Cognitive semantics and image schemas with embodied forces

Peter Gärdenfors
Lund University Cognitive Science
Kungshuset
S-222 22 Lund, Sweden
Peter.Gardenfors@lucs.lu.se

Cognitive semantics

We use language every day without thinking about what we do when we *understand* what is said or written. But when we listen to two persons conversing in a language that is completely unknown to us, we realize that it is not sufficient to hear what somebody is saying – we must also be able to interpret the speech sounds.

But exactly *what* is the meaning of a word? In philosophy, linguistics and psychology, one finds several theories about what meaning is. One theory that has been dominating for a long time says that the meanings of words are found in the external world. A more recent theory, called *cognitive semantics*, claims that the meanings of words are located in our heads.

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. In philosophical ontological terms, this is thus a *realistic* theory. However, a realistic theory does not explain how a person can *grasp* the meanings of different words. Harnad (1987, p. 550) says that

> […] the meanings of elementary symbols must be grounded in perceptual categories. That is, symbols, which are manipulated only on the basis of their form (i.e., syntactically) rather than their "meaning," must be reducible to nonsymbolic, shape-preserving representations.

As a contrast to realist theories, a new semantic theory, called cognitive semantics, has been developed (see e.g. Lakoff 1987, Langacker 1986, 1987, Croft and Cruse 2004, Evans 2006). 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 entities. This paradigm of semantics is thus conceptualistic or cognitivistic.

An important tenet of cognitive semantics is that the structures in our heads that are carrying the meanings of words are of the same nature as those that are created when we *perceive* – when we see, hear, touch, etc, different things. If I see Fido, I see him *as* a dog since the perception I have fits with the cognitive structure in my head that is the concept of a dog. In my mental classification of different animals, there is a "schema" for how a dog looks like (and sounds and smells and feels like). This schema is the very meaning of the word "dog" according to cognitive semantics.

A consequence of the cognitivist position that puts it in conflict with many other semantic theories is that no reference to reality is necessary to determine the meaning of a linguistic expression. Jackendoff (1987, p. 123) says: "The buck stops here: expressions at the level of

conceptual structure simply *are* the meanings of utterances." A related point is that 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.*

Unlike earlier semantic theories, cognitive semantics emphasizes that linguistic meanings do not form an independent system but is closely related to other cognitive mechanisms, in particular perception. Regier (1996, p. 27) expresses the point as follows:

> The idea is that since the acquisition and use of language rest on an experiential basis, and since experience of the world is filtered through extralinguistic faculties such as perception and memory, language will of necessity be influenced by such faculties. We can therefore expect the nature of human perceptual and cognitive systems to be of significant relevance to the study of language itself. One of the primary tasks of cognitive linguistics is the ferreting out of links between language and the rest of human cognition.

This thesis puts cognitive semantics in contact with psychological notions and makes it possible to talk about a speaker "grasping" a meaning (compare Jackendoff 1983).

Because the cognitive structures in our heads, according to cognitive semantics, are connected to our perceptual mechanisms, directly or indirectly, it follows that *meanings are*, at least partly, *embodied*. Jackendoff (1983, pp. 16-18) formulates this as "the cognitive constraint":

> There must be levels of mental representation at which information conveyed by language is compatible with information from other peripheral systems such as vision, nonverbal audition, smell, kinesthesia, and so forth. If there were no such levels, it would be impossible to use language to report sensory input. We couldn't talk about what we see and hear. Likewise, there must be a level at which linguistic information is compatible with information eventually conveyed to the motor system, in order to account for our ability to carry out orders and instructions.

In this article, I shall argue that not only must linguistic information be compatible with information from the perceptual system concerning spatial relations, but our *actions* should also be considered. One distinguishing feature of actions is that they involve *forces* exerted by the agent. Consequently, forces should be among the building blocks of the cognitive semantics – not only spatial relations or other perceptual primitives. I will also make a distinction between first-person and third-person uses of forces, where first-person forces involve *perception of bodily actions.* First person uses of forces are best described in terms of *power*.

Image schemas

These schemas in the head, what do they look like? We do not know much yet about how the brain handles meanings of words.[1] However, cognitive linguists speculate about the structure of schemas and what constituents they have. In any case, the schemas cannot consist of anything that looks like words, since we would then get something that looks like a *translation* from ordinary language to a "mental" language, which would not bring us much closer to what the words mean.

The most important theoretical notion in cognitive semantics is that of an *image schema*. A common assumption is that such schemas constitute the form of representation that is common to perception, memory, and semantic meaning. As an elementary example of an image schema, let us consider the schema for "over" as proposed by Lakoff (1987). "Over" denotes a spatial relation between two objects. The spatial relation concerns the vertical dimension, which is

[1] Although see e.g. Pulvermüller (2001) for some interesting neuroimaging results in this direction.

marked by the axis in figure 1. The object that is in *focus* (i.e. the one being ”over” the other) is called the *trajector*. Following Langacker's (1987) conventions, the focusing is marked by a thick line. The other object is called the *landmark*. In the basic schema for over proposed by Lakoff, the trajectory is supposed to be in horizontal motion in relation to the landmark, for example as in “The bird flew over the yard”.[2] Apart from this, the image schema does not contain any information about which kind of objects the trajector and the landmark are. So the figure is *not a picture* of how the world looks like, but just a schema that can be complemented with more details about its constituents.

