representation, among them non-perceptual properties

(such as, for instance that of being called *Mary*, or being a spinster), and the result will be recognition of the perceived individual as Mary.

This kind of correlated activation of distant cell clusters should be distinguished from another type of coordination which at first sight may seem superficially like it. This second type of coordination is the one that has become familiar as the *binding problem*. In what is arguably its simplest form it arises where there is coordination of two basic sensory features that are perceived as features of the same thing, as for instance in the perception of something as both yellow and sweet, or as both yellow and round: What is it about the coordinated activation of the cell clusters representing two such features that is responsible for the represented information being that the same thing is, say, both round and yellow, and not just that there is something that is round and also something that is yellow? The hypothesis that such cases of coordination are a matter of synchronous firing of the coordinated cells (or the cells in the coordinated clusters) was first put forward by von der Malsburg (1981)[1]. This hypothesis has since been experimentally confirmed by multi-electrode recording from the visual cortex (Eckhorn et al. 1988, Gray et al. 1989).

The kind of binding (henceforth *working memory binding*) just described is distinct from the long term binding mechanisms which are involved in, among other things, the rich representations of individuals that people have of their acquaintances and objects with which they are well familiar. Although the feature- and property-identifying cell clusters that are involved in such long time bindings are the same as those that are linked by synchronous activity in working memory, the nature of the binding is crucially different in the two cases. Long term binding is assumed to be realized through lowering of synaptic thresholds. It gets established by a learning mechanism first postulated by Hebb (1949), who introduced the term *cell assemblies* for the groups of cells (or of cell clusters) that get bound together by synaptic strength modification as the result of such learning. (Hebb's learning rule was confirmed a quarter of a century later by observations on the hippocampus (Bliss and Lomo 1973). This form of learning is called *long term potentiation* (LTP). It and its counterpart, *long term depression* (LTD), are now considered the general neural mechanisms underlying long term learning in the cerebral cortex (Artola et al. 1990).

It should be clear from this brief discussion that binding and activation are not to be conflated. It is true that working memory binding is always a matter of joint (more exactly, synchronous) activation. But long term binding is a synaptically based disposition of the cell assemblies within which such bindings are realized; it is there irrespective of whether the assembly, or any part of it, is active. Long term bindings facilitate joint activation whenever one part of the assembly gets triggered, but in general there isn't even a necessity that activation of one cell cluster automatically triggers activation of the other clusters that are long term bound to it (e.g. if the connected cluster are not excited by other sources). On the other hand, clusters, that are weakly connected can synchronize when they do receive strong input from other sources.

The picture that emerges from these neuroscientific facts and assumptions based on them is one of fairly directly implemented basic concepts (or features) and of complex concepts that are built in some way from simpler concepts as components. This conception is consonant with a well-

[1]The synchronization of neural activity can partly be explained by the ability of neurons to detect temporal coincidence (von der Malsburg 1985), which in turn can be explained by the properties of the cell membrane (Koch 1999). However, while the phenomenon of synchronized neural activity is beyond doubt, the causal mechanisms involved in it remain a topic of dispute (see Wennekers and Palm (1997) for a longer discussion).

established perspective within cognitive psychology, according to which many, most or perhaps even all concepts have a *prototypicality structure*, with the concept's prototype consisting of a weighted set of features. The assumption of prototypicality structure has been used in explaining a range of observations about the way in which concepts are used and about their acquisition (Rosch 1973, Clark 1973, Lakoff 1987).

3 Consequences for Semantics

As already noted in passing, the facts and assumptions of current neuroscience suggest that at the neural level there is no fundamental distinction between representations of individuals and representations of properties - for either representation takes the form of an interconnected family of features, realized by a network of linked feature-representing cells or cell groups. Suppose that this is so. Then we must face the question: How could such a uniform mode of representation for both individuals and properties be compatible with the view of philosophical and linguistic semantics according to which they are entities of fundamentally different kinds? Or - asking the question from a slightly different angle - how can the brain, assuming that it does represent individuals and properties in this uniform manner, execute acts of predication, if it is of the essence to the predication operation that the two entities involved, the predicate and the predicable, are as different as semantics claims they are? It is to this question in particular that the present paper explores some first, tentative answers.

