Conceptual Spaces as a Basis for Cognitive Semantics

Peter Gärdenfors

Cognitive Science
Department of Philosophy
Lund University
S-223 50 Lund, Sweden

E-mail: Peter.Gardenfors@fil.lu.se

1. Program

As an introduction, I want to contrast two general traditions in semantics, one *realistic* and one *cognitive*. According to the realistic approach to semantics the meaning of an expression is something out there *in the world*. In technical terms, a *semantics* for a language is defined as a mapping from the syntactic structures to things in the world (or in several possible worlds). Often meanings are defined in terms of *truth conditions*. A consequence of this approach is that the meaning of an expression is independent of how individual users understand it.

The second paradigm of semantics is conceptualistic or cognitivistic. The central tenet of this approach is that meanings of expressions are *mental entities*. A semantics is seen as a mapping from the linguistic constituents to *cognitive structures*. The external world enters the scene only when the relation between it and the cognitive structure is considered. According to this kind of semantic theory the relation between meanings and the external world is secondary, and only determined after the cognitive structures have been settled. As a consequence, meaning becomes independent of truth.

In this paper, I first give a sketch of a realist semantics in the form of standard intensional semantics, which is formulated in terms of possible worlds. I shall also mention some of the philosophical problems this kind of semantics leads to. Then I present some of the main tenets of what has become known as cognitive semantics. As an ontological framework for a cognitive semantics, I introduce the notion of a *conceptual space* and show how such spaces can be used as a basis for a cognitive semantics. In the final sections, it will be argued that this kind of semantics is useful for understanding *metaphors* and *prototype effects* of concepts.[1]

2. Intensional semantics and its problems

A typical example of a realistic semantic theory is the so called *intensional semantics*. As an analysis of natural language it reaches its peak with Montague (1974). Here a language is mapped onto a *set of possible worlds*. Apart from truth values, possible worlds and their associated sets of individuals are the only primitive semantical elements of the model theory. Other semantical notions are defined as *functions* on individuals and possible worlds. For example, a *proposition* is defined as a function from possible worlds to truth values. Such a function thus determines the *set of worlds* where the proposition is true. According to traditional intensional semantics, this is all there is to say about the meaning of a proposition.

---

[1]Sections 2 and 4-7 in this paper are based on Gärdenfors (1991) and (1992).

As a typical example of the analysis within this tradition let us look at the notion of a *property*. In intensional semantics, a property is something that relates individuals to possible worlds. In general terms, a property can be seen as a many-many relation *P* between individuals and possible worlds such that *iPw* holds just when individual *i* has the property in world *w*. Such a relation is illustrated in Figure 1.

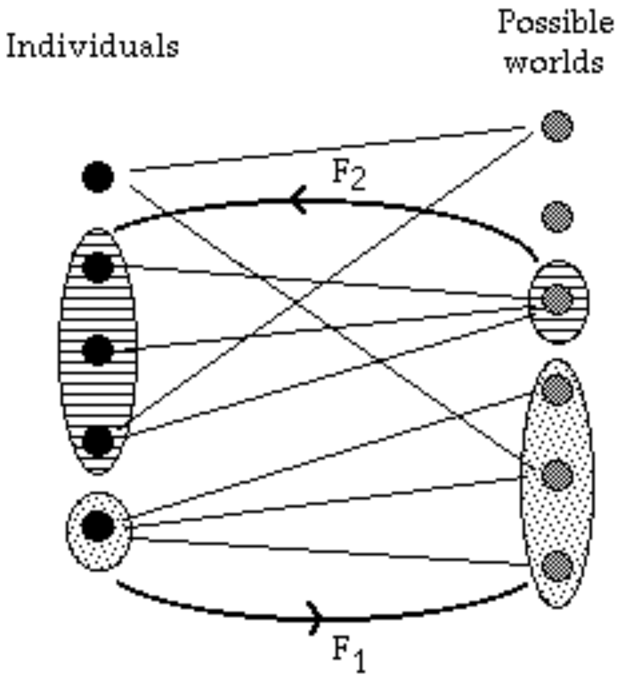

Figure 1.
A property as a many-many relation
between individuals and possible worlds.

In intensional semantics, functions are preferred to many-many relations. There are two ways of turning the relation *P* into a function: Firstly, it may be described as a *propositional function*, i.e. *a function from individuals to propositions*. Since a proposition is identified with a set of possible worlds, this means that a property is a rule which for each individual determines a corresponding set of possible worlds. But we can also turn the table around to get an equivalent function out of *P*: for each possible world *w*, a property will determine a set of individuals which has *w* as an element of the sets of possible worlds the individuals are assigned (cf. Figure 1). This means that an equivalent definition of a property is that it is *a function from possible worlds to sets of individuals*.

In Gärdenfors (1991b), it is argued that the standard definition of a property within intensional semantics leads to a number of serious problems. First of all, the definition is highly counterintuitive since properties become very abstract things. The definition is certainly not helpful for cognitive psychologists who try to explain what happens when a person *perceives* that two objects have the same property in common or, for example, why certain colors look *similar*.

