MEANINGS AS CONCEPTUAL STRUCTURES

Peter Gärdenfors

Lund University Cognitive Science
Kungshuset, Lundagård
S–222 22 LUND

E-mail: peter.gardenfors@fil.lu.se

Abstract: Cognitive semantics relates linguistic expressions to conceptual structures. The purpose of this paper is to present a framework for such a cognitive structure. As a preparation, Putnam's argument against intensional semantics as a theory of meaning is presented. Some of the main tenets of cognitive semantics as it has developed during the last years are outlined. The notion of a conceptual space is proposed as a central tool for representing semantic information. It is outlined how conceptual spaces can be used as a basis for a formal cognitive semantics.The model is then applied to some problems in lexical semantics, such as the effect of varying contrast classes.

1. WHAT IS A SEMANTICS?

A very general answer, that I think everybody can agree on, to the question of what a semantics is, is that it specifies a relation between linguistic expressions and the referents of the expressions. But soon afterwards, opinions diverge. There is, in particular, no agreement on what kind of entities the meanings of various words are. Some say that the referents of language are things in the world, some say they are things, but maybe not in this world, and some say they are mental constructions without any posit that these constructions coalesce with reality.

I want to contrast two general traditions in semantics, one *realistic* and one *cognitivistic*. According to the realistic approach to semantics the meaning of a word or expression is something out there *in the world*. Realistic semantics comes in two flavors: extensional and intensional.

In the *extensional* type of semantics, one starts out from a language L, which may or may not be defined in formal terms, and maps the constitutents of L onto a "world." Names are mapped onto objects, predicates are mapped onto sets of objects or relations between objects, etc. By *compositions* of these mappings sentences end up being mapped onto truth values. The main objective of this kind of semantics is to determine *truth conditions* for the sentences in L. A consequence of this approach is that the meaning of an expression is independent of how individual users understand it. The first developed theory of this type is Frege's semantics, but it gets a more precise form in Tarski's theory of truth.

Schematically, the mappings can be illustrated as in Figure 1.

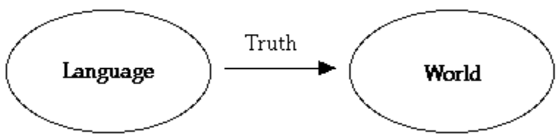

Figure 1: The ontology of extensional semantics

The extensional theory of reference implicit in this kind of semantics was soon found to be wanting on account of several phenomena in natural languages. In order to handle some of these problems, so called *intensional* semantics was developed by logicians and linguists. In this brand, the language L is mapped onto a set of *possible worlds* instead of only a single world. Still, the goal of the semantics is to provide truth conditions for the sentences in L. The meaning

Lund University Cognitive Studies – LUCS 40 1995. ISSN 1101–8453.

of a sentence is taken to be a *proposition*, which is identified with a set of possible worlds – the set of worlds where the sentence is true. The classic form of this semantics is Kripke's (1959) semantics for modal logics. With respect to natural languages, intensional semantics reaches its peak in Montague's (1974) work.

Using the same style as above, intensional semantics can be illustrated as in Figure 2.

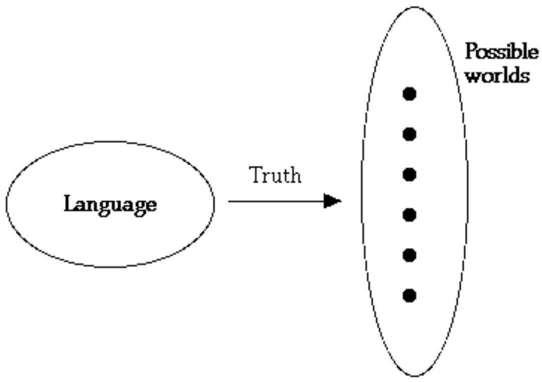

Figure 2: The ontology of intensional semantics

The second paradigm of semantics is *cognitivistic* and this approach will be the focus of the present paper. The core idea is that meanings of expressions are *mental entities*. A semantics is seen as a mapping from the linguistic expressions to *cognitive structures*. Language itself is seen as part of the cognitive structure, and not as an entity with independent standing. Semantics is thus a relation between mental entities.

Within cognitive semantics the emphasis is on *lexical meaning* rather than on the meaning of sentences. In addition, the truth of sentences is not the most important feature, but is replaced with considerations about *acceptance* or *belief*. The external world and truth conditions enter on the scene only when the relation between the world and the cognitive structure is considered.

Interestingly enough, one finds a very similar theory in Aristotles *De Interpretatione*. The following is an excerpt from the first paragraph of E. M. Edghill's translation:

> Spoken words are the symbols of mental experience and written words are the symbols of spoken words. Just as all men have not the same writing, so all men have not the same speech sounds, but the mental experiences, which these directly symbolize, are the same for all, as also are those things of which our experiences are the images.

Aristotle makes a distinction between "mental experiences" and the "things" of which the experiences are "images." Furthermore, spoken or written words refer to the mental experiences, and not to the external reality.

The framework of the kind of cognitive semantics to be discussed here can be illustrated as in Figure 3.

