Playpen: Toward an Architecture for Modeling the Development of Spatial Cognition*

[1,2]Michael Gasser and [1]Eliana Colunga
[1]Computer Science Department
[2]Linguistics Department
[1]Cognitive Science Program
Indiana University
Bloomington, IN 47405, USA
{gasser,ecolunga}@cs.indiana.edu

June 23, 1997

*This report has been prepared in the form of an HTML document, which is available on the World-Wide Web at `ftp://ftp.cs.indiana.edu/pub/gasser/Playpen/TR1/tr/tr.html`. It makes extensive use of color figures, animated applets, and hyperlinks. There is a compressed Postscript version of the report available by anonymous ftp to `ftp.cs.indiana.edu` at `/pub/gasser/Playpen/TR1.ps.Z`. This includes the color figures but not the applets and hyperlinks. Therefore we highly recommend the version on the Web.

Abstract

In this report we argue that the study of the acquisition of word meaning requires taking seriously non-linguistic cognition, in particular human vision and the pre-linguistic development of concepts. We consider the implications of this claim for the acquisition of spatial relations, and we present Playpen, an evolving neural network architecture for modeling the development of spatial language and spatial cognition. Playpen includes modules for high-level vision, the lexicon, and the conceptual space in which vision and lexicon come together, allowing for the mutual influence of all three. This report focuses on the basic building blocks of the network. Feature binding and object segregation are implemented through the use of phase angles, and the learning algorithm is a version of Contrastive Hebbian Learning (Movellan, 1990), adapted for units with phase angles. We argue that to represent and learn the meanings of relational terms, the network also requires units which represent micro-relations explicitly. In Playpen these take the form of **relation units**, hard-wired clusters of simpler units which become activated to the extent that they receive inputs from units representing distinct objects.

1 Introduction

How do words get their meanings? Does this process depend on the details of what the words are about, that is, on the way things actually are in the world? If it does, does it also depend on the mechanisms within learners which allow them to deal with and understand the world, that is, with their sensory/perceptual and motor systems? And if this is the case, does the way the world is understood by learners depend in turn on word meaning and how it is learned?

These are questions that have been around for a long time. They bear on issues as fundamental as what language is and what symbolic cognition in general is. In this paper we will argue briefly that much can be gained by assuming that the answer to all three questions is yes: regularities in the world and the mechanisms of perception and action matter for language, and language in turn matters for cognition. Then we will discuss some methodological consequences of taking this position, and we will consider what this position means for a particular semantic domain, that of spatial relations. Finally, we will present Playpen, an evolving connectionist model designed to simulate the development of spatial cognition and spatial language. We believe that the fundamental questions can only be answered by such a model, one which makes concrete predictions about the behavior that children actually exhibit.

2 Modeling Language Acquisition

2.1 Meaning, Concepts, and Perception in Models of Language Acquisition

Where does linguistic meaning fit into cognition? In this section we consider three possible positions one could take on this question and some of the implications of these positions.

All models must deal with the obvious fact that different languages have different semantic systems, that they divide up the world in different ways. One important distinction between models concerns the way in which language-specific semantics is hooked up to the rest of the cognitive system in such a way that linguistic behavior is roughly appropriate to the non-linguistic context.

Probably the most widely-held view is one in which linguistic meaning is viewed as a symbolic system which maps onto a universal and symbolic conceptual system. This conceptual system may be innate or learned pre-linguistically. For example, this is the view of two very influential frameworks, those of Pinker (1994) and Jackendoff (1992) . We will refer to models that adhere to this view as **symbolic** models of linguistic behavior and acquisition. On the symbolic view, language acquisition involves learning about the particular language's semantic structures and how they map onto universal conceptual structure. Conceptual structure is in turn related to perception and action in ways which are usually left unspecified. An overview of such a perspective is shown in Figure 1.

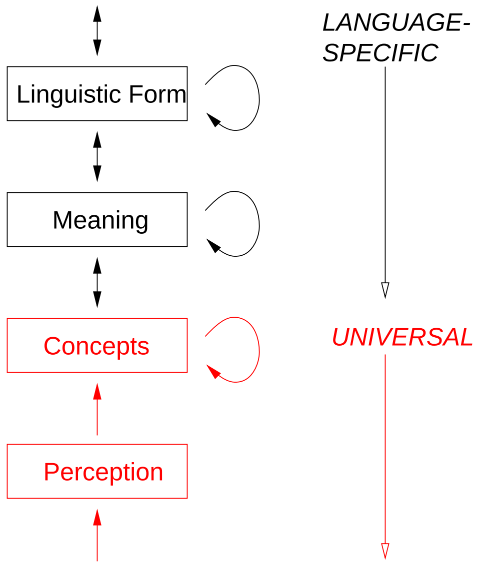

Figure 1: **Symbolic Semantics, Symbolic Concepts**. Meaning and non-linguistic concepts arise independently of each other.

If this position holds, then language acquisition can be studied and modeled without taking perception into account. The acquisition of semantics is a symbolic phenomenon, relating one developing symbolic system, the semantics of the target language, with another, the existing conceptual system. Furthermore, since basic conceptual structure is in place before language is learned, this position does not allow for any significant influence of linguistic categories on concepts. Finally, because semantic development in any language is driven by the same underlying conceptual system, there should be strong similarities in the developmental course of the learning of meaning across diverse languages.

A range of alternatives to this popular view have been set forth in recent years. These positions agree in assigning more significance to the interplay between linguistic and non-linguistic cognition. We will refer to them as **grounded** models of linguistic behavior and/or acquisition. Grounded models are associated with cognitive linguists (Lakoff, 1987; Langacker, 1987a) and with other cognitive scientists who seek to do away with mind-body distinction in one sense or another (Harnad, 1990; Johnson, 1987; Thelen and Smith, 1994; Varela et al., 1991).

We consider here two possible positions within the space of grounded models. The first is that of Regier (1996), whose computational model of the acquisition of spatial relation terms is one of a small number of serious attempts to actually implement the grounding idea. In Regier's model, linguistic meaning is learned directly via perception, and acquisition can only be studied in the context of a model of the vision system. The nature of this system constrains the kinds of possible meanings that languages can encode and the way in which these meanings are learned by children. However, Regier's model makes a clear division between vision and language and has no obvious place for spatial concepts. Presumably the acquisition of spatial terms has little or nothing to do with spatial reasoning, which in any case is not under the influence of linguistic categories. Furthermore, the model only runs in the production direction; it does not tell us how a child learns to comprehend spatial terms. A schematic of Regier's model is shown in Figure 2. The figure does not assign a place to non-linguistic concepts; presumably these would exist in a component of the system parallel to language.

A second, more radical, option within the space of grounded models also has linguistic meaning grounded in perception. The difference here is that there is no distinction between linguistic meaning and non-linguistic concepts and that the model runs in both directions, from language to vision as well

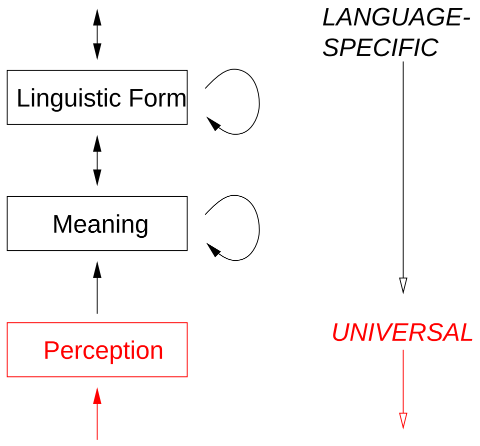

Figure 2: **Grounded Semantics**. Meaning arises directly out of perception but is kept separate from non-linguistic concepts.

as from vision to language. Particular meanings/concepts may be learned in three ways: through non-linguistic perceptual and motoric experience, through a combination of non-linguistic and linguistic experience, and through linguistic input alone. This type of model allows linguistic categories to influence concepts; that is, various forms of linguistic relativism (Gumperz and Levinson, 1996) are possible. Finally such a model predicts that the developmental course of the learning of meaning should depend on the categories inherent in the language being learned. A schematic of this sort of model is shown in Figure 3. This is the view that we favor and the one that is realized in Playpen, the connectionist model we present later in this report.

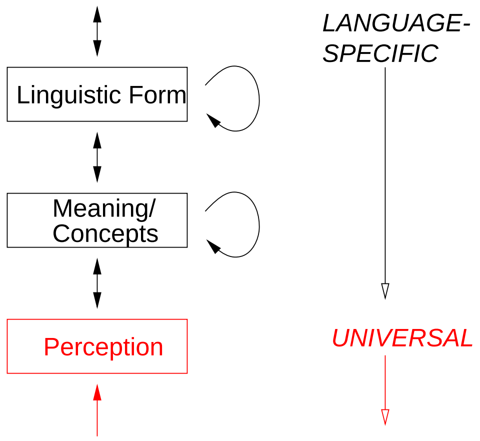

Figure 3: **Grounded Semantics/Concepts**. Meaning and concepts are not distinguished; interactions between language and perception can occur.

2.2 Horizontal and Vertical Approaches to Language

Language is too large a domain to deal with in its entirety, and language scientists must slice up the problem space in some way or another. Normally the slices made are **horizontal**. A body of research covers some aspect of language, for example, syntax or the syntax of relative clauses, or some form of linguistic behavior, for example, syntactic parsing or the parsing of relative clauses. The goal is relatively thorough coverage of the behavior. Contact is often not made with other aspects of language or linguistic behavior; for example, research on syntax may not make reference to phonology or pragmatics and research on parsing may not make reference to production or acquisition. And contact is even less often made with non-linguistic aspects of cognition or with the external world.

Vertical approaches, on the other hand, make explicit contact across different aspects of language or linguistic behavior or across the boundary between linguistic and non-linguistic. While such approaches are "tall," they are of necessity also "thin;" they can cover only very narrow aspects of language or linguistic behavior. Vertical approaches are associated in particular with cognitive linguistics, with anthropological linguistics, and with sociolinguistics. Figure 4 illustrates horizontal and vertical approaches to language.