A related, but more serious, problem for the traditional definition of a property is that it can hardly account for *inductive reasoning*. An inductive inference generally consists in *connecting* two properties to each other. This connection is obtained from a number of instances of individuals exhibiting the relevant properties. If a property is defined as a function from possible worlds to set of individuals, then in order to determine which properties are instantiated by a particular individual (or a set of individuals), one has to determine *which functions* have the individual (or the set of individuals) as value in the actual world. Apart from problems concerning how we determine which is the actual world, this recipe will in general give us *too many* properties. For example, if we are examining a particular emerald it will instantiate a large class of Goodman-type properties like 'grue' apart from standard properties like 'green'. If the only thing we know about properties is that they are some kind of abstract functions, then we have no way of distinguishing natural and inductively projectible properties like 'green' from inductively useless properties like 'grue'. What is needed is a criterion for separating the projectible sheep from the non-projectible goats. However, classical intensional semantics does not provide us with such a criterion.[2]

The final problem that I shall point out for the functional definition of properties is perhaps the most serious one. Putnam (1981) has shown that the standard model-theoretic definition of 'property' which has been given here does not work as a theory of the *meaning* of properties. Putnam concludes that "there are always inifinitely many different interpretations of the predicates of a language which assign the 'correct' truth-values to the sentences in all possible worlds, *no matter how these 'correct' truth-values are singled out*" (1981, p. 35).

Although I have here only mentioned the notion of a property and some of the problems the realist semantics leads to, I believe that there are many other problems for this type of semantics. Some of these problems will be apparent later in this paper. But perhaps the best source is Lakoff's book (1987), which is a lengthy criticism of what he calls "objectivist semantics". The problems for the traditional semantics justifies a search for a fundamentally different kind of semantics.

3. Cognitive semantics

As an alternative approach, I shall give a programmatic presentation of what has become known as *cognitive semantics*. My presentation will be in the form of six slogans, with some comments, where the approach of a cognitively oriented semantics will be contrasted with the more traditional view. Prime examples of works in this tradition are Lakoff's (1987) and Langacker's (1986). Related versions of cognitive semantics can be found in the writings of Jackendoff (1983, 1990), Johnson-Laird (1983), Fauconnier (1985), Talmy (1988), Sweetser (1990) and many others. There is also a French semiotic tradition, exemplified by Desclés (1985) and Petitot-Cocorda (1985), which shares many features with the American (mainly Californian) group.

I. Meaning is *conceptualization* in a cognitive model (not truth conditions in possible worlds).

The prime slogan for cognitive semantics is: *Meanings are in the head*. More precisely, a semantics for a language is seen as a mapping from the expressions of the language to some cognitive or mental entities. A consequence of the cognitivist position that puts it in conflict with many other semantic theories is that no form of truth conditions of an expression is necessary to determine its meaning. The truth of expressions is considered to be secondary since truth concerns the relation between a cognitive structure and the world. To put it tersely: *Meaning comes before truth.*

---

[2]For further discussion of the problems of the traditional account of properties in connection with induction, cf. Gärdenfors (1990) and (1991).

Cognitive semantics should be separated from Fodor's (1981) "Language of Thought" hypothesis. There are similarities, though: Fodor also uses mental entities to represent linguistic information. This is his 'language of thought' which is sometimes also called 'Mentalese'. According to Fodor, this is what speakers use when they compute inferences (according to some internal set of rules) and when they formulate linguistic responses (translated back from Mentalese to some appropriate natural language). However, the mental entities constituting Mentalese form a *language* with syntactic structures goverened by some recursive set of rules. And when it comes to the *semantics* of Mentalese, Fodor still is a realist and relies on references in the external world as well as truth conditions.

II. Cognitive models are mainly *perceptually* determined (meaning is not independent of perception).

Since the cognitive structures in our heads are connected to our perceptual mechanisms, directly or indirectly, it follows that *meanings are*, at least partly, *perceptually grounded*. This, again, is in contrast to traditional realist versions of semantics which claim that since meaning is a mapping between the language and the external world (or several worlds), meaning has nothing to do with perception.

We can talk about what we see and hear. Conversely, we can create pictures, mental or real, of what we read or listen to. This means that we can translate between the visual form of representation and the linguistic code.[3] A central hypothesis of cognitive semantics is that the way we store perceptions in our memories has the *same form* as the meanings of words. Another consequence of the coupling of perceptual representation and meaning is that meaning has *ecological validity*.

III. Semantic elements are based on *spatial* or *topological* objects (not symbols that can be concatenated according to some system of rules).

In contrast to the Mentalese of Fodor and others, the mental structures applied in cognitive semantics *are* the meanings of the linguistic idioms; there is no further step of translating conceptual structure to something outside the mind. Furthermore, instead of being a symbolic system having syntactic structure like Mentalese, the conceptual schemes that are used to represent meanings are often based on *geometric* or *spatial* constructions.

The most important semantic structure in cognitive semantics is that of an *image schema*. Image schemas have an inherent spatial structure. Lakoff (1987) and Johnson (1987) argue that schemas such as 'container', 'source-path-goal' and 'link' are among the most fundamental carriers of meaning. They also claim that most image schemas are closely connected to *kinesthetic* experiences.

IV. Cognitive models are primarily *image-schematic* (not propositional). Image-schemas are transformed by *metaphoric* and *metonymic* operations (which are treated as exceptional features on the traditional view).

Metaphors and metonymies have been notoriously difficult to handle within realist semantic theories. In these theories these linguistic figures have been treated as a deviant phenomenon that has been ignored or incorporated via special stylistic rules. In contrast, they are given key positions within cognitive semantics. Not only poetic metaphors but also everyday 'dead' metaphors are seen

---

[3]For a discussion of the implication sfor semantics of this translatability, cf. Jackendoff (1987).

as central semantic features and are given systematic analyses. One of the first works in this area was Lakoff and Johnson (1980).