We begin by reflecting on certain applications of concepts in the context of perception. For the moment we ignore any possible connections with language, that is, we remain neutral on the question of *lexicalization* (the question whether a concept is connected with a word form which functions as its linguistic label and which makes it possible to verbally express predications in which the concept is involved).

First consider the case where a person A observes a certain object - a banana, say - and then becomes acquainted with some further feature of it. For instance, we may assume that, after having first become aware of it, A takes a bite of the banana (presumably after having removed part of its skin, though that is irrelevant to present concerns) and that he finds it to be (deliciously) sweet. At a minimum, the mental act of ascribing to the banana the additional property of sweetness must involve: (i) a concept of the banana as it is first visually perceived; and (ii) the concept of sweetness that is attributed to the object represented by the first concept on the strength of the tasting experience. In the predicational act in which the banana is judged to be sweet, the second, categorical, concept of sweetness (or of some particular quality of sweetness) is *predicationally linked* with the first concept (the individual concept of the banana). According to the assumptions we have already made, both the individual and the general concept are represented by cell assemblies, and the only plausible assumption that seems open to us with regard to the act of predicating sweetness of the tasted banana is that it takes the form of binding the two concepts together: Some link is established between the network of connected cell groups that constitutes the individual concept and the network, cell group or cell which represents the sweetness concept triggered by the tasting. This is still very vague, and what is missing most saliently, given our earlier remarks, is the question what kind of *linking* is involved. Since we are looking at mental operations that result as immediate reactions to acts of perception, it is reasonable to assume that the linking will at least initially take the form of the synchronization of cell group activity which we assumed to be the neural mechanism behind concept conjunc-

tion in working memory. In fact, we are led to assume this not only with regard to the individual banana concept and the sweetness concept, but also with regard to the different concepts from which the individual concept is composed. If and when the information represented by synchronous cell group activity is retained - that is, when it is transferred from working memory to a memory store from where it can be recalled at some later time - then the synchronization binding will have to be replaced by binding of some other sort. In the light of what we have been saying it should be expected that this new binding takes the form of the kind of synaptic adaptation that was suggested as the binding mechanism of long term memory. Admittedly, in the light of the usual assumptions about the conditions that must prevail in order that synaptic adaptation can occur, it is not yet fully clear how such long time bindings can get established by a single, comparatively short synchronization episode in working memory. This is a difficulty to which we have no solution to offer, but it is a general problem about the transfer from working memory to long term memory to which any account of information processing in the brain will have to find an answer.

Suppose now that our banana taster has seen the banana before - say he observed it when he was in the room at an ealier time, and was struck by a curious distribution of black spots on the side exposed to his view, on account of which it is possible for him to now recognize the banana as the one he saw earlier. In this case the individual concept for the banana which got established during *A*'s first perception of it gets reactivated. Presumably this takes the form of the individual concept being prompted into synchronous firing activity via some of its component features, which are set off by the current perception - once these features are brought to life, the long term, synaptically implemented bindings between them and other cell groups of the individual concept will then cause the others to enter into coordinated activity with them. When our observer then proceeds to taste the already familiar banana and observes it to be sweet, a further synchronization link will occur between the reactivated individual concept and the sweetness feature, just as in the case we considered above. Again it must be possible for the result of this predication to be transferred to long time memory. This will now have the effect that the sweetness feature gets integrated into the already existing individual concept through long time (i.e. synaptically implemented) binding to the other components of the individual concept.

Now let us assume that some of the concepts involved in the mental acts of predication under consideration are lexicalized. As indicated above, we assume that lexicalization makes use of the same binding mechanisms that are responsible for the creation of complex concepts. That is, a concept of the sort that was assumed in the last subsection is linked with another concept of a special sort, a complex concept consisting of features characterizing the word's phonological form together with certain other features which determine its morphological properties, syntactic behavior and, in case the agent is literate, its orthography.[2] It is the lexicalization link between a word form concept and a concept of the kind discussed above which renders the word form *meaningful* in that it enables the agent to associate the word form with certain features which he is in a position to observe (i.e. that can be activated through visual perception).