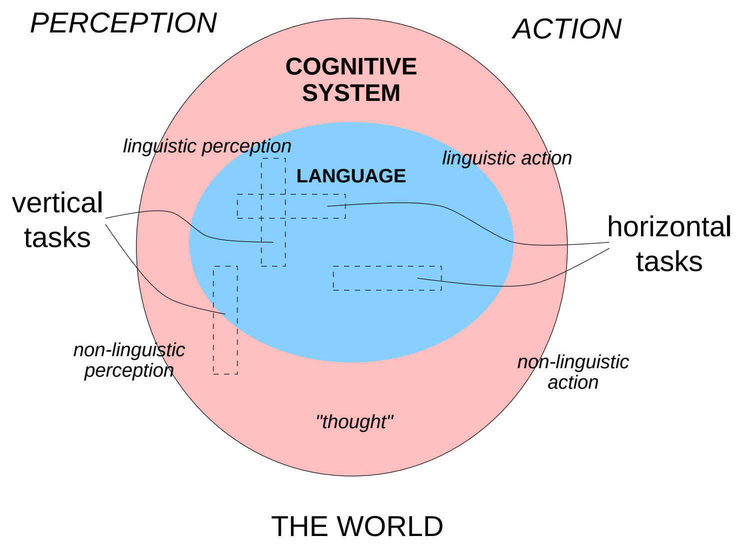

Figure 4: **Vertical and Horizontal Approaches to Language**. Approaches may cover crucial interactions in toy worlds (vertical) or a single domain in a relatively broad fashion (horizontal).

While they have the advantage of broad coverage of a domain, horizontal approaches may miss crucial interactions between domains. A model such as the one we proposed in the last section, one in which language and non-linguistic perception exert mutual influence on one another, obviously requires a vertical methodology. Language, vision, and non-linguistic concepts must all be taken seriously. This is a tall order, enough to daunt even someone who believes in the sorts of interactions we are suggesting. Such an approach can succeed only if

1. the range of linguistic phenomena covered is very narrow
2. there is a body of established results or a coherent theory to guide the modeling in each of the relevant domains.

We have chosen to focus on the **language of spatial relations** and how it emerges in children. Besides the cross-linguistic study of how space is depicted linguistically and the acquisition of spatial language, our modeling will take us into the vision system, in particular, the visual representation of relations, and into the development of concepts of space in children. We believe that progress in each of these areas has reached the point where one may attempt to tie the various pieces together. Our eventual contribution will be an integrated picture of the development of spatial relations, linguistic and otherwise. In the following sections we summarize briefly some relevant facts from these four areas. First we discuss the language of spatial relations and how spatial relations are acquired in different languages. Then we talk about the perceptual end, about vision in particular, and about the acquisition of spatial concepts independent of language. Finally we look at the interactions between the two ends.

3 Constraints on the Model

3.1 Language

3.1.1 The Language of Spatial Relations

Each language provides speakers, hearers, and learners with a finite set of lexical items and structures to apply to a continuous world, and it is convenient to view a language as "slicing up" the world in a particular way. There seem to be both universal and language-specific aspects to the way this happens. All languages apparently make a fundamental distinction between nouns on the one hand and several classes of words on the other, most centrally, verbs. It has been argued (Langacker, 1987b), though not uncontroversially, that this distinction corresponds to a fundamental conceptual distinction between **objects** and **relations**.

The nouns of a language divide the world into categories of objects and substances.[1] Verbs, prepositions, and postpositions break things down in a quite different way from nouns, singling out relations between the objects which the nouns refer to. Spatial relations are an important subcategory, and what is striking here is the relatively small number of discrete spatial relation categories that each language makes available. The **relation term** itself may be a preposition, postposition, verb, or even a noun inflection; morphological details will not concern us further. A complete **spatial relation expression** includes, in addition to the relation term itself, two noun phrases, representing the thing being related (the **trajector**) and the thing it is being related to (the **landmark**). The choice of trajector and landmark matters: *the stick is on the block* does not mean the same thing as *the block is under the stick*. Trajector seems to correlate with the perceptual figure (Herskovits, 1986; Langacker, 1987a).

Even a cursory examination of the spatial relation expressions in a subset of languages reveals that the space of possible relations is sliced up in a variety of ways. Consider some of the possibilities for encoding relations of CONTACT, SUPPORT, and CONTAINMENT between two objects (Landau, 1996). Four possible arrangements of a trajector (black) and landmark (brown) are shown Figure 5. Spanish uses a single word, *en*, for all of them. English uses one word, *on*, for the two situations in which CONTAINMENT does not enter in and another, *in*, for situations in which the trajector is (at least partially) contained in the landmark. German distinguishes two kinds of situations for which English uses *on*: *auf* when the landmark is under the trajector, *an* when the trajector is fixed to a vertical surface of the landmark. Korean distinguishes two kinds of CONTAINMENT (and CONTACT) situations, those in which the trajector fits tightly within the landmark, for which *sok* is used, and those in which there is loose fit, for which *ahn* is used.

But it is not languages which "slice up" the world in particular ways (languages don't actually "do" anything); it is people. In any case our goal is to model individual language learners, not the entire linguistic communities which embody particular languages. Descriptions of language and particular languages are useful to us only insofar as they give us clues about what people must learn to do in order to learn language.

Linguistic descriptions tell us that language is to a large extent about objects; thus a major task for language users and language learners is to find and categorize objects in the world. Within the visual-spatial world, they must be able to (1) segregate a scene into distinct regions associated with distinct objects, (2) cognitively "bind" together the features associated with each distinct object, and (3) assign these cognitive objects to the categories represented by the different nouns of the language. In Section 4.2.3 we discuss a mechanism which satisfies these basic constraints.

Linguistics also tells us that all languages have ways of explicitly encoding relations, so people must be able to find relations in the world and categorize them appropriately. Even if we assume that each scene contains only one salient relation, they must have the ability to (1) segregate a scene into distinct objects and bind their features together ((1) and (2) above), (2) cognitively bind together the relational

[1] Not all nouns refer to physical objects or substances, of course, but all of the early nouns learned by children do.

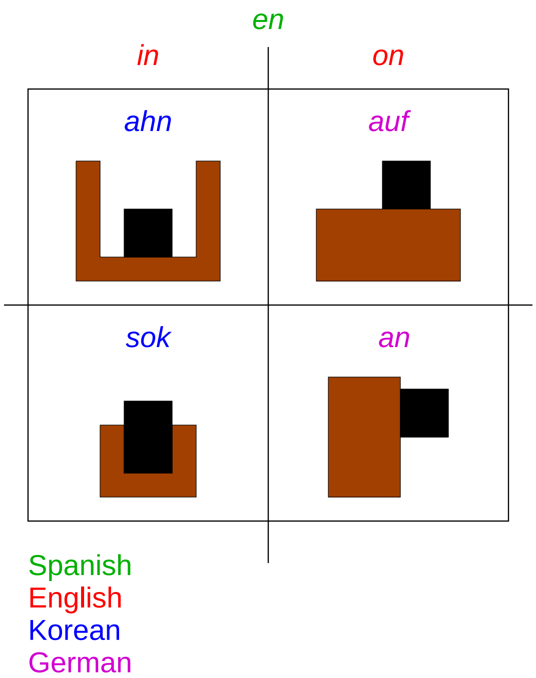

Figure 5: **Spatial Relations Across Languages**. Different languages divide up the space of CONTACT/SUPPORT relations in different ways.

features associated with a given candidate relation, (3) assign trajector and landmark status to the related objects, and (4) assign the cognitive representation of relations to the categories represented by the different relation terms of the language. We believe that these requirements point to an explicit way of representing relational information. Furthermore, since languages differ considerably in the sort of breakdown they make within the space of possible relations, the human capacity to represent and learn spatial relations must be a flexible one. Rather than a set of pre-existing relational categories, what is called for is a set of relational building blocks from which the relational categories of different languages can be assembled. In Section 4.2.4 we describe a representational scheme of this type.

3.1.2 The Acquisition of Linguistic Spatial Relations

As we have just seen, languages look very different from one another with respect to space. We would like to know whether the differences really matter for the acquisition process. Views such as the symbolic position of Pinker and Jackendoff would predict little effect: since all children start out with the same universal spatial categories, they should go through roughly the same stages in acquiring spatial language. In particular where a language does not encode one of the universal categories in a direct way, we would expect over-generalization errors in which a particular form is applied to the universal category. Work by Bowerman and colleagues (Choi and Bowerman, 1992) on the acquisition of English and other languages has shown that this is not the case. Korean children, for example, use no global semantic categories of CONTAINMENT and SURFACE CONTACT/SUPPORT, categories which are not expressed in Korean in a transparent way. Instead they learn the Korean distinction between TIGHT and LOOSE FIT early on. The data seem to support the view that the particular semantics of the lexicon of the target language has a significant effect on the way the language is learned.

We think this view is correct. Thus any model of the acquisition of spatial language must account not only for (1) the developmental path babies follow in learning spatial terms, but also for (2) the interaction between the particular lexicon of the language being learned and the way it is learned. This is an argument for a model of the type shown in Figure 3, one in which linguistic meaning and concepts are not clearly distinguished.

3.2 Non-Language

Spatial cognition in the child emerges from the convergence of perceptual and motoric experience, and a complete characterization of its development certainly requires attention to vision, touch, proprioception, locomotion, and object manipulation, as well as spatial language. We have chosen to start, however, with what is the most important source of perceptual information about space, at least for the seeing infant: vision.

3.2.1 Vision and Imagery

Obviously an overview of the human vision system is beyond the scope of this report. We will only be concerned here with what is most directly relevant to spatial relations. We follow closely Kosslyn's (1994) model of high-level vision and imagery because (1) it takes into account the whole range of subsystems that are involved in high-level vision and (2) it is meant to account for "cognitive graphics" as well as vision; that is, it also runs in the concepts-to-vision direction. The terms we use below are Kosslyn's. The components we are concerned with are illustrated in Figure 6.