Metaphors and metonymies are primarily seen as *cognitive* operations, and their linguistic expression is only a secondary phenomenon. They are analysed as *transformations* of image schemas. As such they are connected to spatial codings of information. In particular, Lakoff (1987, p. 283) puts forward what he calls the '*spatialization of form hypothesis*' which says that conceptual forms are understood in terms of spatial image schemas plus a metaphorical mapping. For example, many uses of prepositions are seen as metaphorical (see e.g. Brugman (1981) and Herskovits (1986))

V. *Semantics* is primary to syntax and partly determines it (syntax cannot be described independently of semantics).

This thesis is anathema to the Chomskyan tradition within linguistics. Within Chomsky's school, grammar is a *formal calculus*, which can be described via a system of rules, where the rules are formulated independently of the meaning of the linguistic expressions. Semantics is something that is added, as a secondary independent feature, to the grammatical rule system. Similar claims are made for pragmatic aspects of language.

Within cognitive linguistics, semantics is the primary component (which, in the form of perceptual representations, existed before language was fully developed). The structure of the semantic schemas put constraints on the possible grammars that can be used to represent those schemas. To give a trivial example of how semantics determines syntax, consider the role of *tenses*. In a Western culture where time is conceived of as a line, it is meaningful to talk about three basic kinds of time: past, present and future. This is reflected in the grammar of tenses in most languages. However, in cultures where time has a circular structure, or where time cannot be given any spatial structure at all, it is not *possible* to make a distinction between, say, past and future. And there are languages which have radically different tense structures, which reflect a different underlying conceptual structuring of time.

VI. Concepts show *prototype* effects (instead of following the Aristotelian paradigm based on necessary and sufficient conditions).

The classical account of concepts within philosophy is Aristotle's theory of *necessary and sufficient conditions* .[4] His view on how concepts are determined has had an enormous influence throughout the history of philosophy. During this century the Aristotelian notions became part of the program of the logical positivists who demanded that all scientific concepts should ideally be *defined* in terms of a limited number of observational terms. If a concept can't be defined by necessary and sufficient conditions, it is not a proper scientific concept, at least according to the early positivist program.

However, one very often encounters problems when trying to apply the Aristotelian theory. As a result of a growing dissatisfaction with the classical theory of concept theory, an alternative theory was developed within cognitive psychology. This is the called *prototype theory* where Eleanor Rosch is one of the main proponents.[5] The main idea of prototype theory is that within a category of

---

[4]See Smith and Medin (1981) for a presentation of this and other theories of concept formation.

[5] See e.g. Rosch (1975), (1978), Mervis and Rosch (1981), Smith and Medin (1981), and Lakoff (1987) for extended discussions of the theory.

objects, like those instantiating a property, certain members are judged to be more representative of the category than others. For example, robins are judged to be more representative of the category 'bird' than are ravens, penguins and emus; and desk chairs are more typical instances of the category 'chair' than rocking chairs, deck-chairs, and beanbag chairs. The most representative members of a category are called *prototypical* members.

Another thesis of prototype theory is that categories are not organized just in terms of simple taxonomic hierarchies. Instead, a 'middle' kind of concepts can be distinguished, which is called the *basic level* of the categorization. Higher levels are called *superordinate* and lower *subordinate*. For example, 'chair' and 'dog' are basic level concepts, while 'furniture' and 'mammal' are superordinate concepts and 'armchair' and 'dachshund' are subordinate. The basic level is characterized by a number of features: (1) It is the highest level at which category member have similarly preceived overall *shapes*, (2) it is the highest level at which a person uses similar actions for handling category members, (3) it is the level at which subjects are fastest at identifying category members, and (4) it is the first level named and understood by children.

Within cognitive semantics, one attempts to account for prototype effects of concepts. A concept is often represented in the form of an image schema and such schemas can show variations just like birds and chairs. This kind of phenomenon is extremely difficult to model using traditional symbolic structures.

4. Conceptual spaces as a framework for a cognitive semantics

After this presentation of some of the central tenets of cognitive semantics, I now want to turn to the *ontology* of such a semantics. As a framework for a cognitive structure used in describing a semantics I want to put forward the notion of a *conceptual space*. A conceptual space consists of a number of *quality dimensions*. As examples of quality dimensions let me mention color, pitch, temperature, weight, and the three ordinary spatial dimensions. The dimensions are taken to be cognitive and infra-linguistic in the sense that we can represent the qualities of objects without presuming an internal language in which these qualities are expressed. Some of the dimensions are closely related to what is produced by our sensory receptors, but there are also quality dimensions that are of an abstract non-sensory character.

The notion of a *dimension* should be understood literally. It is assumed that each of the quality dimensions is endowed with certain topological or metric structures. For example, 'time' is a one-dimensional structure which we conceive of as being isomorphic to the line of real numbers. Similarly, 'weight' is one-dimensional with a zero point, isomorphic to the half-line of non-negative numbers. Some quality dimensions have a *discrete* structure, i.e., they merely divide objects into classes, e.g., the sex of an individual.

A psychologically interesting example of a quality dimension concerns *color perception*. In brief, our cognitive representation of colors can be described by three dimensions (see Figure 2). The first dimension is *hue*, which is represented by the familiar *color circle*. The topological structure of this dimension is thus different from the quality dimensions representing time or weight which are isomorphic to the real line. The second psychological dimensions of color is *saturation*, which ranges from gray to increasingly greater intensities. This dimension is isomorphic to an interval of the real line. The third dimension is *brightness* which varies from white to black and is thus a linear dimension with end points. Together these three dimensions, one with circular structure and two with linear, make up the color space which is a subspace of our perceptual conceptual space.