If the relevant concepts of a person are lexicalized, then he will be able to express a predication verbally, by combining the word forms associated with the concepts that are involved in the predication into an appropriate (i.e. grammatical) string. Exactly how such strings are to be

[2]Neurological evidence implies that such word form concepts are, like the perceptual concepts of which we spoke in the last section, represented in the cerebral cortex. Thus the connections between word form concepts and concepts of the latter kind are once more cases of cortico-cortical association. See Klein and Billard (2001) for a complete model.

chosen as a function of the grammar of the speaker's language is a question which we do not address. We assume for the sake of argument that in the simple cases on which we focus here the string consists of the word form associated with the predicated categorical concept, preceded by either (i) a single word form associated with the individual concept that plays the role of predicable, or else by (ii) a combination of word forms associated with different component concepts of the individual concept. Intuitively, the conventions of the language must be such that the string consisting of the word or word combination for the individual concept and the word for the categorical concept *signifies* that the individual represented by the former concept has the property represented by the latter. (Once more, how this kind of information is encoded in the speaker's representation of his grammar is not of our concern.)

Casting information in words wouldn't be of much use unless there is someone else to whom the words can be addressed and who can do something with them. In order to see a little more clearly what is involved in verbal communication of meaning to others and in the communicational dimension of linguistic meaning itself, let us consider a couple of cases in which a word string of the grammatically simple kind just described is used to communicate the content of a predication to some other agent.

So let us assume that our scenario includes besides the agent A of whom we spoke already also another agent B and that A after having tasted the given banana wants to communicate his experience to B. If this is to be effected by the type of utterance described above, then, as applied by what we have just been saying, A will need a minimum of two lexicalized concepts, one coinciding with the taste quality he has just experienced and at least one connected with his individual concept of the banana. As regards the former we consider two possibilities: (i) the word form that is available to A corresponds directly to the perceptual feature that his tasting of the banana has activated; (ii) the word form corresponds to a more abstract concept which subsumes the given feature as well as a number of other related but distinct features, and which is triggered whenever any one of these is. (In this second case the concept labeled by the word subsumes each of these features.) We suspect that a word like *sweet* as the normal English speaker uses it in classifying his taste experiences is associated with such a superordinated concept, but the matter is not crucial for what we want to say here.[3] Let W_{cat} be the word form which is associated with the categorical concept that enters the predicational judgment that A wants to communicate to B.

Before we turn to the word or words that are available for accessing the individual concept, let us first consider what information must be available to B in order that A's use of W_{cat} can serve his communicative intention. First, of course, W_{cat} must be a word for B. That is, B must have a lexicalized concept with W_{cat} as word form. Secondly, if A's use of W_{cat} is to produce the effect which A intends, then the perceptual concepts that are lexically linked with W_{cat} in B's brain must roughly match the perceptual concepts that are connected with W_{cat} in the brain of A. Exactly how close this match must be in order that A and B can be said to associate the same meaning with W_{cat} is a notoriously difficult question, to which we suspect no conclusive answer is possible.[4] However, when it comes to a word like *sweet*, we may expect that the

[3]The process of abstraction, which leads to concepts which subsume a number of different parts of a single perceptual space or subspace is an important aspect of concept formation and language acquisition and use. There are several ways in which abstraction can be implemented on the neural level.

[4]It has often been claimed, and we think with considerable plausibility, that people who *speak the same language* act on the presumption that they attach the same meanings to the words they use - it is their commitment to use the words with the meanings they have as words of the common language; this commitment has the effect of making word meaning to some extent *transcendent*. Still, this commitment can be upheld only when there is

perceptual concepts which two speakers associate with it are consistent in that by and large they are triggered by the same external stimuli. Let us assume that such is the case for the links between *sweet* and perceptual concepts in the brains of A and B.

