SCCS
TECHNICAL REPORT

Studies in Language Sciences   1

Yuriko Oshima-Takane, Yasuhiro Shirai
and Hidetosi Sirai

The Japanese Society for Language Sciences

August 4, 2000

SCHOOL OF COMPUTER AND COGNITIVE SCIENCES
INSTITUTE FOR ADVANCED STUDIES IN ARTIFICIAL INTELLIGENCE

CHUKYO UNIVERSITY

Contents

Second Language Acquisition and Bilingualism

Perspective-taking and grammar

BRIAN MACWHINNEY, Carnegie Mellon University

Abstract

The comprehension and production of sentences is grounded on a process of perspective-taking that operates on five distinct levels: object perception, actions, spatio-temporal reference frames, predicate chains, and social frames. On each of these levels, speakers and listeners use lexical and grammatical forms to integrate among and shift between competing perspectives. Each of the five perspectival systems is grounded on a constructed, embodied, ego perspective. Relying on preexisting neural support, language promotes the emergence of the integrated human perspective by dynamically linking perspectives across these five systems.

1 Introduction

The basic function of language is communication. When listeners succeed in decoding the message intended by the speaker, the communication has been a success. But how do listeners achieve this feat? How do listeners manage to recreate the original meaning of the utterance, with all its perceptual and emotional richness, just as it existed in the mind of the speaker? We know that successful comprehension depends, in part, on the well-learned mechanics of auditory, lexical, and syntactic processing. However, beyond these mechanical processes, there is a deeper process of interpretation that allows the listener to reorganize the pieces of a message into a meaningful whole.

In this paper, I argue that successful communication is achieved through the activity of perspective-taking. By giving listeners clear cues about which perspectives they should assume and how to move from one perspective to the next, speakers are able to share their perceptions and ideas with their listeners. When language is rich in cues for perspective-taking and perspective-shifting, it awakens the imagination of the listener and leads to a successful consummation or sharing of ideas, impressions, attitudes, and narratives. When the process of perspective-sharing is disrupted by interruptions, monotony, excessive complexity, or lack of shared knowledge, communication tends to break down.

Although we understand intuitively that perspective-taking is central to communication, few psycholinguistic or cognitive models assign it more than a peripheral role. In this paper, we will formulate a hypothesis that places perspective-taking at the core of higher-level cognition. This approach, which we will call the perspective hypothesis, makes the following basic claims:

1. Because of its centrality in communication, perspective-taking has left an indelible imprint on the shape of grammar. The general shape of grammatical forms is specifically designed to support the processes of perspective-taking, perspective-shifting, and perspective-sharing.
2. Comprehension and production of language rely on both depictive and enactive modes to support perspective-taking.
3. Perspective-taking operates through the interaction of five cognitive subsystems that serve to construct object deixis, action perception, space/time, predicate chains, and social roles.
4. Objects are encoded depictively by systems for object recognition in the ventral visual stream. They are encoded attentionally by frontal systems for attentional and visual fixation. These systems create deictic codes for cognition (Ballard, Hayhoe, Pook, & Rao, 1997) which are realized in language through forms of deictic expression.
5. Objects are encoded enactively in terms of the actions or affordances they elicit. This processing is driven by areas of motor and premotor cortex that serve to construct a virtual body image that can perceive the actions of both the self and others, thereby constructing a

basic level of perspective-taking and perspective-shifting. Language reflects these enactive processes by mapping the meanings of closed class words onto the human body image.

6. Space and time are constructed by the projection of the human body image across egocentric, allocentric, and geocentric frames. Language relies on hippocampal and temporal lobe memory systems to shift perspectives in space and time.
7. Predicate chains are constructed by the linking of object, action, and space-time system into sequences of perspective shifts. These shifts are controlled by premotor working memory areas and inferior frontal action planning areas. Grammar is specifically designed to facilitate perspective-shifting during sentence processing.
8. The construction of social and cognitive schemes relies on the linguistic construction of complex forms of perspective-taking. In this way, perspective-taking provides a major vehicle for human socialization. By tracing perspective shifts in language, children are able to learn the paths for practical and social problem-solving that are sanctioned by their culture.
9. The emergence of language as a species-specific human skill was grounded on two basic evolutionary achievements. The first was the development of a system for rapid control of the articulatory apparatus. The second was the linking of the pre-existing perspective-taking system to the new modality of vocal communication.

These nine basic claims of the perspective hypothesis are, of course, extremely programmatic and vague. Pieces of these claims have been articulated in research programs as diverse as Cognitive Grammar (Lakoff & Thompson, 1977; Langacker, 1986; Talmy, 1978, 1988), Vygotskyan developmental psychology (Vygotsky, 1962), Piagetian developmental psychology (Case, 1997; Piaget, 1954; Vosniadou, 1994), ecological psychology (Gibson, 1966, 1977; Michaels & Oudejans, 1992; Neisser, 1976; Warren, 1984; Warren, Young, & Lee, 1986), cognitive neuroscience (Damasio, 1999; Jeannerod, 1997; Mesulam, 1990; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996), mental model theory (Fauconnier, 1994; Miller & Johnson-Laird, 1976; Tabachneck-Schijf, Leonardo, & Simon, 1997), imagery theory (Barsalou, 1999; Kosslyn, Thompson, Kim, & Alpert, 1995; Tversky, 1996), situated cognition (Barwise & Perry, 1983; Glenberg, 1997), and robotics (Bailey, Chang, Feldman, & Narayanan, 1998; Brooks, 1991; Harnad, 1995). The unique contribution of the perspective hypothesis is the attempt to draw a common thread through these diverse approaches to human language and cognition. If an elucidation of these links leads to a deeper understanding of cognition and communication and the generation of new empirical hypotheses, the exploration of the perspective hypothesis will be a success. If this hypothesis cannot be articulated in a coherent and empirically meaningful fashion, then it should be rejected. This paper constitutes an initial attempt to clarify some of the claims and predictions of the general hypothesis.

2 Depictive and Enactive Modes

As an example of the operation of perspective-taking, consider a sentence such as "the skateboarder vaulted over the railing." Here, we take the perspective of "the skateboarder" and imagine the process of crouching down onto the skateboard, snapping up the tail, and jumping into the air, as both rider and skateboard fly through the air over a railing and land together on the other side. Identifying with the skateboarder as the agent, we can evaluate the specific bodily actions involved in crouching, balancing, and jumping. The more we know about skateboarding, the more deeply we can process this utterance. If we know only a little about skateboarding, the perspective we assume will be more monochromatic and superficial. We might simply imagine the skateboard and rider as a relatively non-dynamic unit somehow passing over a railing. This superficial interpretation will not include a real understanding of body movements and physical dynamics. In general, the extent to which we elaborate our understanding of any particular sentence depends on our ability to assume a perspective from which we can enact the entire sentence. The more time

4

that we have available and the richer our knowledge base, the more deeply we can understand the sentence.

Perspective-taking operates in two modalities: a depictive modality and an enactive modality. In the depictive modality, we see the skateboarder vaulting over the railing as a character in a videotape. We adopt an external perspective and the motions of the actor are encoded in a primarily visual and depictive fashion. This type of encoding is the easiest to achieve, because it relies on the phylogenetically ancient mechanisms of visual perception that mammals share with fishes, birds, and reptiles. These mechanisms provide quick, clear encodings of the scene that minimize interpretation through reference to the perceiver. The second mode of perspective-taking is the enactive mode. In this mode, we assume the perspective of the skater and reenact the activities of crouching and vaulting to the degree to which we can covertly replay them in our internal physical experience. Construction of an interpretation in the enactive mode is far more difficult than is construction of an interpretation in the depictive mode. However, this level of processing and elaboration is hypothesized to be deeper (Craik & Lockhart, 1972) and more important for memory, cognitive development, and problem-solving.

Consider another example of the enactive-depictive contrast in the sentence, "the cat licked herself." In the depictive mode, we see a movie of the cat raising her paw to her mouth and licking the fur with her tongue. In the enactive mode, we take the stance of the cat. We refer her paw to our hand and her tongue to our tongue. Most people would say that they are unlikely to employ the enactive mode in this case, as long as the sentence is presented by itself outside of context. However, if we embed the sentence in a larger discourse, we have more incentive to process through the enactive mode. Consider this passage:

> Mary's big grey cat purred softly in her bed. Glancing the first rays of dawn, she stood up, stretching her legs and arching her back. Reaching out her front paws and digging her claws into the carpet, she raised her back haunches for another stretch. Then she walked over to the side of the window and leaned against the glass for warmth. Raising her right paw to her tongue, she licked her fur with undisturbed self contentment.

The more vivid our descriptions, the more they stimulate enactive processes in comprehension. However, both depictive and enactive processes run roughly in parallel. As a result, we can begin to extract some enactive understandings, even as we continue to work out depictive fragments of sentence interpretations. Because our enactive interpretations may not be complete as we listen to ongoing discourse, it is often the case that they are not as fully present to our consciousness.

When we look at sentence production, as opposed to sentence comprehension, the situation is different. In production, we have direct access to memories that are encoded in the enactive mode. Unless we are describing events as we see them, we are retrieving events from memory, and according to the perspective-taking hypothesis, these events are most likely to be encoded enactively.

The contrast between enactive and depictive processing corresponds well with the distinction between processing in the ventral and dorsal visual processing streams (Ungerleider & Haxby, 1994). The ventral stream runs from the primary visual areas in the occipital lobe through the object recognition areas of the temporal lobe. It is this stream that processes images depictively. The ventral stream runs through the parietal and eventually projects to supplementary eye field areas in the premotor cortex. This stream is understood to process objects in terms of links between sensation and action. It is this stream that enables enactive visual processing

3 Subsystems for Support of Perspective-taking

Perspective-taking relies on at least five component subsystems. These subsystems process information in terms of (1) objects, (2) affordances, (3) spatio-temporal reference frames, (4) predicate chains, and (5) social plans. Each of these subsystems organizes perception and action

from the viewpoint of the basic virtual body image that relies on "cognitive simulation" (Feldman et al., 1996) to construct meaning by re-enacting sensory and motoric experiences. Each of these subsystems also maps onto specific lexical and grammatical structures in human language. Let us examine each of these five component systems in turn.

3.1 Objects and Deixis

Our initial mode of interaction with objects is through depictive processing in the ventral stream. However, our use of repeated eye fixations to maintain an object in foveal attention forms the basis for a link between object recognition and perspective. The image that the object projects to the retina binds it to a particular place in visual space. However, as we change our head, this position varies. What is constant is the location of the object in terms of our ability to repeatedly fixate its position. This basic link between perception and eye movements gives rise to the experience of object constancy and the mental construction of deixis. This control of eye movements is governed by areas outside the ventral visual stream such as the supplementary eye fields in premotor cortex. In an early study on this topic, Bossom (1965) adapted monkeys to using special eyeglasses that inverted the visual field. After moving about with these eyeglasses for some days, the monkeys became readapted to the upside down view these glasses provided. When Bossom then lesioned the monkeys at various cortical locations, he found that only lesions to the area of the frontal lobes known as the supplementary eye fields resulted in damage to the readapted visual field. This finding suggests that these frontal structures support the construction of a dynamic and adaptable visual field. Using single-cell recording techniques with macaque monkeys, Olson and Gettner (1995) located cells in the supplementary eye field of prefrontal cortex that respond not to positions in the actual visual field, but to positions on objects in visual memory. These results suggest that the prefrontal visual area works together with parietal areas to facilitate object-centered processing of affordances and spatial representations.

Ballard (1997) views cognition as grounded on the use of "deictic codes." These codes can be used to recover previous visual fixations and body positions. If we move our eyes toward an object and hope to be able to return to that same position in the future, we need to have some way of recreating the eye movements and head position that we used when first looking at that object. By accessing deictic codes, we avoid the many computations that would be involved in having to worry repeatedly about all of the possible objects and locations that exist in the external world. The brain does this by establishing an internal deictic code for each object.