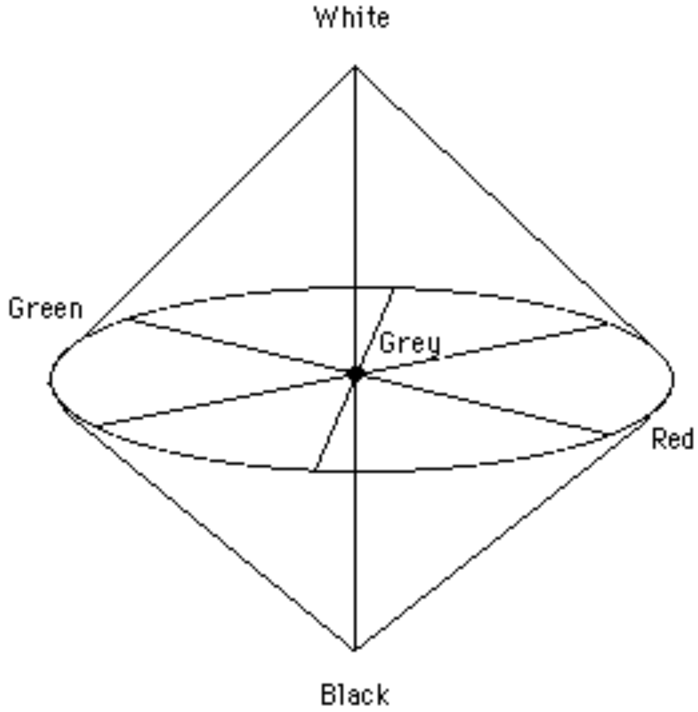

Figure 2.
The full color space

I cannot provide a complete list of the quality dimensions involved in our conceptual spaces. Some of the dimensions seem to be *innate* and to some extent hardwired in our nervous system, as for example color, pitch, and probably also ordinary space. Other dimensions are presumably *learned*. Learning new concepts often involves expanding one's conceptual space with new quality dimensions. *Functional* properties used for describing artifacts may be an example here. Still other dimensions may be *culturally* dependent. 'Time' is a good example – in contrast to our linear conception of time, some cultures conceive of time as circular so that the world keeps returning to the same point in time, and in other cultures it is hardly meaningful at all to speak of time as a dimension. Finally, some quality dimensions are introduced by *science*.

This concludes my general presentation of conceptual spaces.[6] There is a strong similarity between the notion of a conceptual space and the *domains* as used in Langacker's (1986) semantic theory. The following quotation from Langacker (1986, p. 5) concerning his notion of 'domains' strongly supports this thesis:

> "What occupies the lowest level in conceptual hierarchies? I am neutral in regard to the possible existence of conceptual primitives. It is however necessary to posit a number of 'basic domains,' that is, cognitively irreducible representational spaces or fields of conceptual potential. Among these basic domains are the experience of time and our capacity for dealing with two- and three-dimensional spatial configurations. There are basic domains associated with various senses: color space (an array of possible color sensations), coordinated with the extension of the visual field; the pitch scale; a range of possible

[6]For further details of the theory of conceptual spaces, cf. Gärdenfors (1988), (1990), (1991) and (1992), where also different applications of the theory can be found.

> temperature sensations (coordinated with positions on the body); and so on. Emotive domains must also be assumed. It is possible that certain linguistic predications are characterized solely in relation to one or more basic domains, for example time for (BEFORE), color space for (RED), or time and the pitch scale for (BEEP). However, most expressions pertain to higher levels of conceptual organization and presuppose nonbasic domains for their semantic characterization."

The theory of conceptual spaces is a *theory for representing information*, not an empirical psychological or neurological theory, which I believe can be applied to a number of philosophical problems in epistemology and semantics. Here, my primary aim is to show its viability as a foundation for cognitive semantics.

5. Cognitive semantics based on conceptual spaces

I can only outline the first steps in developing a cognitive semantics based on conceptual spaces. According to the cognitive view, semantics is a relation between langauge and a cognitive structure. I submit that the appropriate framework for the cognitive structure is a conceptual space. This means that formulating a semantics for a specific language is to specify the mapping between the lexicon of the language and a conceptual space and to describe the operations on the image schemas defined on the conceptual corresponding to syntactic formation rules.

Slightly more technically, we can define an *interpretation* for a language *L* as a mapping of the components of *L* onto a conceptual space. As a first element of such a mapping, *individual names* are assigned vectors (i.e., points in the conceptual space) or partial vectors (i.e., points with some arguments undetermined). In this way, each name (referring to an individual) is allocated a specific color, spatial position, weight, temperature, etc.[7] If a name is assigned a partial vector, this means that not all the properties are known or have been determined. Following Stalnaker (1981, p. 347), a function which maps the individuals into a conceptual space will be called a *location function*.

As a second element of the interpretation mapping, the *predicates* of the language that denote primary properties are assigned regions in the conceptual space. (In Gärdenfors 1990, 1991 it is argued that the regions correponding to natural predicates are *convex.*) Such a predicate is *satisfied* by an individual just in case the location function locates the individual at one of the points included in the region assigned to a predicate. Some of the so called intensional predicates, like 'tall', 'former' or 'alleged', do not denote primary properties in the sense that their regions can be described independently of other properties. Such secondary predicates, which are 'parasitical' on other properties, can be described in terms of the regions assigned to the primary properties. *Relations* (primary and secondary) can be treated in a similar way.