We now turn to the question what word or words might be available to A to communicate to B which object it is to which he attributes sweetness. In principle there are various options here. One option that is found, it seems, in all natural languages is that of referring to the individual represented by one's individual concept by means of a proper name. We conjecture that names are associated with individual concepts in essentially the same way in which a word like *sweet* is associated with a categorical concept. For the case under discussion, however, a proper name seems an unlikely communicative vehicle. For in order that a proper name can function effectively in communication it is necessary that those who use it are members of a group within which there exists certain common knowledge regarding the entity named. Such knowledge presupposes that there has been considerable interaction within the group pertaining to this entity. Not only must there be enough members who have had an opportunity to acquire an individual concept representing the entity, but they must also have had the chance to establish that they share a concept for it. If those conditions are fulfilled, and moreover members of the group come to associate the same name (i.e. the same phonological string) with the individual concepts they have for the entity, then use of the name by one of them in the presence of another member can be expected to have the effect of triggering in that other member an activation of the individual concept with which the name is linked. In general, however, it takes time for those preconditions to be fulfilled, and it also requires an explicit act of *naming* on the part of somebody or some people within the group to get the name-individual concept association going. It is something that as a rule will happen only if the entity is long-lived and important enough to sustain this fairly complex social process. Thus, in the normal course of things a banana neither has the individual significance nor the longevity needed to become the bearer of a shared name. In particular, it is hard to see how in the case we are discussion a name for the given banana might have been established between A and B so that A can use it when informing B about the banana's taste.

A will thus have to make use of another way of referring to his banana - or, in our terms, another way of activating B's individual concept for this banana. Among the alternative means that English has available for this purpose there are complex definite descriptions and demonstratives, in which a noun (with or without additional adjectives, relative clauses and/or prepositional phrases) is preceded by *the* or by one of the demonstratives *this* and *that* respectively. For present purposes we won't distinguish between these three options - there is much that linguistics has to say about the differences between them, but that is not central to our present concern. Instead, we want to focus on the choice of the other part of the phrase, i.e. of the noun (with or without satellites). How is A to decide on the noun or nouns he is to use?

The choice will have to depend on what A assumes B knows about the individual to which he wants to draw B' attention. There are two cases that should be distinguished here. The first is the one where A assumes - and let us suppose he assumes correctly - that B already has an individual concept for the thing he wants to refer to. (For instance, A may have noted that B was observing him while he was tasting the banana.) In that case A will have to decide on one or more categorical concepts that (i) are part of B's individual concept; (ii) whose activation will trigger the activation of B's individual concept for the individual to which A means to refer; and

enough actual consistency to begin with between the semantic features within which different speakers associate the same word forms.

(iii) for which A and B share a lexicalization (i.e. share that they associate the same word forms with those concepts.

Suppose that A settles on a set of one or more such concepts and utters a definite description or a demonstrative in which the word forms for these concepts are conjoined. (Again, we ignore the question exactly what form this conjunction will take, e.g. whether any of the concept words will turn up in a relative clause and so on.) In order that this phrase (description or demonstrative) has the intended effect on B, the activation of the concepts which B associates with the concept words in this phrase must trigger the activation of B's individual concept for the banana of which they are part. Intuitively, the likelihood that this will happen would seem to depend on two different factors, (i) whether the lexically triggered concepts play a sufficiently salient part in the network which implements B's individual concept, and (ii) whether they play a comparable role in other individual concepts. (In order that the right individual concept be activated by the activation of its part concept or concepts there shouldn't be too much competition from other individual concepts to which the lexically activated concept or concepts also belong.)

Suppose that the descriptive or demonstrative phrase which A uses does activate B's individual concept for the banana in question. Then the joint appearance of this phrase with the categorical concept word *sweet* should have the effect of producing synchronized activation of the individual concept and the categorical concept, representing predication of the property represented by the latter of the object represented by the former. Under appropriate conditions this information will then be transferred into B's long time memory.