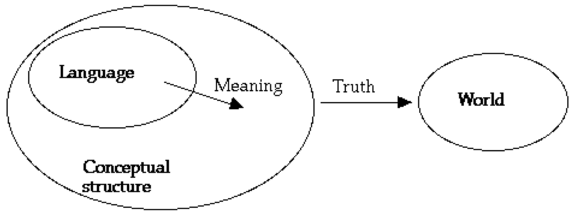

Figure 3: The components of cognitive semantics

My main aim in this paper is to present a model of the cognitive structure presupposed in this model. As a preparation, I will present in the next section Putnam's argument against intensional semantics as a theory of meaning. I will then, in Section 3, outline some of the main tenets of cognitive semantics as it has developed during the last years. I will contrast these with the positions of extensional and intensional semantics, but my primary goal is not to criticize these kinds of semantics.[1] In Section 4, I will introduce the central notion of a *conceptual space* as the main framework for representing semantic information. In the following section, conceptual spaces will be used to outline the bare bones of a formal cognitive semantics. In Sections 6 and 7, the model will then be applied to some problems in lexical semantics, such as the effect of varying contrast classes.

2. PUTNAM'S PREDICAMENT FOR PROPERTIES IN INTENSIONAL SEMANTICS

As a way of motivating the turn to cognitive semantics, I want to present one of the major philosophical problems that the intensional approach faces. In intensional semantics, 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

[1]A rich source for this purpose is Lakoff's book (1987), which is a lengthy criticism of what he calls "objectivist semantics."

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.

In this kind of semantics, a property is seen as 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.

As was mentioned, functions are preferred to many-many relations in intensional semantics. 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 to which the individuals are assigned (cf. Figure 4). This means that an equivalent definition of a property is that it is *a function from possible worlds to sets of individuals*. This alternative definition shows the correspondence between the extensional and the intensional definition of a property.

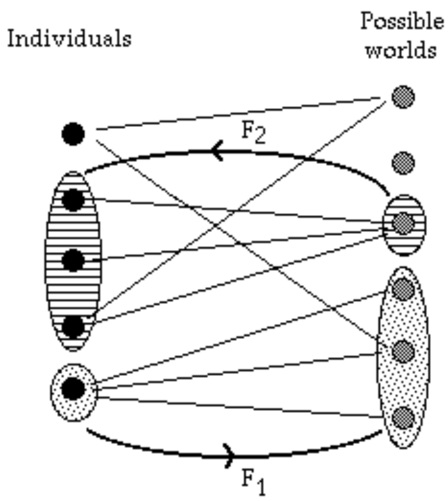

Figure 4: A property as a many-many relation between individuals and possible worlds.
F1: 'Propositional' function mapping individuals on propositions.
F2: 'Extensional' function mapping possible worlds on classes

In Gärdenfors (1991), I have argued that the standard definition of a property within intensional semantics leads to a number of problems. Here, I shall only consider what seems to be the most serious one. Putnam (1981) has shown that the model-theoretic definition of "property" which has been given here does not work as a theory of the *meaning* of properties. In proving this result, Putnam makes two assumptions about "the received view" of meaning: (1) The meaning of a sentence is a function which assigns a truth value to the sentence in each possible world; and (2) the meaning of the parts of a sentence cannot be changed without changing the meaning of the whole sentence.

Putnam's general proof is quite technical, but the thrust of the construction can be illustrated by his example (1981, pp. 33–35). He begins with the sentence

(1) A cat is on a mat.

where "cat" refers to cats and "mat" to mats as usual. He then shows how to give (1) a new interpretation

(2) A cat* is on a mat*.

The definitions of the properties cat* and mat* make use of three cases:

(a) Some cat is on some mat and some cherry is on some tree.
(b) Some cat is on some mat and no cherry is on any tree.
(c) Neither (a) nor (b) holds.

Here are Putnam's definitions:

DEFINITION OF 'CAT*'

x is a cat* if and only if case (a) holds and x is a cherry or case (b) holds and x is a cat; or case (c) holds and x is a cherry.

DEFINITION OF 'MAT*'

x is a mat* if and only if case (a) holds and x is a tree or case (b) holds and x is a mat; or case (c) holds and x is a quark.

Given these definitions it turns out that the sentence (1) is true in exactly those possible worlds where (2) is true. Thus, according to the received view of meaning, these sentences will have the same meaning. In the appendix to his book (1981), Putnam shows that a more complicated reinterpretation of this kind can be constructed for all the sentences of a language. 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). Thus "... *truth-conditions for whole sentences* underdetermine reference" (*ibid*.). The underlying reason is that there are *too many* potential properties if they are defined as functions from individuals to propositions, that is, in terms of possible worlds and truth values. Cat* and mat* are just two examples from this large class.[2]

3. THE COGNITIVE VIEW OF SEMANTICS

In this section, I shall give a programmatic presentation of cognitive semantics with a philosophical slant to it.[3] However, cognitive semantics has mainly developed within linguistics. Prime examples of works in the linguistic tradition are Lakoff's (1987) and Langacker's (1987). Related versions of cognitive semantics can be found in the writings of Jackendoff (1983, 1990), Johnson-Laird (1983), Fauconnier (1985), Talmy (1988), Sweetser (1990), Holmqvist (1993) and many others. There is also a French semiotic tradition, exemplified by Desclés (1985) and Petitot-Cocorda (1985), that shares many features with the American (mainly Californian) group.