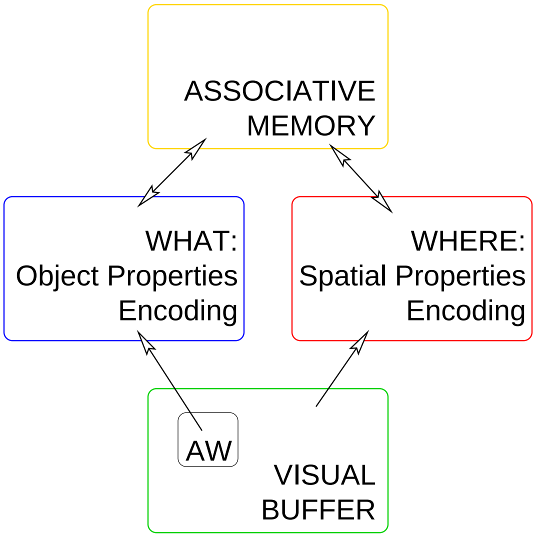

Figure 6: **Basic Components of the Vision System** (based on Kosslyn, 1994). Separate modules are responsible for what is seen in the current Attention Window and where objects are in the current Visual Buffer. Vision makes contact with the rest of cognition in the Associative Memory.

It is generally agreed that the vision system divides into a subsystem responsible for **What** is in an observed scene and a subsystem responsible for **Where** the objects in the scene are. Among the tasks of the What system is the categorization of objects in the scene, a process which permits the assignment of noun labels to the objects. Similarly, the Where system categorizes relations between objects in the scene, a process which permits the assignment of relation terms to the relations in the scene. Thus both subsystems are crucial to the task we are interested in.

For our purposes, visual processing begins in a Visual Buffer (VB), a series of feature-specific maps which have already benefited from edge detection and region filling. The VB's task, among others, is to segregate the scene into regions associated with different objects.

The VB is scanned by an Attention Window (AW), which permits the system to focus on a single object at a time. The AW provides the interface between the VB and the What system, which extracts features spanning more than a "pixel" and ultimately categorizes the contents of the AW. The system operates not only in a bottom-up direction, however; there are top-down influences both on object categorization and on the placement of the AW.

On the Where side, output from the VB is assigned a 3D coordinate system, viewer-centered or object-centered (or both), and the location, size, and orientation of each object in the scene are extracted. Later the Where system is responsible for classifying the relations between objects in the scene. Relations are of two types, categorical relations (such as CONTAINMENT) and continuous relations (such as X-CENTIMETERS-FROM). As on the What side, categorization depends on top-down influences as well as on the strictly visual bottom-up ones.

At its "top," the visual system makes contact with non-visual cognition in an Associative Memory (AM). Both the What and Where systems play a role in the AM and are in turn under its influence when there are top-down effects on object or relation categorization and when the system runs in the imagery direction. It is in this AM that vision and language come together.

Kosslyn (1994) has also amassed considerable evidence that mental images share many of the properties of actual percepts. For example, scanning a mental image takes time proportional to the distance between imaged objects. This evidence suggests that visual mental imagery and visual perception share mechanisms, that imagery amounts in a sense to running the vision system in reverse.

For our purposes, then, two points are important:

1. The vision system has separate modules for handling objects (What) and for handling relations (Where), and these modules have access to different kinds of information in the raw input.
2. Imagery makes use of the same basic mechanisms as vision.

3.2.2 Pre-Linguistic Development of Spatial Relations

Average children spend about a year in the world before producing their first word and it will take them around six months more to start learning words at a fast pace. Babies do not spend this time in idle contemplation of the world; they spend it learning about how their own body works, and more relevant to our current argument, about how the world works. Very young infants display a knowledge of what happens and what doesn't happen in the world; they can predict the consequences of actions and be surprised when their expectations are violated.

For example, babies know about how objects behave in support or occlusion events. Babies as young as 4.5 months realize that objects which are not supported will fall (Needham and Baillargeon, 1993) and 8.5-month-old infants can judge whether an object is being sufficiently supported and be surprised when an insufficiently supported object fails to fall (Baillargeon and Hanko-Summers, 1990). Young infants also know about the impenetrability of objects and the parts of objects that should still be visible given the shape of the occluder (Baillargeon, 1992). Other results show that this knowledge develops. For example, very young babies are not surprised by seemingly unsupported stable objects if there is an occlusion event between the habituation and the test (Spelke et al., 1992). Also, very young babies are not surprised when unsupported objects fail to fall if there is no motion involved in the event (Spelke and Kyeong, 1992).

Babies also have the more abstract notion of "objectness". Young infants are able to use some of the cues adults use to segregate objects, such as relative motion and textural cues (Needham and Baillargeon, 1997). This knowledge also develops: motion comes first, then textural information, and finally gestalt properties (Spelke et al., 1993). Four-month-old infants expect objects to retain characteristics such as size and trajectory even though they are not visible (Baillargeon, 1991).

Infants also seem to be able to categorize spatial relations. Three- to 4-month-old babies categorize LEFT-RIGHT (Behl-Chadha and Eimas, 1995) and ON-UNDER (Quinn, 1994) relations, generalizing over the orientation, the size, and the absolute location of the objects involved in the relation. Seven-month-old infants are able to generalize over different *kinds* of objects involved in the relation. Again, this knowledge develops; younger infants are not able to abstract over different objects involved in the relation, while older infants can (Quinn et al., 1996).

In sum, before they have learned any words, children seem to be forming categories that are useful in representing what the world is like. We think that the learning that occurs during this period is important in setting the basis on which linguistic concepts will be formed; hence our model has a pre-linguistic learning period.

3.3 Interactions Between Language and Non-Linguistic Perception

It seems clear that language and perception have something to do with each other. At the most superficial level we know we can describe what we perceive using words and can also "imagine" what is described to us in words. However, we believe language and perception are deeply interrelated in ways that go beyond these obvious connections and that these inter-relationships should be taken seriously by any model of language acquisition.

Consider first the influence of non-linguistic perception on linguistic behavior and language acquisition. Obviously what is perceived influences the choice of words used to describe it, but our perceptual experience could also directly influence our acquisition of language. There are at least two possibilities: (1) The way in which the world is construed on particular occasions may have an impact on how language is learned. (2) Specific perceptual mechanisms or categories may be prerequisites for the acquisition of specific words or structures.

The first sort of relationship can be shown in an experimental setting by looking at how people generalize nonsense words to novel situations. For example, when shown a block on a box while being told "the block is acorp the box" people interpret *acorp* to mean ON. In contrast, when shown a stick on a box, people interpret *acorp* as ACROSS (Landau, 1996). In the first case, the shape of the trajector is ignored; in the second it matters. We suggest that something like this goes on throughout the acquisition of language. When a child hears a word, the world is generalized according to whatever is perceived in that moment, and contrast and perceptual saliency affect the way in which the situation is construed.

But if cognitive linguists such as Langacker (1987) are right, the influence of perception on the acquisition of language goes beyond this to the second sort of influence. For cognitive linguists, grammar is a mapping between form and function, and they argue that the functional pole of grammatical patterns is concerned with non-linguistic psychological processes such as visual scanning, figure-ground segregation, and imagery as well as with psychological dimensions such as color and depth. For example, the grammatical category TRAJECTOR is defined with respect to the figure in an observed, recalled, or imagined scene. The upshot of this position is that the learner of a language needs access to a relatively direct path from perceptual mechanisms to language learning mechanisms so that such relationships can be acquired, and grammatical structures and words can only be learned once the requisite psychological processes and categories are in place. In other words, since language is, in a very direct way, about perception, language acquisition relies on the perceptual capacities of the learner. Some evidence for this relationship comes from research showing that there is a correlation between conceptual development and linguistic development in semantic domains such as space and time (Weist et al., 1997). While correlations do not establish a causal relationship in one direction or the other, the most plausible explanation for these results seems to be one in which language acquisition presupposes conceptual categories, the sorts of categories that arise out of perceptual learning.

Now consider the influence of language on non-linguistic cognition. There are three possible ways in which such effects could occur. (1) The wording of a particular utterance could influence the way in which the state or event that is referred to is conceptualized or remembered. (2) The language being learned could become associated with the contexts in which it occurs, an effect we could look for in bilinguals. (3) The regularity implicit in the grammar or lexicon of a particular language could favor certain patterns of thought on the part of the speakers of the language.

When we hear a description, we form images in our minds. As noted in Section 3.2.1 above, these images resemble visual percepts and seem to make use of the visual system itself. Different linguistic descriptions of the same scene may evoke quite different images: the noun phrase *half watermelon* is more likely than the noun *watermelon* to cause subjects to include *seeds* in a list of features (Wu, 1995). The way a scene is described may also alter our memory of it. For example, people who see a green car and then have it described as "blue" are more likely to recognize a more bluish car as the one they saw before than people who didn't hear it labeled (Loftus and Palmer, 1974), and people who are asked to label non-prototypical color chips perform worse on a recognition task than people who did not label them during study (Schooler and Engstler-Schooler, 1990).

A deeper relationship between language and thought emerges when we examine the influence of the specific language on how the world is perceived. One possibility is that a language may tie its speakers to particular contexts. Experiments with bilinguals have shown that their two languages evoke different contexts. Chinese-English bilinguals were presented with descriptions of individuals and then asked whether the individuals described were likely to have certain behaviors. Subjects addressed in Chinese extrapolated using Chinese stereotypes and subjects addressed in English used English stereotypes (Hoffman et al., 1986).