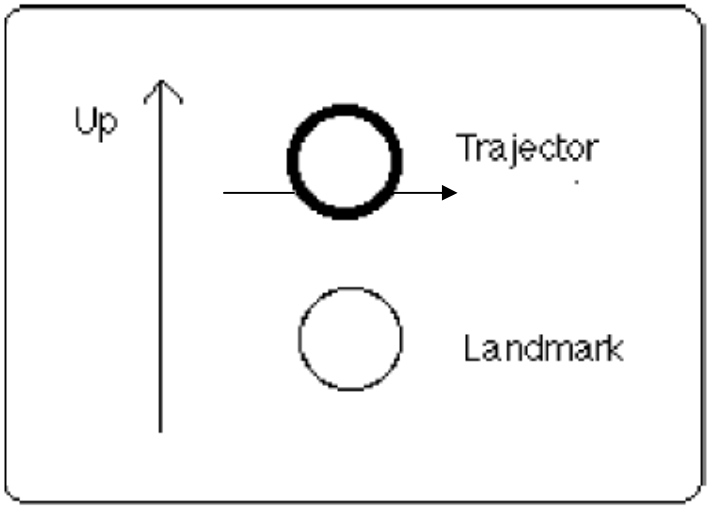


Fig. 1: Image schema for “over.”

Almost exactly the same image schema can be used to represent the meaning of ”under” (see Fig. 2) The only difference is that it is the object that is located *lowest* on the vertical axis that is in focus and thus the trajector, while the other object becomes the landmark. The close relation between the two schemas shows how the meanings of “over” and “under” are connected.

[2] However, there are many variations of meanings of ”over” in Lakoff's analysis that include a stationary trajector. Furthermore, Dewell (1994) argues that in the most natural image schema for ”over” the trajector moves in an arch over the landmark.

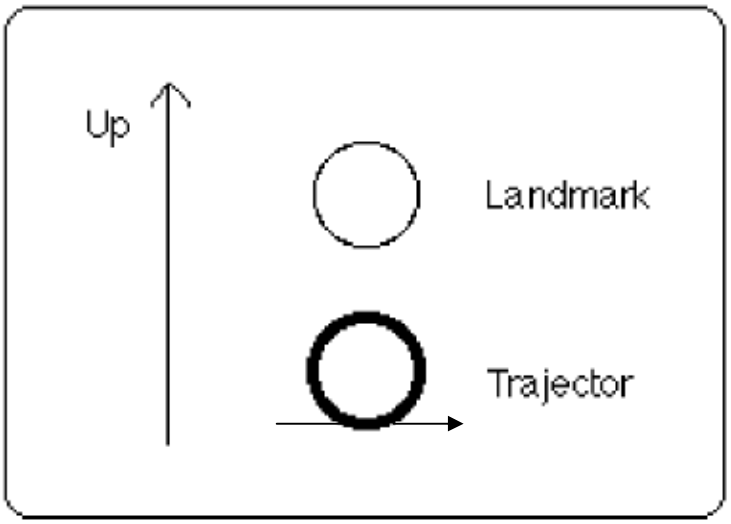


Fig. 2: Image schema for “under.”

There is a close connection between the image schemas of early cognitive semantics and visual processes: The distinction between “trajector” and “landmark” is the same as the distinction between “figure” and “background” in visual perception. The trajector is in the *focus of attention*.

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 *kinaesthetic* experiences.

Every word is supposed to correspond to an image schema. For example, a verb corresponds to a process in time. In a simplified way, such a process can be described as three stages: one schema part for the beginning, one part for the middle and one part for the end of the process. The image schema for ”climb” can look as in Fig. 3 (from Langacker 1987, p. 311).

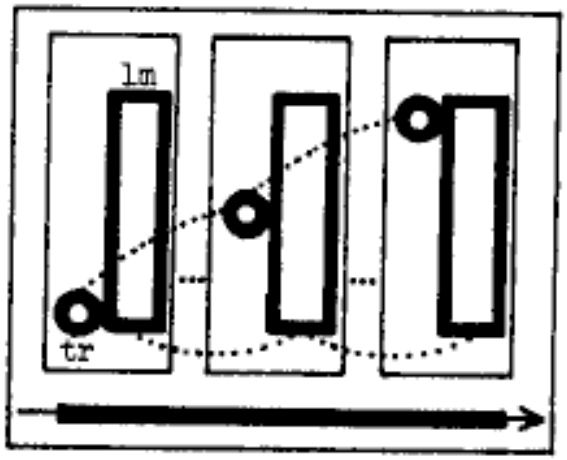


Fig. 3: Image schema for ”climb” (from Langacker 1987).

In this diagram, two dimensions are relevant: the vertical and the temporal. The axis representing the temporal dimension is drawn below the three stages and it is marked as a thick line since the temporal aspect of ”climb” is in focus. The landmark is supposed to be vertically extended and the trajector (the small circle) is assumed to be in physical contact with the landmark.

According to Langacker, the schema for the verb ”climb” can be turned into a schema for ”climber” by using the same dimensions, objects and relations and only changing the *focus* of the schema from the time dimension to the trajector, i.e. the thing doing the climbing (see Fig. 4).

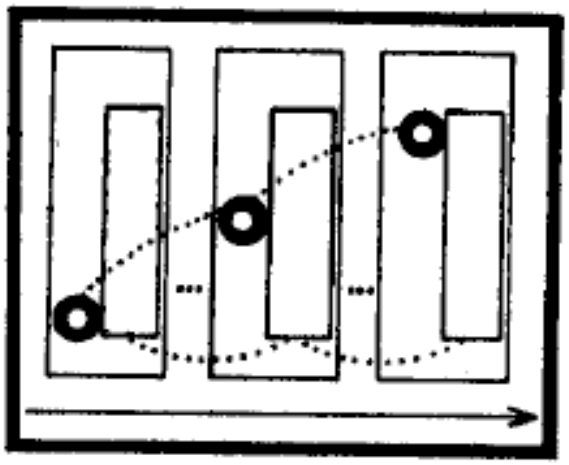


Fig. 4: Image schema for ”climber” (from Langacker 1987).