The prime slogan for cognitive semantics is: *Meanings are mentally encoded*. More precisely, a semantics for a language is seen as a mapping from the expressions of the language to some cognitive or mental entities. This thesis puts cognitive semantics in contact with psychological notions and makes it possible to talk about a speaker "grasping" a meaning (cf. Jackendoff 1983). 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*.

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 is still a realist and relies on references in the external world as well as truth conditions.

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 realistic 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.

But how can we explain that we can talk about what we see and hear? It does not suffice to say that we transfer information from perceptual representations to linguistic representations. As Jackendoff (1987, p. 90) notes, it is a problem of translation:

> "in order to talk about what we see, information provided by the visual system must be translated into a form compatible with the information used by the language system. So the essential questions are: (1) What form(s) of information does the visual system derive? (2) What form of information serves as the input to speech? (3) How can the former be translated into the latter?"

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. 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* (cf. Gibson 1979).

4. CONCEPTUAL SPACES AS A BASIS FOR COGNITIVE SEMANTICS

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

[2] In his comments to this paper, Hans Rott points out that the gist of Putnam's theorem has been formulated earlier, in an informal way, by Quine who does not see the result as a predicament for semantics. On the contrary, he sees it as a support for the thesis that reference is indeterminate.
[3] A fuller exposition can be found in Gärdenfors (to appear b).

like Mentalese, the conceptual schemes that are used to represent meanings are often based on *geometric* or *spatial* constructions.

As a framework for a geometric structure used in describing a cognitive semantics I have proposed (Gärdenfors 1988, 1990, 1991, 1993b, to appear a, to appear b) the notion of a *conceptual space*. A conceptual space consists of a number of *quality dimensions*. As examples, let me mention color, pitch, temperature, weight, and the three ordinary spatial dimensions. I have chosen these dimensions because they are closely connected to what is produced by our sensory receptors (Schiffman 1982). Color and ordinary space are perceived by the visual sensory system, pitch by the auditory system, temperature by thermal sensors, and weight, finally, by the kinesthetic sensors. However, there are also quality dimensions that are of an abstract non-sensory character.

The primary function of the quality dimensions is to represent various "qualities" of objects. They form the "framework" used to assign properties to objects and to specify relations between them.

The dimensions are taken to be independent of language and symbolic representations in the sense that we and other animals can represent the qualities of objects, for example when planning an action, without presuming an internal language in which these qualities are expressed. The quality dimensions should be seen as abstract representations used as a modeling factor in describing mental activities of organisms. They are thus not assumed to have any immediate physical realisation. However, they will hopefully be useful constructs when developing artificial systems.

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. As a first example, I will take the dimension of "time". In science, time is a one-dimensional structure which is isomorphic to the line of real numbers. If "now" is seen as the zero point on the line, the future corresponds to the infinite positive real line and the past to the infinite negative line.

This representation of time is not universal, but is to some extent culturally dependent, so that other cultures have a different time dimension as a part of their cognitive structure. There is thus no unique way of choosing a dimension to represent a particular quality, but in general there is a wide array of possibilities.

It should be noted that some quality "dimensions" only have a *discrete* structure, that is, they merely divide objects into disjoint classes. Two examples are classifications of biological species and kinship relations in a human society. However, even for such dimension one can distinguish a simple topological structure. For example, in the phylogenetic classification of animals, it is meaningful to say that birds and reptiles are more closely related than reptiles and crocodiles.

In order to separate different uses of quality dimensions it is important to introduce a distinction between a *psychological* and a *scientific* interpretation. The psychological interpretation concerns how humans (or other organisms) structure their perceptions. The scientific interpretation, on the other hand, deals with how different dimensions are presented within a scientific theory. The distinction is relevant when the dimensions are seen as cognitive entities, in which case their topological or metrical structure should not be determined by scientific theories which attempt at giving a "realistic" description of the world, but by *psychophysical* measurements which determine the structure of how our perceptions are represented.

A psychologically interesting example of a set of quality dimensions concerns *color perception*. In brief, our cognitive representation of colors can be described by three dimensions. 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. One way of illustrating the differences in topology is by noting that we can talk about psychologically *complementary* colors, that is, colors that lie *opposite* to each other on the color circle. In contrast it is *not meaningful* to talk about two points of time or two weights being "opposite" to each other.

The second psychological dimension of color is *saturation* which ranges from grey (zero color intensity) 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

This space is often illustrated by the so called *color spindle* (see figure 5). Brightness is shown on the vertical axis. Saturation is represented as the distance from the center of the spindle towards its perimeter.