A more controversial possible relation between language and non-linguistic cognition concerns the effects of the regularities inherent in particular languages, what is usually known as **linguistic relativism**. In its strongest form, this position, associated most strongly with the ideas of Benjamin Lee Whorf (1956), holds that categories in the grammars and lexicons of particular languages have a direct impact on the thought patterns of speakers. As we have seen in Section 3.1.1, different languages "slice up" the world in different ways. If a language explicitly codes for a certain distinction, making such a distinction might become relatively easy for speakers of that language. There has been relatively little systematic investigation of relativism (Lucy, 1996), so, despite some intriguing evidence in favor of an influence of language on perception and thought (Lucy, 1992), it is still premature to assume that such an influence is pervasive. Our position is that computational modeling may shed light on the possibility of the language-to-perception/thought relationship in a way that has not been possible before. Given the evidence, we believe that models must remain open to the possibility of such a relationship by maintaining the language-to-concepts-to-vision path in the architecture. Excluding this path precludes any sort of relativism.

Thus we find at least some evidence for all of the following sorts of influences:

1. perception → attention → language acquisition
2. perception → construal → language acquisition
3. wording → attention/memory
4. choice of language → attention/memory
5. linguistic regularity → concepts → perception

To summarize, there seem to be a number of ways in which linguistic and non-linguistic perception interact. Memory, attention, and categorization are influenced by both linguistic and non-linguistic input, and memory, attention, and categorization, in turn, influence both language and perception. These interactions have important effects on the acquisition of language and belong in a model of acquisition. Models of the type shown in Figure 3 incorporate these interactions. Playpen, the model we describe in the next section, is such a model.

4 Playpen

Playpen is an evolving model of the development of spatial cognition and spatial language. We call it "Playpen" because it is meant to deal only with a simple world of blocks, sticks, and containers and because the regularities in the physical external playpen that the model lives in should be incorporated into the network itself as it learns: in a sense the network will "have" whatever it has discovered about the playpen in its connection weights.

4.1 Playpen's Task

Eventually we would like the network to be the brains of a robot which exists in a playpen-like world of blocks and other simple objects. In the short run, we must simulate the world. This requires an account of what regularities there are in the world, including linguistic regularities. Even when Playpen is embedded in a robot, a theory of the development of spatial cognition must include an account of what is out there to be learned. For the purposes of this paper, we will remain vague in this regard; details will come out when we consider particular linguistic relations, as we will do in the next report on Playpen. We do assume, however, that the linguistic input to the child makes a fundamental distinction between expressions for things and expressions for relations. This fact imposes constraints on the network: it must have the capacity to represent objects as collections of features and relations as pairs of objects (which are collections of features).

The world that infants are exposed to includes language, as well as other sorts of stimuli, from the start. But because infants have not yet figured out the phonology of the target language, linguistic input is probably completely irrelevant for the learning of spatial concepts. Once children are capable of recognizing particular words, normally early in the second year, they are in a position to learn the spatial concepts associated with words. Thus it is convenient to divide children's task into two phases, a **pre-linguistic** phase in which they learn about space as they observe and manipulate the things around them, and a **linguistic** phase in which they also receive linguistic input, input within which they are capable of distinguishing particular words. In this latter phase, non-linguistic learning about space continues, but it is supplemented by the co-occurrence of words and phrases with particular scenes.

4.2 Playpen's Architecture and Behavior

Playpen's architecture and behavior are constrained both by the nature of its task and what is known about how language and vision operate.

1. As discussed in Sections 3.2.1 and 3.3 above, processing is interactive and bi-directional. Visual input can yield linguistic output, and linguistic input, along with some visual context, can yield visual output in the form of expectations or imagery.

2. In the pre-linguistic phase, the system receives visual inputs with no explicit labels, and learning consists of somehow extracting regularity from these inputs. In the linguistic phase, some visual inputs are accompanied by linguistic inputs as well. In one sense this just means that the input patterns are more complex; in another it means that there are now new independent grounds for dividing space up into particular categories.

3. Language is explicit about certain categories, in particular objects and relations. Learning language requires that the network have the capacity to represent these categories explicitly.

4. A good deal is known about the human vision system. While we believe there are large gaps that need to be filled in, the model should be in general agreement with the vision facts.

5. A good deal is known about language development and concept development. Playpen is meant to model the behavior of babies, and the most important constraints are those concerned with development: What precedes what? What is hard or easy?

We have paid most attention to the first four constraints. The fifth is likely to play a greater role as we model the development of particular relations.

4.2.1 The Overall Architecture

While we are currently focusing on vision and language and their interaction, the model should eventually bring together input from other perceptual domains — proprioception and touch — and from movement — locomotion, the movement of a limb, and the manipulation of objects with a hand — as well. The overall architecture we envision is shown in Figure 7.

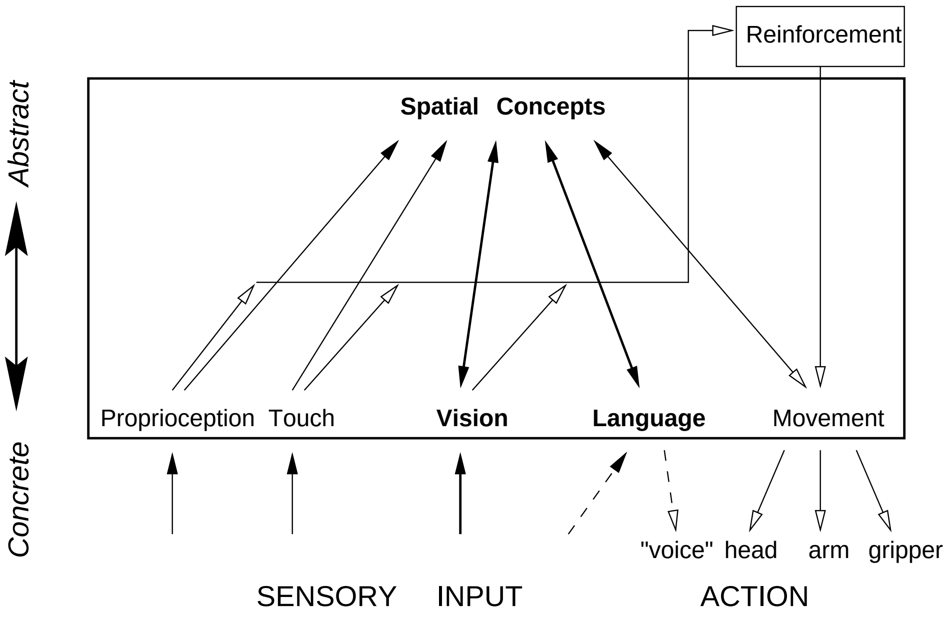

Figure 7: **Playpen Architecture.** Spatial concepts emerge out of the interplay of non-linguistic perceptual input, linguistic input, and action. Portions we focus on are highlighted in boldface or with thick lines.

The visual component of Playpen is based loosely on Kosslyn's model of vision and imagery (Kosslyn, 1994). Figure 8 shows how vision and language interact in the model.

The overall organization of the network is such that higher layers roughly preserve the spatial relations within lower layers; higher representations are more abstract than lower ones. The input visual layers, the Visual Buffer, are topological maps. The Visual Buffer performs bottom-up object segregation. The Attention Window is a mainly stimulus-driven mechanism which zooms in on a part of scene in the Visual Buffer corresponding roughly to a putative object. The Attention Window passes on a region in the Visual Buffer to the What system, which categorizes the object which is in the Attention Window, adding a representation of the object to an **Object Short-Term Memory**, a component not found in Kosslyn's model. The Where system receives the entire scene from the Visual Buffer. The segregation of the scene into regions associated with the different objects is preserved, but lower-level layers in this system are responsible for assigning perspective to the scene, both object-centered and viewer-centered, and higher-level layers extract salient dimensions such as position along the vertical dimension and object size. The representations provide input to the **Spatial Relation Concepts** layer, where the system categorizes spatial relations pre-linguistically. The Object Short-Term Memory

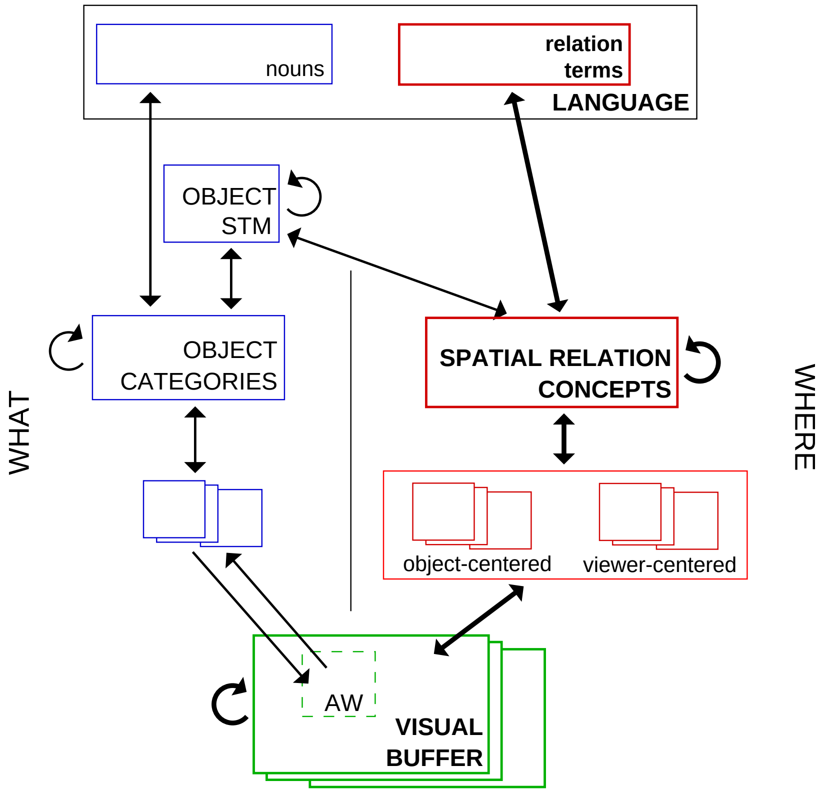

Figure 8: **Playpen Architecture: Vision and Language**. Visual objects and nouns are tied together on the What side of the system. Visual and linguistic relations are tied together on the Where side. Most of what is shown has not been implemented; the regions we have focused on are highlighted in boldface or with thick lines.

and the Spatial Relation Concept layer interact so that it is possible for a relation together with its arguments to be represented, as is required, for example, for the meaning of a spatial relation expression.