The establishment of deictic codes depends on a set mechanisms for the neuronal encoding of the body image and body maps. Damasio (1999) has outlined the ways in which a distributed functional neural circuit involving mid-brain structures and the basal ganglia helps maintain the body image. Motor cortex (Kakei, Hoffman, & Strick, 1999) maintains as many as twelve separate maps of the human body. Additional body maps are located in the cerebellum (Middleton & Strick, 1998). Some of these maps can encode body orientation, head position, and the direction of eye movements. Other may be more linked to the dynamic actions discussed in the next section. The dynamic linkage between alternative encodings on separate maps can be maintained by reverberation in functional circuits.

Connections between posterior and frontal areas (Goldman-Rakic, 1987) provide a method for the temporary storage of deictic codes in premotor working memory areas and the accessing of previous attentional foci. More permanent traces may be established through hippocampal processing (McClelland, McNaughton, & O'Reilly, 1995). The fact that primates have a short-term memory capacity equal to that of humans (Levine & Prueitt, 1989) suggests that the deictic memory system can operate smoothly without additional reliance on verbal memory systems such as the phonological loop (Baddeley, 1990; Gathercole & Baddeley, 1993).

Although deictic processing does not depend on language, linguistic and cognitive processing make heavy use of deixis. Deixis supports contextual grounding in language and cognition (Glenberg & Langston, 1992; Gordon, Grosz, & Gilliom, 1993; Tversky, 1996; Zwaan, 1996). In order to construct narratives (Berman & Slobin, 1994), interpret verbal reference (Gernsbacher,

1990; Silverstein, 1976b), or plan activities (Sacerdoti, 1977), we need to make continual use of deixis. However, the actual use of these deictic codes to support perspective-taking requires additional cognitive processes discussed in the next sections.

3.2 Affordances

Our interactions with objects involve both depiction and enaction. Consider the case of an object like a banana. Enactive interactions with bananas involve motoric gestures such as grabbing, squeezing, sniffing, peeling, breaking, and eating. When we say the word "banana," it is easy to recall the fruit's basic depictive image. However, our actual enactive interactions with a banana trigger a fuller set of perceptions and actions. When we grab a banana, our hands experience the texture of the peel, the ridges along the peel, the smooth extensions between the ridges, and the rougher edges where the banana connects with other bananas into a bunch. These haptic sensations are coordinated with visual sensations such as a perception of the yellow and brown colors of the banana and its curving shape. When we hold or throw a banana, we appreciate its weight and balance. An overripe banana can assault us with its pungent smell. When we peel a banana, we encounter still further sensations involving the action of peeling, as well as the peel itself. With the peel removed, we can access new sensations from the meat of the banana. When we eat a banana, our whole body becomes involved in chewing, swallowing, and digestion. All of these direct interactions in vision, smell, taste, touch, skeletal postures, kinaesthesia, proprioception, and locomotor feedback are provided by a single object which we categorize as a "banana". It is this rich and diverse set of sensations and motor plans that constitutes the fullest grounding for our understanding of the word "banana." Of course, we know other things about bananas. We know that they are rich in potassium and Vitamin E, that they are grown in Central America by United Fruit cooperatives, and so on, but these are secondary, declarative facts (Paivio, 1971; Tabachneck-Schijf et al., 1997).

3.2.1 Fictivity

Researchers in the field of ecological psychology refer to these interactions between the animal and its environment as "affordances." From the viewpoint of the perspective hypothesis, the notion of affordance is informative and insightful. However, it would be best to refer to the imagined actions stimulated by language as fictive affordances, since their use in communication is indirect and fictive. In this regard, it is important to realize that neuronal support for perspective-taking maintains a certain independence from actual perceptual and motoric interactions. Psychophysiologists have often noted a general match between emotions, thoughts, and efferent responses. For example, when we imagine performing bicep curls, there are discharges to the biceps (Jeannerod, 1997). When a trained marksman imagines shooting a gun, the discharges to the muscles mimic those found in real target practice. When we imagine eating, there is an increase in salivation. However, detailed electrophysiological and patient studies (Behrmann, Moscovitch, & Winocur, 1994) underscore the extent to which perspective-taking and imagery are decoupled temporally and cognitively from the actual execution of motoric plans. In this sense, perspective-taking has a fundamentally "fictive" quality. When we imagine curling our biceps or shooting at a target, we are essentially providing instructions to our internalized body image to run through a series of activities without actually executing those actions. In this way, imagery can track actual perception fictively.

For people with sensory or motoric limitations, experiences with a banana may be restricted in certain ways. Blind people can appreciate the smell, touch, and shape of a banana, but cannot process its color. Paraplegics can see, smell, and eat the banana, but their understandings of the actions of peeling or grabbing may be based only on their observing these actions in others. However, the presence of a fundamental body image even in paraplegics (Fourcin, 1975) serves as a solid grounding for learning about bananas, even in this distal way. Recent work using single-cell recording technology with monkeys (Rizzolatti et al., 1996) has pointed to the existence in monkey

premotor cortex of "mirror cells" for actions such as grasping or twisting. By providing links of this type between depictive and enactive perception, the brain provides ways for even the physically handicapped to engage in perspective taking with objects and through language. Linking the activity of these mirror cells to verbal labels may provide further support for constructing cognitive simulations of objects. For example, a word like "twist" can generate both the visual impression of a twisting motion and the motoric impression of a twisting action.

3.2.2 Impact on Language

Sometimes languages express fictive perspective-taking directly in their names for things. For example, in Navajo, a chair is "bikáá'dah'asdáhí" or "on-it-one-sits". Or to take a more familiar example, many languages refer to a cork screw as a "cork puller". In such examples, objects are being characterized in terms of fictive actions or affordances. In their work on procedural semantics, Miller and Johnson-Laird (1976) showed that definitions of nouns in terms of criterial attributes were often not as predictive as definitions in terms of such fictive affordances. For example, they found that attempts to define a "table" in terms of the number or the placement of its legs or the shape of the top often failed to capture the possible variation in the shape of what counts as a table. It works better to define a table instead as an object that provides a space upon which we can place work. In this way, Miller and Johnson-Laird eventually came to the same conclusion that the Navajo reached when they called a table "bikáá'dání" or "at-it-one-works".

Languages often reflect the embodied nature of fictive affordances quite directly. In English, we speak of the hands of a clock, the teeth of a zipper, and the foot of the mountain. In Apache, this penchant for body part metaphors carries over to describing the parts of an automobile. The tires are the feet of the car, the battery is its heart, and the headlights are its eyes. Such perspectival encodings combine with the basic affordances we discussed earlier in the case of "banana" to flesh out the meanings of words, even before they are placed into syntactic combination. The 18th century philosopher Giovanni Batista Vico pointed to exactly this phenomenon:, when he noted that:

> In all languages the greater part of the expressions relating to inanimate things are formed by metaphor from the human body and its parts and from the human senses and passions....for when man understands he extends his mind and takes in the things, but when he does not understand he makes the things out of himself and becomes them by transforming himself into them. (New Science, section 405)

Verbs and adjectives also encode fictive actions. When we hear the word "walk", we immediately activate the basic elements of the physical components of walking. These include alternating motions of the legs, counterbalanced swinging of the arms, pressures on the knees and other joints, and the sense of our weight coming down on the earth. Although all of these fictive actions are eventually available, we only glimpse the tip of this iceberg, unless the sentence calls for the activation of the full set. Adjectives and adverbs also activate fictive perceptions. Consider the combination of the adjective "red" with different nouns. A red tomato is redder than a head of red lettuce. Squeezing a towel and sueezing a sugar cube afford very different sensations. Pointing a football toward something involves different orientations and actions from pointing a tennis ball toward something. In this way, we see how the interpretation of language from deictic centers relies critically on the activity of perspective-taking. In particular, the competitive interactions and polysemic processes discussed in Lakoff (1987) and MacWhinney (1989) are grounded on exactly this type of perspectival interaction.

3.2.3 Impact on Language Development

Children tend to learn new words by matching up their own body concept to the actions and orientations encoded by these words. Lise Menn (personal communication) observed her son looking at a bird and declaring "have no hands". In effect, the child was relating to the bird by assuming its perspective and this led immediately to the observation that the bird had no hands. In

8

other words, the perspective of the bird did not provide the affordance of self-perception of the hands. Similarly, Marilyn Shatz (personal communication) reported the case of a child who, after looking at a tiger's tail, turned around to see if she had a tail too. Such anecdotes reflect the ways in which children attempt to understand distal objects by projecting their own human perspective onto those objects. They then perceive the components of that object in terms of their own body parts. This mode of apperception can also extend to non-animate agents, as we take an embodied approach to understanding the shapes, postures, and positions of trees, cars, houses, and even waterfalls (Werner & Kaplan, 1963).

3.2.4 Experimental Support

The reality of fictive affordances for words can be demonstrated experimentally. The logic behind this type of demonstration involves the notion that compatibility between ideas should facilitate reaction time and memory. The claim is that, if our body is engaged in a particular action, we tend to assimilate words that suggest that action. For example, Klatzky, Pellegrino, McCloskey, and Doherty (1989) asked subjects to judge whether phrases such as "aim a dart" or "close a nail" made sense or not. When the phrases were preceded by hand-shape name primes that matched the action, such as "pinch" for "aim a dart," these judgments were facilitated, as compared to baseline primes such as "push." It appears that subjects used the prime to generate an internal fictive representation of the action which then matched up well with the hand shape described in the target phrase.

Consider another example. Ertel and Bloemer (1975) asked subjects to verify sentences that sometimes contained negative elements. They found that judgments for negative sentences were facilitated when subjects were separating blocks, rather than putting them together. Similarly, Berkowitz and Trocolli (1990) report that when subjects are asked to smile they produce more positive affective judgments about people mentioned in sentences, than when they are asked to frown. These findings for language interpretation match up well with the physiological findings discussed earlier that suggest that people activate complete motoric responses when they imagine actions and objects.

3.2.5 Neuronal Mechanisms

Studies by Parsons et al. (1995) and Martin, Wiggs, Ungerleider, and Haxby (1996) and Cohen et al. (1996) have shown that, when subjects are asked to engage in mental imagery, they use modality-specific cortical systems. Studies using a verb generation task have pointed to an important role for frontal cortex in supporting strategic aspects of meaning access and generation (Petersen, Fox, Posner, Mintun, & Raichle, 1988; Posner, Petersen, Fox, & Raichle, 1988). In this task subjects are shown pictures of objects and asked to think of actions they might perform on these objects. Thus, this task is specifically designed to measure the generation of words on the basis of affordances. In addition, lesion studies (Gainotti, Silveri, Daniele, & Giustolisi, 1995), PET studies (Posner et al., 1988), and fMRI analyses (Menard, Kosslyn, Thompson, Alpert, & Rauch, 1996) have shown that right frontal areas are involved in the generation or retrieval of action terms. Together, these studies point to an important role for frontal cortex in generating access cues for specific actions and the words that express those actions.

This view of frontal functioning fits in well with the characterization of processing in the dorsal visual stream as involving an integration between perception and action. Goodale (1993) notes that patients with lesions to the dorsal stream have problems not only with locating objects in space, but also with forming hand positions that are appropriate for manipulating these objects. Single-cell recording techniques have shown that there are cells in posterior parietal visual areas that only respond to objects when they are being acted upon. These findings support the idea that the dorsal visual stream provides perception-action linkages for processing affordances. These perception-action linkages correspond to what Horowitz and Prytulak (1969) called "reafference", what Teuber (1964) called "corollary discharge", and what Glenberg (1997) called "embodied" perception.