Hue, finally, is represented by the positions along the perimeter of the central circle.[4]

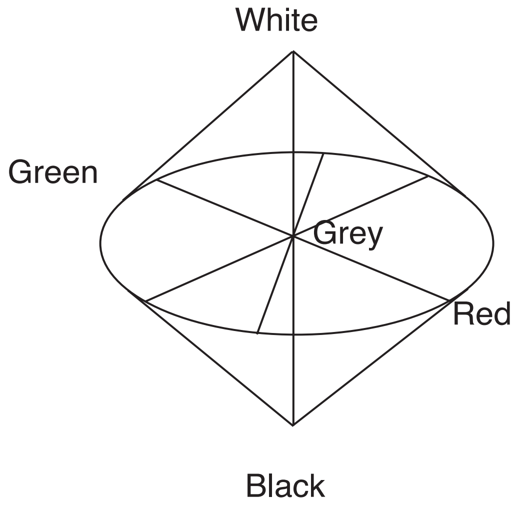

Figure 5: The color spindle.

It is impossible to provide a complete list of the quality dimensions involved in the conceptual spaces of humans. 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. Even if we do not know much about the topological structures of these dimensions, it is quite obvious that there is some such non-trivial structure (see e.g., Vaina's (1983) analysis of functional representation). Still other dimensions may be *culturally* dependent. Finally, some quality dimensions are introduced by *science*. (see Gärdenfors 1993b)

There is a strong similarity between the notion of a conceptual space and the *domains* as used in Langacker's (1987) semantic theory. The following quotation from Langacker (1987, 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 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."

Within the type of cognitive semantics developed by Lakoff and Langacker, the most important structure 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.

Metaphors and metonymies have been notoriously difficult to handle within realist semantic theories. In these theories such linguistic figures have been treated as a deviant phenomena that have 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 as central semantic features and are given systematic analyses. One of the first works in this area was Lakoff and Johnson (1980). Analyses of different kinds of metaphorical expressions have since then become one of the trademarks of cognitive semantics.[5]

Metaphors and metonymies are primarily seen as *cognitive* operations, and their linguistic expression is only a secondary phenomenon. They are analysed as

[4]As Hans Rott correctly points out in his comments on this paper, there are several different models of the perceptual color space in the scientific literature. However, all the examples that he presents uses dimensions, and all of them are three-dimensional. There is a controversy concerning which topology of the color space "best" represents human perception. By focusing on the color spindle, I do not claim that this is the best representation, but only that it is suitable to illustrate some aspects of color perception and of conceptual spaces in general.

[5]ee for example Broström (1994), Brugman (1981), Gärdenfors (to appear a, to appear b), Indurkya (1986), Lakoff (1987, 1992), Sweetser (1990) and Thorangeau and Stenberg (1982).

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).

5. THE FIRST STEPS OF A FORMAL COGNITIVE SEMANTICS

I can only outline the first steps in developing a philosophically oriented cognitive semantics based on conceptual spaces. According to the cognitive view, semantics is a relation between language 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 (that is, points in the conceptual space) or partial vectors (that is, points with some arguments undetermined). In this way each name (referring to an individual) is allocated a specific color, spatial position, weight, temperature, etc. 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. Such a predicate is *satisfied* by an individual only when 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. An example of this will be analysed in Section 7. *Relations* (primary and secondary) can be treated in a similar way (see Holmqvist 1993).

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 determined by a conceptual space, 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.

A consequence of the theory presented here is that if we assume that the meanings of the predicates 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*. 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 green and blue. Analytic-in-*S* can thus be defined on the basis of the topological and metric structure of the conceptual space *S*. However, different conceptual spaces will result in different notions of analyticity. Hence the epistemology underlying the theory of conceptual spaces could be described as a version of neo-Kantianism.

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. Hence, 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. Thus most of the constructions from traditional intensional semantics will be available, should we need them. However, I believe that many of the semantic notions that have been analysed within, for example, the Montague tradition, can be given a cognitively more realistic treatment on the basis of conceptual spaces.

6. USING CONCEPTUAL SPACES IN LEXICAL SEMANTICS

In the previous section, predicates were mapped onto regions of a conceptual space, where "region" should be understood as a spatial notion determined by the topology and metric of the space. This idea can be formulated as a general "contiguity constraint" on concepts, namely that they correspond to a *connected region* of a conceptual space.[6] For example, the point in the time dimension representing "now" divides this dimension, and thus the space of vectors, into two regions corresponding to "past" and "future". In contrast to the traditional definition in intensional semantics that was presented in Section 2, the proposed constraint presumes *neither* the concept of an individual *nor* the concept of a possible world.

A working hypothesis is that most properties expressed by simple words in natural languages correspond to connected regions in the sense specified here. For instance, I conjecture that all *color terms* in natural languages express connected regions with respect to the psychological representation of the three color dimensions. It is well-known that different languages carve up the color circle in different ways, but all carvings seem to be done in terms of connected sets. This means, for example, that there is no language which has a single color word for the hues that are denoted by "green" and "orange" in English (and for no other colors), since such a word would represent two disjoint areas in the color space. Strong support for this conjecture can be found in Berlin and Kay (1969), although they do not treat color words in general but concentrate on basic color terms.