The second case to be considered is that where A assumes that B doesn't yet have an individual concept for the banana and wants him to set up such a concept at the same time as predicating sweetness of the object it represents. In this case the concept words that go into A's description or demonstrative must be able to guide B's attention to the object in question so that he can form an individual concept on the strength of his perception of it. In order for this to work, B must scan his perceptually accessible environment for perceptions which trigger the same concepts that have just been lexically accessed by the concept words in A's descriptive or demonstrative phrase. (Precisely how this detection works is yet another question for which we have no answer as it is.) Once B's perceptual attention has been drawn to the intended referent (by whatever mechanism that achieves this), then the association between it and the categorical concept accessed by *sweet* may be assumed to come about in the same way as in the previous case.

So far, nothing that we have put forward in this section touched on the difference between individual and categorical concepts. That question is still staring us in the face. In fact, after what has been said in this section it is possible to ask it in an even more pointed form: What is it about a concept that makes it into either a categorical or an individual concept for the one whose concept it is? One answer that comes to mind is that an individual concept is one which is so rich in content that it could not be satisfied by more than one thing: In the subsumption lattice of all of a person's concepts at any one time the individual concepts always occupy bottom positions. There seems to be an immediate objection to this suggestion: It appears that many of the individual concepts we actually entertain are quite poor in content. If you tell me: "I saw a man in the street yesterday." and you get called away to the phone before you can say anything else, the effect your utterance will presumably have on me is that I have set up an individual concept for the man you mentioned, but there is very little I know about the individual which this individual concept is supposed to represent. It is only natural for me to assume that there are many men you saw in the street yesterday. So the property of being a man seen by you in

the street yesterday almost certainly will not pick out the individual uniquely. In fact, suppose I happen to know that yesterday you were in the street at least twice, first in the morning on your way to work and then on your way from work in the evening, and that I consider it a practical certainty that you saw men in the street at both times. Then the concepts "man seen by you in the street yesterday on the way to work" and "man seen by you yesterday on the way from work", while each qualifying intuitively as a (non-empty) categorical concept, are both subsumed by the concept I have set up in response to your utterance.

Or so it might seem. But to a closer look the semblance disappears. My individual concept for the man you mentioned before you were called away isn't a concept for just any man you saw in the street yesterday, it is my concept for the person that *you were referring to when you said those words to me*. This is information that won't tell me much about the intrinsic properties of the man in question, but it may nevertheless uniquely identify him, even if it does so in some indirect way which won't help me to find out who he is (unless I rely on you as a source of further information).

We believe that it is this kind of *anchoring* information that in such cases distinguishes individual concepts from categorical ones and that places them necessarily at the bottom of any subsumption hierarchy irrespective of how much or little intrinsic information the concept may contain about the thing it represents. Usually such anchoring information takes the form of an explicit descriptive condition on the represented object (i.e. of a combination of categorical concepts, in the terminology used hitherto) of which it is (practically) certain that only one thing can satisfy it; an example are conditions to the effect of the thing having been located in a given place at a given time. But there are also cases where the anchoring condition is not explicit in the manner of this example. Thus, in the case just discussed, where my individual concept is based on your interrupted speech, the condition presumes there to exist *some* causal relation between me and a particular man, a relation which has been established via (i) your earlier perception of that man and (ii) your referring to that man when speaking to me. In this case the anchoring condition confers uniqueness upon the individual concept by other than straightforwardly descriptive means. There are even cases where this anchoring information cannot really be construed as pertaining to the represented individual in any straightforward sense at all. This happens for instance when an individual is introduced at a particular point in the course of a piece of fiction. The reader will at that point introduce, as part of his representation of the content of the story, an individual concept. And the concept will qualify as an individual concept insofar as one of the concepts it includes will be that the individual was introduced at such and such a point of the text, and/or with such and such words. Of course this information cannot be construed as a property of the (fictional) character described: it is not like his being characterized as a handsome prince, or a toad, or as having lived for most of his life in Baker Street. Nevertheless, as well as the information of having been mentioned by someone on a given occasion serves to uniquely identify a fictional character, it serves to identify an individual in the real world.

By attributing such a crucial role to anchoring concepts in the differentiation of individual from categorical concepts we may seem to have come close to just the kind of fundamental distinction in form which earlier on we we have been saying doesn't exist. Well, close perhaps, but definitely not there. Even if anchoring concepts are concepts of a special sort, they are concepts nevertheless, which a thing can satisfy, or fail to. So it remains true that between individual and categorical concepts there is no absolute difference, such as a difference in logical type.