3.3 Space and Time

The third major component of the perspective-taking system is the one that deals with position and motion in space and time. Researchers have shown that there are three major spatial frames that speakers can use to specify the positions and movements of objects. These frames are organized around egocentric, allocentric, and geocentric perspectives.

3.3.1 The Egocentric Frame

The most basic of these three frames is the ego-centered frame, since it directly encodes the perspective of the speaker. The spatial position of the speaker is encoded by the deictic term "here." Locations away from the speaker are marked by the deictic term "there." Perspective-taking in space and time builds on top of the basic deictic framework constructed by the object recognition and affordance subsystems, since every location in space and time must eventually be referenced to the here and now. Language allows the speaker to modify the initial deictic perspective to include the listener as a part of ego. In that case, "here" can refer to the position of the speaker and the listener, and "there" can refer to a position away from the speaker and listener. Other terms that are grounded in the self's position and perspective include "forward", "backward", "up", "down", "left", and "right". Temporal reference in the egocentric frame uses "now" as the deictic center and constructs the past and future in relation as fictive movements away from the time of speaking.

3.3.2 The Allocentric Frame

The second spatial frame is the object-centered or allocentric frame. This frame is constructed by projecting the perspective of the ego onto an external object. In effect, the speaker assumes the perspective of another object and then judges locations from the viewpoint of that object. The basic activity is still deictic, but it is extended through perspective-taking. For example, "in front of the house" defines a position relative to a house. In order to determine exactly where the front of the house is located, we need to assume the perspective of the house. We can do this by placing ourselves into the front door of the house where we would face people coming to the front door to "interact" with the house. Once its facing is determined, the house functions like a secondary human perspective, and we can use spatial terms that are designed specifically to work with the allocentric frame such as "under", "behind", or "next to". If we use these terms to locate positions with respect to our own bodies as in "behind me" or "next to me", we are treating our bodies as the centers of an allocentric frame. In both egocentric and allocentric frames, positions are understood relative to a figural perspective that has an orientational field like that of the upright human body (Bryant, Tversky, & Franklin, 1992; Clark, 1973).

Shifts in spatial perspective can lead to strange alternations of the perspectival field. For example, if we are lying down on our backs in a hospital bed, we might refer to the area beyond our feet as "in front of me", even though the area beyond the feet is usually referred to as "under me". To do this, we may even imagine raising our head a bit to correct the reference field, so that at least our head is still upright. We may also override the normal shape of the allocentric field by our own egocentric perspective. For example, when having a party in the back yard of a house, we may refer to the area on the other side of the house as "in back of the house", thereby overriding the usual reference to this area as "the front of the house". In this case, we are maintaining our current egocentric position and perspective as basic and locating the external object within that egocentric perspective.

Just as the self may be treated as an object, external objects can be treated as the centers of a complete ego perspective. For example, when we say that the "supermarket is up from the police station" we mean that one can take the perspective of the police station and then use ego-centered deictic reference to describe the position of another object. In this sense, allocentric reference is best viewed as an extension of egocentric reference grounded on a shift of perspective from the self to an external object. Both egocentric and allocentric perspectives are oriented about the basic

human perspective. All that is involved in moving between egocentric and allocentric reference frames is explicit tracking of perspective shift in a way that allows the two frames to be active in parallel.

The marking of allocentric temporal relations parallels the marking of spatial relations. The initial basis for temporal deixis is the time of the speech act in which the ego is present. Reichenbach (1947) has shown how tense can encode three different temporal perspectives: speaking time (ST), event time (ET), and reference time (RT). When event time is judged in relation to speaking time, an egocentric frame is being used. When the event time is judged in comparison to a reference time other than speaking time, an allocentric frame is being used.

3.3.3 The Geocentric Frame

The third system reference system, the geocentric frame, enforces a perspective based on fixed external landmarks, such as the position of a mountain range, the North Star, or a river. These landmarks must dominate a large part of the relevant spatial world, since they are used as the basis for a full-blown Cartesian coordinate system. The Guugu Yimithirr language in northeast Queensland (Haviland, 1996) makes extensive use of this form of spatial reference. In Guugu Yimithirr, rather than asking someone to "move back from the table," one might say "move a bit to the west." We can use this type of geocentric reference in English too, but our uncertainty about whether our listener shares our judgments about which way is "west" makes use of this system far less common. On the other hand, we often make use of Cartesian grids centered on specific local landmarks in English. For example, we can describe a position as being "fifty yards behind the school." In this case, we are adopting an initial perspective that is determined either by our own location (e.g., facing the school) or by the allocentric perspective of the school for which the entry door is the front. If we are facing the school, these two reference frames pick out the same location. When we describe the position as being located "fifty yards toward the mountain from the school", we are taking the perspective of the mountain, rather than that of the speaker or the school. We then construct a temporary Cartesian grid and use a measurement like "fifty yards" to locate a given object.

3.3.4 Impact on Language

As we have already noted, language uses a variety of closed-class forms to express basic spatial relations. In English, much of this work is done through prepositions, pronouns, and tense markers. In other languages, there may be a greater reliance on expressions of topological relations, contact, shape, and enclosure. However, all language provide a rich set of expressions for egocentric and allocentric construction of space and time. These devices can be chained together in expressions such as "in the pond under the log across the stream." Processing of such chains of spatial expressions requires the same perspective-shifting mechanisms needed to process the predicate chains that will be discussed in the next section.

3.3.5 Impact on Language Development

Children begin the process of learning to make spatial reference by developing a basic egocentric understanding of the positions of objects in space. Piaget (1952) has described this development in terms of the development of the object concept and procedures for dealing with invisible displacements. In learning to remember the positions of objects, the preverbal child relies on each of the three spatial reference systems. However, as Piaget has observed, the egocentric frame is primary.

At the end of the second year, when the child comes to the task of learning language, the first locative terms are primarily egocentric and deictic. Allocentric terms such as "in" or "on" are initially processed in terms of affordances and topological relations, rather than through a complete shift of perspective to the distal object. Slowly, the use of the allocentric frame takes on an

independent existence and children learn to shift reference between these frames. Geocentric reference is acquired much later (de Leon, 1994).

Weist (1986) has shown how children begin temporal reference with a tenseless system in which events are simply stated. They then move on to a deictic system in which the event time is coded in reference to speaking time. Finally, they acquire the ability to code event time with respect to reference time in accord with allocentric reference.

There is little research examining the learning of spatial perspective-shifting in natural language use. Research has focused instead on the construction of internalized mental maps from sketchy verbal descriptions. This research has shown that subjects shift between multiple competing frames in accord with task demands ((Carlson-Radvansky & Logan, 1997; Franklin, Tversky, & Coon, 1992; Maki & Marek, 1997). Work with children has shown that learning to manipulate these competing frames develops gradually through the school years (Rieser, Garing, & Young, 1994).

3.3.6 Experimental Support

The effects of perspective-taking have been demonstrated in a variety of studies of discourse processing. The basic logic underlying this research is to show how comprehension is facilitated when language organizes interpretations in terms of the affordances and egocentric perspective of the main character or protagonist. The following example passages adapted from Sanford and Moxey (1995) illustrate these effects:

1. While measuring the wall, Fred laid the sheet of wallpaper on the table. Then he put his mug of coffee on the wallpaper.

2. After measuring the wall, Fred pasted the wallpaper on the wall. Then he put his mug of coffee on the wallpaper.

An embodied perspectival analysis of these passages would require that, when Fred comes to put his mug onto the wallpaper, he would find it glued to the wall. For Fred to carry out the action of putting his mug on the wallpaper would require placing it onto a vertical surface using glue or magnets, which seems like a strange thing to do. Sanford and Moxey used passages of this type to argue that textual coherence depends on the construction of fully enacted representations of passages.

In other studies, the narratives involved describe the motion of a principle character through space. In such passages, listeners come to judge spatial relations from the perspective of the protagonist (Haenggi, Kintsch, & Gernsbacher, 1995; Morrow, 1986; Morrow, Bower, & Greenspan, 1989; Morrow, Greenspan, & Bower, 1987). This same experimental logic has been used to demonstrate the importance of perspective-taking in time. For example, Carrieras, Carriedo, Alonso, and Fernandez (1997) found that, when a protagonist's activities are discussed in the present tense, additional information about that protagonist is quicker to process than when the activities are discussed in the past tense. Zwaan (1996) produced a similar effect by introducing time shifts into narratives such as "a moment later", as opposed to "an hour later." The more that a temporal reference pushes an event into the background, away from the main focus of the perspective of the protagonist, the slower we are to assume that secondary perspective.

3.3.7 Neuronal Mechanisms

Earlier, we discussed ways in which frontal visual fields mechanisms support a flexible system for linking visual spatial fields with actions upon objects. These mechanisms provide a basic construction of the here and now. However, the storage of previous experiences and patterns in space and time is accomplished by other neural mechanisms. In particular, there is evidence from detailed studies of rodents (Redish & Touretzky, 1997) that the hippocampus is involved in the storage of previous spatio-temporal frames. However, this evidence from the study of rodents suggests that the hippocampus is engaged primarily in the encoding of the positions of objects in

12

spatial arrays. A memory of this type is particularly important if your goal is to retrieve a buried not or a piece of cheese. However, language processing in human depends not just on the recall of spatial position, but on the retrieval of whole episodic event codings. Work with patients such as H.M. and others suggests that lesions to the hippocampus may damaged the storage of episodic encodings of this type, but that other brain areas including parts of the temporal lobe may support storage of new events. Both the arcuate fasciculus and direct frontal-temporal pathways serve as a hardware support for working memory. Together, these cortical-cortical and subcortical pathways link frontal premotor cortex to temporal auditory cortex to form a phonological loop (Grasby et al., 1993; Menard et al., 1996; Paulesu, Frith, & Frackowiak, 1993). This loop works to keep alternative verbal messages in working memory. In the next section we explore ways in which the language and the brain support the encoding of events in predicate chains.

3.4 Predicate Chains

The three subsystems we have discussed so far provide grounding for many of the basic units of language. Object recognition and affordance processing provide us with embodied reference for many nouns, verbs, and adjectives. Spatiotemporal relations ground closed-class words such as "now", "behind", and "until".

Spatial relations also serve to link up objects in terms of attachment of prepositional phrases to heads, as in "the bench in the park". However, these affordances and orientations by themselves do not provide us with a relational system that can encode the ways in which objects act upon other objects. Grammar provides a series of devices specifically designed to address this problem.

3.4.1 Impact on Language

At the core of grammatical structure is the ability to form predications. The cognitive prototype for the predication is the linkage of an action to an object through the affordances of that object. For example, the linkage of the action of "climbing" to the object "tree" is given by the fact that a tree provides an affordance for climbing. The basic understanding of the affordances of objects is available to all animals. However, this type of predication is always processed in the here and now from the perspective of ego. What is unique in language is the linkage of predication to perspective-shifting. By storing predications or episodic traces in memory, humans can form chains of predications linked through carefully controlled shifts in perspective. This section will examine some of the devices that express perspective-taking in predicate chains. We will consider the impact of perspective-taking on six aspects of grammar. These include: transitivity, packaging and conflation, fictive actions, c-command, and possession. Although the current treatment looks at only a few selected aspects of a few selected systems, the perspective hypothesis claims that all areas of grammar are motivated by perspective-taking. Of course, the shape of grammatical inflectional marking is strongly influenced by phonological forces and principles for lexical storage of regular and irregular forms (Bybee, 1985). However, perspective-taking and the memory processes upon which it relies are the chief determinants of the functionally-central aspects of grammar.