For more complex concepts like "bird" or "cat" it is perhaps more difficult to describe the underlying conceptual space, and thereby to apply the contiguity constraints. 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.[7]

Their scheme for describing biological forms uses cylinder-like modeling primitives. Each cylinder can be described by two coordinates (length and width). Cylinders are combined by determining the angle between the dominating cylinder and the added one and the position of the added cylinder in relation to the dominating one. The details of the representation are not important in the present context, but it is worth noting that an object can be described by a comparatively small number of coordinates based on lengths and angles. Thus the object can be identified as a structured vector in a multidimensional conceptual space. Figure 6 provides an illustration of the cylinder-based representations.

The proposed constraint that properties correspond to connected regions of a conceptual space also doew away with the problems that Putnam's theorem causes the traditional definition of "property". "Cat" denotes a region of a conceptual space, at least according to Marr and Nishihara's analysis. This region would, at least partly, be determined by the perceptual features of cats. We cannot create a new natural property "cat*" by relating it to what facts are true in various possible worlds: "Cat*" as introduced by Putnam is indeed a propositional function, but definitely not a property that satisfies the contiguity constraint and thus not an eligible candidate for an interpretation function that maps a language into a conceptual space.[8]

Also for concepts on a more abstract level, it seems that some kind of contiguity constraints can be identified. Bickerton (1990, p. 44–46) gives the example of words that are used to express the abstract concepts of "existence," "location," "possession" and "ownership." Different languages use different verbs to express these concepts.

---

[6]A more precise and powerful idea is the following criterion (see Gärdenfors 1990, 1991) where the topological characteristics of the quality dimensions are utilized to introduce a spatial structure on properties:

> *Criterion P:* A *natural property* is a convex region of a conceptual space.

A *convex* region is characterized by the criterion that for very pair of points v_1 and v_2 in the region all points in between v_1 and v_2 are also in the region. The motivation for the criterion is that if some objects which are located at v_1 and v_2 in relation to some quality dimension (or several dimensions) both are examples of the property P, then any object that is located between v_1 and v_2 on the quality dimension(s) will also be an example of P.

[7]This analysis is expanded in Marr (1982), Ch. 5. A related model, together with some psychological grounding, is presented by Biederman (1987).

[8]In a more recent book, Putnam (1988) also discusses "conceptual role" semantics, in particular in relation to *natural-kind terms*. He argues that the meaning of such terms cannot be given in terms of their conceptual roles only, but "once we have identified a word as a natural-kind term, then we determine whether it is synonymous with another natural-kind term primarily on the basis of the extensions of the two words" (p. 50). Here, *extension* is, of course the set of things in the world that the word applies to. So natural-kind terms presume a *realistic* component for their semantics according to Putnam. In Gärdenfors (1993a), I have defended a conceptual role semantics and argued that Putnam's notion of natural kind is misconstrued since his argument relies on a kind of *essentialism*.