The change can be viewed as an example of *refocusing*. This kind of change has an obvious parallel in vision where one by looking at the same scene can be involved in very different cognitive processes by focusing on different aspects of the scene.

As a more advanced example of an image schema, consider Langacker's (1991, p. 22) depiction of “across” in Fig. 5. According to Langacker the meaning of “across” is a “complex atemporal relation” where the trajector (the small circle) is located in different relations to an elongated object, the landmark (the thick rectangle). First, the trajector is outside the landmark, then it is inside, and finally it is on the other side. The image schema contains two domains: a *time dimension*, marked by the horizontal arrow at the bottom, and *two spatial dimension*s, indicated by the rectangle which is repeated five times in different stages of the crossing.

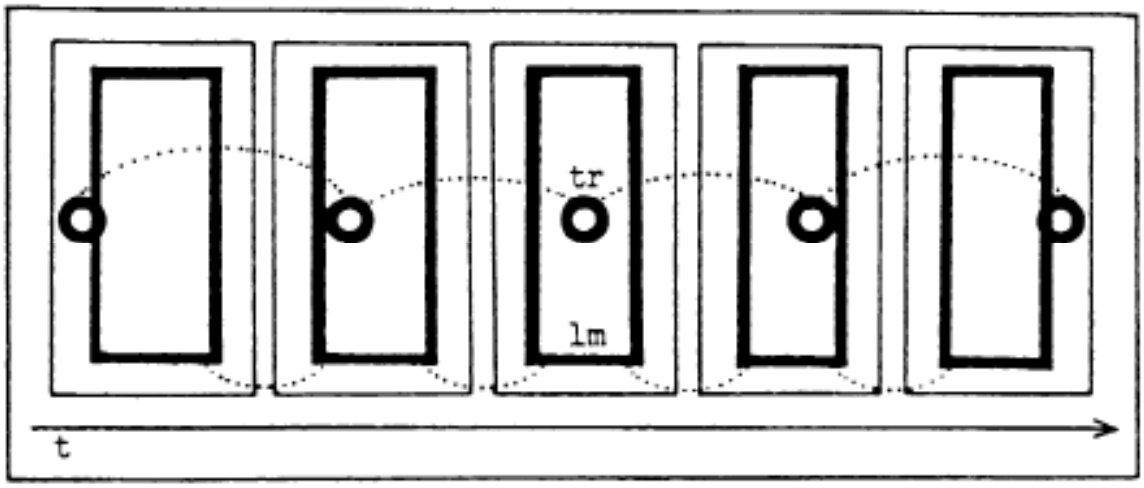


Fig. 5: An image schema for “across” (from Langacker 1991, p. 22).

The researchers within cognitive linguistics present lists of image schemas, but hardly any analysis of which schemas are possible and which are not. A developed theory of images schemas should present a principled account of what constitutes a schema. A fascinating proposal in this direction is that of Thom (1970, p. 232) who claims that any basic phrase expressing an interactive process can be described as one out of sixteen fundamental types. Among these types one finds “begin,” “unite,” “capture” and “cut.” The sixteen types are derived from some deep mathematical results concerning “morphologies” within catastrophe theory. Now, even if there are excellent mathematical reasons why there exist exactly sixteen types of interaction, it is not obvious that they correspond neatly to cognitive representations, even though such a correspondence would be very gratifying.

Neither Lakoff nor Langacker, who use the notion extensively, give a very precise definition of what constitutes an image schema. Zlatev (1997, pp. 40-44) argues that the notion is used in different ways by different cognitive semanticists. Johnson (1987) who was among the first to discuss image schemas is ambivalent between imagery and embodiment. He writes: ”I shall use the terms ”schema,” ”embodied schema,” and ”image schema” interchangeably” (1987, p. 29). Then he specifies the “image” version as a “dynamic pattern that functions somewhat like the abstract structure of an image, and thereby connects up a vast range of different experiences that manifest this same recurring structure” (ibid.). This should be contrasted with the “embodied” version that is described as follows: “[I]n order for us to have meaningful, connected experiences that we can comprehend and reason about, there must be pattern and order to our actions, perceptions, and conceptions. *A schema is a recurrent pattern, shape, and regularity in, or of, these ongoing ordering activities*. These patterns emerge as meaningful structures for us chiefly at the level of our bodily movements through space, our manipulations of objects, and our perceptual interactions” (ibid.).

Holmqvist (1993, p. 31), who tries to develop a formalism for image schemas suitable for computer implementations, defines image schemas as “that part of a picture which remains when all the structure is removed from the picture, except for that which belongs to a single morpheme, a sentence or a piece of text in a linguistic description of a picture […].” However,

the most condensed account I have found comes from Gibbs and Colston (1995, p. 349), who define image schemas as "dynamic analog representations of spatial relations and movements in space." Unlike Lakoff and Langacker, who focus on the spatial structure of image schemas, this definition puts the dynamics of the representations in focus.

In Gärdenfors (2000), I propose that a more precise account of what constitutes an image schema can be given with the aid of the theory of *conceptual spaces*. The spaces can be used to model the *domain* (Langacker 1987, p. 5) that forms the framework for an image schema, such as the spatial and temporal dimensions that have been used in the examples above. Image schemas are often just *geometric* or *topological* structures. For, example a "container" is a closed border that separates space into "inside" and "outside." An object, for example a cup, may be categorized as a container even though is it not physically closed. Cognitively, the rim surface of the cup functions as part of the border (see Herskovits (1986) for a discussion of how the border is determined).