4 Neural Computation

In this paper we have done no more than reflect about neuroscientifically motivated constraints that should be imposed on a theory of conceptual structure, of the formation of concepts and of their use in acts of predication. This is only a first and small step towards *neurosemantics*, as a bridging discipline between the neurosciences on the one hand and linguistic and philosophical semantics on the other, a discipline that we hope will get off the ground properly in the years ahead.

A next step, for which we have made some preparations ourselves and on which we intend to report in a follow-up to the present paper, is the development of a computer model of certain processes of concept formation, activation and combination similar to the ones that were informally discussed in the preceding sections. We conclude with some preliminary remarks on what this model will be like.

First, the neural activity that is to be modelled by our simulation will not simply be the average firing activity of a neuron, or neuron group, over time. Firing averages are a measure of neural activity that has been the focus of many simulation studies in recent times, but we do not think this measure provides enough differentiation to reflect the aspects of human cognition with which we are concerned. To model those it is necessary to simulate the firing behaviour of neurons in a more detailed way. In particular, the model ought to take account of the fact that the synapses which connects a neuron with other neurons are generally of two sorts, (i) excitatory and (ii) inhibitory. A good compromise in modelling this aspect of what causes the propagation of firing patterns through networks of neural cells is, it seems to us, the so-called *leaky-integrator neuron*. The leaky-integrator neuron is a neuron model which simulates the way in which the cell membrane integrates excitatory and inhibitory synaptic activity over time, and which *leaks* (i.e. looses) the integrated electrical potentials when there is no input.

The membrane potential of a brain area can be given by a vector $\mathbf{m}(t)$. The synaptic strength within and between areas can be appoximated by a so called *weight matrix* $\mathbf{W}(t)$. The state of a brain area r at time t will then be modelled as determined by the membrane potential vector $\mathbf{m}_r(t)$ and the area's weight matrix $\mathbf{W}_r(t)$, thus as the pair $< \mathbf{m}_r(t), \mathbf{W}_r(t) >$. The state of the entire brain at t, $A(t)$, is composed of the states of its regions at t together with further *distal* weight matrices $\mathbf{W}_{r,r'}(t)$ which represent the strengths of the connections between the cells in area r and those in area r'.

$$A(t) =< \{< \mathbf{m}_{r1}(t), \mathbf{W}_{r1}(t) >, ..., < \mathbf{m}_{rn}(t), \mathbf{W}_{rn}(t) >\}, \{\mathbf{W}_{r,r'}(t) : r = r_1, ..., r_n; r' = r_1, ., r_n\} > \quad (1)$$

The model assumes that $A(t+1)$, the state of the brain at time $t+1$, depends only on the state at time t, $A(t)$, and on the sensory input $i(t)$ which reaches the system at t. The transition can be schematically expressed as in (2)

$$A(t+1) = S(A(t), i(t)) \quad (2)$$

The function S() can be made explicit. The state of the membrane potentials of the neurons

in a region r can be computed from the input to r and the neural activity within r. The input will consist wholly or partly of sensory input for those regions which lie directly at the sensory periphery, while for regions that are not directly connected with sensory organs it will consist entirely of input from other regions. The neural activity within a region (i.e. the *spike* activity $\mathbf{F}(t)$ at t in each of the different regions is a function of the membrane potential in r at t.

Within this framework the acquisition of individual and categorical concepts can be modelled by the way in which brain areas whose behaviour is governed by (2) react to certain patterns of sensory input. The model assumes that the synaptically determined strengths of links between clusters is changed according to Hebb's rule (3),

$$\Delta w_{ij} = c u_i u_j \tag{3}$$

where u_i is the pre-syntaptic, u_j the post-syntaptic neuron, and c is a variable determining the learning rate.

Here is some of the behaviour which we expect our model will display:

(i) Certain sensory input patterns produce very strong cluster connections. (We connect with this expected behaviour the hypothesis that the capacity for strongly selective reactions to patterns of certain types is the actual basis for the acquisition of categorical and individual concepts in creatures with brains like ours.)