The Language layer has units for two types of words, nouns and relation terms (prepositions in English). The What side of the visual system connects to the nouns, permitting labeling of objects and, in the comprehension direction, the understanding of nouns as visual patterns within the What system and the Visual Buffer. The Where side of the visual system connects to the relation terms, permitting labeling of relations and, in the comprehension direction, together with the What side, the understanding of relation expressions as visual patterns in the Where system and the Visual Buffer.

4.2.2 Processing Units

The network is of the generalized Hopfield type: connections between units are symmetric, and units repeatedly update until the network settles. Each unit has an associated activation function; for most units this is the familiar interactive activation rule (McClelland and Rumelhart, 1981):

If $h_i^t > 0$,

$$\Delta a_i^t = h_i^t(a_i^{max} - (a_i^{t-1} - D_i a_i^{t-1})) \quad (1)$$

Else,

$$\Delta a_i^t = h_i^t((a_i^{t-1} - D_i a_i^{t-1}) - a_i^{min})$$

where a_i^t is the activation of unit i at time t; h_i^t is the input to i at time t; and a_i^{max}, a_i^{min}, and D_i

are the maximum activation, minimum activation, and decay rate associated with i. All units in the network currently have maximum activations of 1 and minimum activations of 0.

As we noted in Section 3.1.1 above, language makes explicit reference to objects and relations, and a model of the acquisition and processing of language requires a means of representing conceptual objects and relations, as well as the associations between lexical/grammatical patterns and conceptual objects and relations. The network is made up of two kinds of units: **Object Units** and **Relation Units**, which serve these two purposes.

4.2.3 Object Units

Object units (OUs) are units with a **relative phase angle** in addition to an activation. As in a number of other recent models (Hummel and Biederman, 1992; Shastri and Ajjanagadde, 1993; Sporns et al., 1989), synchronization functions to bind together the features of distinct objects. Units with the same phase angle are part of the same object, and units with different phase angles belong to different objects.

The connection between each pair of OUs has not only a weight but also an associated **coupling function**, a function of the difference in phase angles of the two units. The coupling function must be symmetric about 0, and its derivative must be anti-symmetric about 0; see the AppendixA for why these constraints must hold. Both the activation and the phase angle of an OU are potentially modified each time a unit is updated, and both depend on the coupling function on the weights into the unit. The input h and change in phase angle $\Delta\varphi$ to an OU i are given by

$$h_i = \sum_{j=1}^{n} a_j \cdot w_{ij} \cdot \Phi_{ij}(\varphi_i - \varphi_j) \tag{2}$$

$$\Delta\varphi_i = \frac{\pi}{\sum_{j=1}^{n}} \sum_{j=1}^{n} a_j \cdot w_{ij} \cdot \Phi'_{ij}(\varphi_i - \varphi_j), \tag{3}$$

where n is number of units in the network, a_j is the activation of unit j, w_{ij} is the weight connecting units i and j, and Φ_{ij} is the coupling function associated with units i and j. A stable state of the network is, then, a state in which neither activations nor phase angles are changing.

The most common coupling function used in the network is $.5 + .5\cos x$. For positive weights, the system consisting of the two units with this coupling function has an attractor at the state where the units are in phase and a repeller at the state where they are out of phase. The two units excite each other at all phase angle differences except π. For negative weights, there is an attractor at the out-of-phase state and a repeller at the in-phase state, and the units inhibit each other except when they are out of phase.

Demo 1 illustrates the behavior of a small network of OUs.

DEMO 1: **Object Units** (only accessible in the WWW version of the report)

The units in Playpen's lexicon representing nouns are also OUs. Learning the meaning of a noun would involve creating positive connections between the noun unit and the associated non-linguistic visual/spatial feature units. These connections would tend to cause the connected units to align their phase angles, so that in the comprehension or production of a phrase the word is "bound" to its meaning. Figure 9 illustrates the relationship between noun OUs and the associated non-linguistic "semantic" OUs.

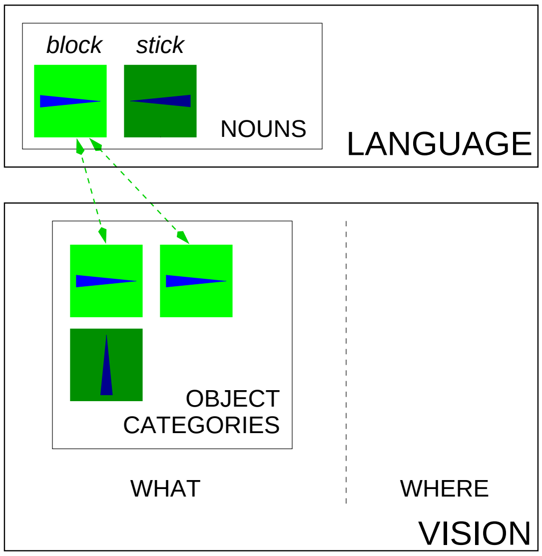

Figure 9: **OUs in the Representation of Noun Meaning**. Nouns and visual objects are both represented by OUs; during production or comprehension synchronized phase angles bind together nouns and their internal "referents". Unit activation is represented by brightness, and phase angles are indicated by the blue triangles. The dashed green arrows represent trainable connections which have developed positive weights. Only a minimum number of units is shown.

4.2.4 Relation Units

The Problem We have seen how OUs permit the network to distinguish clusters of features from one another and how positive connections between noun OUs and non-linguistic "semantic" OUs can implement the binding that is part of understanding and producing nouns.

Now consider what would be required for relations, both relation terms and non-linguistic relations. A binary, non-reflexive relation relates two distinct objects, for example, a stick and a block in SUPPORT(BLOCK, STICK). To implement a relation in a distributed connectionist network, we first need to make two modifications to the standard view from predicate logic. First, rather than treat an expression like this as a predicate with a truth value, we will treat it as a set of correlations or inferences. In this sense, SUPPORT in the example means loosely that stick features and block features in particular relative positions correlate with one another. Viewed in terms of inference, the relation takes the form of a process of pattern completion, for example,

1. Given a supported stick, infer that the supporter is a block.
2. Given a block and stick in a support relationship, infer that the block is the supporter.

Note that the inferences are not absolute; whether they actually hold depends on a number of other factors, for example, whether the stick in question has a flat surface. In the network implementation, whether a particular relation holds would depend on the combined input coming into the relevant units and therefore on the entire network of other relational inferences in the system. Note also that from this connectionist perspective, there is no distinction between a relation between individual objects and a generalized relation over types of objects.

A second modification to the standard view of relations is to treat them as non-atomic. A distributed implementation of a relation involves multiple **micro-relations**, each relating a pair of features, one belonging to each of the two objects. A micro-relation represents a highly specific micro-inference. Thus knowledge about SUPPORT is made up of micro-relations specific to particular relative locations in some abstract representational space.

Within the Playpen framework described so far, we could represent a micro-relation as a negative connection between two OUs. If both units are activated, this connection causes them to repel each other. With different phase angles, they would then correspond to parts of different objects. Thus the micro-inference represented by such a connection would be: *given whatever features are associated with unit A and whatever features are associated with unit B, the A features and the B features belong to two different objects.* Recall, however, that a negative connection represents inhibition as well as repulsion. Thus the desired inference only holds to the extent that the two units are both activated, for example, when both have their activations clamped. When only the phase angles can change, the two units will tend to end up out-of-phase, indicating distinct objects. However, when both activation and phase angle are permitted to change, the negative connection can also result in one unit's inhibiting the other, the precise outcome depending on the initial state of the two units, the coupling function associated with the connection, and of course other weights and inputs into the units.

However, this way of representing relations gets us nowhere when it comes to correlations between relations. For example, we might want to represent the following inference: *if A supports B, then A is probably larger than B.* This involves an association between a SUPPORT relation and a BIGGER-THAN relation, thus in the network minimally four OUs, two for each relation. A further example is the association required to specify the meaning of a relation term: *if "A is on B", then some object A' is on another object B'.* This involves an association between a linguistic *on* relation and a non-linguistic ON relation. Any set of connections we set up among the four units necessary to two relations fails to capture the relationship between the relations that we want. Consider the connections shown in Figure 10. While the individual connections do seem to represent the phase relationships we want for this example, if we consider the connections separately, we see that they miss crucial conditions. Thus the positive connection between the HIGH unit in the LOCATION pair and the SMALL unit in the RELATIVE SIZE pair indicates that objects that are high tend to be small. But what we would like to convey is the more complex fact that objects that are higher than other objects tend to be smaller than the other objects. The relationship we want to represent requires that the micro-relations be treated as units, and the simple network shown in Figure 10 does not permit this.

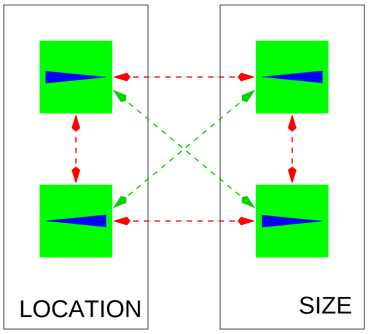

Figure 10: **Attempt at Representing Relation-Relation Correlations.** Two units in each of two different layers on the Where side of the network are shown. Green connections have positive weights, red connections negative weights. This network offers no way to relate the location and size relations to one another directly.

Note that a conjunctive unit connected to both of the OUs in a particular micro-relation does not solve the problem either. While we could arrange for such a unit to turn on to the extent that the two input units are on (using positive connections which affect activation but not phase angle), the unit would fail to represent the relation because it has no way of holding onto the two phase angles of the inputs: the identity of the two objects in the relation is lost.