3.4.1.1 Transitivity.

Verbs are the basic blocks from which predicate chains are constructed. Verbs can be either transitive or intransitive. Intransitive verbs involved a single perspective; transitive verbs involve a shifted perspective. Intransitive verbs can describe actions (e.g. "run" or "jump"), changes of state (e.g. "fall" or "redden"), constant states (e.g. "rest" or "stand"), or processes (e.g. "rain" or "relax"). In each case the verb being described is interpreted from the perspective of the object that is intimately involved in the action, process, state, or change of state. We can talk about corn growing, dominoes falling, geese flying, or a boy standing. In each case, we interpret the growing, falling, flying, or standing from the viewpoint of the nominal perspective. In fact, we can often go beyond simply seeing a distant object undergoing a change. We can actually embody this change through

our own imagined physical activities. When we think about geese flying, we can imagine using our own limbs in this way, even though we cannot really fly. When we talk about a tree standing, we can imagine ourselves as the tree and interpret this ongoing state through the perspective we have when we stand still. When we use intransitive verbs, we are actually performing a cognitive simulation of the activity or process, as we ourselves would execute it. For intransitive verbs such as "twist" or "spread," this requires us to imagine large whole body movements that are often fairly complex.

Intransitive verbs like "twist" or "spread" can also be used transitively. In such extensions, the basic embodied perspective is still that of the object which is twisting or spreading. However, this perspective is then supplemented by that of an external causer. In general, transitive predications can always be decomposed in this way. Consider the simple transitive predication "The farmer grew the corn." In this predication, there are two competing perspectives. The perspective of "the corn" is directly involved with the growing. If we wish to understand the changes that occur in the corn, we would have to assume this perspective. On the other hand, the perspective of the farmer is also relevant, since he is somehow caring for the corn in ways that make it grow. When describing relations of this type, languages have to decide whether to focus on the actor or the patient. Nominative-accusative languages, like English, place focus on the actor by treating it as the perspective for the clause. They then treat the activity of the patient as a secondary perspective contained within the scope of the larger perspective of the subject.

In ergative-absolutive languages, like Basque or Djirbal, the primary focus is on the participant undergoing the change, rather than on the participant causing the change. In the sentence "The farmer grew the corn," the farmer is placed into the ergative case and the corn is in the absolutive case. The absolutive is also the case that is used for the word "corn" in the intransitive sentence "The corn grew." This means that ergative-absolutive languages place default focus on the patient, rather than the agent. They do this in order to focus not on the act of causation, but on the processes of change that occur in the patient. English has some minor constructions that can illustrate the effect of the ergative perspective. In a sentence like "This tent sets up in about ten minutes," we assume the perspective of the tent and imagine it changing shape almost without the intervention of an outside actor, although we realize that this actor is involved on the sidelines.

Just as nominative-accusative language like English can illustrate occasional ergativity effects, languages like Hindi or Samoan can illustrate incomplete or "split" ergativity (Delancey, 1981; Silverstein, 1976a). In Hindi and Gujarati, ergative-absolutive marking is used in the perfective tense, but not the imperfective. This means that a sentence in the imperfective, like "The farmer was growing the corn" is nominative-accusative, but a sentence in the perfective, like "The farmer grew the corn", shifts into ergative-absolutive. In order to understand why this happens, consider the way in which perspective-taking interacts with aspect.

When we describe an event in the present or the imperfective, we equate our perspective with the ongoing perspective of the actor and the action. However, for an event that occurred in the past and which is fully completed, we assume the perspective of the present and describe the past as a separate reality. As a result, we are relatively less involved and less inclined to assume the perspective of the actor. This split allows us to focus on the patient and move the causer into the ergative role.

A second way in which ergative marking can be split is in accord with the person of the agent. When the actor is in third person, nearly all native languages of Australia and many native languages of North America use ergative-absolutive marking. However, when the actor is in first or second person, these languages often use nominative-accusative marking. This split reflects the fact that we are more deeply involved with the first and second person perspectives, for which we can more directly infer causality. For third person actors, we are often on safer ground to defocus their causal activities and focus instead on the perspective of the patient.

Ergative marking can also be used to mark intentionality. Delancey (1981) describes this for the Caucasian language Batsbi, which uses ergative case for the subject of a sentence like (3) when the falling is intentional and absolutive marking of the subject when the falling is unintentional.

14

3.   Bill fell to the ground.

Other factors that can lead to splits in ergative marking include inferential markers and certain discourse structures.

Variations in transitivity can also be used as a way of shifting between various causal perspectives in discourse. In an example like (4), perspective is initially assigned to the first noun (car) as starting point.

4.   The car was struck by a falling limb.

However, in order to fully construct the meaning of the utterance, a secondary perspective is established for the falling limb. The English passive is used when the agent is not the focus. Other languages achieve this effect using topicalization devices (as in Hungarian), verbal conjugations (as in Tagalog), or additional types of ergative marking (as in Jacaltec). Studies by Sachs (1967) and Lombardi and Potter (1993) indicate that passives are eventually reshaped into actives in discourse memory. However, when the discourse is structured in ways that properly support the passive structure, it is more likely to be maintained in delayed recall (Keenan, MacWhinney, & Mayhew, 1977).

Over the course of centuries, languages may tend to shift between relative emphases on ergativity or nominativity. DuBois (1987) has noted that languages tend to omit subjects from transitive clauses, while preserving them for intransitive clauses. Over time, this tendency can lead to a rise in ergativity, since it emphasizes the similarity between the object of the transitive clause and the agent of the intransitive clause. Word order variation can also promote these changes. For example, increased use of structures like the passive or object topicalization in a nominative language like English might tend to move the language slowly in the direction of a more ergative configuration.

3.4.1.2 Packaging and conflation

Individual clauses are the basic links in predicate chains. However, in order to break up larger chains of cause and effect, we have to decide how to package and conflate actions into clauses. Consider the following alternative ways of viewing a situation:

5.   The beam fell.
6.   The beam fell when the crane operator released a lever.
7.   The crane dropped the beam.
8.   The crane operator released the beam.
9.   The crane operator pulled a lever to release the beam.
10.   The crane operator dropped the beam by pulling a lever.

The selection of one of these ways of depicting the action over another depends on the perspective we take. If we take the perspective of the falling beam, we will select either of the first two sentences. If we choose (5), we restrict our perspective entirely to "the beam". If we select (6), we take "beam" as our first perspective, but then add "the crane operator" as a secondary perspective. In (8) through (10) we adopt the perspective of "the crane operator" and include or exclude the way in which the operator dropped the beam.

Perspectives can be conflated in a variety of ways. Consider the contrast between these four sentences describing the movements of small toys:

11. The lion pushed the giraffe, and the giraffe bumped into the table.
12. The lion hit the giraffe and it bumped into the table.
13. The lion bumped the giraffe into the cube.
14. The lion bumped the giraffe, sending it flying toward the table.

In (11), the two actions are packaged into separate full clauses. First, we assume the perspective of the lion hitting the giraffe. Then perspective shifts to the giraffe which bumps into the table. Sentence (12) has the same structure, but the pronoun "it" refers ambiguously to either the initial perspective of "lion" or the new perspective of "giraffe." In (13) and (14) the two separate actions are conflated into one with "lion" as the dominant perspective and "giraffe" as the subordinate perspective. The conflation of multiple actions into a single verb is also exemplified in (15):

15. So far, the people of this small textile town in northwestern Carolina have been unable to pray Mrs. Smith's two little boys home again.

In this example, the verb "pray" conflates the action of praying and the action of bringing home the two little boys. As in example (14), the perspective of the subject controls two actions at once. When verbs conflate actions in this way, they are forced to accommodate to all the meanings being combined (MacWhinney, 1989). Consider these examples:

16. The light bulb flashed.
17. The light bulb flashed until morning.

In example (16), the default reading is that the light bulb flashed no more than once or twice. However, in (17) we have to assume that the light bulb flashed iteratively until the morning. In other words, when we take the perspective of the light bulb and then evaluate the temporal frame "until morning", we have to understand the action of flashing as occurring repeatedly across a period.

3.4.1.3 Fictive action

Language provides various devices and forms to support the interpretation of causal action from the viewpoint of a dynamic perspective. Often, that perspective is an animate, human perspective. However, even when the perspective is non-animate, it still carries the full force of a causal perspective. Sentence (18) illustrates this effect.

18. The library contains three major collections.

Here, the library is viewed as an agent that actively holds collections of books. As Talmy (1988) has noted, this is only a fictive agency, since the library is not a real agent and the act of "holding collections" is not a real dynamic action. Nonetheless, the syntax of the sentence invokes a causal action frame with an agential perspective. To further illustrate this, consider the contrast between sentence (19) and (20).

19. She walked down through the cornfields, out to the river.
20. The path winds down through the cornfields, out to the river.

In (19) the human agent moves over a real path. In (20), on the other hand, the path moves fictively over the same path. Similarly, in (22), the screws are selected as the perspective and this tends to elevate the static force they are exerting to the level of a full causal action.

21. The carpenter holds the four legs firmly against the center pedestal.
22. The screws hold the four legs firmly against the center pedestal.

These examples of fictive motion and fictive causation illustrate the extent to which perspective-taking dominates our general view of causal relations in the physical universe. These same forces work for fictive social causation. For example, in (23) the activities of the initial perspective (Tim) trigger a series of activities in the secondary perspective (Mary).

23. Tim's failure to reply to her invitation led to Mary's breaking off their relation.

Here the notion is that one social action leads to another. In fact, both actions are really non-actions. Although this causation is on the social level, we apply a basic physical causal model to even these social effects.

3.4.1.4 C-command.

Perspective-taking in predicate chains impacts certain key aspects of the grammar of pronominal coreference. This effect results from a basic fact about language use, which is that starting points must be fully referential (MacWhinney, 1977). Gernsbacher (1990) has discussed this requirement in terms of her theory of "structure building". The idea here is that listeners attempt to build up a sentence's interpretation incrementally. To do this, they need to have the starting point fully identified, since it is the basic building block upon which they will build the rest of the interpretation. In dozens of psycholinguistic investigations, Gernsbacher has shown that the initial nominal phrase has the predicted "advantage of first mention." This advantage makes the first noun more memorable and more accessible for further meaningful processing.

When the first noun is low in referentiality (Ariel, 1990), the foundation is unclear and the process of comprehension through structure building is thwarted. If the starting point is a full nominal, referentiality is seldom at issue. However, if the starting point is a pronoun, then there must be a procedure for making it referential by finding an antecedent. One way of doing this is to link the pronoun up to an entity mentioned in the previous discourse. In a sequence like (24), it is easy to link up "he" in the second sentence with "John" from the first sentence, since John has already been established as an available discourse referent. However, in (25), there is nothing to link "he" to and the second sentence seems awkward without this previous link.

24. John was trying to list the Ten Commandments. He was unable to get past the first six.
25. Only a few of the guests arrived on time. He says Bill came early.

The theory of perspective-taking attributes these effects to the fact that starting points serve as the basis for the construction of the embodied mental model conveyed by the clause.

The generative theory of Government and Binding (Chomsky, 1982; Reinhart, 1981) treats this phenomenon in terms of structural relations in a phrase-marker tree. According to this theory, the problem in the second sentence in (25) is that Bill does not "c-command" the pronoun "he" and cannot "bind" the pronoun anaphorically. Here, the topmost S node dominates both "he" and "Bill." However, the VP node and the lower S node only dominate "Bill" and not "he." Therefore the noun "Bill" does not c-command the pronoun and cannot be coreferential with the pronoun. However, if we shift "Bill" and "he" we get "Bill says he came early." In this sentence, "Bill" c-commands "he", because the only node dominating "Bill" also dominates "he." Because "Bill" c-commands "he", it can bind the pronoun and the noun and the pronoun can be coreferential. As a result, there is no problem with the second sentence in (26).

26.   Only a few of the guests arrived on time. Bill says he came early.

This effect is not a simple matter of linear order, since coreference between a pronoun and a following noun is perfectly good when the pronoun is in an initial subordinate clause. Consider the contrast between (27) and (28), where the asterisk on (28) indicates that "he" cannot be coreferential with "Lester."