The dynamic embodiment of image schemas

As can be seen from this brief presentation of image schemas, they are described as being tightly connected to the *perceptions* of the language users. However, on similar grounds as above, it can be argued that much of the meaning of words connects to the *actions* that the language users perform. As we shall see, there is a certain tension between these two positions.

In the tradition of Lakoff and Langacker, focus has been on the *spatial* structure of the image schema (the very name "image" schema indicates this). Lakoff (1987, p. 283) goes as far as putting forward what he calls the "spatialization of form hypothesis" which says that the meanings of linguistic expressions should be analyzed in terms of *spatial* image schemas plus metaphorical mappings. For example, many uses of prepositions, which primarily have a spatial meaning, are seen as metaphorical when applied to other domains (see, for example, Brugman 1981 and Herskovits 1986). Words like "in," "at," "on," "under," etc, primarily express spatial relations and when combined with non-locational words they create a "spatially structured" mental representation of the expression. These spatially structured representations are, naturally, also used when we are interpreting visual information. 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.

More recently, one can distinguish a more *dynamic* and *embodied* view on the image schemas (which thus no longer are mere "images"). As we shall see below, an early proponent for this view is Talmy (1988), who emphasizes the role of *forces* and dynamic pattern in image schemas in what he calls "force dynamics". This will be the topic of the following section.

There is a new tradition within cognitive science that emphasizes the role of *mental simulation* in cognitive processes (or mental emulation to follow Grush's (2004) terminology). Barsalou (1999) argues that concepts should be understood as *perceptual symbols* that are dynamic patterns of neurons functioning as simulators that combine with other processes to create conceptual meaning. However, in Barsalou's theory, these meaning carriers are still closely related to perceptual processes (hence their name). Another tradition focuses on the *motor* aspects of meaning schemas. For example, Pulvermüller (2001) has shown that when you, for example, read the word "kick", the same part of motor cortex is activated as when you actually

kick. It thus seems that the brain simulates the action it reads about. Another fascinating result is presented by Glenberg and Kaschak (2002), who demonstrate that processing sentences describing hand movements involves activating motor programs for such movements. In their experiment, subjects were asked to judge the acceptability of sentences describing movement to or from the body, e.g. "Put your finger under your nose" vs. "Put your finger under the faucet". To respond yes or no, they had to push buttons that were either close or further away from their own body. Glenberg and Kaschak found that subjects took longer time to respond when the direction on the action in the sentence was *opposite* to the direction of their hand movement when pressing the button. This indicates that the sentence had already activated a motor schema that contravened the correct response movement.[3]

Actions, forces and embodied schemas

With this dynamic and embodied perspective in mind, one can note that even for typical spatial prepositions such as "in," there are elements of their meaning that depend on *force relations*. For example, Herskovits (1986) notes that the topmost pear in figure 6a is considered to be "in" the bowl even though it is not spatially inside the bowl. If the other pears are removed, but the topmost is left in exactly the same spatial position as in figure 6b, then the pear is no longer "in" the bowl. So spatial location is not sufficient to determine whether an object is "in" a bowl. In figure 6a, the reason why the topmost pear is "in" the bowl is that it is physically *supported* by the other pears, while in 6b it has no such support. The notion of "support" clearly involves forces.

(a) (b)

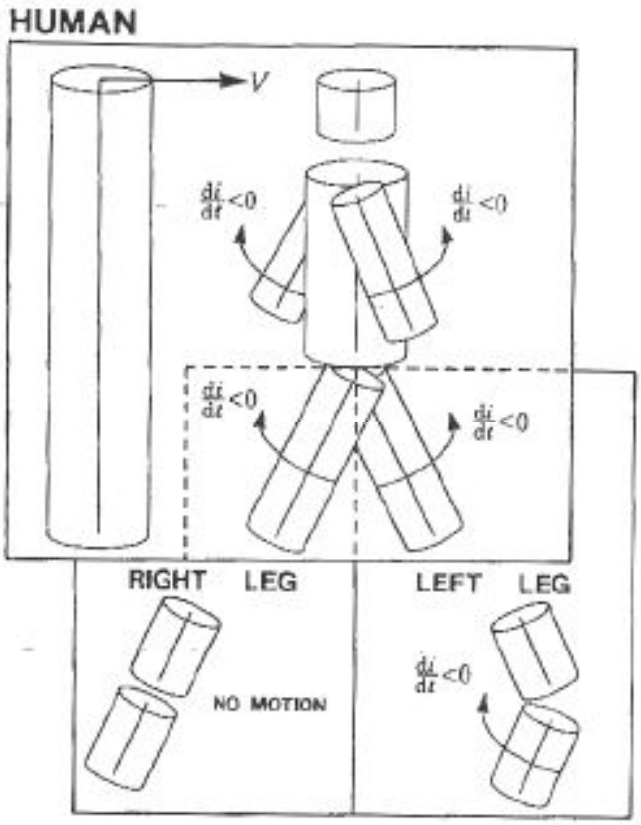
Figure 6. (a) The topmost pear is in the bowl. (b) The topmost pear is not in the bowl.

Also Bowermann's (1996) analysis of "aan," "op" and "in" in Dutch as well as the Korean prepositional verbs "nehta" (put loosely in or around) and "kkita" (fit tightly) involve force components that cannot be reduced to spatial relations.