If we assume that an individual is completely determined by its set of properties, then all points in the conceptual space can be taken to represent *possible individuals.* On this account, a possible individual is a *cognitive* notion that need not have any form of reference in the external world. This construction will avoid many of the problems that have plagued other philosophical accounts of possible individuals. A point in a conceptual space will always have an internally consistent set of properties – since, for example, 'blue' and 'yellow' are disjoint properties in the color space, it is not possible that any individual will be both blue and yellow (all over). There is *no need for meaning postulates* or their ilk in order to exclude such contradictory properties.

---

[7]Abstract entities may be assigned values on a different set of quality dimensions.

One important contrast to the traditional intensional semantics is that the one outlined here does not presume the concept of a *possible world*. However, different location functions describe alternative ways that individuals may be located in a conceptual space. Thus, these location functions have the same role as possible worlds in the traditional semantics. This means that we can *define* the notion of a possible world as a possible location function, and this can be done without introducing any new semantical primitives to the theory.

If we assume that the meanings of the predicates, among other things in a language *L* , are determined by a mapping into a conceptual space *S*, it follows from the topological structure of different quality dimensions that certain statements will become *analytically* true (in the sense that they are independent of empirical considerations). For example the fact that comparative relations like 'earlier than' are *transitive* follows from the linear structure of the time dimension and is thus an analytic feature of this relation (analytic-in-*S*, that is). Similarly, it is analytic that everything that is green is colored (since 'green' refers to a region of the color space) and that nothing is both red and green. Analytic-in-*S* is thus defined on the basis of the topological and metric structure of the conceptual space *S*. A consequence of this definition is that an analytic statement will be satisfied for all location functions. However, different conceptual spaces will yield different notions of analyticity.

With the aid of the notion of a conceptual space, I have tried to show how the slogans of cognitive semantics presented in Section 3 can be given some substantial content. That meaning is conceptualization (slogan I) is pretty obvious, given that the framework of a semantics is a conceptual space. Since many of the dimension of a conceptual space are directly connected to preceptual mechanisms, this also shows that the relevant cognitive models are perceptually determined (slogan II). And conceptual spaces are, by definition spatial, and not symbolic (slogan III). Of course, developing the mappings from an actual natural language to a cognitively realistic conceptual space is a Herculean task. Some first steps towards the completion of this task has been taken by the linguists in this tradition, as e.g. in the works by Langacker (1986), Lakoff (1987), Jackendoff (1990), and Talmy (1988).

6. Metaphors

As a way of filling out slogan IV within a cognitive semantics based on conceptual spaces, let us look at the way metaphors work. This is a problem which has been notoriously difficult to handle within realist semantic theories. In these theories, metaphors have been treated as a deviant phenomenon that should be ignored or incorporated via special rules. The view within cognitive semantics is that metaphors should be treated on par with all other semantic processes, or perhaps even as one of the central semantic features of language. Here I will present a summary of the theory of metaphors outlined in Gärdenfors (1992).

The core hypothesis is that *a metaphor expresses a similarity in topological or metrical structure between different quality dimensions*. A word that represents a particular structure in one quality dimension can be used as a metaphor to express a similar structure about another dimension. In this way one can account for how a metaphor can *transfer knowledge* about one conceptual dimension to another.

As a simple example, let us consider words that refer to the length dimension, like 'longer', 'distant', 'in front of', and 'forward'. This dimension refers to the most salient direction of the two-dimensional surface we are normally moving on. Unless altered by the communicative context, the default direction of this dimension is determined by the speaker's front and back. The spatial length dimension is represented by a topological structure that is isomorphic to the real line, where we, for the present purposes, can take the zero point to represent 'here'.

In our conceptual space (modern Western), the time dimension has the same structure as the real line. According to the hypothesis about how metaphors work we can then use some of the words we use to talk about length when we want to say something about time. In support of the hypothesis it can be noted that we speak of 'longer' and 'shorter' intervals of time, a 'distant' future; and we say that we have some tasks 'in front of' us, that some events are 'behind' us, and that we are looking 'forward' to doing something. Here the structure underlying the length dimension is transferred to the time dimension and we know what the words mean as expressions about time since we can identify the corresponding structure on the conceptual time dimension.

I propose that the length dimension is the more fundamental one and these expressions are thus used metaphorically for the time dimension. This may be difficult to see since these expressions about time are so idiomatic in our language that we no longer think of them as metaphors. However, their origin as metaphors can be highlighted by comparing our time expressions to those of other cultures. We need not go very far from the standard Western view of time; a particularly revealing example can be found in the ancient Greek conception of time. The Greeks thought of time as a river flowing past us. We sit in the river with our backs towards the future and see the events pass by, become distant and eventually disappear in oblivion. But we do not see what is coming. Indeed, one Greek word for 'future' (οπισφε) is the same as the word for 'behind'! In contrast, we think of time rather as a road we are travelling along in which the past is 'behind' us and we are looking 'forward' to the future. Both the ancient Greeks and we view time as a one-dimensional line (in contrast to a circular conception of time as in some cultures) – the only difference is the 'direction' of the line.