(ii) At the same time the model will also be able to form individual concepts involving comparatively weak connections between the cell groups involved. Such concepts will be hard to activate and their activation will require joint activaton either of a large number of the features that are part of them or of one or more features which are distinctive of the concept in that they have even weaker (if any) links to other concepts.

(iii) In line with what has been suggested in this paper, lexicalisation will be modelled as the association of non-verbal concepts and lexical items (or *word forms*). An additional task we see for the model is that of accounting for the phenomenon of being *linguistically switched on* (or *off*): Much of the time we make use of lexicalized perceptual concepts without the corresponding words actually coming to our mind. If one our basic assumptions - that the very same cell clusters that are involved in perception are also part of the corresponding lexicalised concepts - is right then the question arises: How come that the word which lexicalises a perceptual concept is triggered sometimes, when the concept is perceptually accessed, but not at other times? We hope to elicit from the model behaviour which simulates this phenomenon - sometimes the lexical item is activated when the corresponding concept is, sometimes it is not. To this end we will add some threshold mechanism which prevents lexical access unless communicative intent provides the additional input potential that the representation of the lexical item needs on top of the input it gets from the perceptual concept which it lexicalizes.

(iv) The model should demonstrate the conjoined activity of *individual* and *categorical* concepts in response to simulated sensory input from the represented individual and together with input that qualifies as representing the given property. In this way it will, if our expectations concerning it are confirmed, provide support for the thesis that at the neural level there is no difference in type between the partners in predication.

A computer model of this kind, though very much more concrete than the informal considerations of the present paper, will still only be a modest step in the direction of a neuroscientifically tenable account of the immensely complex and sophisticated cognitive processes of which human beings are capable, and of which their ability to put thoughts into words and get thoughts out of words are perhaps the most impressive manifestations. But it would be some step, and we are determined to take it.

References

Artola, A., Brocher, S., and Singer, W. (1990). Different voltage dependend thresholds for inducing long-term depression and long-term potentiation in slices of rat visual cortex. *Nature*, 347:69–72.

Bliss, T. V. P. and Lomo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthesized rabbit following stimulation of the perforatant path. *Journal of Physiology*, 232:331–356.

Clark, E. V. (1973). What's in a word. In Moore, T., editor, *Cognitive Development and the Acquisition of Language*. New York: Academic Press.

Eckhorn, R., Bauer, R., Jordan, W., Bosch, M., Kruse, W., Munk, M., and Reitboeck, H. J. (1988). Coherent oscillations: A mechanism of feature linking in the visual cortex? *Biological Cybernetics*, 60:121–130.

Frege, G. (1879). *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Nebert, Halle.

Fuster, J. M. (1995). *Memory in the Cerebral Cortex*. MIT Press.

Gray, Koenig, Engel, and Singer (1989). Oscilatory responses in cat visual cortex exibit intercolumnar synchronization which reflects global stimulus properties. *Nature*.

Hebb, D. O. (1949). *The Organization of Behaviour*. Wiley New York.

Hubel, D. and Wiesel, T. (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160:106–154.

Klein, M. and Billard, A. (2001). Words in the cerebral cortex - predicting fmri-data. In *Proceedings of the 8th Joint symposium on neural computation - The brain as a dynamical system, San Diego*.

Koch, C. (1999). *The Biophysics of Computation: Information Processing in Single Neurons*. Oxford University Press.

Lakoff, G. (1987). *Woman, Fire, and Dangerous Things*. University of Chicago Press.

Rosch, E. (1973). Natural categories. *Cognitive Psychology*, 4:328–350.

von der Malsburg, C. (1981). The correlation theory of brain function. Technical Report 81-2, Abteilung fuer Neurobiologie, MPI fuer biophysikalische Chemie, Goettingen.

von der Malsburg, C. (1985). Nervous structures with dynamical links. *Ber. Bunsenges. Phys. Chem*, 89:703–710.

Wennekers, T. and Palm, G. (1997). On the relation between neural modelling and experimental neuroscience. *Theory in Bioscience*, 116:273–289.