In general, the role of *forces* is underrated within cognitive semantics. In Piaget's theory of sensory-motor schemas, which were developed for modeling cognitive development and not semantics, motor patterns are central. These can be seen as a special case of the dynamic patterns that form our fundamental understanding of the world. I believe that many ideas from the schemas of developmental psychologists can fruitfully be incorporated in the construction used by cognitive semanticists. "Image" schema is thus partly a misnomer. Schemas based on forces, embodied or not, should also be elements in the toolbox used for describing semantic structures.

[3] See Tseng and Bergen (2005) for a similar result concerning sign language,

If we consider the basic building blocks in natural languages, *verbs* normally express actions. An important question is how the meaning of such verbs can be expressed with the aid of images schemas. One idea comes from Marr and Vaina (1982), who extend Marr and Nishihara's (1978) cylinder models of objects to an analysis of *actions*. An action is, in Marr and Vaina's model, described via differential equations for movements of the body parts of, say, a walking human (see Fig. 6).[4]



4 To be accurate, Marr and Vaina (1982) only use differential inequalities, for example, expressing that the derivative of the position of the upper part of the right leg is positive in the forward direction during a particular phase of the walking cycle.

Fig. 6: Dynamical cylinder model representing "walk" (from Marr and Vaina 1982)

It is clear that these equations can be derived from the *forces* that are applied to the legs, arms, and other moving parts of the body. As we shall see below, Talmy (1988) convincingly demonstrates that a great deal of our understanding of verbs depends on the forces that are involved in the actions expressed by the verbs.

Similar considerations apply to *functional* concepts. For example, Vaina (1983) notes that when deciding whether an object is a "chair," the perceptual dimensions of the object, like those of shape, colour, texture and weight, are largely irrelevant, or at least extremely variable. I propose to analyse functional concepts by reducing them to the actions that the objects "afford." To continue with the example, a chair is prototypically an object that affords back-supported *sitting* for one person, that is, an object that contains a flat solid surface at a reasonable height from the ground and another flat surface that supports the back. In support of this analysis, Vaina (1983, p. 28) writes: "[T]he requirement for efficient use of objects in actions induces strong constraints on the form of representation. Each object must first be categorized in several ways, governed ultimately by the range of actions in which it can become involved."

In more general terms, I propose that function concepts be interpreted in terms of an *action space*. Even though we still know very little about how action space is structured (see Gärdenfors, to appear), it seems clear that an analysis of actions should be based on forces; and for a large class of actions on embodied forces.

Langacker's analysis of image schemas is done mainly in spatial terms. For example, his description of "climb" involves only the vertical dimension, together with the time dimension, where the latter is ubiquitous with the verbs. No forces are involved in his analysis. Thus the schema does not differentiate between "pull up," "push up," and "climb." What is missing is that the meaning of "climb" involves that the trajectory exerts a vertically directed force.

The upshot is that by adding force dimensions to an image schema, we may obtain the basic tools for analyzing dynamic properties. The forces involved need not only be physical forces, but can also be *social* or *emotional* forces. I will return to this distinction below.

One who very early emphasized the role of forces in image schemas is Mark Johnson, who in his 1987 book *The Body in the Mind* argues that forces form perceptual Gestalts that serve as image schemas (even though the word "image" may be misleading here).

He writes:

> Because force is *everywhere*, we tend to take it for granted and to overlook the nature of its operation. We easily forget that our bodies are clusters of forces and that *every* event of which we a part consists, minimally, of forces in interaction. […] We *do* notice such forces when they are extraordinarily strong, or when they are not balanced off by other forces. (1987, p.42)

Johnson presents a number of "preconceptual Gestalts" for force. These Gestalts function as the correspondence to image schemas but with forces as basic organizing feature rather than spatial relations. The force Gestalts he presents are "compulsion," "blockage," "counterforce," "diversion," "removal of restraint," "enablement" and "attraction" (1987, pp. 45-48).

Another exception is Talmy (1988), who emphasizes the role of *forces* and dynamic pattern in image schemas in what he calls "force dynamics". He recognizes the concept of force in such

expressions such as (1) *The ball kept (on) rolling along the green* and (2) *John can't go out of the house*. He also notices the possibility in language to choose between what he calls force-dynamically neutral expressions and ones that do exhibit force-dynamic patterns, like in (3) *He didn't close the door* and (4) *He refrained from closing the door*.[5] Forces are furthermore taken as governing the linguistic causative, extending to notions like letting, hindering, helping, etc. He develops a schematic formalism that, for example, allows him to represent the difference in force patterns in expressions like "The ball kept rolling because of the wind blowing on it" and "The ball kept rolling despite the stiff grass."

Talmy's dynamic ontology consists of two directed forces of unequal strength, the focal called "Agonist" and the opposing element called "Antagonist", each force having an intrinsic tendency towards either action or rest, and a resultant of the force interaction, which is either action or rest.

> All of the interrelated factors in any force-dynamic pattern are necessarily copresent wherever that pattern is involved. But a sentence expressing that pattern can pick out different subsets of the factors for explicit reference – leaving the remainder unmentioned – and to these factors it can assign different syntactic roles within alternative constructions. (Talmy 1988, p. 61)

Power vs. force

After having argued for the importance of forces in cognitive semantic analyses, I now turn to an ambiguity in the very notion of "force". In the Western academic society, Newtonian physics has become a role model for science and when we speak of "force" it is natural to think of them and *represent* them as Newtonian forces. But when it comes to everyday human thinking, I believe it is important to distinguish between a *first-person* and a *third-person* perspective of forces.