Apparently relations between relations require explicit units representing micro-relations and taking OUs and other relation units as inputs to their two role "arms." The relation units (RUs) should

1. Become activated to the extent that both input (groups of) units are activated and the two (groups of) units are out-of-phase with each other.

2. Preserve the phase angles of the inputs, passing them on to other RUs.
3. Associate with other RUs via a mapping of the arms which preserves the appropriate phase angle relationships and also reflects the strength of the correlation between the two micro-relations.

The Solution RUs, which are the major innovation of the Playpen architecture, are used to represent relational information; they are "about" two different objects. Each RU is made up of a cluster of five units hard-wired in such a way that the unit as a whole is activated to the extent that it is receiving input from two distinct objects. Figure 11 shows a pair of RUs along with four OUs illustrating the connectivity within the RUs and between them and other units of both types. An RU has two **interfaces**, each consisting of a pair of OUs: one to handle interaction with other RUs, the other to handle interaction with OUs. The two interfaces are connected in such a way that the corresponding arms tend to be in-phase and the opposing arms tend to be out-of-phase. The RU is considered activated when all four of its interface units are activated. Each RU also has a simple unit with no phase angle and negative connections to all four of the OUs. This **bias unit** has a resting activation of 1.0 and turns off only when its input falls below a threshold. The bias unit functions to prevent the OUs on one interface from turning on those on the other interface unless they are both sufficiently activated. Without the bias unit, "one-armed" relations, those in which one arm only is activated on each interface, would be possible.

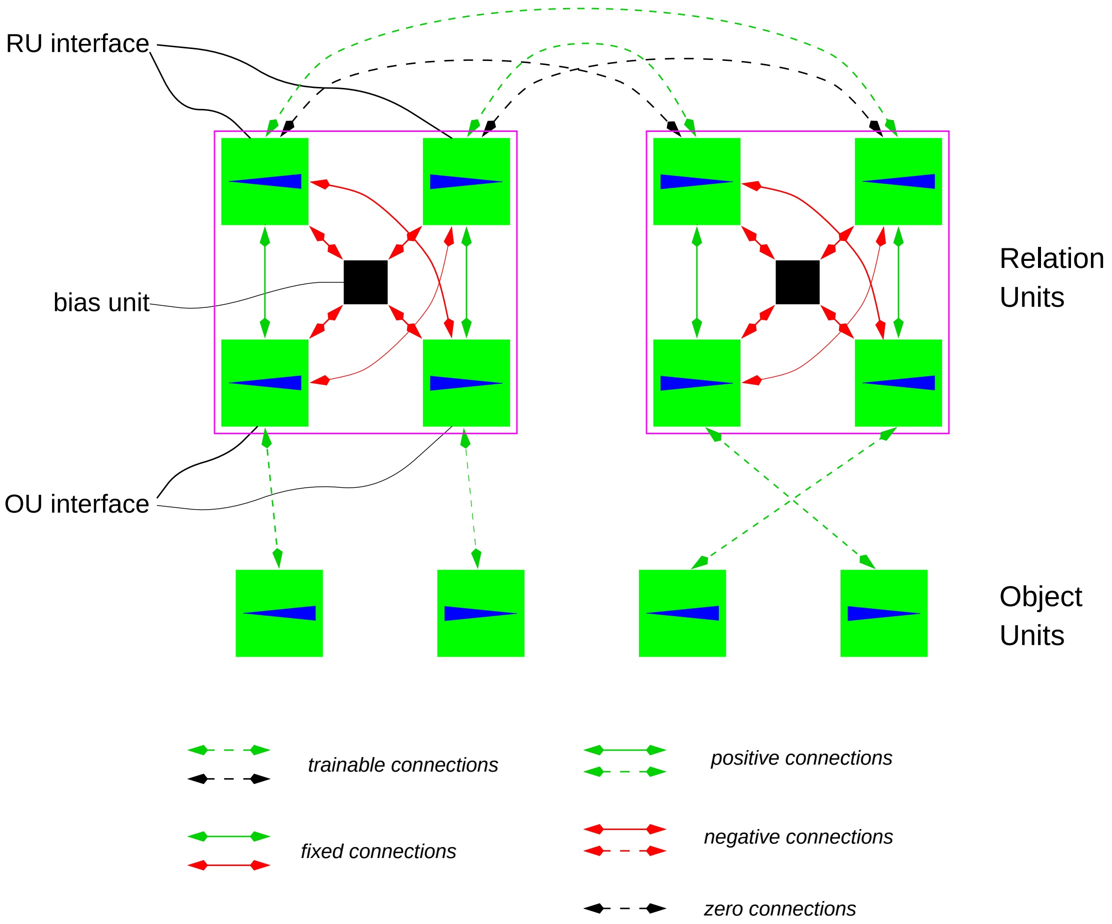

Figure 11: **Relation Units.** Two RUs are shown, each surrounded by a magenta border. Both are in their active state: all four arms are maximally activated and the bias units are inhibited. The two RUs are connected to each other on their RU interfaces, and each RU is also connected to two OUs on its OU interface.

Connectivity between RUs is also constrained, though the weights themselves are trainable. There are four connections joining each pair of connected RUs on their RU interfaces, but only two distinct weights, one for the connections joining corresponding arms, the other for the connections joining opposing arms. The coupling function on these connections is $.5 + .5 \cos x$. These connections implement the four possible relationships that can exist between micro-relations:

1. The two micro-relations have no effect on each other. In this case both weights are 0.
2. The two micro-relations correlate negatively with one another; that is, the presence of one leads one to expect the absence of the other. In this case both weights are negative.
3. The two micro-relations correlate positively with one another, and corresponding arms are bound to the same object. In this case the weight for the corresponding arms is positive and that for the opposing arms 0.
4. The two micro-relations correlate positively with one another, and opposing arms are bound to the same object. In this case the weight for the corresponding arms is 0 and that for the opposing arms positive.

In the resting state of an RU, the four OUs have activations of 0 and the bias unit an activation of 1. When an RU is activated on one or the other or both of its interfaces in such a way that the two OUs on an interface are highly activated and out-of-phase, the bias unit is inhibited, and the OUs on the interface (if not already activated) become active and take on the phase angles of the corresponding OUs in the other interface. When the RU is completely activated, the four OUs have activations of 1 and the bias unit an activation of 0. Demo 2 illustrates a simple network of two RUs, one of which is activated by input from a single OU.

DEMO 2: **Relation Units** (only accessible in the WWW version of the report)

Now consider how RUs would permit us to represent the meaning of a spatial relation term. The word itself has an associated RU with a trajector and a landmark arm on both interfaces. The RU interface of this RU connects to appropriate RUs in the Spatial Relation Concepts layer, which in turn connect to OUs specifying the locations associated with features of the two related objects. The nouns representing the trajector and landmark of the relation must be in phase with the corresponding arms of the relation term RU. We are far from a complete account of how this takes place in either production or comprehension. For now, we simply assume that a trajector OU implements this binding process. Figure 12 illustrates these relationships.

4.2.5 Sequential Time

Though our concern is with *static* relations, we cannot ignore movement. In order to categorize relations, children must be able to segregate the scene into distinct objects. However, before seven months, they are generally unable to segregate static configurations of objects (Spelke, 1990), and Thelen and Smith (1994) have proposed that they may learn to use static properties to perform object segregation by observing objects move and then come to rest in space.

If we are to deal with movement, we must first have a means of dealing with time. In a generalized Hopfield network, we face the problem that settling itself requires time. The response of the network to temporal inputs or the network's generation of temporal output must take place at a time scale beyond that necessary for settling. Most approaches to time in neural networks incorporate some sort of short-term memory, a means by which units respond not only to the current state of the network but also to a limited record of its previous states. Within the generalized Hopfield framework, this can be accomplished through the inclusion of **delays** on connections, an idea due originally to Kleinfeld (1986). Connecting any two units in the network there may be any number of connections, each with its own delay. Input to a unit along a connection is a function of the activation and phase angle of the destination unit at the time before the delay. Equation 4 shows this relationship.