27.   After he drank the vodka, Lester started to feel dizzy.
28.   *He started to feel dizzy, after Lester drank the vodka.

Contrasts of this type create problems for the simplest versions of the formalist approach, since they involve identical structural relations. However, Reinhart (1983) addresses this type of problem by invoking discourse-level constraints which prefer to have full NPs introduced before pronouns. This discourse-based account succeeds in explaining the problems with (28) but it also predicts problems with sentences like (29), whereas no problems occur in fact.

29.   She had just gotten back from vacation, when Mary saw the stack of unopened mail
      piled up at her front door.

The principle that requires that subjects be referential also applies in a somewhat weakened form to the direct and indirect objects of verbs. Van Hoek (1997) shows how availability for coreference is determined by position in the argument chain (Givón, 1976). Although attention is first focused on the subject or trajector, it then moves secondarily to the object or other complements of the verb which are next in the "line of sight" (Langacker, 1995). This secondary perspectival effect is illustrated in (30) through (35).

30.   *John told him that Bill was crazy.
31.   *I'm willing to give him fifty dollars for Ben's bike.
32.   *Him, John's mother likes.
33.   *The idea that Sarah might get poisoned frightened her.

As the object moves into a prepositional phrase, this constraint weakens further:

34.   ?People often said to her that Mary was a lunatic.
35.   ?John said to him that Bill was crazy.

By the time we reach elements that are no longer in the main clause, as in (36), coreference back to the main clause is not blocked, since elements in a subordinate clause are not crucial perspectives for the structure building process.

36.   The students who studied with him enjoyed John.

This gradient pattern of acceptability for increasingly peripheral clausal participants matches up quite well with the view that the process of perspective-taking during structure building requires core participants to be referential.

Further evidence for the gradient nature of the constraint against non-referential perspectives comes from sentences with special aspect markings, as in (37) and (38):

37. She had just gotten back from vacation, when Mary saw the stack of unopened mail piled up at her front door.

38. *She got back from vacation, when Mary saw the stack of unopened mail piled up at her front door.

Because of the presence of aspectual markers like "had" and "just" in (38), the initial main clause is made relevant for the interpretation of later material. As a result, the possibility is left open that the perspective "she" will be coreferential with later material. Sentences (39) and (40) based on Reinhart (1983) provide further examples of these aspectual effects on perspective-taking.

39. In Carter's home town, he is still considered a genius.

40. *In Carter's home town, he is considered a genius.

These aspectual forces can also interact with further subtle point of view effects (van Hoek, 1997), as in (41) and (42).

41. In Tim's play, he offers Mary a mansion.

42. *In Tim's play, he promised Mary a role.

Just as relevance markers can increase the openness of a main clause pronoun to coreference, so indefinite marking can decrease the openness of a subordinate clause noun for coreference, as in (43).

43. While Ruth argued with the man, he cooked dinner.

44. ?While Ruth argued with a man, he cooked dinner.

45. While Ruth was arguing with a man, he was cooking dinner.

The addition of an aspectual marker of current relevance in (45) overcomes the effect of indefiniteness in (44), again making "man" available as a coreferent for "he". Gradient patterning of this type provides good evidence that pronominal coreference is under the control of pragmatic factors (Kuno, 1986). Rather than deriving from autonomous formal constraints, we see that the patterns of possible pronominal coreference we have surveyed emerge directly from the forces of perspective and embodiment.

Wh-words introduce a further uncertainty into the process of structure building. In a sentence like (46), the initial wh-word "who" indicates the presence of information that needs to be identified.

46. *Who does he hate most?

In this case, it is the pronoun "he", rather than the initial word "who" which serves as the starting point for structure building. Since this sentence has no noun to which the pronoun can be bound, it must be bound to some external discourse referent. In any case, the wh-word is not a good candidate for the binding of the crucial subject pronoun. However, when there is a pronoun that is not in the crucial subject role, coreference between the wh-word and the pronoun is often possible, as in (47) through (51).

47. Who is hated by his brother most?

48. Who thought that Mary loved him?

49. Who hates his mother most?

50. Who said Mary kissed him?

51. Who hates himself most?

In these examples, the wh-word can be coreferent with non-central components, such as objects and elements from embedded clauses. Only coreference with subjects, as in (46), is blocked.

This brief discussion of constraints on coreference has only sampled a few of the most interesting patterns that emerge from a perspective-taking approach to grammar. Kuno (1986) presents a great deal of additional evidence for the importance of pragmatic and functional patterns for additional areas such as reflexive marking and constraints on repeated nominalizations.

3.4.1.5 Possessives.

We can distinguish structures that require the maintenance of multiple perspectives from those that simply require repeated perspective shifting. Sentence (52) illustrates how the possessive construction can require repeated perspective shifting.

52. My mother's brother's wife's sister's doctor's friend had a heart attack.

In order to determine the identity of this "friend," we have to trace through a series of social relationships starting with "my mother". However, once we have traversed one link in this chain, we can drop the initial perspective and shift to the new one. As a result, as long as we can correctly identify the relations involved, this structure is not impossible to process.

Algonquin languages like Cree and Fox allow for a particularly clear marking of the types of perspective shifting that go on in possessive constructions. These languages make a distinction between a proximal third person and a distal or obviated third person. In a Plains Cree sentence like (53) (Aissen, 1997), the object of the verb "dog" is farther from the starting point than is the subject "man":

53. Pakamahwe:w na:pe:w atimwa.

 Hit (3-3')   man (3) dog(3')

 "The man hit his dog."

This same relation between a closer and more distant third person is also found in possessive constructions, such as the one in (54):

54. Ote:ma

 his (3) horse (3')

The choice of obviative for the possessed in these examples indicates that the possessor is closer to the perspective of the ego. To understand these constructions we assume the perspective of the possessor and then imagine possessing the object being possessed.

3.4.2 Language Development

Slobin (1985) has noted that early child grammar is heavily dominated by prototypical frames. These frames are organized about the perspective of the child as actor. The manipulative activity frame which corresponds closely to the basic causal frame of nominative/accusative languages. In this frame, the child views herself as the causer who is acting transitively on objects that interest her or which she wants to transform. The second frame is the result frame which corresponds with the way in which ergative/absolutive languages process the results of causal actions. In addition to

these two basic frames, children also rely on figure-ground relations to code possession, sources, positions, and goals. From our current perspective, what is interesting about these prototypical frames is the extent to which each is organized from the perspective of the child as actor. The fundamental quality of the egocentric perspective has its impact not only on the learning of spatial relations, but also on the acquisition of causal action expressions.

Perspective maintenance has also been implicated in studies of children's imitations and productions of conjoined sentences (Ardery, 1979; Lust & Mervis, 1980; Slobin & Welsh, 1973). These studies have shown that young children find it easier to imitate a sentence like (55), as opposed to one like (56).

55. Mary cooked the meal and ate the bread.
56. Mary cooked and John ate the bread.

In (55) there is no perspective shift, since the perspective of Mary is maintained throughout. In (56), on the other hand, perspective shifts from Mary to John. Moreover, in order to find out what Mary is cooking, we have to maintain the perspective of both Mary and John until the end of the sentence.

3.4.3 Experimental Support

Psycholinguistic experimentation has provided a wide array of empirical support for the perspective hypothesis. It is not possible to completely review this extensive array of evidence here. However, we can mention results from two areas of study. The first is the processing of restrictive relative clauses. The account of relative clause processing presented here was first proposed in MacWhinney (1982) and further elaborated in MacWhinney and Pléh (1988). The predictions of this account can be illustrated by looking at four basic type of relative clauses, given in (57) through (60).

57. SS: The dog that chased the cat kicked the horse.   0 switches
58. OO: The dog chased the cat the horse kicked.   1 switch
59. OS: The dog chased the cat that kicked the horse.   1 switch
60. SO: The dog the cat chased kicked the horse.   2 switches

In the SS sentence type, the perspective of the main clause is also the perspective of the relative clause. This means that there are no perspective switches in the SS relative type. In the OO type, the object of the main clause is not the subject of the relative clause. Instead, perspective switches once from the main clause subject (dog) to the relative clause subject (horse). In the OS type, perspective also switches once. However, in this case, it switches to the main clause object, which then continues as the perspective of the relative clause. In the SO relative clause type, there is a double perspective shift. Perspective begins with the main clause subject (dog). When the next noun (cat) is encountered, perspective shifts once. However, at the second verb (kicked) perspective has to shift back to the initial perspective (dog) to complete the construction of the interpretation. Sentence that have further embeddings have even more switches. For example, sentence (61) has five perspective switches (dog -> cat -> boy -> cat -> dog).

61. The dog the cat the boy liked chased snarled.   5 switches

Sentences that have as much perspective-shifting as (61), without including any additional lexical or pragmatic support, are incomprehensible at first hearing.

Studies of the acquisition of relative clauses by children largely support the order of difficulty predicted by the perspective-taking account (MacWhinney, 1982). This predicted order is: SS > OO = OS > SO. This order appears to predict results across a wide variety of experimental paradigms including imitation, comprehension, and sentence memory. In addition, a study of online sentence processing effects in Hungarian (MacWhinney & Pléh, 1988) with adult subjects further supported a role for the perspective account. However, there was also evidence in Hungarian for the importance of additional parallel structure effects. In Hungarian, all six orders of the subject, object, and verb are grammatical. In three of these orders (SOV, SVO, and VSO), the subject is the topic; in three other orders (OSV, OVS, and VOS), the object is the topic. When the main clause subject is the topic, the English pattern of difficulty appears (SS > OO = OS > SO). However, when the main clause object is the topic, the order of difficulty is OO > OS = SO > SS. Sentences (62) and (63) illustrate this contrast in Hungarian, using English words:

62. SOV SS: The boy who liked the girl (he) the bike hit.
63. OSV OO: The boy who the girl liked, the bike hit (him).

Sentence (62) illustrates the sentence that is easiest in the SOV word order, when the subject is the topic; whereas (63) illustrates the sentence that is easiest in the OSV word order when the object is the topic. In (63) the initial noun is marked for accusative case in Hungarian. This means that it functions as a patient perspective for both the main and relative clauses.

As a second example of the ways in which psycholinguistic experimentation has supported the perspective hypothesis, let us consider the process of reformulation during speech production. MacWhinney and Bates (1978) asked English, Hungarian, and Italian children and adults to describe triplets of pictures involving simple transitive actions. For example, one picture showed (64) and another showed (65).

64. A cat gives flowers to a bunny.
65. A cat gives flowers to a boy.

MacWhinney (1977) found that, for pictures like these, subjects sometimes produced retraces like (66) or (67), but never produced retraces like (68) or (69).

66. A bu # a kitty's giving a flower to a bunny.
67. A boy # the cat's giving a boy a flower.
68. A ca # a bunny gets flowers from a cat.
69. A ki # a boy gets flowers from a kitty.

In other words, retracing always moves toward the unmarked perspective of the actor who did the giving, rather than the actor who did the receiving. When we start to describe a picture, we often choose a perspective on the basis of nongrammatical factors such as salience or positioning in the picture (Flores d'Arcais, 1975; Flores d'Arcais, 1987; Johnson-Laird, 1968a, 1968b; MacWhinney, 1977; Osgood & Bock, 1977; Pinker & Birdsong, 1979; Sridhar, 1988). However, once we have started to formulate a verb and the rest of the utterance, we may realize that the perspective that we selected was not the best. In such cases, we retrace and begin again with a new, more appropriate perspective.

3.4.4 Neuronal Mechanisms

Perspective shifting requires a system for representing and accessing competing plans. It appears that the frontal lobes are uniquely adapted to perform exactly this function. Dorsolateral

prefrontal cortex plays a fundamental role in the storing of alternative representations in working memory (Barch et al., 1997; Braver et al., 1997; Cohen et al., 1997; Goldman-Rakic, 1987; Owen, Downes, Sahakian, Polkay, & Robbins, 1990). The ability to store the traces of recent event perceptions in working memory is crucial for the construction of connected discourse. If we could not recall previously mentioned characters and actions, we would be unable to follow even the most basic descriptions and narratives.