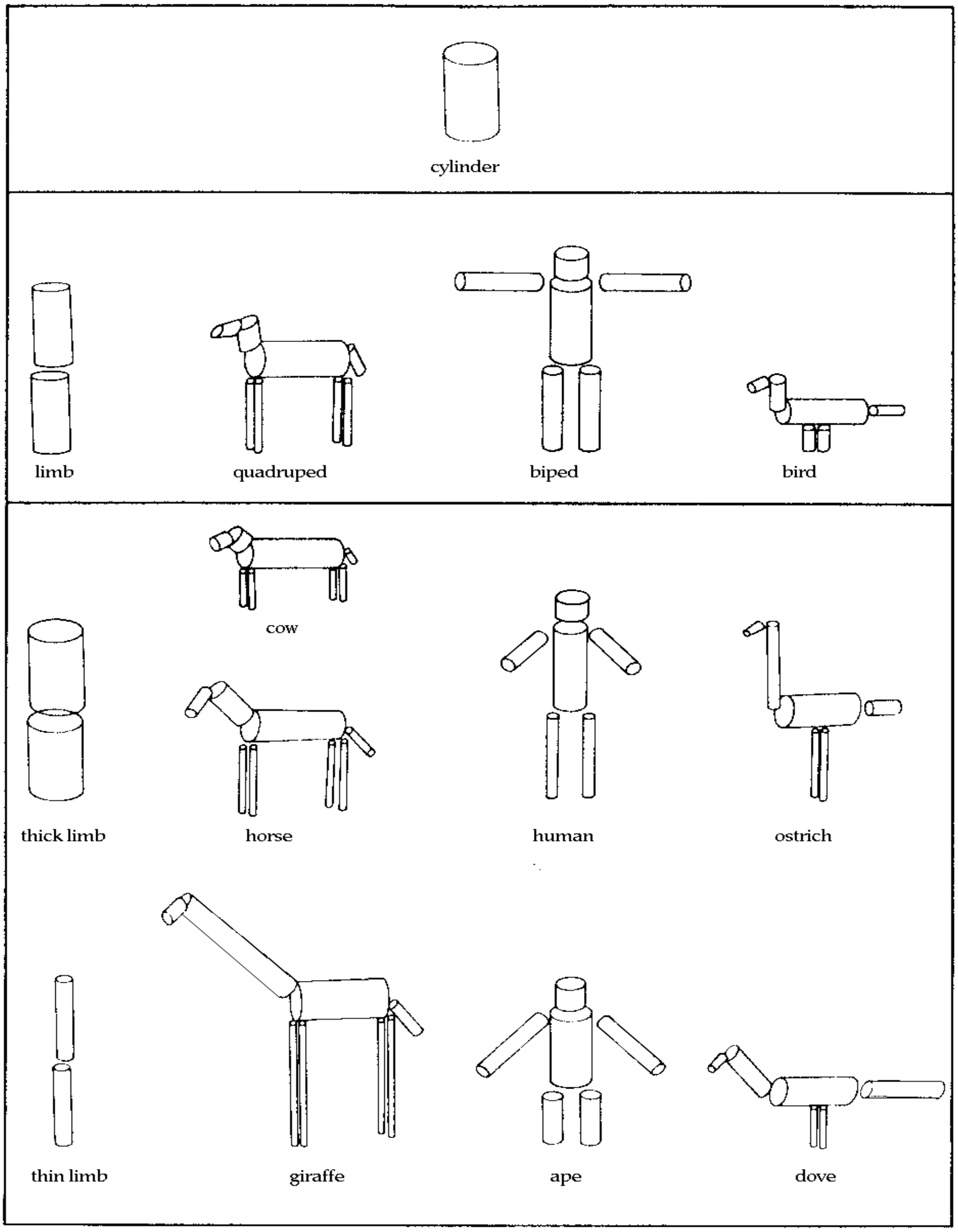

Figure 6: Representation of animal shapes using cylinders as modelling primitives (from Marr (1982)).

In English "be" is used for "existence," "location" and "ownership," while "have" is used for "possession." Bickerton suggests that the contiguity relations of the four abstract notions can be represented as in Figure 7.

Existence	Location
Possession	Ownership

Figure 7: Spatial relation between some abstract lexemes.

He claims (p. 45) that "no language has turned up that uses the same verb for 'location' and 'possession' but a different verb (or verbs) for 'existence' and 'ownership', or that has the same verb for 'existence' and 'ownership' but a different verb (or verbs) for 'location' and 'possession'." This is thus further evidence for the contiguity constraint.

7. THE EFFECT OF CONTRAST CLASSES

I have proposed that properties correspond to connected regions of a conceptual space. However, the situation is more complicated than this since the meaning of a word is often determined by the *context* in which it occurs. Consider the word "red". In the *Advanced Learner's Dictionary of Current English*, it is defined as "of the colour of fresh blood, rubies, human lips, the tongue, maple leaves in the autumn, post-office pillar boxes in Gt. Brit." This definition fits very well with letting "red" correspond to the normal region of the color spindle. Now consider "red" in the following contexts:[9]

- Red book
- Red wine
- Red hair
- Red skin
- Redwood

In the first example, "red" corresponds to the dictionary definition, while it would be purple when predicated of wine, orange when used about hair, tawny when of skin, and pinkish brown when of wood. Thus the other uses don't fit with the standard region assigned to red. How can we then explain that the same word is used in so many different contexts?

One way to account for the phenomenon would be to say that it is only in "red book" that the word is used in its proper meaning, while in the other cases it is used as a *metaphor*. Broström (1994, pp. 101–102) discusses this solution. She says:

> "Given that the same color term clearly has a different reference in each domain, that would seem to give us every ingredient of metaphor. Still we hesitate to call this metaphor. Why? The most reasonable answer is that the color terms aren't used so much to refer to particular colors as to maintain the color *contrasts* between different referents. Every "domain" is thus a contrast class, to which we apply color terms of maximal distinctiveness."

I don't see how this phenomenon can be analysed in a simple way using possible worlds or some other tools from intensional semantics. However, the idea of a contrast class can quite easily be given a general interpretation with the aid of conceptual spaces. For each domain, for example skin color, we can map out the class of possible colors on the color spindle. This mapping will determine a subset of the full color space. The shape of this subset may be rather irregular. However, if the subset is filled out to form a space with the same topology as the full space we obtain a picture that looks like Figure 8.

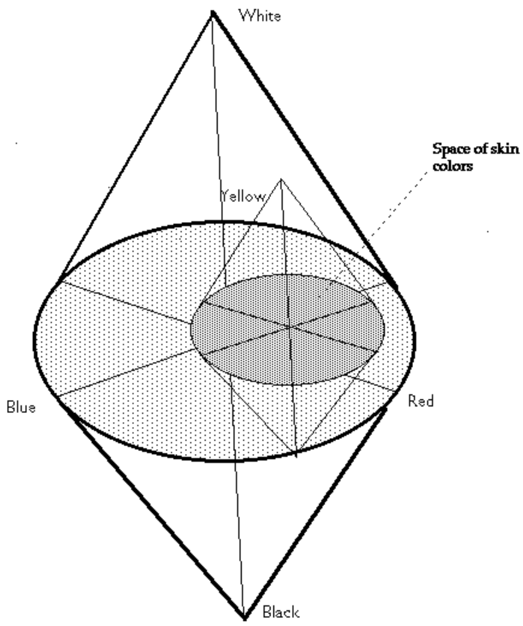

Figure 8: The subspace of skin colors embedded in the full color spindle

In this smaller spindle, the color words are then used in the same way as in the full space, even if the hues of the color in the smaller space don't match the hues of the complete space. Thus, "white" is used about the lightest forms of skin, even though white skin is pinkish, "black" refers to the darkest form of skin, even though black skin is brown, etc.