From the first-person perspective, it is the forces that act directly on you that are considered. These "forces" are not just the physical Newtonian forces, but more importantly also the *social* or *emotional* forces that affect you. It is perhaps more appropriate to call forces seen from a first-person perspective "powers." First-person powers are *experienced* either as physical forces or as emotional or social pressures that make you move in a particular direction. A happy smile of a person that makes you behave more gently and a threat of a monetary fine if you park in front of the cathedral that prevents you from doing so are examples of emotional and social powers that cannot be reduced to physical forces. Thus, from the first-person perspective, the category of powers is broader than that of physical.

From the third-person perspective, one sees forces acting upon an object from the outside, so in this case you do not experience the forces directly, but your perceptual mechanisms infer them. Hence, they are not *embodied* in the same way as in the first-person perspective. From the first-person perspective, powers act directly on you, while from the third-person perspective forces act at a distance (*pace* Newton).[6]

---

5 Examples from Talmy (1988, p. 52).

6 There is also a second-person perspective where the subject can "put himself in the shoes of the other". This perspective is what is involved in empathy, joint attention and other aspects of a "theory of mind" (see Gärdenfors 2003, ch. 4). Some researchers put forward "mirror neurons" as a possible mechanism behind this perspective (e.g. Arbib and Rizolatti (1998) and Gallese (2000))

One reason for why this distinction is seldom made is that we are extremely good at perceiving forces acting upon other objects.[7] In the Uppsala school of psychology, where Gunnar Johansson and Sverker Runesson are the main proponents, it is claimed that we can *directly perceive* the forces that control different kinds of motion. According to their Gibsonian perspective, the information that our senses, primarily vision, receive about the movements of an object or an individual is sufficient for our brains to be able to extract, with great precision, the underlying forces. Furthermore, the process is automatic – we cannot help but seeing the forces. Runesson (Runesson 1994, Runesson and Frykholm 1983) calls this process ”kinematic specification of dynamics.” Of course, the perception of forces is not perfect – we are prone to illusions, just as we are in all types of perception.

The importance of this distinction is that our understanding of the third-person perspective *derives* from the first-person person perspective. (This is why Newton had such problems in convincing his contemporaries about forces acting at a distance). The upshot is that meanings of words that are based on first-person powers should be seen as more fundamental than meanings based on third-person forces. In brief, I submit that the meanings of the force elements of image schemas are grounded in the first-person perspective.

There is much in Johnson’s (1987) book that indicates the centrality of the first-person “power” perspective. For one thing, he focuses on the role of interaction: “force is always experienced through *interaction*. We become aware of a force as it affects us or some object in our perceptual field” (p. 43). Interaction is primarily seen from a first-person perspective, while forces are abstractions that are seen from a third-person view.

Furthermore, in his description of the “enablement” Gestalt or schema, he explicitly focuses on first-person “powers”:

> If you choose to focus on your acts of manipulation and movement, you can become aware of a felt sense of power (or lack of power) to perform some action. You can sense that you have the power to pick up the baby, the groceries, and the broom but not to lift the front end of your car. While there is no actualized force vector here, it is legitimate to include this structure of possibility in our common gestalts for force, since there are potential force vectors present and there is a definite “directedness” (or potential part of motion) present. (Johnson 1987, p. 47)

This quotation is a perfect example of what I mean by the first-person perspective of forces.

Interpersonal power relations play an important role in our everyday social interactions. Parents use their power over their children, employers over their employees, teachers over their pupils, etc. These power relations are as ubiquitous as they are implicit: it is very rare to see the subordinate *explicitly* challenging the one in power, although the tacit power structure is well known to all of us. The power structure is constantly being stabilized, confirmed, questioned, and sometimes challenged through our interactions. Linguistic utterances are one ways of expressing these relations, as are actions, body language, etc.

A brief case study: modal verbs

As an illustration of the role of power relations in semantic analyses, I shall briefly account for the analysis of *modal verbs* presented in Winter and Gärdenfors (1995) and Gärdenfors (1998).

[7] However, it seems that other animal species may not have this capacity to the same extent (see e.g. Povinelli 2000).

First of all, it should be noted that is very unnatural to ask for the *reference* of modal verbs such as "want," "must," "may" and "shall" (let alone their reference in other "possible worlds," which have been the dominating semantic entities in modal logic). The modal verbs have a role not primarily with respect to the sentence where they occur, but in relation to the *speech act* in which they are uttered (Austin 1962, Searle 1979). Consequently, the main linguistic objects of the analysis are speech acts. The meaning of modal verbs is thus more a matter of *pragmatics* than of traditional (referential) semantics.

The semantics for modal expressions that I outline here is both *cognitive* and *social*. Unlike most works in the area of cognitive semantics (Lakoff 1987, Langacker 1987), I do not use spatially based image schemas to describe the meanings of modals, but the semantic primitives are *power relations* and *expectations*.[8] In brief, it is proposed that the primary meaning of modal verbs is to express power relations and exactly what is said depends on the expectations of the speaker. The most elementary power relations are to be dominant or to be subordinate – one is in power and the other is obedient. The speaker's expectations may, for example, concern what the speaker believes about what the listener wants, what the listener believes, or what the listener believes about what the speaker wants.[9]

The main justification for my choice of semantic primitives derives from the fact that language is not autonomous. It is a tool to be used in a *social context*. The social context is partly determined by the power relations between the agents. The objects of power are actions to be performed, for example the action of killing a wasp. I can kill it myself, but if I have power over you, I can also command you to do it by saying "You *must* kill the wasp!" Another important factor of a speech situation is the agents' *attitudes* to the relevant actions.[10] For example, I may want to kill the wasp, while you may want this action not to be performed.