This is an example of how a study of the basic metaphors of a language can reveal the structure of the underlying conceptual dimensions. Another linguistic category that is essentially metaphorical is the class of *prepositions*. Words like 'in', 'at', 'on', 'under' etc. originate in spatial metaphors and when combined with non-locational words they create a 'spatially structured' mental representation of the expression. Herskovits (1986) presents an elaborated study of the fundamental spatial meanings of prepositions and she shows how the spatial structure is transferred in a metaphoric manner to other contexts. A sentence like "We meet *at* six o'clock" provides a further illustration to the dependence of temporal language on spatial dimensions (Herskovits (1986), p. 51). Here "six o'clock" is conceived as a point on a travel trajectory, and the locational preposition 'at' is used in exactly the same way as in "The train is at the bridge".

The theory presented here seems to go along the same lines as the one developed by Lakoff and Johnson (1980). They analyse several networks of metaphors used to talk about special topics. Among other things, they argue, in line with the description above, that the introduction of a new metaphor *creates* similarities of a new kind. These similarities are not 'objective', but, once one quality dimension has been connected to another via a metaphor, this connection may serve as a generator for new metaphors based on the same kind of similarity.

A closely related point is raised by Tourangeau and Sternberg (1982). Their 'domains-interaction' view is based on the observation that "metaphors often involve seeing in a new way not only two particular things but the domains to which they belong as well. ... Metaphors can thus involve whole systems of concepts" (p. 214). In other words, *a metaphor does not come alone* – it is not only a comparison between two single concepts, but involves an identification of the structure of two quality dimensions. Black (1979, p. 31) makes essentially the same point by the phrase "Every metaphor is the tip of a submerged model."

Tourangeau and Sternberg's analysis is obviously congenial to the present one. They even use the notion of 'dimension' when spelling out their view. Also Indurkhya (1986) interprets metaphors in terms of mappings between different domains. However, the topological structure of the domains is not exploited.

These arguments can only indicate the general direction of a systematic analysis of metaphors. Further examples and an extended analysis can be found in Gärdenfors (1992). I hope they show that an analysis of metaphors in terms of similarities of topological structures between dimensions is a promising program.

7. Prototype theory

Finally, in order to say something about how slogan VI can be treated within a cognitive semantics, I want to show that describing properties as convex regions of conceptual spaces fits very well with the so called *prototype theory* of categorization developed by Rosch and her collaborators (Rosch 1975, 1978, Mervis and Rosch 1981, Lakoff 1987). The main idea of prototype theory is that within a category of objects, like those instantiating a property, certain members are judged to be more representative of the category than others. The most representative members of a category are called *prototypical* members.

Now, if a traditional definition of a property is adopted, it is very difficult to explain such prototype effects. Either an object is a member of the class assigned to a property (relative to a given possible world), or it is not, and all members of the class have equal status as category members. Rosch's research has been aimed at showing asymmetries among category members and asymmetric structures within categories. Since the traditional definition of a property neither predicts nor explains such asymmetries, something else must be going on.

In contrast, if properties are described as convex regions of a conceptual space, prototype effects are indeed to be expected. In a convex region one can describe positions as being more or less *central*. For example, if color properties are identified with convex subsets of the color space, the central points of these regions would be the most prototypical examples of the color. In a series of experiments, Rosch has been able to demonstrate the psychological reality of such 'focal' colors.

For more complex categories like 'bird' it is perhaps more difficult to describe the underlying conceptual space. However, if something like Marr and Nishihara's (1978) analysis of shapes is adopted, we can begin to see how such a space would appear.[8] Their scheme for describing biological forms uses hierarchies of cylinder-like modelling primitives. Each cylinder is described by two coordinates (length and width). Cylinders are combined by determining the angle between the dominating cylinder and the added one (two polar coordinates) and the position of the added cylinder in relation to the dominating one (two coordinates). The details of the representation are not important in the present context, but it is worth noting that on each level of the hierarchy an object is described by a comparatively small number of coordinates based on lengths and angles. Thus the object can be identified as a hierarchially structured vector in a (higher order) conceptual space.

---

[8]This analysis is expanded in Marr (1982), Ch. 5.

Figure 3 provides an illustration of the hierarchial structure of their representations.

Furthermore, a 'prototypical' vector for an object category like 'bird' identifies a spatial structure that can serve as an image schema for that category. Such an image schema represents a *basic level* category (cf. Section 3), while subordinate categories like 'ostrich' are represented by subregions of the convex region associated with the prototypical object. Superordinate categories like 'animal' do not have any associated image schemas. This way of representing object categories can form a foundation for an explanation of many of the characteristics of basic level categories.