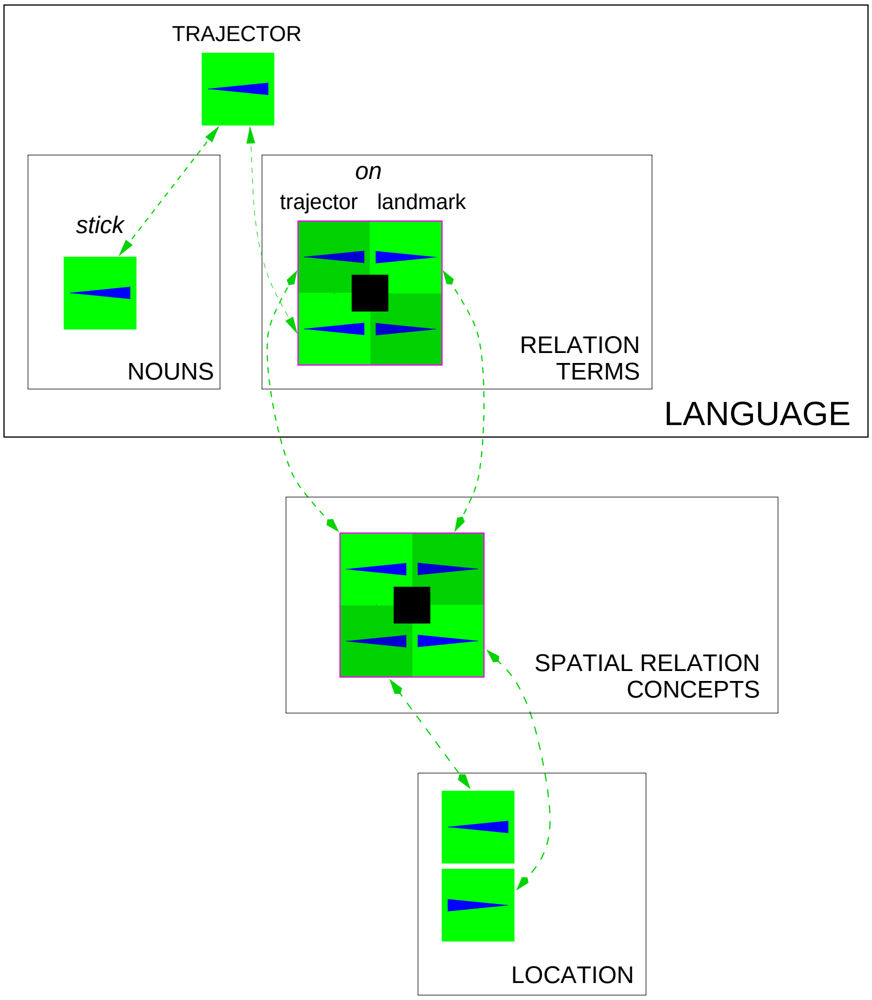

Figure 12: **Relation Units in the Representation of Word Meaning.** Only one of many possibly relevant units is shown in the Spatial Relation Concepts and Location layers. The RU interfaces of the RUs appear on top, the OU interfaces on the bottom. Only the positive connections between the two RUs are shown; there would be also be connections with weight 0 joining the opposing arms of the two RU interfaces. The trajector OU implements the phase relationship between the trajector arm of the word RU and the corresponding noun OU.

$$h_i^t = \sum_{j=1}^{n} \sum_{k=0}^{m} a_j^{t-k} \cdot w_{ij}^k \cdot \Phi_{ij}^k(\varphi_i^t - \varphi_j^{t-k}) \tag{4}$$

where input, activation, and phase angle are a function of time (t) and m is the maximum delay.

Demo 3 shows the behavior a simple network with two hard-wired connections between each pair of units, one with no delay and one with a delay of one "primitive time step."

DEMO 3: **Delay Connections** (only accessible in the WWW version of the report)

4.2.6 Learning

The network is trained using a variant of Contrastive Hebbian Learning (CHL) (Movellan, 1990), modified to accommodate unsupervised learning (auto-association) and phase angles. In Contrastive Hebbian Learning, weight updates take place in positive (Hebbian) and negative (anti-Hebbian) phases. During the positive phase, the input and output units are clamped to training patterns, and the network is then allowed to settle. Weight updates during this phase are proportional to the product of the activations of the units on either end of the connection. During the negative phase, only the input units are clamped, and the network is then allowed to settle. Weight updates during this phase are proportional to the negative of the product of the activations on either end of the connection. For OUs, which have phase angles, the coupling function applied to the phase angle difference also enters in. The net weight update on the connection joining units i and j following the presentation of a training pattern is

$$\Delta w_{ij} \propto \breve{a}_i^+ \cdot \breve{a}_j^+ \cdot \Phi_{ij}(\breve{\varphi}_i^+ - \breve{\varphi}_j^+) - \breve{a}_i^- \cdot \breve{a}_j^- \cdot \Phi_{ij}(\breve{\varphi}_i^- - \breve{\varphi}_j^-), \tag{5}$$

where ($\breve{}$) over a symbol refers to that quantity when the network has stabilized and the $+$ and $-$ superscripts refer to the positive and negative phases of learning. When the network generates the appropriate output for each training input, the changes from the positive and negative phases cancel each other out, and the weights no longer change. See the Appendix for the derivation of the the CHL rule for OUs.

CHL was originally described for hetero-associative learning. For auto-associative learning, there is no distinction between input and output units, so we must decide which units are clamped and which not clamped during the negative phase. One way to proceed is to randomly select input/output units to clamp during the negative phase, and it is this approach that we have followed with Playpen.

Demo 4 illustrates the two phases of learning in a simple network of OUs.

DEMO 4: **Learning** (only accessible in the WWW version of the report)

5 Conclusions and Future Work

In the first part of this report we tried to convince you that in order to understand where the meanings of words come from, it makes sense to look outside as well as within language, especially at human vision and at pre-linguistic conceptual development. But how is one to take on the complexity of these diverse domains, each of which occupies an entire research community? We believe that these research communities have given us much to go on, that we can take enough off the shelf to get us started, and that by slicing our domain of interest very thin, specifically by confining ourselves to simple static spatial relations, it is possible to build an insightful model of the development of word meaning.

In the second part of the report we outlined the beginning of the Playpen project, whose goal is such a model. We described a simple neural network architecture based on features of the human vision system which allows for the emergence of spatial concepts from the interaction of vision and language. We emphasized three basic building blocks which we believe are required: object units, relation units, and delay connections.

In a subsequent technical report (available in September 1997), we describe how Playpen models the relative difficulty of the acquisition of the words *on*, *under*, *left*, and *right*.

Future work includes

1. implementation of other components of the architecture, especially those in the What system

2. application of the model to other specific relations, especially IN and BEHIND.
3. taking the visual world seriously, eventually using input from a camera.

A Mathematical Details of the Model

In this section we show how Contrastive Hebbian Learning (CHL) (Movellan, 1990) needs to be modified to accommodate units with relative phase angles. We follow the derivation in Movellan closely.[2]

Movellan defines a continuous Hopfield Energy function

$$F = E + S \tag{6}$$

where E reflects the constraints imposed by the weights in the network and S the tendency to drive the activations to a resting value. For our network S is the same as for a network with no phase angles:

$$S = \sum_{i=1}^{n} \int_{rest_i}^{a_i} f_i^{-1}(a)da \tag{7}$$

where n is the number of units in the network, a_i is the activation of unit i, f_i is the activation function for unit i, and $rest_i = f(0)$.

However, E becomes

$$E = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_i \cdot w_{ij} \cdot a_j \cdot \Phi_{ij}(\varphi_i - \varphi_j) \tag{8}$$

where w_{ij} is the weight connecting units i and j and Φ_{ij} is the coupling function associated with units i and j. In what follows we will abbreviate $\Phi_{ij}(\varphi_i - \varphi_j)$ as $\Phi_{j\to i}$.

The coupling function must be differentiable and satisfy the following:

$$\Phi_{i\to j} = \Phi_{j\to i} \tag{9}$$

$$\Phi'_{i\to j} = -\Phi'_{j\to i} \tag{10}$$

When the network is stable, the inverse of the activation function for each unit is equal to the input into that unit:

$$f_i^{-1}(\breve{a}_i) = \breve{h}_i = \sum_{j=1}^{n} \breve{a}_j w_{ij} \breve{\Phi}_{j\to i} \tag{11}$$

where ($\breve{}$) represents equilibrium and h_i is the input to unit i. Furthermore, when the network is stable, the phase angle of each unit no longer changes:

$$\Delta\breve{\varphi}_i = \frac{\pi}{\sum_{j=1}^{n} w_{ij}} \sum_{j=1}^{n} \breve{a}_j w_{ij} \breve{\Phi}'_{j\to i} = 0 \tag{12}$$

Movellan defines the contrastive function J as

$$J = \breve{F}^{(+)} - \breve{F}^{(-)} \tag{13}$$

[2]We ignore the possibility of delay connections, but they do not affect the derivation of the learning rule.

and shows that the CHL rule minimizes J. We follow his derivation for the case where units have phase angles.

The energy of the network E at equilibrium is

$$\breve{E} = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\breve{a}_i w_{ij}\breve{a}_j\breve{\Phi}_{j\to i} \tag{14}$$

Extracting the terms with a w_{ij} term,

$$\breve{E} = -\frac{1}{2}\left(2\breve{a}_i w_{ij}\breve{a}_j\breve{\Phi}_{j\to i} + \sum_{k=1}^{n}\sum_{\substack{l=1\\ k,l\neq i,j;k,l\neq j,i}}^{n}\breve{a}_k w_{kl}\breve{a}_l\breve{\Phi}_{l\to k}\right) \tag{15}$$

Differentiating with respect to a single weight w_{ij} and considering that w_{ij} is the only weight depending on w_{ij},

$$\begin{aligned}\frac{\partial\breve{E}}{\partial w_{ij}} \quad = \quad & -\frac{1}{2}\Bigg[2\breve{a}_i\breve{a}_j\breve{\Phi}_{j\to i} + 2w_{ij}\breve{a}_i\breve{\Phi}_{j\to i}\frac{\partial\breve{a}_j}{\partial w_{ij}} + 2w_{ij}\breve{a}_j\breve{\Phi}_{j\to i}\frac{\partial\breve{a}_i}{\partial w_{ij}} + \\ & \breve{a}_i w_{ij}\breve{a}_j\left(\breve{\Phi}'_{j\to i}\left(\frac{\partial\breve{\varphi}_i}{\partial w_{ij}} - \frac{\partial\breve{\varphi}_j}{\partial w_{ij}}\right) + \breve{\Phi}'_{i\to j}\left(\frac{\partial\breve{\varphi}_j}{\partial w_{ij}} - \frac{\partial\breve{\varphi}_i}{\partial w_{ij}}\right)\right) + \\ & \sum_{k=1}^{n}\sum_{\substack{l=1\\ k,l\neq i,j;k,l\neq j,i}}^{n} w_{kl}\left(\breve{a}_k\breve{\Phi}_{l\to k}\frac{\partial\breve{a}_l}{\partial w_{ij}} + \breve{a}_l\breve{\Phi}_{l\to k}\frac{\partial\breve{a}_k}{\partial w_{ij}} + \breve{a}_k\breve{a}_l\breve{\Phi}'_{l\to k}\left(\frac{\partial\breve{\varphi}_k}{\partial w_{ij}} - \frac{\partial\breve{\varphi}_l}{\partial w_{ij}}\right)\right)\Bigg]\end{aligned} \tag{16}$$