The ability to shift between perspectives requires a neural system for representing alternative perspectives, as well as a method for inhibiting one or more of the competing perspectives. For example, in the Stroop task (Cohen, Dunbar, & McClelland, 1990), the reader must inhibit the perception of the color of the word in order to quickly read the name of the word. In processing SO relative clauses, we need to move quickly from the viewpoint of the subject of the main clause to the viewpoint of the subject of the subordinate clause. In processing social relations, we need to quickly assess the viewpoints of other people, particularly as they conflict with our own views.

The rich pattern of interconnectivity between frontal areas and from frontal areas to posterior, thalamic, and cingulate areas (Fuster, 1989; Kolb & Whishaw, 1995) supports the view of the frontal system as a system that integrates a variety of mental facilities, all in the service of perspective-taking. Mesulam (1990) asks, "Why does (prefrontal) area PG project to so many different patches of prefrontal cortex? Why are the various areas of prefrontal cortex interconnected in such intricate patterns?" The perspective hypothesis suggests that the answer to this question lies in the fact that the frontal cortex is attempting to integrate perspective-taking and perspective-shifting across causal chains.

Areas of inferior frontal cortex around Broca's area play an important role in controlling action sequences in speech, gesture, and oral functions (Fuster, 1989; Greenfield, 1991). To the degree that language comprehension and production depend on the construction and processing of action sequences, it makes sense that both Broca's area and supplementary motor areas should be involved in supporting language processing. From the viewpoint of the perspective hypothesis, it would be easy to suspect that these frontal areas are particularly involved in the construction and support of predicate chains. In fact, recent fMRI work has linked the processing of syntactically complex sentences to Broca's area (Booth et al., 1999; Just, Carpenter, Keller, Eddy, & Thulborn, 1996).

3.5 Social frames

Perspective-taking in social and interpersonal frames has its impact not on the grammar of the clause, but rather on the structure of discourse as it is represented through coordination, subordination, predicate chains, and the elaboration of certain lexical and rhetorical structures. Social and interactional frames determine the ways we negotiate points-of-view, disagreements, and shared understanding between different social agents. The elaboration of cognitive structures to support complex social interactions is certainly not unique to man or to human language. The roots of social perspective-taking lie in the basic process of imitation. Young dogs and tigers learn to hunt and kill through imitation. Young beavers learn to build dams through imitation. Young human children learn to walk, talk, and sing through imitation. Imitation involves a particularly direct form of social perspective-taking. By taking on the perspective of the parent, the child learns to construct the parent's actions, emotions, and perspectives. Eventually, the child comes to act like the parent. Through observational learning, the young of many species watch adult interactions and acquire age-appropriate role relations. By watching how group members interact, and by assuming alternative perspectives of group members during interactions, a child can learn a great deal about the social world. Finally, perspective is also useful in organizing nonlinguistic plans for group activities, such as hunting, fleeing, or foraging.

Complex social structures depend on both externalized and internalized linguistic expression. Scholars have long understood the extent to which inner speech supports human cognition. As Plato put it so eloquently in his Theaetetus, "The soul in thinking appears to be just talking - asking questions of herself and answering, affirming, and denying. And when she has arrived at a decision, this is called her opinion. I say therefore that to form an opinion is to speak, and opinion is the word spoken - I mean to oneself in silence and not aloud to others." Vygotsky (1962) extended this

basic insight by stressing the extent to which inner speech (Sokolov, 1972) derives from the social use of language. In effect, we come to speak with ourselves in ways that we have learned through speaking with others. In concert with Luria (1959; 1975) and others, Vygotsky elaborated a view of mental functioning that linked inner speech to planning within a social context. The notion of inner speech plays a pivotal role in the account currently being developed. Without access to linguistic expression, animals are able to construct a basic ego-centered social frame. However, inner speech empowers man with ways of operating upon the system of social frames in a fuller, and more symbolic fashion. Let us examine a few of these systems of linguistic support for social frames.

3.5.1 Social scenarios

Individual lexical items like "libel", "Internet", or "solidarity," encode social scenarios organized about the perspective of social actors. Let us take the noun "libel" as an example. When we speak of some communication as being "libel" or "libelous," we mean, roughly, that speaker A has declared that speaker B has engaged in some illegal or immoral activity, and that speaker B has convinced a general audience C that speaker A's claims are false and designed to make audience C think poorly of speaker A in ways that influence speaker A's ability to function in public life with audience C. In fact, the full legal characterization of libel is more complex than this, but the everyday use of the word "libel" has roughly this basic form. This single word conveys a complex set of interacting and shifting social perspectives. To evaluate whether or not a statement is libelous, we have to assume the perspective of speaker A, speaker B, and audience C to evaluate the various claims and possible counterclaims. All of this requires continual integration and shifting of social roles and perspectives.

3.5.2 Implicit causality

Verbs like "promise", "forgive", "admire", and "persuade" encode multiple relations of expectation, benefit, evaluation, and prediction between social actors. To evaluate the uses of these verbs requires flexible perspective-taking and coordination. Within this larger group of mental state verbs, one dimension of contrast is known as "explicit causality". Sentence (70) illustrates the use of the experiencer-stimulus verb "admire"; whereas sentence (71) illustrates the use of a stimulus-experiencer verb like "apologize."

70. John admired Mary, because she was calm under stress.
71. John apologized to Mary, because he had cracked under stress.

McDonald and MacWhinney (1995) asked subjects to listen to sentences like (70) and (71) while making a cross-modal probe recognition judgment. The probes were placed at various points before and after the pronoun ("he" and "she"). McDonald and MacWhinney found that stimulus-experiencer verbs like "apologize" in (71) tend to preserve the advantage of first mention for the first noun (John) as a probe throughout the sentence. However, experiencer-stimulus verbs like "admired" in (70) tend to force a shift in perspective away from the initial perspective (John) to the stimulus (Mary) right at pronoun. The fact that these perspective shifts are being processed immediately on-line is good evidence in support of the perspective-taking account of sentence processing.

3.5.3 Expectations and Hypotheticals

Verbs and nouns often characterize complex configurations of social relations within individual clauses. Conjunctions and adverbs are used more to express ways in which clauses interact in terms

24

of presuppositions and perspective. Consider the conjunctions "but" and "although" in sentences like (72) and (73).

72. Mary wanted to win the race, but she felt a need to maintain her allegiance to Helen.

73. Mary wanted to win the race, although she felt a need to maintain her allegiance to Helen.

To understand (72), we have to figure out why Mary's winning of the race would weaken her allegiance to Helen. To understand (75) we additionally have to figure out how Mary thinks she is going to be able to balance her desire to win with her allegiance to Helen.

Language also provides devices for explicit constructions of hypothetical situations. The conjunction "if" is used to establish fictive mental states that very much echo the fictive motion and fictive causality we discussed earlier. Example (74) illustrates this.

74. If I were you, I would share the cookie with me.

To extract the meaning of (74), we need to take the perspective of the speaker and then imagine taking the perspective of the listener. Having done this, we need to understand why the speaker claims that the listener would want to share a cookie.

3.5.4 Mutual Reference

Within larger discourse frames, the establishment of reference for previously mentioned objects and actions relies on devices such as articles and pronouns. The study of these devices has been a major topic in functional linguistics (Haviland & Clark, 1974; Hawkins, 1977a, 1977b; Li & Thompson, 1979; MacWhinney, 1985). These analyses have shown that, in order to make proper use of pronouns and definite articles, we have to assume the perspective of our listener. If we choose to produce a sentence like (75), we need to be sure that our listener knows who "him" is, which car is being mentioned, which key unlocks that car, where the key is located, and where the car is located.

75. Please give him the key to the car.

In order to guarantee successful use of these forms, we have to track our listener's state of knowledge about the objects and positions being mentioned. This requires us to keep track of the conversation from the listener's perspective. As the conversation or narrative progresses, we have to continually update our assumed state of mutual reference to objects and spatial locations, since these form the backbone of a great deal of oral communication.

3.5.5 Theory of Mind

Together, these various devices allow us to talk about a wide range of social perspectives. Within developmental psychology, the study of the ability to take other mental and social perspectives has been discussed in terms of a "theory of mind" (Bartsch & Wellman, 1995; Flavell & Miller, 1997). The idea behind theory of mind is that we have to construct a mental model of the knowledge state of other people in order to solve certain problems and communicate successfully. Sentence (76) illustrates the type of embedded representations computed through a theory of mind.

76. Knowing what you expect me to know about what you promised me allows me to surmise that you will not be surprised if I turn down your offer.

3.5.6 Contrasting Perspectives

The various social conventions and forms we have mentioned so far have been confined to the lexical level. However, the construction of alternative social perspectives extends far beyond this level to encompass the whole of discourse. To illustrate how these various devices work together to build up larger perspectives, consider example (77) from Fauconnier and Turner (1996). In this example, a contemporary philosopher is imagining a dialog with Kant.

77. I claim that reason is a self-developing capacity. Kant disagrees with me on this point. He says it's innate, but I answer that that's begging the question, to which he counters, in <u>Critique of Pure Reason</u>, that only innate ideas have power. But I say to that, what about neuronal group selection? And he gives no answer.

Fauconnier and Turner note that this brief dialog established three mental spaces -- one for the speaker, one for Kant, and one for the projection of the two into a comparison space where the debate occurs.

Example (77) illustrates how persuasion involves negotiation between competing perspectives. On the one hand, speakers must demonstrate an understanding of the listeners' perspectives. At the same time, speakers want to be able to move listeners to move closer to their perspective. They do this by creating a hypothetical set of intermediary propositions that all can agree to. Then they show that this intermediate perspective could be reconceptualized as being exactly what the speaker believes in the first place. In this way, speakers and listeners move back and forth negotiating perspectives and social frames. Along the way, they rely on lexical, clausal, and discourse structures to cast their viewpoints into the most favorable perspectives.

3.5.7 Multifocal Chains

In order to build up persuasive and entertaining discourse, we need to control the shifting of perspective between social actors. Sometimes we can organize a narrative chain from a single perspective. For example, Bill could describe his travels through the Florida Everglades totally through the first person. This might work if he were travelling alone through the swamps. However, at his first encounter with another actor, be it an alligator or an egret, there could be a temporary shift in perspective. Although discourses are full of digressions to the perspectives of secondary actors, they typically maintain coherence by relating these excursions back to an ongoing basic chain.

A second type of perspectival organization structures a discourse as a juxtaposition of two or more simultaneous perspectives. This form of organization can involve comparisons and contrasts, or it can simply develop two alternative views of the same set of events. For example, we could describe the events surrounding the Battle of Stalingrad from the perspective of Hitler and the Wehrmacht, on the one hand, and Stalin and the Red Army on the other hand. A third form of organization involves the nesting of one full perspective chain within another. For example, within the story of Macbeth, we find nested the play that echoes the planning of the murder of Duncan.

Together, these various methods for maintaining and shifting perspective allow us to construct narratives and conversations that express and develop multifocal perspectives. This multifocality produces memories that are also organized about alternative perspectives. As a result, we can access our knowledge about people and places from alternative viewpoints. Our memories of Rome could be organized around restaurants in which we had eaten, events in Roman history, or ways to get around Rome in bus. The more we know about Rome and the Romans, the more multifocal our memories. Eventually, we can learn to view the city from the viewpoint of people who live in different districts or who have different occupations. This multifocality of representations reflects our expertise in dealing with any subject that we understand well. The more multifocal our representations, the more flexible the thinking and problem-solving that depends upon them.