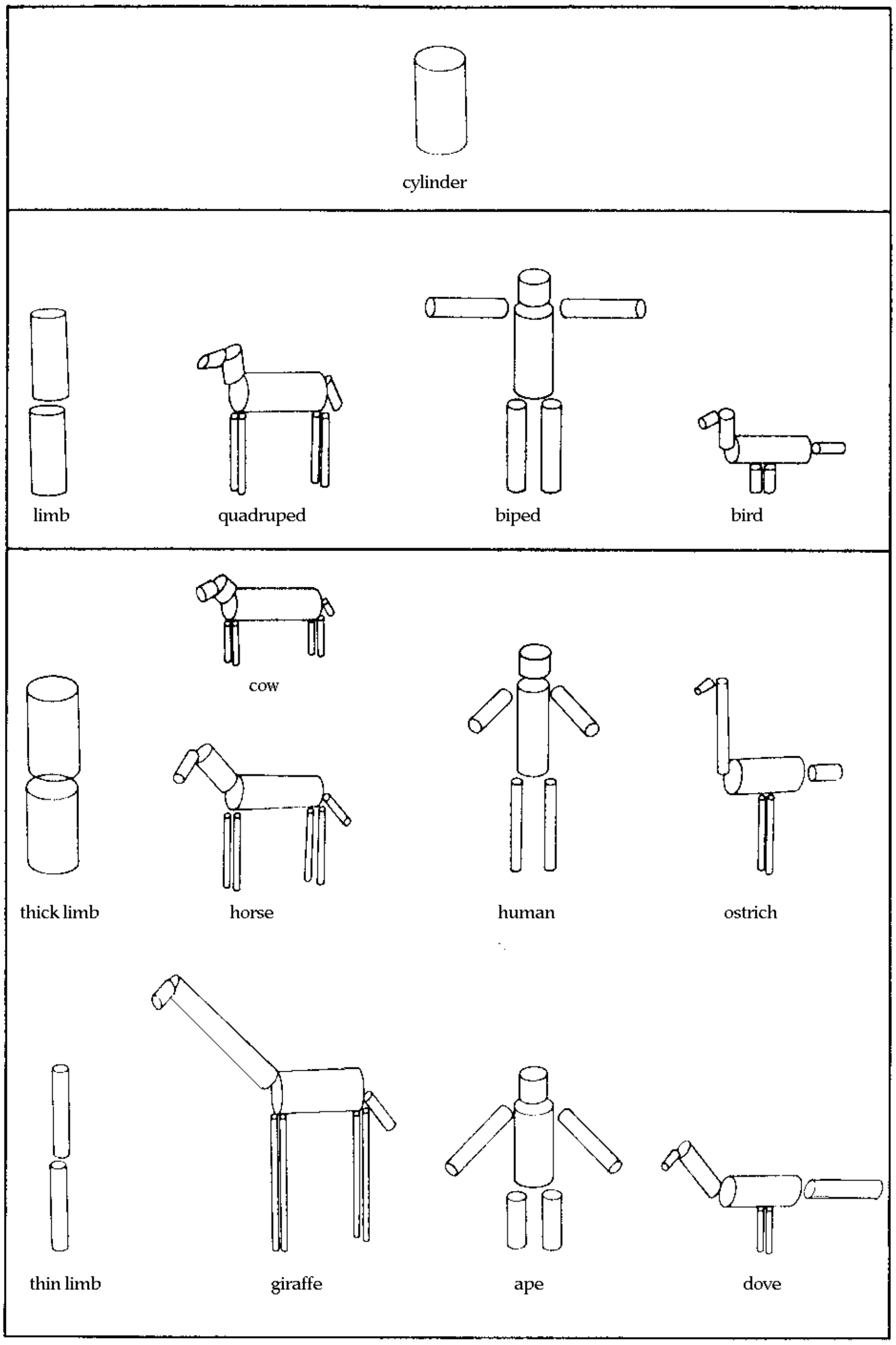

Figure 3.
Hierarchical representation of animal shapes using cylinders as modelling primitives (from Marr (1982)).

It should be noted that even if even if different members of a category are judged to be more or less prototypical, it does not follow that some of the existing objects must represent 'the prototype'. If a category is viewed as a convex region of a conceptual space, this is easily explained, since the central member of the region (if unique) is a possible individual in the sense discussed above (if all its properties are specified) but need not be among the existing members of the category. Such a prototype point in the region need not be completely described as an individual, but is normally represented as a partial vector, where only the values of the dimensions that are relevant to the category have been determined. For example, the general shape of the prototypical bird would be included in the vector, but its color or age would presumably not.

It is possible to argue in the converse direction too and show that if prototype theory is adopted, then the representation of properties as convex regions is to be expected. Assume that some quality dimensions of a conceptual space are given, for example the dimensions of color space, and that we want to partition it into a number of categories, for example color categories. If we start from a set of prototypes $p_1, ..., p_n$ of the categories, for example the focal colors, then these should be the central points in the categories they represent. One way of using this information is to assume that for every point p in the space one can measure the *distance* from p to each of the p_i's. If we now stipulate that p belongs to the same category as the *closest* prototype p_i, it can be shown that this rule will generate a partitioning of the space that *consists of convex areas* (convexity is here defined in terms of the assumed distance measure). This is the so called *Voronoi tessellation*.

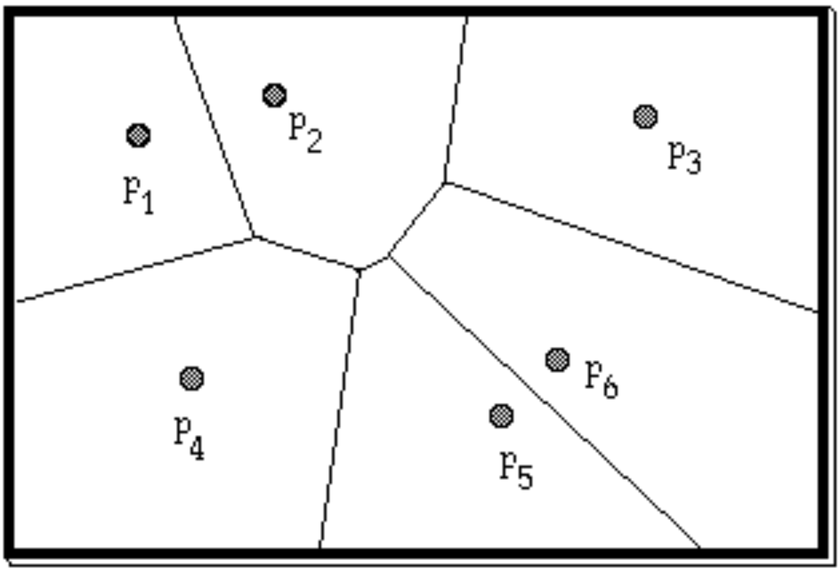

Figure 4

Voronoi tessellation of the plane into convex sets.