For an example of how these factors are used in analyzing the meaning of modal verbs, consider a man saying *I ought to go now*. In doing this he indicates that the listener has some power over him, and that he believes that the listener expects him to stay, but there is some other, morally (or physically) stronger power that forces him to leave. If the speaker did not believe that the listener expected him to stay, he would have said nothing. Or if listener has some power of the speaker and expects him to stay, but there is no external stronger power, then the speaker would have said "May I go now."

For a second example, consider a situation when the speaker, who has the power, expects that the listener is negative to eating the oatmeal, orders him to do it by uttering "You *shall* eat your oatmeal!" By using the modal verb "shall," the speaker indicates her expectation about the listener's reluctance. If the speaker believed that the listener wanted to eat the oatmeal, she would have said nothing, and if the speaker believed that the listener wanted to eat the oatmeal and that the listener also believed that the speaker did not want him to eat it, she would have said "You *may* eat the oatmeal."

---

[8] For a more detailed exposition of power relations and expectations, see Winter and Gärdenfors (1995).
[9] For a general treatment of the role of expectations in reasoning, see Gärdenfors (1994).
[10] Attitudes to actions concern the agents' *preferences*, and should not be confounded with so called *propositional attitudes*, e.g. believing or hoping.

As a comparison, Talmy (1988) uses his "force dynamics" to generate an analysis of modal expressions. In his analysis, *physical* forces are seen as more fundamental than the social. By metaphorical extension, the expressions used to express physical forces are used in the "psychological, social, inferential, discourse, and mental-model domains of reference and conception" (1988, p. 49).

One difference between the two analyses is that Talmy views utterances in the first or second person involving modals as parallel to sentences in the third person with the same modals while the position presented here is that first and second person utterances involving modals are primary. Modal verbs are pragmatic operators used in speech acts. In support of this, it can be noted that questions and imperatives are typical moods for modals while declarative uses of modals will occur only in certain dialogue situations. Speech acts can be seen as *moves* in a "language game" with the *stakes* given by the social relations that include the assignment of power, while third person expressions are secondary *reports* of such moves (cf. Sweetser 1990, p. 65).

In contrast to Talmy (and Johnson), I view social power relations as semantically fundamental, and physical forces as derived. Focusing on speech acts brings forward the power play of cognitive agents. I readily admit a great influence from Talmy's force dynamics, but I have found that the *attitudes* to actions, and above all expectations about these attitudes, are more adequately accounted for in terms of social power than in terms of physical forces.[11]

In line with this position, Talmy (1988, p. 79) concedes that "[a] notable semantic characteristic of the modals in their basic usage is that they mostly refer to an Agonist that is sentient and to an interaction that is psychosocial, rather than physical, as a quick review can show." I completely agree, but see this as an argument for the primary meaning of the modals being determined by social power relations, while the (few) uses of modals in the context of physical forces are derived meanings.

Conclusion

This article has started out from cognitive semantics and the notion of an image schema as the basic carrier of meaning. In cognitive semantics, the focus has been on spatial and visual aspects of image schemas (as the word "image" indicates). But the meanings of many words depend on functional properties, which, in turn, relate to the actions involved. Therefore, I have proposed that the perceptual dominance of cognitive semantics should be complemented by an account of actions and the forces or powers on which they depend.

My main thesis has been that the meanings of many expressions cannot be properly analyzed unless one accounts for the forces that are involved in a process. By adding force dimensions to an image schema, we obtain the basic tools for analyzing dynamic and functional properties. I have also emphasized the distinction between embodied first-person forces, which may be called powers, and third-person external forces. Among the first-person powers one not only physical forces but also social and emotional factors. In most cases, such force patterns are embodied in the sense that they are grounded in our bodily experiences.

11 Talmy is not by any means unaware of some of the wider uses of forces/power, and he writes: "In addition, FD [force dynamic] principles can be seen to operate in discourse, preeminently in directing patterns of argumentation, but also in guiding discourse expectations and their reversal." (Talmy 1988, p. 50)

The pragmatic analysis of the meaning of modal verbs that I have outlined is based on social power relations. This analysis is one indication of the rich possibilities that open up if forces or powers, in particular first-person powers, are added to the toolbox of semantic analysis in the cognitive tradition. It also shows that it is difficult, and to a large extent artificial, to keep a sharp border between semantics and pragmatics.

The emotional and social powers that become relevant from a first-person perspective bring forth a connection to the debate on the role of "theory of mind" for semantics and for language learning (see Bloom 2000 for a survey). When children learn the meaning of a new word, they also learn something about the thoughts of others. Semantic development and the growth of a "theory of mind" are strongly correlated. The semantic learning of children builds very strongly on the (implicit) assumption that every word has a function. Therefore, when a child hears a new word it presumes that there is some concept corresponding to the word and it makes an effort to identify this concept. This view presumes that concepts come before the words that denote them. This view has been contested by philosophers who claim that thinking requires words, but Bloom (2000) argues that there is no support for this claim (see also Gärdenfors 2003). I totally agree with his position that language is primarily a tool for communicating concepts and ideas, and not a mechanism for generating the concepts.

To give a deeper understanding of how children grasp the meaning of words, it would therefore be fruitful to further develop the connection between cognitive semantics and an embodied "theory of mind". For example, the work of Tomasello and his colleagues (1999, to appear) on the role of joint attention and joint intentions is highly relevant here. There is also recent work on different levels of mimetic communication (Zlatev, Persson and Gärdenfors, to appear) that brings in embodied aspects of communication that would be important for a semantic/pragmatic analysis.

References

Austin, J. L. (1962), *How to Do Things with Words*, London: Oxford U. P.

Bloom, P. (2000), *How Children Learn the Meaning of Words*, MIT Press, Cambridge, MA.