3.5.8 Neuronal Mechanisms

We have already considered ways in which frontal working memory structures can store perspectives and frontal attentional processes can encourage perspective-shifting. However, we also know that lesions to areas such as the orbital gyrus can lead to a loss of normal motivation and social orientation. Little is known yet about the nature of the social processes supported by frontal cortex. Moreover, current models of frontal functioning typically emphasize low-level attentional processes (Shallice & Burgess, 1996). The perspective hypothesis would suggest that a pivotal component of social functioning is the ability to construct models of other people's minds and to use these models to control our own behavior and interactions with others. Although many aspects of these social functions could occur without the use of language, they can become greatly elaborated through communication and the co-construction of social structures.

4. Conclusion

The perspective hypothesis offers a new way of understanding the linkage between language, society, and the brain. In this new formulation, communication is viewed as a social interaction that activates mental processes of perspective-taking. Because perspective-taking is fundamental to communication, grammar has developed a set of devices designed specifically to mark shifts in perspectives. The process of perspective-shifting relies on at least five major neuronal systems that involve large areas of the cortex. Together, these systems allow us to store and retrieve previous impressions, actions, and affordances and link them into predicate chains and complex social plans. The perspective hypothesis makes a series of empirically testable claims about cognitive processing, language processing, language structure, and neuronal processing.

References

Aissen, J. (1997). On the syntax of obviation. *Language, 73,* 705-750.

Ardery, G. (1979). The development of coordinations in child language. *Journal of Verbal Learning and Verbal Behavior, 18,* 745-756.

Ariel, M. (1990). *Accessing noun phrase antecedents.* London: Routledge.

Baddeley, A. D. (1990). *Human memory: Theory and practice.* Needham Heights, MA: Allyn & Bacon.

Bailey, D., Chang, N., Feldman, J., & Narayanan, S. (1998). Extending embodied lexical development. *Proceedings of the 20th Annual Meeting of the Cognitive Science Society,* 64-69.

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences, 20,* 723-767.

Barch, D. M., Braver, T. S., Nystrom, L. E., Forman, S. D., Noll, D. C., & Cohen, J. D. (1997). Dissociating working memory from task difficulty in human prefrontal cortex. *Neuropsychologia, 35,* 1373-1380.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22,* 577-660.

Bartsch, K., & Wellman, H. (1995). *Children talk about the mind.* New York: Oxford University Press.

Barwise, J., & Perry, J. (1983). *Situations and attitudes.* Cambridge, MA: MIT Press.

Behrmann, M., Moscovitch, M., & Winocur, G. (1994). Intact visual imagery and impaired visual perception in a patient with visual agnosia. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 1068-1087.

Berkowitz, L., & Troccoli, B. T. (1990). Feelings, direction of attention, and expressed evaluations of others. *Cognition and Emotion, 4,* 305-325.

Berman, R. A., & Slobin, D. I. (1994). *Relating events in narrative: A crosslinguistic developmental study.* Hillsdale, NJ: Lawrence Erlbaum.

Booth, J. R., MacWhinney, B., Thulborn, K. R., Sacco, K., Voyvodic, J., & Feldman, H. (1999). Functional organization of activation patterns in children: Whole brain fMRI imaging during three different cognitive tasks. *Progress in Neuropsychopharmocology and Biological Psychiatry, 23*, 669-682.

Bossom, J. (1965). The effect of brain lesions on adaptation in monkeys. *Psychonomic Science, 2*, 45-46.

Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., & Noll, D. C. (1997). A parametric study of prefrontal cortex involvement in human working memory. *Neuroimage, 6*, 49-62.

Brooks, R. A. (1991). How to build complete creatures rather than isolated cognitive simulators. In K. VanLehn (Ed.), *Architectures for intelligence* (pp. 225-240). Hillsdale, NJ: Lawrence Erlbaum.

Bryant, D. J., Tversky, B., & Franklin, N. (1992). Internal and external spatial frameworks for representing described scenes. *Journal of Memory and Language, 31*, 74-98.

Bybee, J. (1985). *Morphology: A study of the relation between meaning and form.* Amsterdam: John Benjamins.

Carlson-Radvansky, L. A., & Logan, G. D. (1997). The influence of reference frame selection on spatial template construction. *Journal of Memory and Language, 37*, 411-437.

Carreiras, M., Carriedo, N., Alonso, M. A., & Fernández, A. (1997). The role of verb tense and verb aspect in the foregrounding of information during reading. *Memory and Cognition, 25*, 438-446.

Case, R. (1997). The development of conceptual structures. In W. Damon & D. Kuhn & R. Siegler (Eds.), *The handbook of child development* (pp. 745-800). New York: Wiley.

Chomsky, N. (1982). *Some concepts and consequences of the theory of government and binding.* Cambridge, MA: MIT Press.

Clark, H. H. (1973). Space, time, semantics, and the child. In T. E. Moore (Ed.), *Cognitive development and language acquisition* (pp. 28-63). New York: Academic Press.

Cohen, D. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., & Smith, E. (1997). Temporal dynamics of brain activation during a working memory task. *Nature, 386*, 604-608.

Cohen, J., Dunbar, K., & McClelland, J. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review, 97*, 332-361.

Cohen, M. S., Kosslyn, S. M., Breiter, H. C., DiGirolamo, G. J., Thompson, W. L., Anderson, A. K., Bookheimer, S. Y., Rosen, B. R., & Belliveau, J. W. (1996). Changes in cortical activity during mental rotation. A mapping study using functional MRI. *Brain, 119*, 89-100.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-684.

Damasio, A. (1999). *The feeling of what happens: Body and emotion in the making of consciousness.* New York: Harcourt Brace.

de Leon, L. (1994). Exploration in the acquisition of geocentric location by Tzotzil children. *Linguistics, 32*, 857-884.

Delancey, S. (1981). An interpretation of split ergativity and related patterns. *Language, 57*, 626-658.

Du Bois, J. (1987). The discourse basis of ergativity. *Language, 63*, 805-856.

Ertel, S., & Bloemer, W. (1975). Affirmation and negation as constructive action. *Psychologische Forschung, 37*, 335-342.

Fauconnier, G. (1994). *Mental spaces: Aspects of meaning construction in natural language.* Cambridge: Cambridge University Press.

Fauconnier, G., & Turner, M. (1996). Blending as a central process of grammar. In A. Goldberg (Ed.), *Conceptual structure, discourse, and language* (pp. 113-130). Stanford, CA: CSLI.

Feldman, J., Lakoff, G., Bailey, D., Narayanan, S., Regier, T., & Stolcke, A. (1996). L$_0$ -- The first five years of an automated language acquisition project. *AI Review, 10*, 103-129.

Flavell, J. H., & Miller, P. H. (1997). Social cognition. In W. Damon & D. Kuhn & R. Siegler (Eds.), *Handbook of child psychology* (pp. 851-898). New York: Wiley.

Flores d'Arcais, G. B. (1975). Some perceptual determinants of sentence construction. In G. B. F. d'Arcais (Ed.), *Studies in perception* (pp. 343-373). Milan: Martello-Giunti.

Flores d'Arcais, G. B. (1987). Syntactic processing during reading for comprehension. In M. Coltheart (Ed.), *Attention and Performance VII: The psychology of reading* (Vol. 12, pp. 619-654). Hillsdale, NJ: Lawrence Erlbaum.

Fourcin, A. J. (1975). Language development in the absence of expressive speech. In E. H. Lenneberg & E. Lenneberg (Eds.), *Foundations of language development: A multidisciplinary approach* (Vol. 2, pp. 263-268). New York: Academic Press.

Franklin, N., Tversky, B., & Coon, V. (1992). Switching points of view in spatial mental models. *Memory and Cognition, 20,* 507-518.

Fuster, J. M. (1989). *The prefrontal cortex.* New York: Raven Press.

Gainotti, G., Silveri, M. C., Daniele, A., & Giustolisi, L. (1995). Neuroanatomical correlates of category-specific semantic disorders: A critical survey. *Memory, 3,* 247-264.

Gathercole, V., & Baddeley, A. (1993). *Working memory and language.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Gernsbacher, M. A. (1990). *Language comprehension as structure building.* Hillsdale, NJ: Lawrence Erlbaum.

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Boston: Houghton Mifflin.

Gibson, J. J. (1977). The theory of affordances. In R. E. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 67-82). Hillsdale, NJ: Lawrence Erlbaum.

Givón, T. (1976). Topic, pronoun, and grammatical agreement. In C. Li (Ed.), *Subject and topic* (pp. 149-188). New York: Academic Press.

Glenberg, A. (1997). What memory is for. *Behavioral and Brain Sciences, 20,* 1-55.

Glenberg, A., & Langston, W. (1992). Comprehension of illustrated text: Pictures help to build mental models. *Journal of Memory and Language, 31,* 129-151.

Goldman-Rakic, P. S. (1987). Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In V. B. Mountcastle & F. Plum & S. R. Geiger (Eds.), *Handbook of Physiology, vol. 5* (pp. 373-417). Bethesda, MD: American Physiological Society.

Goodale, M. A. (1993). Visual pathways supporting perception and action in the primate cerebral cortex. *Current Opinion in Neurobiology, 3,* 578-585.

Gordon, P., Grosz, B., & Gilliom, L. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science, 17,* 311-317.

Grasby, P. M., Frith, C. D., Friston, K. J., Bench, C., Frackowiak, R. S. J., & Dolan, R. J. (1993). Functional mapping of brain areas implicated in auditory-verbal memory function. *Brain, 116,* 1-20.

Greenfield, P. (1991). Language, tools and brain: The ontogeny and phylogeny of herarchically organized sequential behavior. *Behavioral and Brain Sciences, 14,* 531-595.

Haenggi, D., Kintsch, W., & Gernsbacher, M. A. (1995). Spatial situation models and text comprehension. *Discourse Processes, 19,* 173-199.

Harnad, S. (1995). Grounding symbolic capacity in robotic capacity. In L. Steels & R. Brooks (Eds.), *The "artificial life" route to "artificial intelligence": Building situated embodied agents* (pp. 277-286). Mahwah, NJ: Lawrence Erlbaum.

Haviland, J. (1996). Projections, transpositions, and relativity. In J. Gumperz & S. Levinson (Eds.), *Rethinking linguistics relativity* (pp. 271-323). New York: Cambridge University Press.

Haviland, S., & Clark, H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior, 13,* 512-521.

Hawkins, J. (1977a). The pragmatics of definiteness: Part 1. *Linguistische Berichte, 48,* 1-25.

Hawkins, J. (1977b). The pragmatics of definiteness: Part 2. *Linguistische Berichte, 48,* 1-27.

Horowitz, L., & Prytulak, L. (1969). Redintegrative memory. *Psychological Review, 76,* 519-531.

Jeannerod, M. (1997). *The cognitive neuroscience of action.* Cambridge, MA: Blackwell.

Johnson-Laird, P. (1968a). The choice of the passive voice in a communicative task. *British Journal of Psychology, 59*, 7-15.

Johnson-Laird, P. (1968b). The interpretation of the passive voice. *Quarterly Journal of Experimental Psychology, 20*, 69-73.

Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., & Thulborn, K. R. (1996). Brain activation modulated by sentence comprehension. *Science, 274*, 114-116.

Kakei, S., Hoffman, D. S., & Strick, P. L. (1999). Muscle and movement representations in the primary motor cortex. *Science, 285*, 2136-2139.

Keenan, J., MacWhinney, B., & Mayhew, D. (1977). Pragmatics in memory: A study in natural conversation. *Journal of Verbal Learning and Verbal Behavior, 16*, 549-560.

Klatzky, R. L., Pellegrino, J. W., McCloskey, B. P., & Doherty, S. (1989). Can you squeeze a tomato? The role of motor representations in semantic sensibility judgments. *Journal of Memory and Language, 28*, 56-77.