It can now be seen why this process of different contrast classes for the same word is closely related to that of metaphorical uses of a word. What characterizes a metaphor is that it expresses a similarity in topological or metrical structure between different quality dimensions (see Gärdenfors to appear a). 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. Now, in the case of contrast classes, one set of dimensions is not really

[9]Clark (1992, pp. 369–372) uses this example to make a somewhat different albeit related point. Also see Broström (1994, pp. 101–102)

mapped onto *another* set, but the conceptual space is mapped onto a *subspace* of itself retaining the same topological structure.

Another thing to notice about color terms is that not all words will be used, but only those that are "basic" in the sense of Berlin and Key (1969). On this point, Broström (1994, p. 102) writes:

> "The "late" color terms will seldom be used since they are only needed to contrast with the early ones – "lilac" is only needed if "blue" has already been used (thus the rarity of the categorization "lilac flowers"). If we regard the reference and meaning of color terms as relative rather than absolute, we avoid the conclusion that we are dealing with metaphor. There is no understanding in the prototypical metaphorical sense involved. We do not understand Caucasian skin as though it were paint white, we just call it "white" to distinguish it from other ethnic skin colors, such as "black", "yellow", or "red"."

The rule that it is always the most basic color terms that are first employed is not without exceptions. For example, the basic distinction among wine colors in English and French is between "white" and "red" (which is really pale yellow and purplish red). "Black" wines are almost unheard of. There is "rosé" wine and even a "vin jaune" in France. Note that the Portuguese "green" wine (vinho verde) does not refer to the color of the wine, but to the fact that the grapes are not quite ripe when they are picked (there is a "green" red wine as well as a "green" white wine). However, the terminology is far from universal. In Spanish and Portuguese, the basic distinction is between "colored wine" and "white" wine. In Catalan, the names are actually "black" wine and "white" wine, in full conformity with Berlin and Kay's scheme.[10]

It is, of course, not only color terms that appear in different contrast classes. The same phenomenon can appear with most adjectives. For example, the same stream of tap water can be described as "hot" if seen as water in general, but as "cold" if seen as bath water; a "big" chihuahua is a "small" dog, etc.

Do problems relating to the variety of contrast classes really belong to semantics? Since contrast classes are often determined by the context rather than by linguistic markers, should these kinds of problems perhaps be classified as part of *pragmatics*? I don't think the proper classification of the contrast class phenomenon is a serious problem since I don't believe that one can draw a sharp borderline between semantics and pragmatics. Sometimes, the contrast class that is relevant for a word like "red" is marked semantically in the language, and sometimes it is implicit in the context of the speech act (see e.g. Clark 1992, pp. 370–371). I thus agree with Jackendoff (1987 p. 97) that there is no formal distinction of level between semantics and pragmatics. Also Langacker (1987, Section 4.2) argues that semantics is just conventionalized pragmatics.

8. CONCLUSION: IN CHASE OF SPACE

In this paper, I have tried to summarize the foundations of cognitive semantics and I have presented the skeleton of a formal cognitive semantics based on conceptual spaces. This kind of semantics has been contrasted with the more traditional extensional and intensional types of semantics. I have given an analysis in terms of conceptual spaces of the role of contrast classes in lexical semantics.

My analysis presumes the central canon of cognitive semantics: meanings are mental entities. The referents of words are identified with conceptual structures in people's heads. This position has been attacked by several researchers, notably Putnam (1981, 1988) and Burge (1979) who claim that a conceptualistic approach to semantics, *mentalism* as they call it, is doomed to fail. Putnam's main reason for this malediction is summarized by the slogan "meanings ain't in the head." For example, he claims that he cannot distinguish oaks from elms, but he knows that the meaning of the words "oak" and "elm" are different. Putnam also claims that, as a consequence of this, meanings must be determined by reference to the external world.[11]

I believe that this claim is wrong. In Gärdenfors (1993a), I argue that the social meanings of the expressions of a language are indeed *determined from their individual meanings together with the structure of linguistic power* that exists in the community. In contrast to Putnam, I claim that no reference to the external world is needed to handle the problem he presents.

My position can be summarized as follows: Meanings are not in the head of a single individual, but they *emerge* from the conceptual schemes in the heads of the language users together with the semantic power structure. Even if Putnam cannot

[10] I am grateful to Enric Vallduvi for this useful piece of information.

[11] And, as I have argued in Gärdenfors (1993a), this is also implicit in Burge's argument.

distinguish oaks from elms, they are distinguished in the emergent social semantics. So when he says that he knows that the meaning of "elm" and "oak" are distinct, he knows that the *social* meanings differ. In his individual conceptual space, however, they are undistinguishable.