Bowermann, M. (1996). "How much space gets into language?" in *Language and Space*, P. Bloom, M. A. Peterson, L. Nadel and M. F. Garrett, eds., MIT Press, Cambridge, MA, 385 – 436.

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

Croft, W. and Cruse, D. A. (2004), *Cognitive Linguistics*, Cambridge University Press, Cambridge.

Dewell, R. B. (1994), "*Over* again: Image-schema transformations in semantic analysis," *Cognitive Linguistics 5*, 351-380.

Evans, V. (2006), "Lexical concepts, cognitive models and meaning-construction", *Cognitive Linguistics 17*, 491-534.

Gallese V. (2000), "The inner sense of action: Agency and motor representations," *Journal of Consciousness Studies* 7(10), 23-40.

Gärdenfors, P. (1994), "The role of expectations in reasoning", in *Knowledge Representation and Reasoning Under Uncertainty*, M. Masuch and L. Polos (eds.), Berlin: Springer-Verlag, 1-16.

Gärdenfors, P. (1998), "The pragmatic role of modality in natural language", pp. 78-91 in *The Role of Pragmatics in Contemporary Philosophy*, ed. by P. Weingartner, G. Schurz and G. Dorn, Hölder-Pichler-Tempsky, Vienna.

Gärdenfors, P. (2000), *Conceptual Spaces: The Geometry of Thought*, MIT Press, Cambridge, MA.

Gärdenfors, P. (2003), *How Home Became Sapiens*, Oxford University Press, Oxford.

Gärdenfors, P. (to appear), "Representing actions and functional properties in conceptual spaces", to appear in *Body, Language and Mind*, ed. by T. Ziemke and J. Zlatev.

Gibbs, R. W. and Colston, H. L. (1995), "The cognitive psychological reality of image schemas and their transformations," *Cognitive Linguistic 6,* 347-378.
Glenberg, A. M. and Kaschak, M. P. (2002), "Grounding laguage in action", *Psychonomic Bulletin and Review 9*, 558-565.
Grush, R. (2004), "The emulator theory of representation: motor control, imagery and perception," *Behavioral and Brain Sciences* **27**, 377-442
Harnad, S., ed., (1987), *Categorical Perception*, Cambridge University Press, Cambridge.
Herskovits, A. (1986), *Language and Spatial Cognition: An Interdisciplinary Study of the Prepositions in English*, Cambridge University Press, Cambridge.
Holmqvist, K. (1993) *Implementing Cognitive Semantics*. Lund University Cognitive Studies 17, Lund.
Jackendoff, R. (1983), *Semantics and Cognition*, MIT Press, Cambridge, MA.
Jackendoff, R. (1987), *Consciousness and the Computational Mind*, Cambridge, MA: MIT Press.
Johnson, M. (1987), *The Body in the Mind: The Bodily Basis of Cognition*, University of Chicago Press, Chicago, IL.
Lakoff, G. (1987), *Women, Fire, and Dangerous Things*. The University of Chicago Press, Chicago, IL.
Langacker, R. W. (1986), "An introduction to cognitive grammar," *Cognitive Science 10*, 1-40.
Langacker, R. W. (1987), *Foundations of Cognitive Grammar, Vol. I,* Stanford University Press, Stanford, CA.
Langacker, R. W. (1991), *Concept, Image, and Symbol: The Cognitive Basis of Grammar*, Mouton de Gruyter, Berlin.
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*: 269-294.
Marr, D. and Vaina, L. (1982), "Representation and recognition of the movements of shapes," *Proceedings of the Royal Society in London, B214*, 501-524.
Povinelli, D. (2000), *Folk Physics for Apes*, Oxford University Press, Oxford.
Pulvermüller, F. (2001), "Brains reflections of words and their meanings," *Trend in Cognitive Science*, vol. 5, no. 12, 512-524.
Regier, T. (1996), *The Human Semantic Potential: Spatial Language and Constrained Connectionism*, MIT Press, Cambridge, MA.
Rizolatti, G. and Arbib, M. A. (1998), "Language within our grasp," *Trends in Neuroscience 21*, 188-94.
Runesson, S. (1994), "Perception of biological motion: The KSD-principle and the implications of a distal versus proximal approach," in Gunnar Jansson, Sten-Sture Bergström, and William Epstein, (eds.), *Perceiving Evens and Objects*: 383-405. Hillsdale, NJ: Lawrence Erlbaum Associates.
Runesson, S. and Frykholm, G. (1983), "Kinematic specification of dynamics as an informational basis for person and action perception. Expectation, gender recognition, and deceptive intention," *Journal of Experimental Psychology: General 112*:585-615.
Searle, J. R. (1979), *Expression and Meaning: Studies in the Theory of Speech Acts*, Cambridge: Cambridge U. P.
Sweetser, E. (1990), *From Etymology to Pragmatics*, Cambridge University Press, Cambridge.
Talmy, L. (1988), "Force dynamics in language and cognition," *Cognitive Science 12*, 49-100.
Thom, R. (1970), "Topologie et linguistique," in *Essays on Topology*, Springer-Verlag, Berlin, 226-248.
Tseng, M. J. and Bergen B. K. (2005), Lexical processing drives motor simulation", in *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society*.
Vaina, L. (1983), "From shapes and movements to objects and actions," *Synthese 54*: 3-36.
Winter, S. and Gärdenfors, P. (1995), "Linguistic modality as expressions of social power", *Nordic Journal of Linguistics 18*, 137-166.
Zlatev, J. (1997), *Situated Embodiment: Studies in the Emergence of Spatial Meaning*, Gotab Stockholm.