Thus, assuming that a metric is defined on the subspace that is subject to categorization, a set of prototypes will, by this method, generate a unique partitioning of the subspace into convex regions. Hence there is an intimate link between prototype theory and analysing properties as described by convex regions in a conceptual space.

8. Conclusion

It is becoming more and more obvious that the classical view of semantics leads to serious problems when one tries to apply it to various features of natural languages. I have presented a brief sketch of an alternative position that has become known as cognitive semantics.

As an ontological foundation for a cognitive semantics, I propose, that conceptual spaces be used as a framework for representing information. I have outlined the first steps of a cognitive semantics based on conceptual spaces. Furthermore, I have argued that the conceptual spaces are useful for understanding some semantic areas that have been particularly problematic for the classical view, namely metaphors and prototype effects in concepts.

Acknowledgement

Research for this paper has been supported by the Swedish Council for Research in the Humanities and Social Sciences.

References

Black, M. (1979). "More about metaphor," in *Metaphor and Thought*, ed. by A. Ortony, Cambridge University Press, Cambridge, pp. 19-45.

Brugman, C. (1981). *Story of Over*, Indiana Linguistics Club, Bloomington.

Declés, J.-P. (1985). *Représentation des connaissances*, Actes Semiotiques - Documents, VII, 69-70, Institut National de la Langue Française, Paris.

Fauconnier, G. (1985). *Mental Spaces*, MIT Press, Cambridge, MA.

Fodor, J. A. (1981). *Representations*, MIT Press, Cambridge, MA.

Gärdenfors, P. (1988). "Semantics, conceptual spaces and the dimensions of music," in *Essays on the Philosophy of Music*, ed. V. Rantala, L. Rowell, and E. Tarasti, (*Acta Philosophica Fennica*, vol. 43), Helsinki, pp. 9-27.

Gärdenfors, P. (1990). "Induction, conceptual spaces and AI," *Philosophy of Science 57*, pp. 78-95.

Gärdenfors, P. (1991). "Frameworks for properties: Possible worlds vs. conceptual spaces," *Language, Knowledge and Intentionality (Acta Philosophica Fennica, vol. 49)*, ed. by L. Haaparanta, M. Kusch, and I. Niiniluoto, Helsinki, pp. 383-407.

Gärdenfors, P. (1992). "Mental representation, conceptual spaces and metaphors," to appear in *Synthese*.

Herskovits, A. (1986). *Language and Spatial Cognition: An Interdisciplinary Study of the Prepositions in English*, Cambridge University Press, Cambridge.

Indurkhya, B. (1986). "Constrained semantic transference: A formal theory of metaphors," *Synthese 68*, pp. 515-551.

Jackendoff, R. (1983). *Semantics and Cognition*, MIT Press, Cambridge, MA.

Jackendoff, R. (1987). "On Beyond Zebra: The relation of linguistic and visual information," *Cognition 26*, pp. 89-114.

Jackendoff (1990). *Semantic Structures*, MIT Press, Cambridge, MA.

Johnson-Laird (1983). *Mental Models*, Cambridge University Press, Cambridge.

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

Lakoff, G. and Johnson, M. (1980). *Metaphors We Live By*, University of Chicago Press, Chicago, IL.

Langacker, R. W. (1986) *Foundations of Cognitive Grammar, Vol. 1)*, Stanford University Press, Stanford, CA.

Marr, D. and Nishihara, H. K. (1978). ”Representation and recognition of the spatial organization of three-dimensional shapes,” *Proceedings of the Royal Society in London, B 200*, pp. 269-294.

Marr, D. (1982). *Vision*, Freeman, San Francisco.

Mervis, C. and Rosch, E. (1981). ”Categorization of natural objects,” *Annual Review of Psychology* 32, pp. 89-115.

Montague, R. (1974). *Formal Philosophy*, Ed. by R. H. Thomason, Yale University Press, New Haven.

Petitot-Cocorda, J. (1985). *Morphogenèse du Sens I*, Presses Universitaires de France, Paris.

Putnam, H. (1981). *Reason, Truth, and History*, Cambridge University Press: Cambridge.

Rosch, E. (1975). ”Cognitive representations of semantic categories,” *Journal of Experimental Psychology: General* 104, pp. 192-233.

Rosch, E. (1978). ”Prototype classification and logical classification: The two systems,” *New Trends in Cognitive Representation: Challenges to Piaget's Theory*, ed. E. Scholnik, Lawrence Erlbaum Associates: Hillsdale, NJ, pp. 73-86.

Smith, E. och Medin, D. L. (1981). *Categories and Concepts*, Harvard University Press, Cambridge, MA.

Stalnaker, R. (1981). ”Antiessentialism,” *Midwest Studies of Philosophy 4*, pp. 343-355.

Sweetser, E. (1990). *From Etymology to Pragmatics*, Cambridge University Press, Cambridge.

Talmy, L. (1988). "Force dynamics in language and cognition," *Cognitive Science 12*, pp. 49-100.

Tourangeau, R. and Sternberg, R. J. (1982). "Understanding and appreciating metaphors," *Cognition 11*, pp. 203-244.