Considered as a theory about the meaning of linguistic expressions, however, cognitive semantics is still rather undeveloped. Its most detailed applications have been areas where language is tightly connected to perception, as for example in spatial prepositions. Cognitive semantics has also offered new insights in the mechanisms of metaphors. Its strength lies mainly in the analysis of lexical items, even though there are interesting attempts to explain syntactic features by cognitive means (e.g., Langacker 1987, Holmqvist 1993, 1994, to appear). However, there are areas where traditional semantics is strongly developed and where cognitive semantics is still weak, for example, in analyses of negation and quantifiers.[12]

The main factor preventing a rapid development of a cognitive semantics based on conceptual spaces is the lack of knowledge about the relevant quality dimensions. It is almost only for perceptual dimensions that psychophysical research has succeeded in identifying the underlying topological structures (and, in rare cases, the psychological metric). For example, we only have a very sketchy understanding of how we perceive and conceptualize things according to their shapes. The models developed by Marr and Nishihara (1978), Pentland (1986), Biederman (1987), and Tversky and Hemenway (1984) among others, seem to point in the right direction, but there still remains a lot to learn about "shape space."

Thus, those who want to contribute to the research program in cognitive semantics should start hunting for the conceptual spaces since they, to a large extent, determine the meanings of linguistic expressions. Even if results may not be easily forthcoming, they are sure to have repercussions in the other cognitive sciences as well.

ACKNOWLEDGEMENTS

Research for this paper has been supported by the Swedish Council for Research in the Humanities and Social Sciences. The paper draws partly on material from Gärdenfors (1991) and Gärdenfors (to appear b). Hans Rott's helpful comments on the earlier version led to some improvements. I also want to thank Clyde Hardin, William Lycan and Johanna Seibt for their discussions.

[12]However, for a cognitively oriented analysis of quantifiers, see Moxey and Sanford (1993).

REFERENCES

Berlin, B. and Kay, P. (1969), *Basic Color Terms: Their Universality and Evolution*. Berkeley, CA: University of California Press.

Bickerton, D. (1990), *Language and Species*, Chicago: The University of Chicago Press.

Biederman, I. (1987), "Recognition-by-components: A theory of human image understanding," *Psychological Review 94*: 115–147.

Broström, S. (1994), *The Role of Metaphor in Cognitive Semantics*, Lund: Lund University Cognitive Studies 31.

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

Burge, T. (1979), "Individualism and the mental," in *Midwest Studies in Philosophy, vol. 4: Studies in Metaphysics*, Minneapolis: University of Minnesota Press, pp. 73–121

Clark, H. H. (1992), *Arenas of Language Use*, Chicago: The University of Chicago Press.

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*, Cambridge, MA: MIT Press.

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

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*: 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. (1993a), "The emergence of meaning," *Linguistics and Philosophy 16*: 285–309.

Gärdenfors, P. (1993b), "Induction and the evolution of conceptual spaces," in *Charles S. Peirce and the Philosophy of Science*, ed. by E. C. Moore, Tuscaloosa: The University of Alabama Press, pp. 72–88.

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

Gärdenfors, P. (to appear b), "Conceptual spaces as a framework for cognitive semantics," to appear in the proceedings of Second International Colloquium on Cognitive Science, Universidad del Pais Vasco, 1991.

Gibson, J. J. (1979), *The Ecological Approach to Visual Perception*, Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Holmqvist, K. (1993), *Implementing Cognitive Semantics*, Lund: Lund University Cognitive Studies 17.

Holmqvist, K. (1994), "Conceptual enginieering I: From morphemes to valence relations," Lund: Lund University Cognitive Studies 28.

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

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

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

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

Johnson, M. (1987), *The Body in the Mind: The Bodily Basis of Reason and Imagination*, Chicago: The University of Chicago Press.

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

Kripke, S. (1959), "A completeness theorem in modal logic," *Journal of Symbolic Logic 24*: 1–24

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

Lakoff, G. (1992), "The contemporary theory of metaphor," to appear in Ortony, A. (ed.) *Metaphor and Thought* (2nd edition), Cambridge: Cambridge University Press.

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

Langacker, R. W. (1987), *Foundations of Cognitive Grammar, Vol. 1*, Stanford, CA: Stanford University Press.

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

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

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

Moxey, L. M. and A. J. Sanford (1993), *Communicating Quantities: A Psychological Perspective*, Hove: Lawrence Erlbaum Associates.

Pentland, A.P. (1986), "Perceptual organization and the representation of natural form," *Artificial Intelligence 28*: 293–331.

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

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

Putnam, H. (1988), *Representation and Reality*, Cambridge, MA: MIT Press.

Schiffman, H. R. (1982), *Sensation and Perception, 2nd edition*, New York: John Wiley & Sons.

Stalnaker, R. (1981), "Antiessentialism," *Midwest Studies of Philosophy 4*: 343–355

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

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

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

Tversky, B. and Hemenway, K. (1984), "Objects, parts, and categories," *Journal of Experimental Psychology: General 113*: 169–191.

Vaina, L. (1983), "From shapes and movements to objects and actions," *Synthese 54*: 3–36.