Kolb, B., & Whishaw, I. Q. (1995). *Fundamentals of Human Neuropsychology. Fourth Edition.* New York: W. H. Freeman.

Kosslyn, S. M., Thompson, W. L., Kim, I. J., & Alpert, N. M. (1995). Topographical representations of mental images in primary visual cortex. *Nature, 378*, 496-498.

Kuno, S. (1986). *Functional syntax.* Chicago: University of Chicago Press.

Lakoff, G. (1987). *Women, fire, and dangerous things.* Chicago: Chicago University Press.

Lakoff, G., & Thompson, H. (1977). Introducing cognitive grammar, *Papers from the Third Annual Meeting of the Berkeley Linguistics Society.* Berkeley: Berkeley Linguistics Society.

Langacker, R. (1995). Viewing in grammar and cognition. In P. W. Davis (Ed.), *Alternative linguistics: Descriptive and theoretical models* (pp. 153-212). Amsterdam: John Benjamins.

Langacker, R. W. (1986). An introduction to cognitive grammar. *Cognitive Science, 10*, 1-40.

Levine, D. S., & Prueitt, P. S. (1989). Modelling some effects of frontal lobe damage -- Novelty and perseveration. *Neural Networks, 2*, 103-116.

Li, C. N., & Thompson, S. A. (1979). Third-person pronouns and zero-anaphora in Chinese discourse. In T. Givón (Ed.), *Syntax and semantics: Discourse and syntax* (Vol. 12, pp. 311-336). New York: Academic Press.

Luria, A. R. (1959). The directive function of speech in development and dissolution. *Word, 15*, 453-464.

Luria, A. R. (1975). Basic problems of language in the light of psychology and neurolinguistics. In E. H. Lenneberg & E. Lenneberg (Eds.), *Foundations of language development: A multidisciplinary approach* (Vol. 2, pp. 49-73). New York: Academic Press.

Lust, B., & Mervis, C. A. (1980). Development of coordination in the natural speech of young children. *Journal of Child Language, 7*, 279-304.

MacWhinney, B. (1977). Starting points. *Language, 53*, 152-168.

MacWhinney, B. (1982). Basic syntactic processes. In S. Kuczaj (Ed.), *Language acquisition: Vol. 1. Syntax and semantics* (pp. 73-136). Hillsdale, NJ: Lawrence Erlbaum.

MacWhinney, B. (1985). Grammatical devices for sharing points. In R. Schiefelbusch (Ed.), *Communicative competence: Acquisition and intervention* (pp. 325-374). Baltimore, MD: University Park Press.

MacWhinney, B. (1989). Competition and lexical categorization. In R. Corrigan & F. Eckman & M. Noonan (Eds.), *Linguistic categorization* (pp. 195-242). Philadelphia: Benjamins.

MacWhinney, B., & Bates, E. (1978). Sentential devices for conveying givenness and newness: A cross-cultural developmental study. *Journal of Verbal Learning and Verbal Behavior, 17*, 539-558.

MacWhinney, B., & Pléh, C. (1988). The processing of restrictive relative clauses in Hungarian. *Cognition, 29*, 95-141.

Maki, R. H., & Marek, M. N. (1997). Egocentric spatial framework effects from single and multiple points of view. *Memory and Cognition, 25*, 677-690.

Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature, 379*, 649-652.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102*, 419-457.

McDonald, J. L., & MacWhinney, B. J. (1995). The time course of anaphor resolution: Effects of implicit verb causality and gender. *Journal of Memory and Language, 34*, 543-566.

Menard, M. T., Kosslyn, S. M., Thompson, W. L., Alpert, N. M., & Rauch, S. L. (1996). Encoding words and pictures: A positron emission tomography study. *Neuropsychologia, 34*, 185-194.

Mesulam, M.-M. (1990). Large-scale nuerocognitive networks and distributed processing for attention, language, and memory. *Annals of Neurology, 28*, 597-613.

Michaels, C. F., & Oudejans, R. R. D. (1992). The optics and actions of catching fly balls: Zeroing out optical acceleration. *Ecological Psychology, 4*, 199-222.

Middleton, F. A., & Strick, P. L. (1998). Cerebellar output: Motor and cognitive channels. *Trends in Cognitive Sciences, 2*, 348-354.

Miller, G., & Johnson-Laird, P. (1976). *Language and perception.* Cambridge, MA: Harvard University Press.

Morrow, D. G. (1986). Places as referents in discourse. *Journal of Memory and Language, 25*, 676-690.

Morrow, D. G., Bower, G. H., & Greenspan, S. L. (1989). Updating situation models during narrative comprehension. *Journal of Memory and Language, 28*, 292-312.

Morrow, D. G., Greenspan, S. L., & Bower, G. H. (1987). Accessibility and situation models in narrative comprehension. *Journal of Memory and Language, 26*, 165-187.

Neisser, U. (1976). *Cognition and reality.* San Francisco: Freeman.

Olson, C. R., & Gettner, S. N. (1995). Object-centered direction selectivity in the macaque supplementary eye field. *Science, 269*, 985-988.

Osgood, C. E., & Bock, K. J. (1977). Salience and sentencing: Some production principles. In S. Rosenberg (Ed.), *Sentence production: Developments in research and theory* (pp. 89-140). Hillsdale, NJ: Lawrence Erlbaum Associates.

Owen, A. M., Downes, J. D., Sahakian, B. J., Polkay, C. E., & Robbins, T. W. (1990). Planning and spatial working memory following frontal lobe lesions in man. *Neuropsychologia, 28*, 1021-1034.

Paivio, A. (1971). *Imagery and verbal processes.* New York: Rinehart and Winston.

Parsons, L. M., Fox, P. T., Downs, J. H., Glass, T., Hirsch, T. B., Martin, C. C., Jerabek, P. A., & Lancaster, J. L. (1995). Use of implicit motor imagery for visual shape discrimination as revealed by PET. *Nature, 375*, 54-58.

Paulesu, E., Frith, C. D., & Frackowiak, R. S. J. (1993). The neural correlates of the verbal component of working memory. *Nature, 362*, 342-345.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature, 331*, 585-589.

Piaget, J. (1952). *The origins of intelligence in children.* New York: International Universities Press.

Piaget, J. (1954). *The construction of reality in the child.* New York: Basic Books.

Pinker, S., & Birdsong, D. (1979). Speaker's sensitivity to rule of frozen word order. *Journal of Verbal Learning and Verbal Behavior, 18*, 497-508.

Posner, M., Petersen, S., Fox, P., & Raichle, M. (1988). Localization of cognitive operations in the human brain. *Science, 240*, 1627-1631.

Potter, M. (1993). Very short term conceptual memory. *Memory and Cognition, 21*, 156-161.

Redish, D., & Touretzky, D. S. (1997). Cognitive maps beyond the hippocampus. *Hippocampus, 7*, 15-35.

Reichenbach, H. (1947). *Elements of symbolic logic.* New York: The Free Press.

Reinhart, T. (1981). Definite NP anaphora and c-command domains. *Linguistic Inquiry, 12*, 605-635.

Reinhart, T. (1983). *Anaphora and semantic interpretation.* Chicago: University of Chicago Press.

Rieser, J. J., Garing, A. E., & Young, M. F. (1994). Imagery, action, and young children's spatial orientation: It's not being there that counts, it's what one has in mind. *Child Development, 65*, 1262-1278.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3*, 131-141.

Sacerdoti, E. (1977). *A structure for plans and behavior*. New York: Elsevier Computer Science Library.

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics, 2*, 437-442.

Sanford, A. J., & Moxey, L. M. (1995). Aspects of coherence in written language: A psychological perspective. In M. A. Gernsbacher & T. Givón (Eds.), *Coherence in spontaneous text* (pp. 218-248). Philadelphia: John Benjamins.

Shallice, T., & Burgess, P. (1996). The domain of supervisory processes and temporal organization of behavior. *Philosophical Transactions of the Royal Society of London B, 351*, 1405-1412.

Silverstein, M. (1976a). Hierarchy of features and ergativity. In R. Dixon (Ed.), *Grammatical categories in Australian Languages* (pp. 112-171). Canberra: Australian Institute of Aboriginal Studies.

Silverstein, M. (1976b). Shifters, linguistic categories and cultural description. In K. H. Basso & H. A. Selby (Eds.), *Meaning in Anthropology* (pp. 11-55). Albuquerque: University of New Mexico Press.

Slobin, D. (1985). Crosslinguistic evidence for the language-making capacity. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition. Volume 2: Theoretical issues* (pp. 1157-1256). Hillsdale, N. J.: Lawrence Erlbaum.

Slobin, D. I., & Welsh, C. A. (1973). Elicited imitation as a research tool in developmental psycholinguistics. In C. A. Ferguson & D. I. Slobin (Eds.), *Studies of child language development* (pp. 485-497). New York: Holt, Rinehart and Winston.

Sokolov, A. (1972). *Inner speech and thought*. New York: Plenum Press.

Sridhar, S. N. (1988). *Cognition and sentence production: cross-linguistic study*. New York: Springer Verlag.

Tabachneck-Schijf, H. J. M., Leonardo, A. M., & Simon, H. A. (1997). CaMeRa: A computational model of multiple representations. *Cognitive Science, 21*, 305-350.

Talmy, L. (1978). The relation of grammar to cognition - a synopsis. In C. S. Laboratory (Ed.), *Theoretical Issues in Natural Language Processing-2* (pp. 14-24). Urbana, IL: University of Illinois.

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science, 12*, 59-100.

Teuber, H.-L. (1964). The riddle of frontal lobe function in man. In J. M. Warren & K. Akert (Eds.), *The frontal granular cortex and behavior* (pp. 410-477). New York: McGraw-Hill.

Tversky, B. (1996). Spatial perspective in descriptions. In P. Bloom & M. A. Peterson & L. Nadel & M. F. Garrett (Eds.), *Language and space* (pp. 463-491). Cambridge, MA: MIT Press.

Ungerleider, L. G., & Haxby, J. V. (1994). 'What' and 'where' in the human brain. *Current Opinion in Neurobiology, 4*, 157-165.

van Hoek, K. (1997). *Anaphora and conceptual structure*. Chicago: University of Chicago Press.

Vosniadou, S. (1994). Universal and culture-specific properties of children's mental models of the earth. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind* (pp. 176-212). New York: Cambridge University Press.

Vygotsky, L. (1962). *Thought and language*. Cambridge: MIT Press.

Warren, W. H. (1984). Perceiving affordances: Visual guidance of stair climbing. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 683-703.

Warren, W. H., Young, D. S., & Lee, D. N. (1986). Visual control of step length during running over irregular terrain. *Journal of Experimental Psychology: Human Perception and Performance, 12*, 259-266.

Weist, R. (1986). Tense and aspect. In P. Fletcher & M. Garman (Eds.), *Language acquisition (2nd Ed.)* (pp. 356-374). Cambridge: Cambridge University Press.

Werner, H., & Kaplan, B. (1963). *Symbol formation: An organismic-developmental approach to language and the expression of thought*. New York: Wiley.

Zwaan, R. A. (1996). Processing narrative time shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1196-1207.

視点設定と文法

ブライアン・マックウィニー（カーネギーメロン大学）

要旨

　文の理解と産出は５つに区別されるレベル（物体知覚・動作・時空間参照フレーム・述部の連鎖・社会的フレーム）において作用する視点設定の過程にその基礎をおく。これらのそれぞれのレベルにおいて、話し手と聞き手は言語の持つ語彙や文法という形式を使って、二つ以上の競合する視点を統合したり、あるいは視点をシフトさせる。５つの視点の体系は構成され身体化された自己の視点にその基礎をおいている。言語は既存の脳神経系を利用してこれらの５つのシステムの視点ををダイナミックに結びつけ、統合された人間の視点を発現させる。