From Equation 10, we have

$$\begin{aligned}&\breve{a}_i w_{ij}\breve{a}_j\left(\breve{\Phi}'_{j\to i}\left(\tfrac{\partial\breve{\varphi}_i}{\partial w_{ij}} - \tfrac{\partial\breve{\varphi}_j}{\partial w_{ij}}\right) + \breve{\Phi}'_{i\to j}\left(\tfrac{\partial\breve{\varphi}_j}{\partial w_{ij}} - \tfrac{\partial\breve{\varphi}_i}{\partial w_{ij}}\right)\right) \\ &= 2w_{ij}\breve{a}_i\breve{a}_j\breve{\Phi}'_{j\to i}\tfrac{\partial\breve{\varphi}_i}{\partial w_{ij}} + 2w_{ij}\breve{a}_i\breve{a}_j\breve{\Phi}'_{i\to j}\tfrac{\partial\breve{\varphi}_j}{\partial w_{ij}}\end{aligned} \tag{17}$$

and

$$\sum_{k=1}^{n}\sum_{\substack{l=1\\ k,l\neq i,j;k,l\neq j,i}}^{n} w_{kl}\left(\breve{\Phi}'_{l\to k}\left(\frac{\partial\breve{\varphi}_k}{\partial w_{ij}} - \frac{\partial\breve{\varphi}_l}{\partial w_{ij}}\right)\right) = 2\sum_{k=1}^{n}\sum_{\substack{l=1\\ k,l\neq i,j;k,l\neq j,i}}^{n} w_{kl}\breve{a}_k\breve{a}_l\breve{\Phi}'_{l\to k}\frac{\partial\breve{\varphi}_k}{\partial w_{ij}} \tag{18}$$

Substituting these into Equation 16,

$$\frac{\partial\breve{E}}{\partial w_{ij}} = -\frac{1}{2}\left(2\breve{a}_i\breve{a}_j\breve{\Phi}_{j\to i} + 2\sum_{k=1}^{n}\frac{\partial\breve{a}_k}{\partial w_{ij}}\sum_{l=1}^{n} w_{kl}\breve{a}_l\breve{\Phi}_{l\to k} + 2\sum_{k=1}^{n}\breve{a}_k\frac{\partial\breve{\varphi}_k}{\partial w_{ij}}\sum_{l=1}^{n} w_{kl}\breve{a}_l\breve{\Phi}'_{l\to k}\right) \tag{19}$$

From 11 and 12, we have the following for the case where $i \neq j$. Since there are no self-recurrent connections in our network, we need only consider this case.

$$\frac{\partial\breve{E}}{\partial w_{ij}} = -\breve{a}_i\breve{a}_j\breve{\Phi}_{j\to i} - \sum_{k=1}^{n}\breve{h}_k\left(\frac{\partial\breve{a}_k}{\partial w_{ij}}\right) - \sum_{k=1}^{n}\Delta\breve{\varphi}_k\breve{a}_k\frac{\partial\breve{\varphi}_k}{\partial w_{ij}} \tag{20}$$

From 12, the last term is 0, and we have

$$\frac{\partial \breve{E}}{\partial w_{ij}} = -\breve{a}_i \breve{a}_j \breve{\Phi}_{j \to i} - \sum_{k=1}^{n} \breve{h}_k \left(\frac{\partial \breve{a}_k}{\partial w_{ij}} \right) \tag{21}$$

From Equation 7,

$$\frac{\partial \breve{S}}{\partial w_{ij}} = \sum_{k=1}^{n} f_k^{-1}(\breve{a}_k) \frac{\partial \breve{a}}{\partial w_{ij}} \tag{22}$$

and from Equation 11, we have

$$\frac{\partial \breve{F}}{\partial w_{ij}} = -\breve{a}_i \breve{a}_j \breve{\Phi}_{j \to i} \tag{23}$$

making

$$\frac{\partial \breve{J}}{\partial w_{ij}} \propto \breve{a}_i^{(+)} \breve{a}_j^{(+)} \breve{\Phi}_{j \to i}^{(+)} - \breve{a}_i^{(-)} \breve{a}_j^{(-)} \breve{\Phi}_{j \to i}^{(-)} \tag{24}$$

which shows that the modified CHL rule

$$\Delta w_{ij} \propto \breve{a}_i^{(+)} \breve{a}_j^{(+)} \breve{\Phi}_{j \to i}^{(+)} - \breve{a}_i^{(-)} \breve{a}_j^{(-)} \breve{\Phi}_{j \to i}^{(-)} \tag{25}$$

descends in the J function.

References

Baillargeon, R. (1991). Reasoning about the height and location of a hidden object in 4.5- and 6.5-month-old infants. *Cognition*, *38*, 13–42.

Baillargeon, R. (1992). The object concept revisited: New directions in the investigation of infants' physical knowledge. In C. E. Granrud (Eds), *Visual perception and cognition in infancy.* Hilsdale, NJ: Lawrence Erlbaum Associates.

Baillargeon, R. & Hanko-Summers, S. (1990). Is the top object adequately supported by the bottom object? young infants' understanding of support relations. *Cognitive Development*, *5*, 29–53.

Behl-Chadha, G. & Eimas, P. (1995). Infant categorization of left-right spatial relations. *British Journal of Developmental Psychology*, *13*, 69–79.

Choi, S. & Bowerman, M. (1992). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition*, *41*, 83–121.

J. J. Gumperz & S. C. Levinson (Eds.) (1996). *Rethinking linguistic relativity.* Cambridge: Cambridge University Press.

Harnad, S. (1990). The symbol grounding problem. *Physica D.*, *42*, 335–346.

Herskovits, A. (1986). *Language and spatial cognition: An interdisciplinary study of the prepositions in english.* Cambridge: Cambridge University Press.

Hoffman, C., Lau, I., & Johnson, D. R. (1986). The linguistic relativity of person cognition. *Journal of Personality and Social Psychology, 51*, 1097–1105.

Hummel, J. E. & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review, 99*, 480–517.

Jackendoff, R. (1992). *Languages of the mind.* Cambridge, MA: MIT Press.

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason.* Chicago: Chicago University Press.

Kleinfeld, D. (1986). Sequential state generation by model neural networks. *Proceedings of the National Academy of Science, 83*, 9469–9473.

Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate.* Cambridge, MA: MIT Press.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind.* Chicago: Chicago University Press.

Landau, B. (1996). Multiple geometric representations of objects in languages and language learners. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), *Language and space.* Cambridge, MA: MIT Press.

Langacker, R. (1987a). *Foundations of cognitive grammar.* Stanford: Stanford University Press.

Langacker, R. (1987b). Nouns and verbs. *Language, 63*, 53–94.

Loftus, E. F. & Palmer, J. C. (1974). Reconstruction of automovile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior, 13*, 585–589.

Lucy, J. A. (1992). *Grammatical categories and cognition: a case study of the linguistic relativity hypothesis.* Cambridge: Cambridge University Press.

Lucy, J. A. (1996). The scope of linguistic relativity: an analysis and review of empirical research. In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity.* Cambridge: Cambridge University Press.

McClelland, J. & Rumelhart, D. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88*, 375–107.

Movellan, J. (1990). Contrastive Hebbian learning in the continuous Hopfield model. In D. Touretzky, J. Elman, T. Sejnowski, & G. Hinton (Eds.), *Proceedings of the 1990 Connectionist Models Summer School.* San Mateo, CA: Morgan Kaufmann.

Needham, A. & Baillargeon, R. (1993). Intuitions about support in 4.5-month-old infants. *Cognition, 47*, 121–148.

Needham, A. & Baillargeon, R. (1997). Object segregation in 8-month-old infants. *Cognition, 62*, 121–149.

Pinker, S. (1994). *The language instinct: How the mind creates language.* New York: W. Morrow.

Quinn, P. (1994). The categorization of above and below spatial relations by yound infants. *Child Development, 65*, 58–69.

Quinn, P., Cummins, M., Kase, J., Martin, E., & Weissman, S. (1996). Development of categorical representations for above and below spatial relations in 3- to 7-month old infants. *Developmental Psychology, 32*(5), 942–950.

Regier, T. (1996). *The human semantic potential: Spatial language and constrained connectionism.* Cambridge, MA: MIT Press.

Schooler, J. & Engstler-Schooler, T. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology*, *22*(1), 36–71.

Shastri, L. & Ajjanagadde, V. (1993). From simple associations so systematic reasoning: A connectionist representation of rules, variables, and dynamic bindings using temporal synchrony. *Behavioral and Brain Sciences*, *16*, 417–494.

Spelke, E. S. (1990). Principles of object segregation. *Cognitive Science*, *14*, 29–56.

Spelke, E. S., Breinlinger, K., Jacobson, K., & Phillips, A. (1993). Gestalt relations and object perception: A developmental study. *Perception*, *22*(12), 1483–1501.

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, *99*, 605–632.

Spelke, E. S. & Kyeong, K. I. (1992). Infants' sensitivity to effects of gravity on visible object motion. *Journal of Experimental Psychology Human Perception and Performance*, *18*(2), 385–393.

Sporns, O., Gally, J. A., George N. Reeke, J., & Edelman, G. M. (1989). Reentrant signaling among simulated neuronal groups leads to coherency in their oscillatory activity. *Proceedings of the National Academy of Sciences*, *86*, 7265–7269.

Thelen, E. & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action.* Cambridge, MA: MIT Press.

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind.* Cambridge, MA: MIT Press.

Weist, R. M., Lyytinen, P., Wysocka, J., & Atanassova, M. (1997). The interaction of language and thought in children's language acquisition: a crosslinguistic study. *Journal of Child Language*, *24*, 81–121.

Whorf, B. L. (1956). *Language, thought and reality: Selected writings of Benjamin Lee Whorf.* Cambridge, MA: MIT Press.

Wu, L. (1995). *Perceptual representation in conceptual combination.* PhD thesis, University of Chicago, Chicago, IL.